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

VOLUME XXIII SEPTEMBER 1962 NUMBER 3

CONTENTS

© 1962 by Radio Corporation of America
All rights reserved

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

RCA REVIEW

BOARD OF EDITORS

Chairman

R. S. HOLMES
RCA Laboratories

E. I. ANDERSON
Home Instruments Division

A. A. BARCO
RCA Laboratories

G. L. BEERS
Radio Corporation of America

G. H. BROWN
Radio Corporation of America

A. L. CONRAD
RCA Service Company

E. W. ENGSTROM
Radio Corporation of America

D. H. EWING
Radio Corporation of America

A. N. GOLDSMITH
Honorary Vice President, RCA

J. HILLIER
RCA Laboratories

E. C. HUGHES
Electron Tube Division

E. A. LAPORT
Radio Corporation of America

H. W. LEVERENZ
RCA Laboratories

G. F. MAEDEL
RCA Institutes, Inc.

W. C. MORRISON
Defense Electronic Products

L. S. NERGAARD
RCA Laboratories

G. M. NIXON
National Broadcasting Company

H. F. OLSON
RCA Laboratories

J. A. RAJCHMAN
RCA Laboratories

D. S. RAU
RCA Communications, Inc.

D. F. SCHMIT
Radio Corporation of America

L. A. SHOTLIFF
RCA International Division

S. STERNBERG
Astro-Electronics Division

W. M. WEBSTER
RCA Laboratories

I. WOLFF
Radio Corporation of America

Secretary

C. C. FOSTER
RCA Laboratories

REPUBLICATION AND TRANSLATION

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

ELECTRONIC TECHNIQUES IN A SYSTEM OF HIGHWAY VEHICLE CONTROL

By

L. E. Flory,* G. W. Gray,* R. E. Morey,* W. S. Pike,*
and C. O. Caulton†

Summary—*In the early 1950's, a project was initiated at RCA Laboratories to determine how electronic techniques might best be applied to problems of highway vehicle control. From these studies an over-all system concept has evolved. This paper describes some of the important basic elements of the system which have been built and demonstrated.*

Introduction

THE PURPOSE OF TRAFFIC CONTROL is to facilitate the quick, safe movement of vehicles. All existing traffic controls are based on a single concept—communication with the driver to advise him as to a proper course of action. The driver then applies the necessary controls. Nearly all of the signal inputs to the driver are visual—warning signs, traffic lights, stop signs, brake lights, turn indicators, hand signals and the like.

Mechanical improvements in cars and road systems have been such that the driver is now the least reliable link in the control system. Among the more common causes of driver failure are fatigue, inattentiveness induced by protracted driving on high-speed roads, indecision, and faulty judgment. Even in the case of an alert driver, reaction time alone can account for a delay of an appreciable part of a second in applying controls.

The Role of Electronics

It is proposed to employ electronic methods to

(1) improve normal communication with the driver,

* RCA Laboratories, Princeton, N. J.

† RCA Service Company, Camden, N. J.

(2) introduce emergency warning systems which will communicate warning to the driver if he neglects to act promptly or acts in a manner which jeopardizes the safety of himself or of other drivers.

(3) eventually take over direct control of a portion of the driving procedure.

There has been considerable difference of opinion as to the best approach to electronic traffic control. There are, in fact, many possible choices. Before a complete "ultimate" system can be devised, numerous decisions must be made. Many of these decisions cannot be made in advance because the decisions will affect future highway and vehicle development in ways which are now unpredictable. The approach, therefore, must be to determine where and how electronic controls might be used under present conditions, build and test the basic elements of these controls, and then, with experience gained from this beginning, proceed step-by-step toward a complete system.

It has been said that electronic techniques exist for performing any degree of control which traffic engineers might call for. This is basically true, but many complications arise, not the least of which is the matter of economics. Thus continuous cooperation between electronics people and highway and traffic engineers will be necessary. Many probing attempts will have to be made to determine which electronic techniques are most suitable, and it is quite possible that traffic engineers may make radical changes in their thinking as a result of these probing experiments.

The Proposed System

In considering the requirements of an over-all system, it soon became apparent that devices installed in cars would be of no more than interim value unless they could eventually be integrated into the system. Thus, anticollision radar, supersonic devices, acceleration-deceleration warning systems, and the like were excluded from consideration.

Consider travel on a limited-access highway, not because it is where electronic controls are most needed, but rather because the problems there are relatively simple. A single lane of traffic with no passing requires only two functions from the driver. First, he must stay in his lane, and second, he must avoid colliding with a car in front of him. As a matter of fact, a high percentage of accidents on turnpikes occur as a result of rear-end collisions. In this over-simplified situation, if automatic means are provided for keeping a car in its lane and

for preventing collision with a car ahead, we have the elements of a complete system.

One more operational requirement must be met. This is that any system must be capable of being introduced on a gradual basis and must at all stages be compatible with existing traffic. This is an obvious but very important requirement. It would be impossible to convert many thousands of miles of highway and tens of millions of vehicles to automatic control overnight. It would be highly impractical to build a controlled road and immediately require that all cars traveling this road be equipped with the necessary controls. Thus, equipped and unequipped vehicles will have to use the same highways intermixed for years to come. Within the framework of the proposed system, a measure of control of unequipped cars is available in the form of traffic-actuated warning signals which will benefit cars with no special equipment. In any case, as a minimum requirement, the introduction of controls in the roadway and the presence of controlled cars must not increase the danger to unequipped cars. It is further obvious that with intermixed traffic, a car equipped with controls must not only be protected against collision with another of its kind, but also must be prevented from colliding with unequipped vehicles.

Thus, the following sequence of implementation might be effected: first, a stage of improved communication with the driver to enable him to make decisions better and faster; second, a stage where some of the decisions are made for him and he is given a warning if he is approaching a dangerous condition; and third, a stage where the system actually takes control of the car either on a continuous basis or as a limit-type operation which would take over only if the driver failed to respond properly to the warning system. All of these stages can be realized within the framework of one system, so that each added feature will contribute toward the ultimate system, thus realizing the feature of compatibility and still allowing the system to grow. This, of course, does not exclude the use of other electronic aids in the interim which may independently contribute to the immediate overall safety but which will not become a permanent part of the final control system.

For purposes of discussion it is convenient to consider the completely automatic stage first. In the simple case of a single lane of traffic on a controlled-access highway, the two functions to be performed are lane guidance and collision prevention. Consider first the question of lane guidance, since it is the simplest of the two functions from an electronic standpoint. It is obvious that to accomplish guidance there must be some interaction between the vehicle and the highway lane.

This can be accomplished by laying down in or on the roadway a trail which a vehicle can follow. The railroads do this with steel rails, but a mechanical system seems too expensive and inflexible for private vehicles. Airlines follow a radio beam, but the precision is not sufficiently high for highway use. A radioactive strip has possibilities, but it appears that the most practical means is the insertion of a cable carrying an alternating current down the center of the lane. Simple detectors mounted on the car can then provide control signals to the steering mechanism which will keep the car centered over the wire. This method permits the use of different frequencies for different lanes and a means for route selection by change in frequency.

In 1953, Dr. V. K. Zworykin demonstrated the basic principles of the system using 1/5 scale model cars. In 1957, a demonstration was given in Nebraska in which full-size cars were used to demonstrate improved control techniques. Shortly thereafter, a project was initiated to apply these techniques to actual fully automatic control of suitably equipped cars. For this purpose a quarter-mile oval track was built at RCA Laboratories and equipped with the necessary electronic installations. In 1960, in conjunction with General Motors Corporation, which furnished cars with the necessary servo systems, a demonstration was made to the press and interested highway officials.

Demonstration Equipment

Figure 1 is a photograph of the two steering antennas mounted on the front of a car. Figure 2 is a block diagram of the automatic steering circuits used with these antennas. The signal from each antenna is amplified and the two amplitudes are compared in order to produce an error signal showing the car's position relative to the wire in the roadway. This signal is applied to the steering servo and turns the wheels to steer the vehicle back over the guide wire so that the signals from the two antennas are equal and balanced. The overall loop gain is a function of the vehicle speed thus requiring the electrical gain in the steering servo to vary as vehicle speed varies in order to prevent oscillations in the system.

The problem of maintaining spacing between vehicles is much more complex since it involves the interaction of two vehicles moving independently, only one of which (the trailing car) can be assumed to have any special equipment. This situation suggests that the road must be the agent for transmitting signals from the lead vehicle to the following one. Also, signals from the lead vehicle must originate without any action from it, so that the system will function with non-

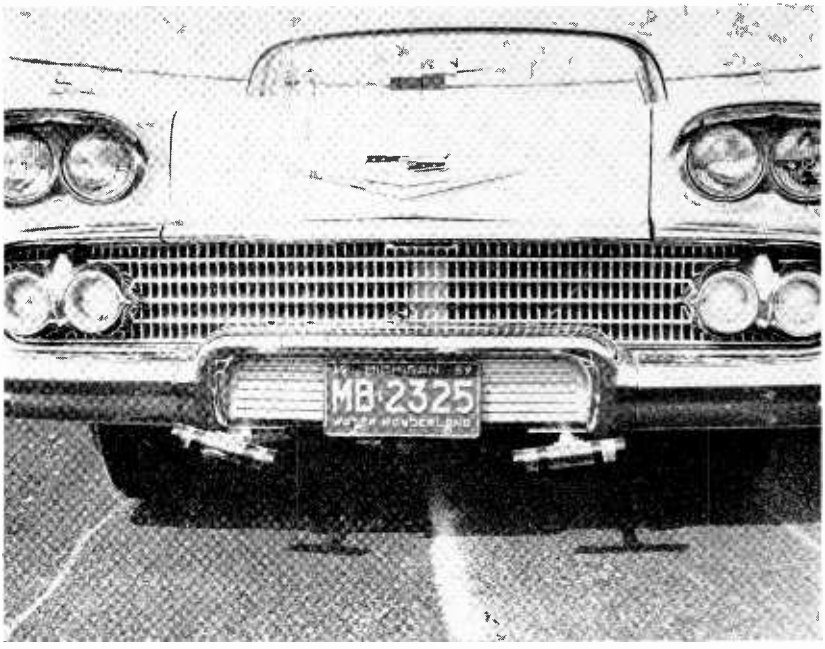

Fig. 1—Steering and spacing antennas on "equipped" car.

equipped vehicles present. Thus, the first requirement for the automatic road is a vehicle detector which will determine the position of all vehicles on the road whether they have special equipment or not. The road, knowing the location of all vehicles on it, can then transmit to following vehicles information that the vehicle may use if it is to be automatically controlled. Figure 3 shows the block diagram of the vehicle detector. The detecting element is a loop of wire buried in the road and approximately the size of the block to be protected. The loop is resonated at some low radio frequency, say 300 kc, and driven with a current of that frequency from the driving amplifier. The vehicle passing over the loop, will cause a slight decrease in the in-

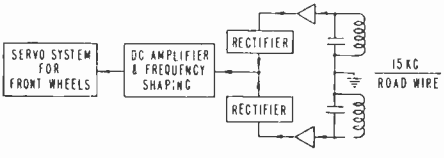

Fig. 2—Block diagram of automatic steering system.

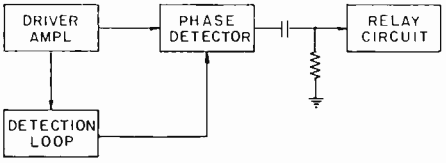

Fig. 3—Block diagram of vehicle detector.

ductance of the loop and thereby a change in the phase angle of the voltage across the loop. This phase change is detected by the phase detector and produces an output voltage pulse in response to the passing of the vehicle.

The output of the phase detector is a-c coupled to a relay circuit, one pole of the relay being utilized as the output of the vehicle detector for indicating the presence of a vehicle over the detection loop. The purpose in a-c coupling the relay circuit is to filter out slow changes in tuning of the loop due to weather variations. This is necessary since variations in tuning of the loop due to weather may be several times as large as the signal from a vehicle. However, this raises a complication in the relay circuit since the relay indication must not drop out if the vehicle stops on the loop, and with simple a-c coupling this would be the case. Figure 4 shows the relay circuit in more

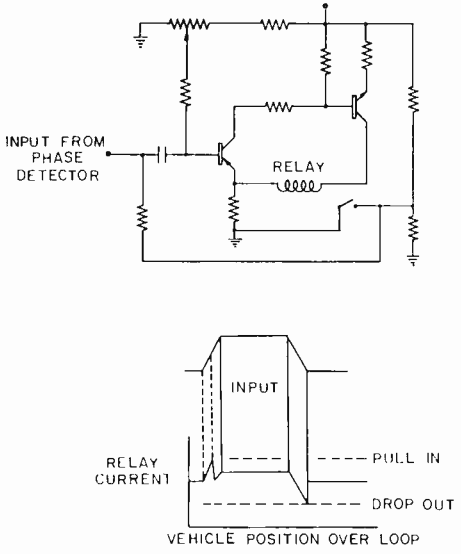

Fig. 4—Vehicle detector relay operation.

detail. The circuit is normally adjusted so that the relay current in the absence of a vehicle is half-way between the pull-in and drop-out current. Thus, if a vehicle should stop on the detection loop, it will cause the relay to pull in, and as time passes, the relay current will drop back to its normal value between pull-in and drop-out. However, the relay still holds in and indicates the vehicle's presence. Then, when the vehicle leaves the loop the relay current will drop enough to cause the relay to drop out and thus indicate that the vehicle has left the loop.

However, the case of a vehicle which passes over the detection loop rather rapidly is not quite so simple. Here, the relay would pull in, but there would not be any reverse signal which would cause the relay to drop out. The current would just drop back to the normal value as the vehicle left the loop, and never fall low enough to cause the relay to open. In order to overcome this difficulty there is a feedback circuit from a second pole of the relay back to the input. As shown in the diagram of relay current versus vehicle position, when the relay pulls in, this circuit feeds in a small relay-open signal so that the current drops back a little more than half the pull-in current. Thus, as the vehicle leaves the detection loop the relay circuit is so biased that the phase detector output returning to normal value will be sufficient to cause the relay to open.

A lane equipped with such detectors will know the location of all vehicles in the lane. The road can then be equipped with circuits which will provide signals carrying information as to the speed and position. Any vehicle equipped with an appropriate receiver can then receive the signals and know the distance to and speed of any vehicle in front of it. In order to simplify the circuitry in the road, the speed information is not transmitted directly but instead, a pulse is transmitted each time the lead vehicle activates a detector and then the vehicle receiving the pulses can compute the speed from the time between pulses.

Figure 5 shows a diagram of the wiring installed in and alongside the roadway. The detection loops are installed about every 20 to 30 feet; also, there is a second set of wires consisting of antennas to radiate information to any vehicles equipped to receive the information. These tail-warning antennas consist of a wire running down the center of the lane, one wire for each detection loop or block. One end of the antenna is brought out to the side of the road and grounded; the other end is brought out to where it will be connected to a tail-warning signal generator. A third wire for steering runs down the center of the lane for a mile or more between signal sources.

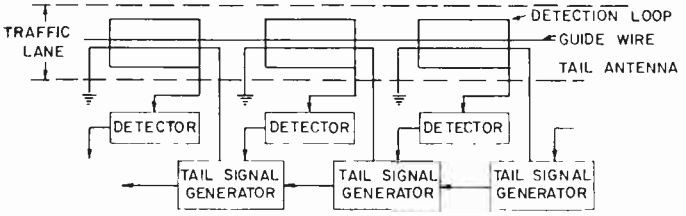

Fig. 5—Block diagram of road installation.

The output of each detector is connected to the adjacent tail-warning signal generator; these generators are connected to each other along the length of the road. Thus, information from one detector is carried along the road and may be transmitted to a following vehicle by means of the tail-warning signal generators. The basic means by which the information is carried along the road is shown in Figure 6. This consists of an attenuating line made up of series diodes and shunt resistors. The detector applies a d-c voltage to the diode line in the block immediately behind the location of the vehicle which is causing the detector to activate. The diodes act upon this voltage in two ways. First, they prevent the transmission of the voltage in a forward direction, that is, in a direction in which traffic is moving. Second, in the rearward direction, the diodes cause a linear attenuation of the voltage, block by block, until many blocks behind the detector originating the signal, the voltage has returned to ground level. Shunt diodes to ground are added to prevent the voltage going positive. The diodes used are silicon junction diodes so that in the direction of propagation there is a uniform drop of about ½ volt per diode. Also, the junction diode provides a very low a-c impedance in the direction of propaga-

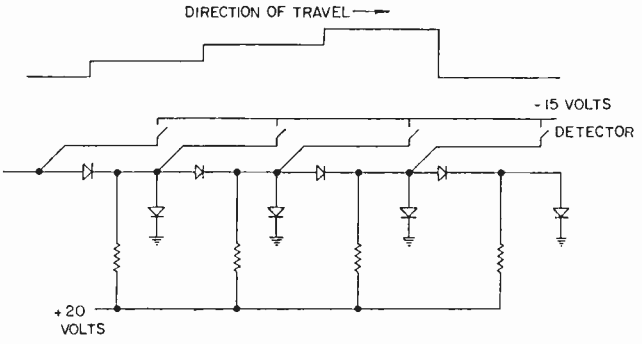

Fig. 6—Basic circuit of diode tail.

tion. Thus, a measurement of the d-c voltage on the diode line will indicate how far in advance of that position there is another vehicle. One of the functions of the tail-warning signal generator is to transmit this d-c voltage to any vehicle which may be at that particular block and equipped to receive the information. Figure 7 shows a block diagram of the tail-warning signal generators which indicates how this d-c voltage is transmitted to vehicles on the road. From a central source which may supply miles of roadway, a 100-cycle sawtooth with a small amount of 4.5-kc sinewave added to it, is supplied. At each generator the sawtooth peak is clamped to ground, and the

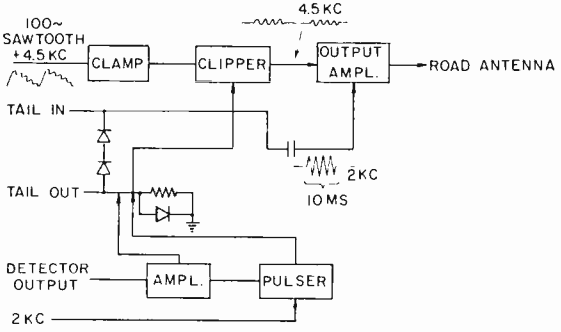

Fig. 7—Block diagram of tail-warning signal generators.

portion of the sawtooth whose d-c level is above that of the d-c voltage existing on the diode line is clipped off. From this clipped portion of the sawtooth the 4.5 kc is filtered off and the 100-cycle component of the sawtooth rejected. Thus, bursts of 4.5 kc at a 100-cycle repetition rate are produced, the length of these bursts being controlled by the d-c level on the diode line. The 4.5 kc is present 100 per cent of the time if the diode line voltage is zero, indicating no car within range of detection. Levels are adjusted so that with about 10 per cent duty cycle of the 4.5 kc, the presence of a vehicle immediately in front is indicated. These bursts of 4.5 kc then go to an output amplifier which is connected to the road antennas at each block and thus can be picked up by any vehicle at that point on the road.

The diode line is also used to transmit information as to the speed of a leading vehicle. When the detector applies the d-c voltage to the diode line it also activates a pulse circuit, which applies a 10-millisecond pulse of 2 kc to the diode line. At each tail-warning signal generator this 10-millisecond pulse is detected and fed to the output

amplifier and thus, to the road antenna. Both the d-c voltage for distance measurement and the 10-millisecond pulse cannot transmit forward along the road because of encountering an open diode. Likewise, if there is a vehicle within the activated diode line a signal can only transmit backward along the road as far as that vehicle because it will also cause another diode to appear with reverse voltage so as to render it nonconductive. If there is no vehicle within the diode tail length the speed pulses will be rapidly attenuated beyond the last conducting diode of the diode line because of the conducting shunt diode.

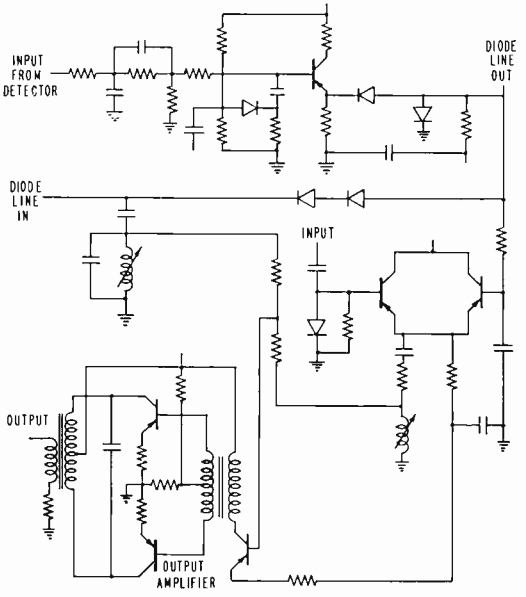

Fig. 8—Speed and distance modulator circuit.

Figure 8 is a schematic diagram of the tail-warning signal generator. The output amplifier is a conventional audio amplifier with a class-B push-pull stage. The clipper is a fairly conventional two-transistor, emitter-coupled circuit and the pulser is a simple diode circuit utilizing the pulse from the detector to momentarily gate on the 2 kc. However, the amplifier, which is a buffer between the detector and the diode line, has an additional function to perform. Usually when a vehicle activates a detector, the previous detector is still activated. Thus, for a period there will be two detectors on at one time. In this case, the

10-millisecond pulse from the activated detector will not propagate back along the road since the previous detector, which is still activated, is also applying the same d-c voltage to the diode line, and, therefore, no potential exists across the diode between the two detectors. In order to overcome this difficulty when the detector first comes on, the voltage applied to the diode line is higher than normal, thus assuring that the first diode back does have voltage across it to render it conductive during the time the 10-millisecond pulse is being generated. This is accomplished simply by a two-resistor attenuator with the series resistor bypassed by a capacitor so that the voltage starts off

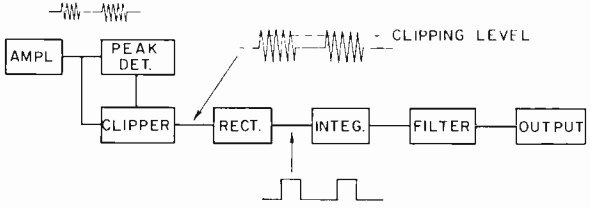

Fig. 9—Block diagram of distance receiver.

higher than that to which it will decay. The diode line is shown here with two diodes in series in order to increase the voltage drop per block, if it were desired to make a longer tail, one diode per block could be used. By changing the amplitude of the 100-cycle sawtooth and the d-c voltage level that is applied to the generators by the detector, it is possible to change the characteristic of the distance information that a following car will pick up. Since these signals are generated at a central location, a controller may easily change these at will. Thus, if the road conditions are poor due to rain or ice, vehicles can be spaced further apart.

The detectors and tail-warning signal generators are the main part of the highway installation. Once these are installed, any "equipped" vehicle may pick up information from the road as to any traffic in front. Since the system tells a vehicle the distance and effectively the speed of any vehicle in front, this is all the following vehicle needs to determine the proper course of action.

In order that a vehicle may automatically maintain proper speed and spacing, it must have three components: a distance receiver, a speed receiver, and a computer to take the information from the receivers and determine the proper course of action. Figure 9 shows a block diagram of the distance receiver. An antenna under the front

bumper is connected to a 4.5-kc amplifier. The top 10 per cent amplitude of the 4.5-kc bursts is clipped off by means of the peak detector and clipper. The pulses from the peaks of the 4.5-kc bursts are rectified so as to form the envelope of the signal. This signal is then integrated, filtered to remove the 100-cycle component, and fed to an output amplifier. Zero output represents zero input and is thus an indication of a vehicle directly in front, while minus 2 volts indicates there is no vehicle within detectable range. Since most failures in the receiver will result in an output indication of zero, that is, a car immediately in front, the receiver is essentially fail-safe. The peak detector and

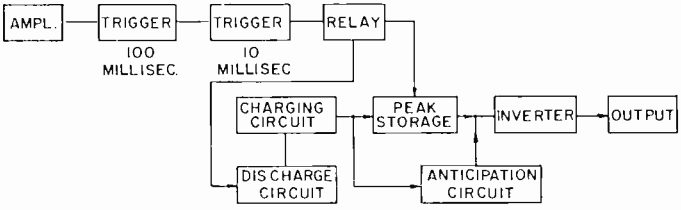

Fig. 10—Block diagram of speed receiver.

clipper are needed in order to prevent errors in distance measuring occurring due to changes in level of the signal from the road or change of gain in the amplifier stage. The bursts of 4.5 kc have finite rise and decay times due to limitations in bandwidths of the antenna and receiving amplifier, so that the apparent width of the burst will be different depending on the level at which the measurement is made. Therefore, the receiver measures the widths 10 per cent down from the peak, independent of the overall amplitude.

Figure 10 is a block diagram of the speed receiver. This receiver determines the speed of a leading vehicle by means of the 2-kc pulses transmitted each time the leading vehicle advances one block length. The easiest way to make this measurement would be to integrate the pulses and thus obtain a measurement of the average speed while the vehicle was traveling several blocks. This system results in a loss of a great deal of time however, since any change in speed of the leading vehicle will not be detected until the vehicle has traveled a few blocks. Instead, the receiver measures the time interval between each pair of pulses and computes the speed for that particular block. Thus, within the time interval with which the lead vehicle travels one block length, any change in speed will be detected. Even more can be done since

the receiver will know when to expect the next 2-kc pulse if the lead vehicle maintains constant speed. However, if the vehicle slows up, the pulse will be late in arriving. As soon as the receiver expects to receive the pulse, but does not, it knows that the lead vehicle has started to slow down and can start to take action even before the next 2-kc pulse is received.

The signal is picked up by an antenna similar to that for the distance receiver and likewise, under the front bumper. The signal from the antenna is amplified and the output 2-kc pulse triggers a 100-millisecond multivibrator. The purpose in this 100-millisecond delay is simply to render the receiver insensitive to further pulses in case the detector relay contact which has initiated the pulse bounces and produces a second pulse.

The leading edge of the pulse from the first multivibrator circuit triggers a second one which produces a 10-millisecond pulse. This 10-millisecond pulse is just long enough to cause a relay to pull in and then immediately drop back to normal condition. When the relay closes it causes the peak voltage, which has been building up in a charging circuit, to be stored in a peak storage circuit. Then, when the relay drops back to normal condition, it momentarily activates a discharge circuit which discharges the charging circuit so as to allow it to start charging again. Thus, the voltage stored in the peak storage circuit is a measure of the time between the last two pulses. Since this voltage is inversely related to speed, it is inverted and fed to an output amplifier. The inverter circuit is somewhat nonlinear and this, coupled with the exponential characteristics of the charging circuit, approximately matches the hyperbolic function between speed and period between pulses, so that the output voltage is a fairly linear function of speed. Thus, the output voltage is the speed of a lead car measured between the last two pulses and, immediately a new pulse is received, the speed is recomputed and a new output is produced. Any time the voltage on the charging circuit rises higher than that in the peak storage, the charging circuit voltage is immediately transferred to the inverter and the output by means of the anticipation circuit. Thus, whenever a pulse fails to arrive at the proper time for the lead vehicle having maintained constant speed, the output immediately starts indicating the lead vehicle's deceleration.

The speed of an equipped following vehicle is obtained from a tachometer attached to the transmission. This signal, along with that from the speed and the distance receivers is all that is needed to determine the proper course of action as to throttle and brakes. Figure 11 is a block diagram of the spacing computer for the automatic

vehicle. Basically the following vehicle maintains the same speed as the leading vehicle except that, when the distance between the two is great, the following vehicle speeds up somewhat and thus closes the gap. For the brake operation, the basic signal is the closing speed which is determined by comparison of the leading and following vehicle speed signals. The closing speed signal is modified by the distance between the two vehicles so that when there is a great distance be-

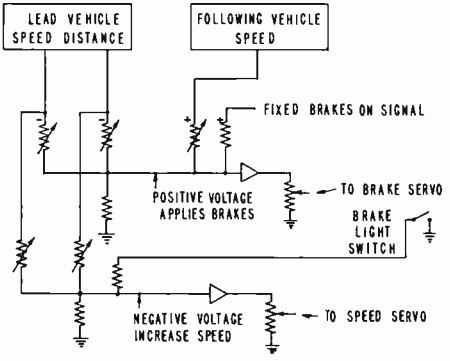

Fig. 11—Block diagram of spacing computer for automatic vehicle.

tween the vehicles a small closing speed signal will not cause the brakes to operate. However, if there is little distance between, a small closing speed signal will cause the brakes to go on. Since all three input signals to the computer are zero under the condition of being stopped close behind a stopped vehicle, a small fixed signal which actuates the brake mechanism is required in order to prevent crawling into the lead vehicle very slowly under these conditions. As soon as the lead vehicle produces a speed signal and moves off a bit, the fixed signal is overcome and the equipped vehicle is allowed to continue. Also, to insure that the throttle never fights the brake action, whenever the brakes go on, the action of the brake light switch modifies the speed computation so as to insure a closed throttle.

As described, the system will control the speed and spacing of an automatic vehicle following another vehicle. However, if there is no leading vehicle there are no speed pulses produced for the following vehicle to receive and use in order to maintain its own speed. Figure 12 shows a block diagram of a process amplifier used to overcome this difficulty. This amplifier is inserted between the distance and speed receivers and the computer. This amplifier looks at the distance

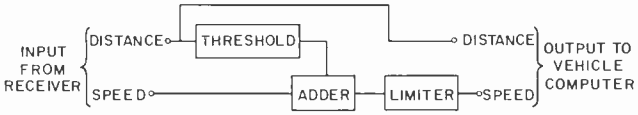

Fig. 12—Block diagram of process amplifier.

signal and whenever the distance indication is near the maximum, it adds a large component to the speed signal. The speed signal then goes through a limiter which sets the maximum speed indication which can be fed to the computer. The limiter is controlled by the operator of the vehicle and serves two functions: (1) it is the means by which the operator of the vehicle sets in the speed which he desires to go in the absence of any traffic necessitating a slower speed, and (2) when following another vehicle, if that vehicle speeds up and goes faster than the operator of the automatic vehicle wishes to go, the limiter limits the speed signal from the leading vehicle, and thus the automatic vehicle will maintain the desired speed, and let the leading vehicle pull away.

Fig. 13—"Equipped" car following a standard car on the RCA Laboratories test track. When the equipped vehicle is under manual control, steering, throttle control, and braking are performed with the "joy stick" control shown to the left of the driver's seat.

Figure 13 is a photograph taken from the backseat of a fully automated car following another car on the test track at RCA Laboratories.

After installation of the equipment in the road, it is to be expected that several years or more may elapse before a large percentage of the vehicles are equipped to make full use of the signals provided by the road. Therefore, it is desirable that as much assistance as possible be given to unequipped cars, or partially equipped cars. For instance, on highways at night, it is not always apparent when coming up behind another vehicle that the closing speed may be very high. By simply equipping the vehicle with a distance receiver and differentiating the signal out of the distance receiver, an approximate closing speed signal would be obtained, which when large, could sound a warning to the driver and alert him that there might be a dangerous situation arising. Likewise, not all vehicles are equipped with stop lights that give a very clear indication of when the brakes are applied. By having a vehicle equipped with a speed receiver and an alarm to sound whenever the speed indication dropped substantially, a driver would be warned whenever the vehicle in front of him slowed up substantially.

These two examples show how equipment which is relatively simple could be used to alert a driver to possible dangerous situations. Other systems can be devised in which the road warns drivers without their needing any special equipment at all. For example, one situation which has caused many accidents is that of fog, or other poor visibility on turnpikes. This condition has many times caused multiple collisions involving ten or more vehicles at once. Figure 14 shows a block diagram of a highway system to aid standard vehicles in fog areas. The roadway is equipped with vehicle detectors as described for the fully automatic system and the vehicle detectors activate a diode tail in the same manner as for the automatic system. However, in addition to operating radio signals, the amplitude of the diode tail is used to activate colored lights imbedded in the center line of the lane. For example, when the amplitude is high, a red light would be activated; when the amplitude is medium, an amber light would be activated; and when the amplitude is low, a green light would be activated. With this system installed in the roadway a driver would be alerted to the fact that he was approaching another vehicle by the approaching green light, even before he could see the vehicle, if the visibility were poor. If the string of green lights were not "moving" in the same direction he was traveling, then he would know there was a stopped car ahead.

If the green lights were moving as the driver approached them,

he would get a measure of his closing speed from the speed with which he was overtaking the last green light. He again gets a measure of the closing speed at the junction between the amber and green, and again from the junction between the red and amber lights. Thus, a driver might elect to drive over the green lights, up to the junction of the amber, and hold his position there. If the lead car slowed down, or speeded up, he would see this from the manner in which the lights operated and would be able to change his own speed accordingly.

Under very foggy conditions the lights would even be useful to aid the driver in steering and keeping himself centered in the lane. Therefore, it would be desirable to provide such a guidance light even

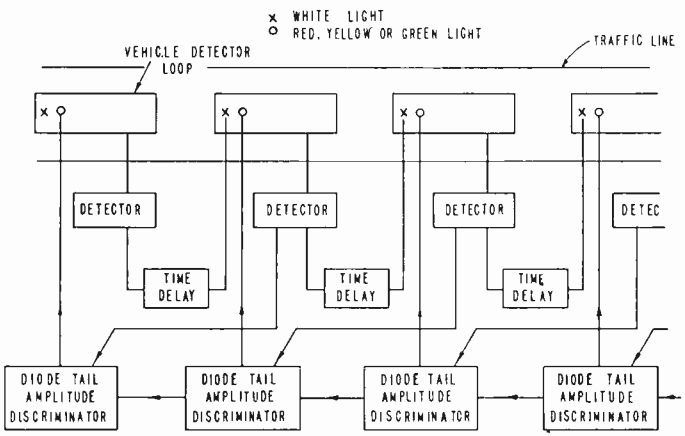

Fig. 14—Highway system to aid unequipped vehicle in a fog.

though there were no vehicle in front. One way of accomplishing this would be to have the detector activate a white light a block or so in front, and thus provide the driver with a white light to steer toward as an indication where the roadway is and that there is no vehicle in front of him.

This method can be refined by including a time delay with which the light comes on in front of the driver so that if he is exceeding some speed limit, he will arrive over the light before it comes on, and thus not see it, and thereby be forced to slow up and maintain whatever speed was set into the time delay by the highway officials.

Conclusions

An experimental system has been described which demonstrates how electronic techniques can be applied to certain types of highways

For the sake of simplicity a one-dimensional analysis is carried out. Newton cooling of the window surfaces (taken into account in References (4) and (5)) is neglected. Close quantitative agreement with practical data can hardly be expected from such a simple model. However, the theory demonstrates how the various material constants and the frequency affect the maximum transmitted power of a window with a geometry chosen so as to optimize the thermal conductance.

Integration of the Stationary Heat-Flow Equation

The stationary (time independent) heat-flow equation states that the divergence of the heat flow $-k(\theta)\ (d\theta/dx)$, is equal to the yield of a heat source, p,

$$-\frac{d}{dx}\left(k(\theta)\frac{d\theta}{dx}\right)=p. \tag{1}$$

Here θ is the temperature above room temperature, $k(\theta)$ is the thermal conductivity, and p is the power dissipated per cubic centimeter. Equation (1) is nonlinear because of the temperature dependence of the thermal conductivity. The equation can be linearized by the transformation of θ to a new temperature ϕ which is measured in a nonlinear scale;

$$k_0\phi=\int_0^{\theta}k(\theta)\,d\theta, \quad k_0=k(0). \tag{2}$$

From Equations (1) and (2),

$$\frac{d^2\phi}{dx^2}=-\frac{p}{k_0}. \tag{3}$$

For most ceramics the temperature dependence of the thermal conductivity can be approximated by

$$k(\theta)=\frac{k_0}{1+\alpha\theta}. \tag{4}$$

From Equations (2) and (4), the true temperature θ is related to ϕ by

$$\theta=\frac{1}{\alpha}\left(e^{\alpha\phi}-1\right). \tag{5}$$

The power dissipated per volume element by dielectric loss heating is proportional to the dielectric loss factor of the material, δ, which depends on temperature. Thus p in Equation (3) will depend on ϕ. To facilitate the computation, it is assumed that δ depends linearly on ϕ,

$$\delta = \delta_0 (1 + \vartheta\phi), \tag{6}$$

where δ_0 is the dielectric loss factor at room temperature. Equations (5) and (6) imply a logarithmic dependence of the loss factor on temperature. In many ceramics the temperature dependence is exponential rather than logarithmic. It can be shown from the results of this paper, however, that waveguide windows break under thermal stress for temperature differences of the order of 100°C between the center of the window and the waveguide wall. In this temperature range, the logarithmic approximation was found to be satisfactory in all practical cases.

With Equation (6) one can rewrite the heat-flow Equation (3);

$$\frac{d^2\phi}{dx^2} + \beta^2\phi = -\frac{p_0}{k_0}, \tag{7a}$$

$$\beta^2 = \frac{p_0\vartheta}{k_0}. \tag{7b}$$

Here p_0 is the dielectric power loss per volume element at room temperature.

Consider now the model of a microwave window depicted in Figure 1. The ceramic slab is placed in an ideal strip line with negligible fringe fields, which supports pure TEM waves. The metallic strips are assumed to be kept at room temperature, $\phi = 0$. Heat-radiation losses and convection cooling of the window surfaces (Newton cooling) are neglected. Thus the problem is reduced to one dimension, and the only heat-transport phenomenon to be considered is that of conduction to the cool waveguide walls. If the load side of the window is terminated by the characteristic impedance of the strip line,

$$Z = \frac{d}{b}\sqrt{\frac{\mu_0}{\epsilon_0}}, \tag{8}$$

and if the window is very thin compared to the wavelength. one can

write for the average power density dissipated in the window at room temperature

$$p\theta = \frac{1}{2}\omega\,\epsilon_0\,\epsilon_r\,\delta_0\,\hat{E}^2\tau = 2\pi\,\epsilon_r\,\delta_0\,\frac{P}{db\lambda}. \tag{9}$$

Here ϵ_r is the relative dielectric constant of the window material, λ the free-space wavelength, $\hat{E}$ the amplitude of the electric field strength, τ the duty cycle, and P the transmitted average power. The solution

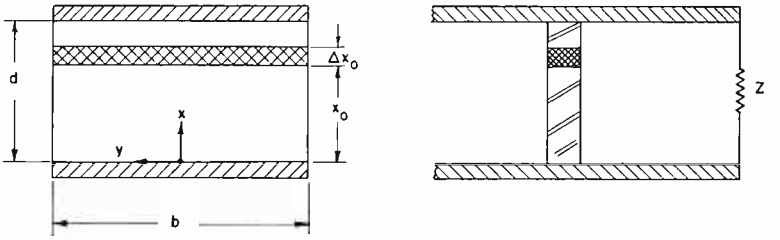

Fig. 1.

of Equation (7) which meets the boundary condition of $\phi = 0$ at $x=0$ and $x=d$ is

$$\phi = \frac{1}{\vartheta}\left(\tan\frac{\beta d}{2}\sin\beta x - 2\sin^2\frac{\beta x}{2}\right). \tag{10}$$

It is seen that for $\beta d = \pi$, a *thermal catastrophe* occurs and the temperature "runs away" due to the increase of the dielectric loss factor with temperature. Note that the dependence of the thermal conductivity on temperature (Equation (4)) does not lead by itself to a temperature runaway. Although Equation (10) was derived for a highly idealized model, it bears out the fact that in practice the temperature dependence of the loss factor is an important parameter and βd must be kept well below the critical value of π. The maximum allowable value of βd where the window breaks due to thermal stresses is calculated in the next section.

Power-Handling Capability

The following thermal stress analysis is based on the assumptions that the expansion coefficient α, Young's modulus E, and the thermal conductivity are independent of temperature. Due to the last assump-

tion, $\theta = \phi$. Suppose the window depicted in Figure 1 is heated to a temperature $\theta(x)$. As a consequence the material will expand. Suppose now the expansion is prevented by application of a fictitious volume force. The pressure in y direction due to this fictitious force is given by Hooke's law;

$$-\sigma_1 = \alpha E \theta (x). \tag{11}$$

This pressure also exists at the window surfaces $y = \pm(b/2)$ which is incompatible with the boundary condition that no external force exists. To comply with the boundary condition, a fictitious surface force (tension) is applied at $y = \pm(b/2)$ so as to cancel the compression (Equation (11)) at the surface. This surface force is transformed into the body in such a way that the resulting tension in the middle of the window, $y = 0$, is with very good approximation independent of x and equal to the average of the value at the surface;[6]

$$\sigma_2 = \frac{\alpha E}{d} \int_0^d \theta \, dx. \tag{12}$$

The sum of the compression (Equation (11)) and tension (Equation (12)) yields the thermal stress at $y = 0$;

$$\sigma = \alpha E \left(\frac{1}{d} \int_0^d \theta dx - \theta \right). \tag{13}$$

A detailed discussion of the derivation and validity of Equation (13) is given by Timoshenko.[6]

In all practical window materials, the allowable compression far exceeds the tensile strength. Maximum tension occurs at $x=0$ and $x=d$, where $\theta = 0$ and $\sigma_1 = 0$. With $\theta = \phi$, given by Equation (10), one obtains, from Equation (12),

$$\sigma_2 = \frac{\alpha E}{\vartheta} \left[\frac{\tan \frac{\beta d}{2}}{\frac{\beta d}{2}} - 1 \right] < \sigma_{TS}. \tag{14}$$

[6] S. Timoshenko, *Theory of Elasticity*, McGraw-Hill Book Company, Inc., New York, 1934.

The maximum allowable value of σ_2 is the tensile strength of the material, σ_{TS}.

With the material constants α, E, ϑ, and σ_{TS} given, the maximum value of βd can be calculated from Equation (14). From Equations (7b) and (9) the following expression for the average power is obtained.

$$P = \frac{(\beta d)^2 k_0 \lambda b}{2\pi \epsilon_r \delta_0 \vartheta d}. \tag{15}$$

From Equation (15), b should be as large as possible, and d as small as possible, which obviously increases the thermal conductance and thus the maximum transmitted power of the window. However, b has an upper limit given by

$$b < \frac{\lambda}{2} \tag{16}$$

to avoid the excitation of nontransverse waveguide modes in the transmission line (it should be recalled that the fringe field of the strip line is neglected). The value of d has a lower limit due to r-f breakdown. One can write for the peak power transmitted

$$\hat{P} = \frac{\hat{E}^2 d^2}{2Z}. \tag{17}$$

With the characteristic impedance from Equation (8) one obtains

$$d > \frac{2\hat{P}\sqrt{\frac{\mu_0}{\epsilon_0}}}{\hat{E}_{\max}^2 b} = \frac{2P\sqrt{\frac{\mu_0}{\epsilon_0}}}{\hat{E}_{\max}^2 b\tau}, \tag{18}$$

where $\hat{E}_{\max}$ is the r-f breakdown-field strength. If the limiting values are taken for both b and d, one finds from Equation (15) the maximum average power transmitted by the window;

$$P < \frac{1}{4} \beta d \hat{E}_{\max} \lambda^{3/2} \left(\frac{k_0 \tau}{\pi \epsilon_r \delta_0 \vartheta \sqrt{\frac{\mu_0}{\epsilon_0}}} \right)^{1/2}. \tag{19a}$$

βd must be calculated from Equation (14) for $\sigma_2 = \sigma_{TS}$. It is interesting to see how the tensile strength would affect the maximum power if the dielectric loss factor were independent of temperature ($\vartheta = 0$). From Equations (7b), (14), and (19a)

$$P < \frac{1}{2} \hat{E}_{\max} \lambda^{3/2} \left(\frac{3k_0 \tau \sigma_{TS}}{\pi \epsilon_r \delta_0 \alpha E \sqrt{\frac{\mu_0}{\epsilon_0}}} \right)^{1/2}, \qquad (19b)$$

and $\vartheta = 0$. For $\vartheta = 0$, it is thus equally important to improve the tensile strength, thermal conductivity, loss factor, dielectric constant, expansion coefficient, or Young's modulus. However, it is seen from Equation (14) that for increasing ϑ, the value of $\sigma_{TS}\vartheta/(\alpha E)$ affects the value of βd less and less. Thus the tensile strength, the expansion coefficient, and Young's modulus decrease in importance as compared to the thermal conductivity and all dielectric constants (ϵ_r, δ_0, ϑ). In the examples shown in Table I, it is assumed that the wavelength $\lambda = 11$ centimeters, the duty cycle $\tau = 10^{-3}$, and the breakdown-field strength $\hat{E}_{\max} = 30$ kilovolts per centimeter. The material constants were supplied by various manufacturers of window materials. The maximum power is calculated from Equations (14) and (19a). It should be noted that Equation (19b) leads to incorrect results even for very small values of ϑ. Take, for example, the case of alumina, where $\vartheta = 7 \times 10^{-4}/°C$. Equation (19b) yields a maximum power of about 70 kilowatts, whereas the correct Equation (19a) gives about 40 kilowatts. It should also be noted that many manufacturers now claim much smaller loss factors for high-density alumina and beryllia than given above; such reductions in loss factor would increase the theoretical power limit appreciably. There is little doubt in the author's mind that the maximum power for a high-density beryllia window will exceed that of a sapphire window appreciably.

Temperature Runaway under Window Bombardment

Thus far, dielectric loss heating has been assumed to be the only heat source. In practical high-power tubes, however, the window is exposed to bombardment[3,7] by various particles, mostly electrons, which provide local heat centers. In the following, the influence of such local

---

[7] J. R. M. Vaughan, "Some High-Power Window Failures," *Trans. I.R.E. PGED*, Vol. 8, p. 302, July, 1961.

Table I

	Alumina	Synthetic Sapphire*	Beryllia
$\frac{\sigma_{TS}}{10^4 \text{ psi}}$	2.5	5.8	2.1
$\frac{E}{10^7 \text{ psi}}$	4	5	5
$\frac{\alpha}{10^{-6}/°C}$	6.7	5.8	5.8
$\frac{k_0}{10^{-2} \text{ cal/cm sec } °C}$	4.8	10	63
ϵ_r	9	9.4	6
$\frac{\delta_0}{10^{-4}}$	9.8	0.33	4.4
$\frac{\vartheta}{10^{-4}/°C}$	7	25	7
$\frac{P}{10^3 \text{ watts}}$	40	360	235

* The electric field is assumed to be vertical to the optical axis.

heat centers on the temperature distribution and on the temperature runaway will be discussed.

Consider the strip-line model depicted in Figure 1. The window is heated uniformly by dielectric loss heating corresponding to the parameters k_0, ϑ, and β defined by Equations (4), (6), and (7b). In addition, it is assumed that in a very thin layer located at $x = x_0$, the parameters have different values k_0', ϑ', and β', which may be due to bombardment and/or material defects. To account for bombardment in this manner, the window must of course be thin. The solutions of Equation (7a) for the three regions indicated in Figure 1 are

$$\vartheta\phi_I = C_1 \sin \beta x - 1 + \cos \beta x, \tag{20a}$$

$$\vartheta'\phi_{II} = C_2 \sin \beta' x + C_3 \cos \beta' x - 1, \tag{20b}$$

$$\vartheta\phi_{III} = C_4 \sin \beta (d - x) - 1 + \cos \beta (d - x). \tag{20c}$$

Equations (20a) and (20c) satisfy the boundary condition of zero temperature, i.e., $\phi = 0$, at $x = 0$ and $x = d$, respectively. The constants C_1 to C_4 are determined by the boundary conditions at the surfaces of the layer, $x = x_0$ and $x = x_0 + \Delta x_0$, which require continuity of temperature and heat flow. If it is assumed that the material constant a defined by Equation (4) is the same in all three regions of Figure 1, the boundary conditions can be expressed by

$$C_1 \sin \beta x_0 - 1 + \cos \beta x_0 = \frac{\vartheta}{\vartheta'} (C_2 \sin \beta' x_0 + C_3 \cos \beta' x_0 - 1), \tag{21a}$$

$$C_1 \cos \beta x_0 - \sin \beta x_0 = \frac{\beta' k_0' \vartheta}{\beta k_0 \vartheta'} (C_2 \cos \beta' x_0 - C_3 \sin \beta' x_0), \tag{21b}$$

$$C_2 \sin \beta'(x_0 + \Delta x_0) + C_3 \cos \beta'(x_0 + \Delta x_0) - 1$$

$$= \frac{\vartheta'}{\vartheta} [C_4 \sin \beta(d - x_0 - \Delta x_0) + \cos \beta(d - x_0 - \Delta x_0) - 1], \tag{22a}$$

$$C_2 \cos \beta'(x_0 + \Delta x_0) - C_3 \sin \beta'(x_0 + \Delta x_0)$$

$$= \frac{\beta k_0 \vartheta'}{\beta' k_0' \vartheta} [- C_4 \cos \beta(d - x_0 - \Delta x_0) + \sin \beta(d - x_0 - \Delta x_0)]. \tag{22b}$$

Now let $\Delta x_0 \to 0$ but $\beta' \Delta x_0 \neq 0$. This is tantamount to the assumption that a finite power per square centimeter cross section, $p_0' \Delta x_0$, is dissipated in an infinitesimally thin layer of finite thermal conductance per square centimeter cross section, $k_0'/\Delta x_0$. The solution of Equations (21) and (22) then yields

$$C_1 = \left[\tan \frac{\beta d}{2} - (1 - \cos 2\sqrt{\rho p_B}) \left(\left(1 - \frac{\vartheta}{\vartheta'} \right) \frac{\cos \beta(d - x_0)}{\sin \beta d} - \cot \beta d \right) \right.$$

$$+ \frac{\sin 2\sqrt{\rho p_B}}{2\sqrt{\rho p_B}} \left(\beta d \rho \frac{\cos \beta(d - x_0) \sin \beta x_0}{\sin \beta d} \right.$$

$$\left. \left. - \frac{4 p_B}{\beta d} \frac{\sin \beta(d - x_0) \left(1 - \frac{\vartheta}{\vartheta'} - \cos \beta x_0 \right)}{\sin \beta d} \right) \right]$$

$$\left[\cos 2\sqrt{\rho p_B} + \frac{\sin 2\sqrt{\rho p_B}}{2\sqrt{\rho p_B}}\left(\beta d\rho \frac{\cos\beta(d-x_0)\cos\beta x_0}{\sin\beta d} - \frac{4p_B}{\beta d}\frac{\sin\beta(d-x_0)\sin\beta x_0}{\sin\beta d}\right)\right]^{-1} \tag{23}$$

where

$$p_B = \lim_{\Delta x_0 \to 0} \frac{p_0' \Delta x_0 \vartheta' d}{4k_0} \tag{24}$$

is the normalized bombarding power, and

$$\rho = \lim_{\Delta x_0 \to 0} \frac{k_0}{d}\frac{\Delta x_0}{k_0'} \tag{25}$$

is the normalized thermal resistance induced by the bombardment. The case $p_B = 0$, $\rho \neq 0$ is also of some interest because a finite localized thermal resistance may exist in the absence of bombardment due to material defects. The constant C_4 is given by the same formula as C_1 (Equation (23)) with x_0 replaced by $d - x_0$. The temperature distributions in regions I and III can then be calculated from Equations (5), (20), and (23). The temperature "distribution" in the infinitely thin region II is of no particular interest for the present investigation and will therefore not be discussed.

Equation (23) shows that a thermal-stress analysis taking the bombardment into account would be rather complicated. A simpler way of estimating the influence of p_B is to investigate the runaway process mentioned earlier. The temperature is predicted to run away at certain "critical" values of transmitted power, measured in terms of $(\beta d)^2$ (Equation (7b)) and bombarding power, p_B, (Equation (24)). To get the maximum permissible values of $(\beta d)^2$ and p_B for a given thermal resistance ρ (Equation (25)), one must consider bombardment at the "most vulnerable spot" on the window: x_0 must be chosen so that the critical value of $(\beta d)^2$ is at its minimum. An inspection of the denominator of Equation (23) shows that for

$$(\beta d)^2 \leq 4\frac{p_B}{\rho}, \tag{26a}$$

the most vulnerable spot occurs at $x_0 = d/2$, and the temperature runs away if

$$\frac{\tan\dfrac{\beta d}{2}}{\dfrac{\beta d}{2}} = \rho \frac{\cot\sqrt{\rho p_B}}{\sqrt{\rho p_B}}. \tag{26b}$$

However, if

$$(\beta d)^2 \geqq 4\frac{p_B}{\rho}, \tag{27a}$$

the most vulnerable spot occurs at $x_0 = 0$ or $x_0 = d$, and the temperature runs away if

$$\frac{\tan \beta d}{\beta d} = -\rho \frac{\tan 2\sqrt{\rho p_B}}{2\sqrt{\rho p_B}}. \tag{27b}$$

This behavior can be well-understood physically because for small thermal resistance, ρ, and large bombardment power, p_B, the most vulnerable spot must obviously be where the temperature due to dielectric loss heating is highest, i.e., in the center of the window. For large thermal resistance and small bombarding power it is also obvious that the thermal resistance will be most effective where the heat flow is at its maximum, i.e., at the waveguide walls. It is surprising, however, that the most vulnerable spot moves *abruptly* from $x_0 = d/2$ to $x_0 = 0$, d, when the limit of Relation (26a) is reached. This odd behavior is believed to lie in the one-dimensional nature of the model under investigation.

Figure 2 shows the value of the critical normalized transmitted power where the temperature runs away versus the normalized bombarding power for various values of the induced thermal resistance, as calculated from Equations (26) and (27). As an example, consider a square window ($d = b$) with a thickness of 0.3 cm, $k_0 = 4.8 \times 10^{-2}$ cal/cm sec°C, $\vartheta' = 7 \times 10^{-4}$ 1/°C, which is bombarded with 100 watts. From Equation (24) one obtains $p_B = 0.3$. Figure 2 shows that the critical power transmitted through the window is reduced by the bombardment to about 74 per cent if $\rho = 0$ and the bombardment occurs at the center. If a thermal resistance of $\rho = 0.5$ is induced by the bombarding particles and the bombardment occurs at the waveguide wall, the critical power transmitted is reduced to about 49 per cent.

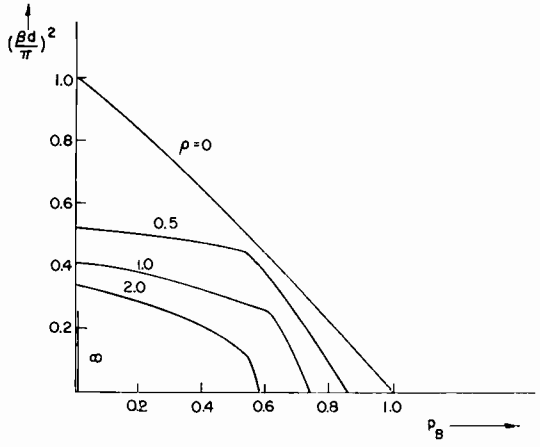

Fig. 2.

Conclusions

The temperature dependence of the dielectric loss angle is an important factor in the process of dielectric loss heating. The theoretical limit of output power of a microwave tube due to thermal destruction of the window lies well above the present state of the power-tube art, at least for the best materials available. Window bombardment may greatly reduce the power-handling capability, particularly if a thermal resistance is induced by the bombarding particles.

NONLINEAR ELECTRODYNAMICS OF SUPERCONDUCTORS WITH A VERY SMALL COHERENCE DISTANCE

By

R. H. Parmenter

RCA Laboratories
Princeton, N. J.

Summary—*A study is made of the electrical properties of a homogeneous superconductor having a Pippard coherence distance much smaller than the London penetration depth, a situation which will occur if the normal-electron mean free path is made sufficiently small as a result of either disorder or boundary scattering. The appropriate nonlinear generalization of the London theory of superconductivity is obtained and compared with the phenonenological Ginzburg–Landau theory. An investigation is made of critical currents and magnetic fields. It is shown that an intermediate state will exist in external magnetic fields much higher than the thermodynamic critical magnetic field.*

Introduction

HERE we wish to study the electrical properties of homogeneous superconductors having a very small coherence distance. Specifically, we assume that the Pippard coherence distance, ξ, is much smaller than the London penetration depth, λ. As is well known, in this limit the electrodynamics of superconductivity becomes local in the sense that the current density **J** at any point in space is a function of the magnetic vector potential **A** at that same point. When this condition is not satisfied, **J** at any point will be a function of the value of a weighted average of **A** over a region surrounding the point,[1] this region being of the size of ξ. It is conventional to consider only the low-field limit where **J** is a linear function of **A**; in the local case this leads to London's equation, and in the nonlocal case to Pippard's equation. In this paper we wish to consider problems involving high fields where **J** is a nonlinear function of **A**. We thus need to obtain a nonlinear generalization of London's equation.

Except for temperatures very close to the superconducting transition temperature, the penetration depth (about 500 Å) is usually at least an order of magnitude smaller than the coherence distance for an "ideal" superconductor. (By "ideal" is meant complete absence of

[1] A. B. Pippard, "An Experimental and Theoretical Study of the Relation between Magnetic Field and Current in a Superconductor," *Proc. Roy. Soc.* (London), Vol. A216, p. 547, February 24, 1953.

impurities or geometrical imperfections.) However, it is well known[1] that homogeneous disorder will simultaneously increase λ and decrease ξ. In the limit of strong disorder (small electronic mean free path l associated with the normal state), ξ will be proportional to l, λ to $l^{-1/2}$. Therefore, there is physical motivation for studying the properties of a homogeneous superconductor having $\xi \ll \lambda$. Furthermore, there are both theoretical[2] and experimental[3] reasons for believing that strong homogeneous disorder does not negate the applicability of the model of a superconductor used in the microscopic theory of superconductivity.[4] On the contrary, the suppression of crystalline anisotropy by the disorder appears to render the Bardeen–Cooper–Schrieffer (BCS) model more applicable to a homogeneous alloy than it is to an ideal superconductor.

If one or more dimensions of a superconducting specimen are much smaller than the penetration depth, then boundary scattering can play the same role as disorder in causing $\xi \ll \lambda$. Here also there is reason to believe that the BCS model of a superconductor is still applicable.[5]

Cooper Pairs with a Finite Drift Velocity

The first stage of the present calculation involves generalizing the BCS theory of superconductivity to the case where there is a finite drift velocity $\mathbf{v}_0$ of the Cooper pairs (pairs of antiparallel-spin electrons, every pair having the same center-of-mass, or drift, velocity). This is done assuming a position-independent $\mathbf{v}_0$. This is justified (despite the fact that the results obtained need to be applied to a situation where $\mathbf{v}_0$ varies with position) by the fact that variations of $\mathbf{v}_0$ with position will usually be negligible over distances of the order of ξ as long as $\xi \ll \lambda$. Bardeen[6] was the first to consider the case of finite $\mathbf{v}_0$, but he was primarily interested in the limit of small $\mathbf{v}_0$ (currents much smaller than the critical current). We shall set the problem up in a manner analogous to that of Bardeen, but we shall not restrict ourselves to small currents.[7]

---

[2] P. W. Anderson, "Theory of Dirty Superconductors," *Jour. Phys. Chem. Solids*, Vol. 11, p. 26, September, 1959.

[3] D. M. Ginsberg and J. D. Leslie, "Far-Infrared Absorption in a Lead-Thallium Superconducting Alloy," *IBM Journal of Research*, Vol. 6, p. 55, January, 1962.

[4] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Theory of Superconductivity," *Phys. Rev.*, Vol. 108, p. 1175, December 1, 1959. This theory will be referred to as the BCS theory.

[5] D. M. Ginsberg and M. Tinkham, "Far-Infrared Transmission through Superconducting Films," *Phys. Rev.*, Vol. 118, p. 990, May 15, 1960.

[6] J. Bardeen, "Two-Fluid Model of Superconductivity," *Phys. Rev. Letters*, Vol. 1, p. 399, December 1, 1958.

[7] After completing this calculation, the writer discovered that many of the results presented in this section were obtained previously, but apparently not published, by K. T. Rogers, *Superconductivity in Small Systems*, thesis, University of Illinois, 1960.

The problem is specified mathematically by saying that we wish to minimize the free energy of the system (in the laboratory coordinate system) subject to the constraint that there be a net current density

$$\mathbf{J}_0 = n_0 e \mathbf{v}_0 \tag{1}$$

due to the superconducting electrons, i.e., Cooper pairs. Here n_0 is the total conduction-electron density, and is therefore temperature independent. In addition to the superconducting electrons, at a finite temperature there will be quasi-particle excitations, i.e., normal electrons and holes. Under the steady-state conditions considered in this paper, there will be equal numbers of normal carriers of each sign of charge. This is consistent with the fact that the total conduction-electron density is associated with the superconducting electrons.

$\mathbf{J}_0$ is not necessarily the total current density in the laboratory coordinate system. Once we have constrained the Cooper pairs to move with a net drift velocity $\mathbf{v}_0$, it will generally happen that the lowest free energy of the system is obtained by allowing the quasi-particles to have a net current density $\mathbf{J}_Q$ associated with them,

$$\mathbf{J}_Q = 2e \sum_k (\hbar \mathbf{k}/m) f_k. \tag{2}$$

Here f_k is the distribution function for a quasi-particle of wave vector $\mathbf{k}$. The factor of two accounts for the spin degeneracy. The total current can be written

$$\mathbf{J} = \mathbf{J}_0 + \mathbf{J}_Q. \tag{3}$$

$\mathbf{J}_Q$ will always tend to oppose $\mathbf{J}_0$, so that $\mathbf{J}$ is smaller than $\mathbf{J}_0$. Whenever the superconducting electrons have a distribution in k-space not centered on the origin, as illustrated in Figure 1, the normal holes will tend to be on the forward side of the distribution and normal electrons on the back side. Thus $\mathbf{J}_Q$ opposes $\mathbf{J}_0$. Since the normal carriers are in the distribution which minimizes the free energy (for a given $\mathbf{v}_0$), there is *no dissipation* or ohmic heating associated with $\mathbf{J}_Q$. It should be emphasized that the breakup of $\mathbf{J}$ into $\mathbf{J}_0$ and $\mathbf{J}_Q$ does *not* correspond to the superfluid and normal components, respectively, in the two-fluid theory of superconductivity.[8] $\mathbf{J}$ as a whole, being dissipationless, rep-

[8] C. J. Gorter, *Progress in Low Temperature Physics*, Vol. I, Chap. I, North-Holland Publishing Co., Amsterdam, 1955.

resents the supercurrent. The calculation of a normal current will be deferred until later.

Carrying out the minimization of the free energy, just as is done in BCS, we find the integral equation for the energy gap. It is more convenient to write this equation in terms of wave vectors and energies measured with respect to a coordinate system moving with the velocity $\mathbf{v}_0$, since in such a coordinate system the Cooper pairs have no net drift velocity. Thus $\mathbf{k}$, $\hbar\mathbf{k}$, and $\epsilon_k = (\hbar^2/2m)(k^2 - k^2{}_F)$ are, respectively, the wave vector, momentum, and Bloch energy (relative to the Fermi level) in the moving coordinate system. The corresponding

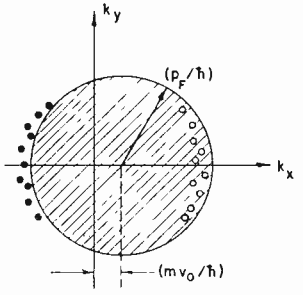

Fig. 1—Schematized distribution in k-space of superconducting electrons (shaded region), normal electrons (filled circles), and normal holes (open circles).

quantities in the laboratory coordinate system will be $\mathbf{k} + (m\mathbf{v}_0/\hbar)$, $\hbar\mathbf{k} + m\mathbf{v}_0$, and

$$\frac{1}{2m}\left[(\hbar\mathbf{k} + m\mathbf{v}_0)^2 - \hbar^2\mathbf{k}_F{}^2\right] \cong \epsilon_k + \hbar\mathbf{k}\cdot\mathbf{v}_0 \cong \epsilon_k + p_F v_0 \cos\theta_k \tag{4}$$

where θ_k is the angle between $\mathbf{v}_0$ and $\mathbf{k}$, and p_F is the Fermi momentum. Here we have used the fact that $mv_0 \ll p_F$. Our integral equation is

$$\frac{1}{2}\sum_{k'} V_{kk'}\left(\frac{\epsilon_{0k'}}{\epsilon_{0k}}\right)\frac{(1 - f_{k'} - f_{-k'})}{\sqrt{\epsilon_{k'}{}^2 + \epsilon_{0k'}{}^2}} = 1, \tag{5}$$

while the distribution function for normal carriers

$$f_k = [\exp\{\beta(E_k + p_F v_0 \cos\theta_k)\} + 1]^{-1}, \tag{6}$$

$$E_k = \sqrt{\epsilon_k{}^2 + \epsilon_{0k}{}^2}, \tag{7}$$

$$\beta = 1/kT, \quad (k = \text{Boltzmann's constant}) \tag{8}$$

is exactly the same function of the normal-carrier energy *in the laboratory coordinate system* as is found in BCS. Now we make the assumption that the Cooper-pair electron–electron interaction potential $V_{kk'}$ is independent of the center-of-mass motion of the pairs;[9] i.e., $V_{kk'}$ in the moving coordinate system is the same as that used by BCS. Thus we shall assume

$$\begin{aligned} V_{kk'} &= V, \text{ if } |\epsilon_k|, |\epsilon_{k'}| < \hbar\omega \\ &= 0, \text{ otherwise,} \end{aligned} \tag{9}$$

$\hbar\omega$ being a mean phonon energy. Thus

$$\frac{1}{2} V \sum_{k'}{}' \left(\frac{\epsilon_{0k'}}{\epsilon_{0k}} \right) \frac{(1 - f_{k'} - f_{-k'})}{\sqrt{\epsilon_{k'}^2 + \epsilon_{0k'}^2}} = 1, \tag{10}$$

the prime on the summation designating the restriction $|\epsilon_{k'}| < \hbar\omega$. The right-hand side of Equation (10) is independent of $\mathbf{k}$, while the left-hand side depends on $\mathbf{k}$ only through ϵ_{0k} (as long as $|\epsilon_k| < \hbar\omega$). This means that actually ϵ_{0k} is independent of $\mathbf{k}$ as long as $|\epsilon_k| < \hbar\omega$, so that we may drop the subscript from ϵ_{0k}. This is important, since it indicates that no anisotropy of ϵ_0 in k-space is introduced by the finite drift velocity v_0, despite the fact that the distribution of electrons in k-space is anisotropic.

In summing over $\mathbf{k}'$ in Equation (10), the factor $(1 - f_{k'} - f_{-k'})$ will now give the same result as would $(1 - 2f_{k'})$. Making this substitution, and replacing the summation by the equivalent integration,

$$N(0) V \int_0^{\hbar\omega} \frac{d\epsilon}{\sqrt{\epsilon^2 + \epsilon_0^2}} I(\beta\sqrt{\epsilon^2 + \epsilon_0^2}, \beta p_F v_0) = 1, \tag{11}$$

where

$$I(a,b) \equiv \frac{1}{2} \int_{-1}^{1} d\mu \tanh \frac{1}{2}(a + b\mu) = \frac{1}{b} \ln \left[\frac{\cosh\left(\frac{a+b}{2}\right)}{\cosh\left(\frac{a-b}{2}\right)} \right]. \tag{12}$$

[9] This is not strictly correct, but it is an excellent approximation as long as v_0 is much smaller than the velocity of sound, a condition well satisfied under all ordinary circumstances. See R. H. Parmenter, "High-Current Superconductivity," *Phys. Rev.*, Vol. 116, p. 1390, December 15, 1959.

Here $N(0)$ is the density of one-electron states of a given spin in the normal metal at the Fermi level (just as in BCS). The integration in Equation (12) is over orientations of $\mathbf{k}'$, the hyperbolic tangent being $(1-2f_{k'})$. In the limit $v_0=0$, Equation (11) correctly reduces to the BCS equation for ϵ_0.

$\mathbf{J}_Q$ was defined to be the current density, due to quasi-particles, as measured in the laboratory coordinate system. Since there is no net electric charge associated with the quasi-particles (just as many normal electrons as normal holes), the current density due to these carriers will be invariant to Galilean transformations. Thus we can calculate $\mathbf{J}_Q$ in the moving coordinate system. Substituting Equation (6) into Equation (2), and replacing the summation by the equivalent integration,

$$J_Q = 2N(0)\,e(p_F/m)\int_0^\infty d\epsilon\, Q(\beta\sqrt{\epsilon^2+\epsilon_0^2},\ \beta p_F v_0), \tag{13}$$

where

$$Q(a,b) \equiv \frac{1}{2}\int_{-1}^{1} \mathrm{d}\mu\,\mu\left[1-\tanh\frac{1}{2}(a+b\mu)\right] \tag{14}$$

The upper limit of the energy integration in Equation (13) should be $\hbar\omega$ rather than infinity. For the weak-coupling case where $\hbar\omega \gg \epsilon_0$ (to which we restrict ourselves in this paper), the integrand of Equation (13) is effectively zero for $\epsilon \gtrsim \hbar\omega$, so that setting the limit of integration equal to infinity is an excellent approximation. Unlike $I(a,b)$, the integral defining $Q(a,b)$ cannot be evaluated analytically (because of the factor μ in the integrand). In analogy with Equation (1), we define a mean velocity of quasi-particles, $\mathbf{v}_Q$ by the equation

$$\mathbf{J}_Q = n_0 e\mathbf{v}_Q \tag{15}$$

where, as before, n_0 is the total conduction-electron density. By defining $\mathbf{v}_Q$ in this way, we can now write

$$\mathbf{J} = n_0 e\mathbf{v}, \tag{16}$$

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_Q, \tag{17}$$

$\mathbf{J}$ being the total current density, and $\mathbf{v}$ thus necessarily being the drift velocity averaged over all charge carriers. Making use of the facts that

$$N(0) = \frac{3}{4}\frac{n_0}{E_F},$$

$$E_F = \frac{p_F^2}{2m},$$

E_F being the Fermi energy, we can rewrite Equation (13) as

$$p_F v_Q = 3 \int_0^\infty d\epsilon \, Q(\beta\sqrt{\epsilon^2 + \epsilon_0^2}, \; \beta p_F v_0). \tag{18}$$

By interchanging the integrations with respect to μ and ϵ, we can obtain the following identity,

$$-p_F v_0 = 3 \int_0^\infty d\epsilon \, Q(\beta\epsilon, \beta p_F v_0). \tag{19}$$

The physical significance of this is the following: In the limit as $\epsilon_0 \to 0$, $v_Q \to -v_0$ or $J \to 0$. In other words, as $\epsilon_0 \to 0$ our superconducting state goes over into the normal state of zero current in the laboratory coordinate system.

We wish to calculate, in the laboratory coordinate system, the energy density W_0 associated with the superconducting state. This will be done by first calculating the energy density W_0' in the coordinate system moving with velocity $\mathbf{v}_0$, and then performing a Galilean transformation. Consider the energy density W_0'' in the coordinate system moving with velocity $\mathbf{v}$, the coordinate system in which there is no net electron current. We can write W_0 and W_0' in terms of W_0'', i.e.,

$$W_0 = W_0'' + \frac{1}{2} n_0 m (0 - \mathbf{v})^2$$

$$W_0' = W_0'' + \frac{1}{2} n_0 m (\mathbf{v}_0 - \mathbf{v})^2.$$

Thus

$$W_0 = W_0' + \frac{1}{2} n_0 m (2\mathbf{v}_0 \cdot \mathbf{v} - v_0^2). \tag{20}$$

Now, just as in BCS,

$$W_0' = 2\sum_k f_k\epsilon_k(1-2h_k) + 2\sum_{k>k_F}\epsilon_k h_k + 2\sum_{k<k_F}|\epsilon_k|(1-h_k)$$

$$-\sum_{kk'} V_{kk'}\sqrt{h_k(1-h_k)\ h_{k'}(1-h_{k'})}\,(1-2f_k)\,(1-2f_{k'}), \tag{21}$$

where

$$h_k \equiv \frac{1}{2}\left[1-\frac{\epsilon_k}{\sqrt{\epsilon_k^2+\epsilon_0^2}}\right]. \tag{22}$$

Replacing the sums by the corresponding integrals,

$$W_0' = 2N(0)\int_0^{\hbar\omega}\left[1-\frac{\epsilon}{\sqrt{\epsilon^2+\epsilon_0^2}}I(\beta\sqrt{\epsilon^2+\epsilon_0^2},\beta p_F v_0)\right]\epsilon d\epsilon$$

$$-V\left[N(0)\epsilon_0\int_0^{\hbar\omega} I(\beta\sqrt{\epsilon^2+\epsilon_0^2},\beta p_F v_0)\frac{d\epsilon}{\sqrt{\epsilon^2+\epsilon_0^2}}\right]^2 \tag{23}$$

As an example, ϵ_0, v, and W_0 will be calculated for the one case where all the necessary integrations can be performed analytically, this being the case of the absolute zero of temperature ($\beta=\infty$). We have

$$\lim_{\beta\to\infty} I(\beta\sqrt{\epsilon^2+\epsilon_0^2},\beta p_F v_0)$$

$$=\lim_{\beta\to\infty}\left[2Q(\beta\sqrt{\epsilon^2+\epsilon_0^2},\beta p_F v_0)+1\right]^{1/2}$$

$$=\frac{\sqrt{\epsilon^2+\epsilon_0^2}}{p_F v_0}, \quad \text{if} \quad \frac{\sqrt{\epsilon^2+\epsilon_0^2}}{p_F v_0}<1,$$

$$=1, \quad \text{if} \quad \frac{\sqrt{\epsilon^2+\epsilon_0^2}}{p_F v_0}>1. \tag{24}$$

We define

$$q=\left[1-\left(\frac{\epsilon_0}{p_F v_0}\right)^2\right]^{1/2} \tag{25}$$

$$\epsilon_{00} = 2\hbar\omega \exp\left\{\frac{-1}{N(0)\ V}\right\} \tag{26}$$

$$v_{00} = \frac{\epsilon_{00}}{p_F}. \tag{27}$$

ϵ_{00} is the BCS value of ϵ_0 at $T = 0$, $v_0 = 0$ (in the weak-coupling limit where $\hbar\omega \gg \epsilon_{00}$). Equation (11), the integral equation for the energy gap ϵ_0, becomes

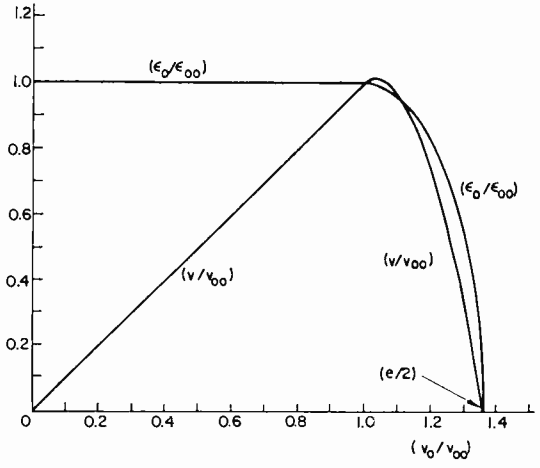

Fig. 2—Plot of ϵ_0/ϵ_{00} and v/v_{00} versus v_0/v_{00} at $T = 0$. Maximum v/v_{00} of 1.011 occurs at $v_0/v_{00} = 1.029$, $\epsilon_0/\epsilon_{00} = 0.994$.

$$\epsilon_0 = \epsilon_{00}, \tag{28}$$

for $v_0 < v_{00}$, while for $v_0 > v_{00}$ it becomes

$$\frac{e^q}{1+q} = \frac{v_0}{v_{00}} \tag{29}$$

Equations (25), (28), and (29) serve to determine ϵ_0 as a function of v_0. A plot of ϵ_0 versus v_0 is shown in Figure 2. From Equations (25) and (28), it is seen that ϵ_0 vanishes at

$$v_0 = \left(\frac{1}{2}\ e\right)\ v_{00} = 1.359\ v_{00}. \tag{30}$$

Substituting Equation (24) into (18), $v_Q = 0$ if $v_0 < v_{00}$, while if $v_0 > v_{00}$,

$$v_Q = -v_0 q^3. \tag{31}$$

Thus, the net drift velocity v is given by

$$\begin{aligned} v &= v_0 \quad , \qquad v_0 < v_{00}, \\ &= v_0 (1 - q^3), \qquad v_0 > v_{00}. \end{aligned} \tag{32}$$

Differentiating the expression

$$v = \frac{e^q (1 - q^3)}{(1 + q)} v_{00} \tag{33}$$

with respect to q,

$$\frac{dv}{dq} = \frac{qe^q}{(1 + q)^2} [2 - (1 + q)^3] v_{00} \tag{34}$$

Thus the maximum value of v occurs at

$$q = 2^{1/3} - 1,$$

where

$$\begin{aligned} \left(\frac{v_0}{v_{00}} \right) &= \frac{e^q}{(1 + q)} = 1.0293 \\ \left(\frac{v}{v_{00}} \right) &= 3qe^q = 1.0112, \\ \left(\frac{\epsilon_0}{\epsilon_{00}} \right) &= e^q \sqrt{\frac{1 - q}{1 + q}} = 0.9939. \end{aligned} \tag{35}$$

A plot of v versus v_0 is shown in Figure 2.

It is seen that v_Q tends to oppose v_0 such that when ϵ_0 goes to zero (i.e., $q = 1$), $v = v_0 + v_Q$ also goes to zero, and the state of the system becomes identical to the normal state of zero current flow. For v_0 less than v_{00}, $v_Q = 0$ and $\epsilon_0 = \epsilon_{00}$. This is a consequence of the fact that the system is at the absolute zero of temperature and has a positive thermal

energy gap[10] $2(\epsilon_0 - p_F v_0)$. For v_0 greater than v_{00} however, the thermal gap has gone to zero, so that normal carriers are present even at $T = 0$, causing v to be less than v_0, ϵ_0 less than ϵ_{00}. Despite the fact that the thermal gap has vanished, thermodynamic metastability still obtains, in the sense that an arbitrary modification of the normal-carrier distribution function f_k will always *increase* the free energy of the system. We will return to this question of stability when we consider critical currents and magnetic fields in later sections.

Substituting Equation (24) into (23), the integrals for W_0' at $T = 0$ can be evaluated;

$$\begin{aligned} W_0' &= -\frac{1}{2} N(0) \epsilon_{00}^2 \qquad , \quad v_0 < v_{00} \\ &= -\frac{1}{2} N(0) \epsilon_0^2 + \frac{1}{3} N(0) p_F^2 v_0^2 q^3, \quad v_0 > v_{00} \end{aligned} \tag{36}$$

Making use of Equation (20) and the fact that

$$\frac{1}{2} n_0 m = \frac{1}{3} N(0) p_F^2, \tag{37}$$

we can now write the energy density in the laboratory coordinate system as

$$\begin{aligned} W_0 &= -\frac{1}{2} N(0) \epsilon_{00}^2 \left[1 - \frac{2}{3} \left(\frac{v_0}{v_{00}} \right)^2 \right] \quad , \quad v_0 < v_{00} \\ &= -\frac{1}{6} N(0) \epsilon_{00}^2 \frac{e^{2q}}{(1+q)^2} (1 - 3q^2 + 2q^3), \quad v_0 > v_{00}. \end{aligned} \tag{38}$$

W_0 properly vanishes in the limit $\epsilon_0 = 0$ $(q = 1)$. It is easy to show that, *for all values of* $\mathbf{v}_0$,

$$\frac{dW_0}{dv_0} = n_0 m v. \tag{39}$$

The significance of this relationship will be made apparent in the next section.

---

[10] In an analogy with semiconductor physics, we refer to $2\epsilon_0$ as the *optical* gap and $2(\epsilon_0 - p_F v_0)$ as the *thermal* gap, these being the minimum energies (in the laboratory coordinate system) required to produce a normal hole–electron pair with or without, respectively, the restraint of momentum conservation.

ELECTRODYNAMICS

Consider a group of Cooper pairs (superconducting electrons) moving with drift velocity $\mathbf{v}_0$.

$$\frac{d\mathbf{v}_0}{dt} = \frac{\partial \mathbf{v}_0}{\partial t} + \mathbf{v}_0 \cdot \nabla \mathbf{v}_0 = -\frac{e}{m}\left(\mathbf{E} + \frac{1}{c}\mathbf{v}_0 \times \mathbf{H}\right). \tag{40}$$

Here $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic fields, respectively. $d\mathbf{v}_0/dt$ is the rate of change of drift velocity of a given Cooper pair, while $\partial \mathbf{v}_0/\partial t$ is the rate of change of drift velocity of the Cooper pairs at a given point in space. Here we are making use of the fact that the spatial variation of E and H is negligible over distances of the order of the size of a pair (the coherence distance), also that all pairs at a given point in space are moving with the same velocity. Since

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

$$\mathbf{v}_0 \times \nabla \times \mathbf{v}_0 = 0,$$

we can rewrite Equation (40) as

$$\frac{\partial \mathbf{v}_0}{\partial t} + \nabla\left(\frac{v_0^2}{2}\right) + \frac{e}{m}\mathbf{E} = \mathbf{v}_0 \times \left(\nabla \times \mathbf{v}_0 - \frac{e}{mc}\mathbf{H}\right)$$

$$= \mathbf{v}_0 \times \nabla \times \left(\mathbf{v}_0 - \frac{e}{mc}\mathbf{A}\right) \tag{41}$$

Making use of Maxwell's equation,

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{c}\nabla \times \frac{\partial \mathbf{A}}{\partial t},$$

we have

$$\frac{\partial}{\partial t}\nabla \times \left(\mathbf{v}_0 - \frac{e}{mc}\mathbf{A}\right) = \nabla \times \mathbf{v}_0 \times \left[\nabla \times \left(\mathbf{v}_0 - \frac{e}{mc}\mathbf{A}\right)\right] \tag{42}$$

Thus if $\nabla \times [\mathbf{v}_0 - (e/mc)\,\mathbf{A}]$ vanishes at any time (e.g., $t = 0$), then it necessarily vanishes at all times. We assume this to be the case. (We shall later attempt to justify the assumption.) Thus

$$\nabla \times \mathbf{J}_0 = -\frac{c}{4\pi\lambda^2}\mathbf{H}, \tag{43}$$

where $\mathbf{J}_0 = -n_0 e \mathbf{v}_0$, and

$$\lambda = \sqrt{\frac{mc^2}{4\pi n_0 e^2}}. \tag{44}$$

Note the minus sign in the definition of J_0, which we have not previously included (e.g., Equation (1)). This inconsistency will lead to no difficulties. It is necessary to include the minus sign here to maintain consistency with Maxwell's equations. The vector potential **A** is undefined to an arbitrary term $\nabla\phi$, ϕ being the gauge, since **H** is unchanged by a change in ϕ. We choose **A** such that

$$\mathbf{J}_0 = -\frac{c}{4\pi\lambda^2}\mathbf{A}, \tag{45}$$

or

$$\mathbf{v}_0 = \frac{\mathrm{e}}{mc}\mathbf{A}. \tag{46}$$

Assuming the $\nabla (v_0^2/2)$ term in Equation (41) is negligible with respect to $\partial \mathbf{v}_0/\partial t$,

$$\frac{\partial \mathbf{J}_0}{\partial t} = \frac{c^2}{4\pi\lambda^2}\mathbf{E}. \tag{47}$$

Under low-frequency conditions

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \rightarrow \frac{4\pi}{c}\mathbf{J}. \tag{48}$$

Since $\nabla \cdot \mathbf{J}_0 = 0$,

$$-\nabla \times \nabla \times \mathbf{J}_0 = \nabla^2 \mathbf{J}_0 - \nabla (\nabla \cdot \mathbf{J}_0) = \nabla^2 \mathbf{J}_0. \tag{49}$$

Combining Equations (43), (48), and (49),

$$\lambda^2 \nabla^2 \mathbf{J}_0 = \mathbf{J}(\mathbf{J}_0). \tag{50}$$

The right-hand side of this equation is written in this fashion in order to emphasize that $\mathbf{J}$ is a function of $\mathbf{J}_0$. We can equally well write

$$\lambda^2 \nabla^2 \mathbf{v}_0 = \mathbf{v}(\mathbf{v}_0). \tag{51}$$

The foregoing is a slight modification of the usual derivation of the London theory.[11] Applied to free electrons in a metal, the derivation is invalid because of the distribution of velocities.[12] Applied to Cooper pairs, all of which have the same center-of-mass velocity, the derivation is valid. It is, however, incomplete, in the sense that we had to assume that curl $\{\mathbf{v}_0 - (e/mc)\mathbf{A}\}$ vanished at time $t = 0$. Now

$$2m\left(\mathbf{v}_0 - \frac{e}{mc}\mathbf{A}\right) = \mathbf{P} \tag{52}$$

is the momentum of a Cooper pair. Feynman[13] has given arguments indicating that the momentum flow of Bose particles is indeed irrotational whenever $\mathbf{P}$ is a slowly varying function of position on the atomic scale (such as is the case here). But in order to insure that a Cooper pair may be treated as a Bose particle, it is necessary that the mean distance between neighboring pairs be large compared with the mean distance ξ separating the two Fermi particles (superconducting electrons) which constitute the pair. At first sight, it appears difficult, if not impossible, to satisfy this condition, because of the high density of conduction electrons in a metal.

The resolution of this difficulty lies in the following observation. The positive two-particle spatial correlation between antiparallel-spin electrons is not due equally to all Cooper pairs, but rather due exclusively to pairs composed of electrons whose energies lie close to the Fermi energy. As far as electrons lying well below the Fermi energy are concerned, there is no difference between the superconducting and the normal state. It is convenient to take the normal state (at $T = 0$) as our reference, or vacuum, state. We may then describe the superconducting state in terms of electron Cooper pairs lying above the Fermi energy and hole Cooper pairs lying below. Both types will lie near the Fermi level. It is the Cooper pairs defined in this manner which are entirely responsible for the antiparallel-spin two-particle correlation in a superconductor. At $T = 0$, the density of such pairs is

[11] See e.g., J. Bardeen, *Handbuch der Physik*, Vol. XV, p. 274, Springer-Verlag, Berlin, 1956.

[12] J. Lindhard, "Resistance-Independent Absorption of Light Waves by Metals, and the Properties of Ideal Conductors," *Phil. Mag.*, Vol. 44, p. 916, August, 1953.

[13] See reference 8, Chap. II.

$$\sum_{k>k_F} h_k + \sum_{k<k_F} (1-h_k) = N(0) \int_0^{\hbar\omega} \left[1 - \frac{\epsilon}{\sqrt{\epsilon^2+\epsilon_0^2}} \right] d\epsilon \cong N(0)\,\epsilon_0. \tag{53}$$

R is defined as the radius of the sphere whose volume equals the amount of space for each pair of this type, i.e.,

$$\frac{4}{3}\pi R^3 = [N(0)\,\epsilon_0]^{-1}.$$

We now can write

$$\frac{R}{\xi} = \left(\frac{a_0}{\lambda_F} \right)^{1/3} \left(\frac{2\pi\alpha\lambda}{\xi} \right)^{2/3} \tag{54}$$

Here λ is the penetration depth, $\xi = (\hbar v_F/\pi\epsilon_0)$ is the coherence distance of the BCS theory, a_0 is the Bohr radius (for electrons of the appropriate effective mass), $\lambda_F = 2\pi\hbar/p_F$ is the Fermi wavelength of electrons at the Fermi level, and $\alpha = 1/137.04$ is the fine-structure constant. Since a_0/λ_F is typically about 0.1 and $2\pi\alpha = 0.046$, it follows that $R \gg \xi$ when $\lambda \gg \xi$. This conclusion will also hold at finite temperatures. Thus we are justified in treating the Cooper pairs as Bose particles having an irrotational momentum flow.

Equation (50) represents the nonlinear generalization of the usual London equation, which is linear in $\mathbf{J}$. Knowing $\mathbf{J}$ as a function of $\mathbf{J}_0$ (remember that this function changes with temperature), Equation (50) can be integrated to find $\mathbf{J}_0$ and $\mathbf{J}$ (or $\mathbf{v}_0$ and $\mathbf{v}$) as functions of position. Knowing the energy gap $2\epsilon_0$ as a function of $\mathbf{v}_0$, we now have ϵ_0 as a function of position. In the limit of small current densities, we can recover the London equation, since

$$\lim_{\mathbf{J}_0 \to 0} \mathbf{J}(\mathbf{J}_0) = \gamma(T)\,\mathbf{J}_0. \tag{55}$$

$\gamma(T)$ is a temperature-dependent coefficient which goes to unity at $T = 0$ and zero at the transition temperature. Substituting Equation (55) into (50) results in the ordinary London equation with an effective penetration depth given by

$$\lambda_{\text{eff}} = \frac{\lambda}{\sqrt{\gamma(T)}}. \tag{56}$$

This is the conventional temperature-dependent London penetration depth. In the two-fluid theory of superconductivity,[8] $n_0\gamma(T)$ is the temperature-dependent density of superconducting electrons.

Von Laue[14] has shown that the form of the London equation implies the existence of the Meissner effect, in the sense that the components of **J** and **H** cannot have a *maximum* at an *interior* point of a superconductor. Von Laue's proof can be easily modified to suit Equation (50). Consider a small region in the neighborhood of a point P where $\mathbf{J}_0 = \mathbf{J}_{0P}$ and $\mathbf{J} = \mathbf{J}_P$. For the values of $\mathbf{J}_0$ and $\mathbf{J}$ at points near P, we can write

$$(\mathbf{J} - \mathbf{J}_P)_{\parallel} = \left(\frac{dJ}{dJ_0} \right)_P (\mathbf{J}_0 - \mathbf{J}_{0P})_{\parallel} \tag{57}$$

$$(\mathbf{J} - \mathbf{J}_P)_{\perp} = \left(\frac{J_P}{J_{0P}} \right) (\mathbf{J}_0 - \mathbf{J}_{0P})_{\perp}. \tag{58}$$

Here we have used the fact that $\mathbf{J}_0$ and $\mathbf{J}$ at any point are parallel to each other. The subscript $\parallel$ denotes the vector component parallel to $\mathbf{J}_{0P}$ or $\mathbf{J}_P$; the subscript $\perp$, the perpendicular component. These two equations mean that in the neighborhood of P, we can linearize Equation (50), i.e.,

$$\lambda^2 K_j^{-1} \nabla^2 J_j = J_j, \tag{59}$$

where

$$K_{\parallel} = \left(\frac{dJ}{dJ_0} \right)_P , \qquad K_{\perp} = \left(\frac{J_P}{J_{0P}} \right) \tag{60}$$

J_j is the j'th component of **J** ($j = \parallel$ or $\perp$). Let $\mathcal{J}_j(r)$ be the average of J_j, the averaging to be done over the surface of a sphere of radius r centered on point P. Equation (59) becomes

$$\lambda^2 K_j^{-1} \frac{d^2}{dr^2} (r\mathcal{J}_j) = r\mathcal{J}_j. \tag{61}$$

[14] M. von Laue, *Theory of Superconductivity*, Chap. 7, Academic Press, New York, 1952.

The solution noninfinite at $r = 0$ can be expanded as

$$\mathcal{J}_j(r) = A_j \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} \left(K_j \frac{r^2}{\lambda^2} \right)^n. \tag{62}$$

Since by definition $J_\perp$ vanishes at P, necessarily $\mathcal{J}_\perp = 0$ at $r = 0$. But this happens only by taking $A_\perp = 0$. Thus $\mathcal{J}_\perp$ vanishes in the neighborhood of P. This means, not that $J_\perp$ vanishes in the neighborhood, but rather that the spherical average of $J_\perp$ vanishes. Thus $J_\perp$ has no extremum at P. Necessarily $A_\parallel = J_P$. Making the convention that J_P is a positive number, we have, for $r > 0$ but suitably small

$$\begin{aligned} \mathcal{J}_\parallel(r) &> J_P, \quad \text{if } \left(\frac{dJ}{dJ_0} \right)_P > 0, \\ \mathcal{J}_\parallel(r) &< J_P, \quad \text{if } \left(\frac{dJ}{dJ_0} \right)_P < 0. \end{aligned} \tag{63}$$

All this means that the components of $\mathbf{J}$ have no maximum at P when $(dJ/dJ_0)_P$ is positive, have no minimum when $(dJ/dJ_0)_P$ is negative. No matter what the sign of $(dJ/dJ_0)_P$, we can say that the components of $\mathbf{J}_0$ and $-\mathbf{A}$ have no maximum at P. This follows from the fact that a maximum of J is a maximum of J_0 when $(dJ/dJ_0)_P$ is positive; a minimum of J is a maximum of J_0 when $(dJ/dJ_0)_P$ is negative.

Let F be the internal free energy density (i.e. that part of the free energy not containing the magnetic field $\mathbf{H}$ explicitly) associated with the superconducting phase. In general

$$\mathbf{J} = -c \left(\frac{dF}{d\mathbf{A}} \right). \tag{64}$$

Making use of Equation (46), this can be rewritten

$$n_0 m v = \frac{dF}{dv_0}. \tag{65}$$

This is the generalization to finite temperatures of Equation (39). We see that it is a direct consequence of the proportionality between $\mathbf{A}$ and $\mathbf{v}_0$.

In case J_0, J, etc. are functions of one dimension only, say x, we can perform a first integration of Equations (50) or (51) analytically

with the aid of Equation (65). Thus

$$\lambda^2 \frac{d^2 v_0}{dx^2} = v, \tag{66}$$

integrates to

$$\frac{1}{2} n_0 m \lambda^2 \left(\frac{dv_0}{dx} \right)^2 = \frac{1}{8\pi} H^2 = F(v_0) + C, \tag{67}$$

C being an integration constant independent of x. The problem is now reduced to quadratures, since

$$\frac{x}{\lambda} = \int \left[\frac{\frac{1}{2} n_0 m}{F(v_0) + C} \right]^{1/2} dv_0, \tag{68}$$

the limits of the integral being chosen to satisfy boundary conditions.

Critical Current and Magnetic Field

In this section, the previous results are applied to the problems of critical currents and magnetic fields. The discussion is restricted to the case $T = 0$, since only for this case have we obtained ϵ_0, v, and $F = W_0$ as known functions of v_0. The discussion could easily be extended to finite temperatures if we had solved Equation (11) for ϵ_0 as a function of v_0 and T and had thereupon evaluated v_Q. [Once v_Q is obtained, v and F can be found with the aid of Equations (17) and (65), respectively.]

Consider a normal metal with a uniform applied magnetic field H_a. Introduce a superconducting inclusion in the region $-X \leqslant x \leqslant +X$, as illustrated by Figure 3. We choose as the boundary condition the requirement that ϵ_0 go to zero at $x = \pm X$ (i.e., continuity of ϵ_0 at the normal–superconducting interfaces). This serves to determine X for a given H_a. At $x = X$, since $\epsilon_0 = 0$, we must have $v_0 = (1/2)\, e\, v_{00}$ (see Figure 2) and $F(v_0) = W(v_0) = 0$. Equation (67) thus implies

$$C = \frac{1}{8\pi} H_a^2. \tag{69}$$

At $x = 0$, the center of the superconducting inclusion, $v_0 = v = 0$. Equations (68) and (69) give

$$\frac{X}{\lambda} = \int_0^{1/2\, e\, v_{00}} \left[\frac{\frac{1}{2} n_0 m}{W_0(v_0) + \frac{1}{8\pi} H_a^2} \right]^{1/2} dv_0. \tag{70}$$

H_0 is defined to be the $T = 0$ thermodynamic critical magnetic field of the BCS theory, i.e.,

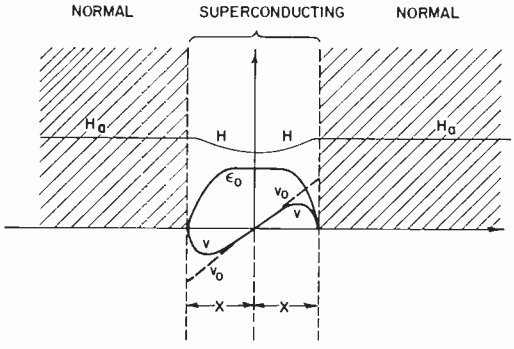

Fig. 3—Plot of H, ϵ_0, v_0, and v associated with a superconducting inclusion of width $2X$ in a normal metal.

$$\frac{H_0^2}{8\pi} = \frac{1}{2} N(0)\, \epsilon_{00}^2. \tag{71}$$

Making use of the explicit expressions for $W_0(v_0)$,

$$\frac{X}{\lambda} = \sqrt{\frac{2}{3}} \int_0^1 \left[\left(\frac{H_a}{H_0} \right)^2 - 1 + \frac{2}{3} z^2 \right]^{-1/2} dz$$

$$+ \sqrt{\frac{2}{3}} \int_0^1 \left[\left(\frac{H_a}{H_0} \right)^2 (1+q)^2 e^{-q} - \frac{1}{3} (1 - 3q^2 + 2q^3)\, e^{+q} \right]^{-1/2} q dq. \tag{72}$$

The first integral represents that portion of X for which $v_0/v_{00} \leqslant 1$;

the second, that portion for which $1 \leqslant v_0/v_{00} \leqslant e/2$. It is computationally convenient to set

$$X = X_1 + X_2. \tag{73}$$

where

$$\frac{X_1}{\lambda} \equiv \sqrt{\frac{2}{3}} \int_0^{e/2} \left[\left(\frac{H_a}{H_0} \right)^2 - 1 + \frac{2}{3} z^2 \right]^{-1/2} dz$$

$$= \operatorname{arcsinh} \left[\frac{1.10973}{\sqrt{(H_a/H_0)^2 - 1}} \right]. \tag{74}$$

$$\frac{X_2}{\lambda} \equiv \sqrt{\frac{2}{3}} \int_0^1 \left[\left(\frac{H_a}{H_0} \right)^2 (1+q)^2 e^{-q} - \frac{1}{3} (1 - 3q^2 + 2q^3) e^{+q} \right]^{-1/2} q dq$$

$$- \sqrt{\frac{2}{3}} \int_1^{e/2} \left[\left(\frac{H_a}{H_0} \right)^2 - 1 + \frac{2}{3} z^2 \right]^{-1/2} dz$$

$$= 0.0089728 \left(\frac{H_0}{H_a} \right)^3 - 0.0004659 \left(\frac{H_0}{H_a} \right)^5 + \ldots \tag{75}$$

We see that X_1, and thus X, diverges logarithmically as H_a approaches H_0 from above. This means that for the applied magnetic field H_a less than the thermodynamic critical magnetic field H_0, the superconducting region completely fills the volume of the superconductor, i.e., a complete Meissner effect. On the other hand, for H_a greater than H_0, a superconducting inclusion of finite width $2X$ can exist. We wish to determine whether or not the free energy of the system is lowered by the presence of the inclusion. The appropriate free energy density of the inclusion is

$$G = W_0 + \frac{1}{8\pi} H^2 - \frac{1}{4\pi} H_a B, \tag{76}$$

where

$$B = \frac{1}{X} \int_0^X H dx \tag{77}$$

is the magnetic induction of the inclusion. The term $(8\pi)^{-1} H^2$ in G accounts for the magnetic field energy; the term $-(4\pi)^{-1} H_a B$ accounts for the constraint that the total magnetic flux be fixed during a transformation from superconducting to normal phase or vice versa. Making use of Equations (67) and (69),

$$G = \frac{1}{8\pi}\left[2H^2 - H_a^2 - 2H_a X^{-1} \int_0^X H dx \right] , \tag{78}$$

and

$$\int_{-X}^{X} G dx = \frac{1}{4\pi} \int_0^X [2H(H - H_a) - H_a^2]\, dx. \tag{79}$$

The corresponding quantities for the normal phase, where $W_0 = 0$, $H = B = H_a$, are

$$G_n = -\frac{1}{8\pi} H_a^2, \tag{80}$$

$$\int_{-X}^{X} G_n\, dx = -\frac{1}{4\pi} \int_0^X H_a^2\, dx, \tag{81}$$

so that

$$\int_{-X}^{X} [G - G_n]\, dx = \frac{1}{4\pi} \int_0^X 2(H - H_a)\, H dx. \tag{82}$$

Since $H < H_a$ in the superconducting region, it follows that the free energy of the system is lowered by inserting the superconducting inclusion; this holds for all values of $H_a > H_0$. Thus an intermediate state can be formed by filling the superconductor with superconducting regions of thickness $2X$ separated by normal regions of negligible thickness, as shown in Figure 4. The normal regions should be made as small as possible in order to minimize the free energy; the presence of the normal regions serves only to enforce the boundary condition $\epsilon_0 = 0$ at the boundaries of each superconducting region.

The results of the previous paragraph indicate that the intermediate state is stable, relative to the normal state, for all values of $H_a > H_0$,

no matter how large H_a may be. This is certainly not true. The difficulty lies in the fact that when

$$H_a \approx \left(\frac{\lambda}{\xi} \right) H_0 \tag{83}$$

X becomes comparable to ξ, and it is no longer true that the variation of v_0 with position is negligible over lengths of the order of the coherence distance. Thus, when Equation (83) holds, the present theory is no longer applicable. Such a rapid variation of v_0 and ϵ_0 with position introduces additional positive kinetic energy into the free energy. Presumably, this causes the intermediate state to become energetically unstable with respect to the normal state.

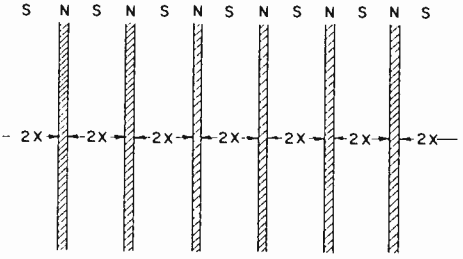

Fig. 4—Structure of the intermediate state.

Our picture of the intermediate state is geometrically similar to that of Goodman,[15] namely superconducting laminars separated by normal regions of negligible thickness. Goodman described the superconducting regions by means of the London theory, but he introduced the additional *ad hoc* assumption of a positive surface energy at each normal–superconducting interface. Unlike the present theory, Goodman's results indicated that the intermediate state set in at applied fields less than H_0.

Another model of the intermediate state has been proposed by Abrikosov,[16] who made use of the Ginzburg–Landau phenomenological theory of superconductivity.[17] Like Goodman, Abrikosov obtained mag-

[15] B. B. Goodman, "Simple Model for the Magnetic Behavior of Superconductors of Negative Surface Energy," *Phys. Rev. Letters*, Vol. 6, p. 597, June 1, 1961.

[16] A. A. Abrikosov, "On the Magnetic Properties of Superconductors of the Second Group," *Jour. Exptl. Theoret. Phys.* (USSR), Vol. 32, p. 1442, June, 1957; translation in *Soviet Physics JETP*, Vol. 5, p. 1174, December 15, 1957.

[17] V. L. Ginzburg and L. D. Landau, "On the Theory of Superconductivity," *Jour. Exptl. Theoret. Phys.* (USSR), Vol. 20, p. 1064 (1950).

netic field penetration into bulk samples at fields less than H_0. This occurred with the aid of vortex filaments of supercurrent, analogous to the vortex filaments of superfluid flow in liquid helium.[13] It is easy to show that vortex filaments cannot occur in the present theory. The solution of Equation (50) corresponding to a filament gives rise to a magnetic field diverging as the reciprocal of the distance from the center of the filament. Such a solution is unsatisfactory in that the magnetic energy is infinite.

We wish to consider the magnetization M of a bulk superconductor in the present model.

$$-4\pi M = H_a - B. \tag{84}$$

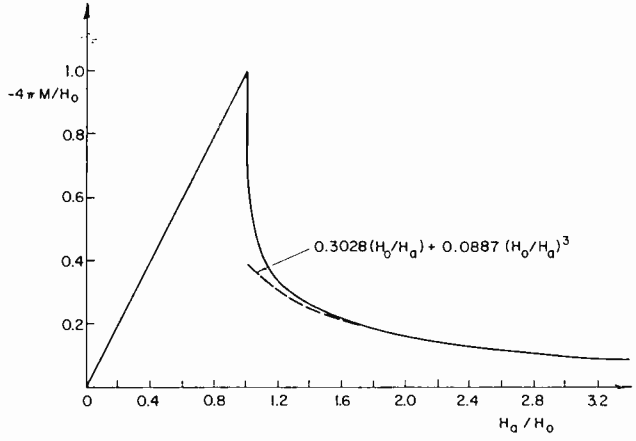

Fig. 5—Magnetization as a function of applied magnetic field.

For $H < H_0$, $B = 0$ and $-4\pi M = H_a$. For $H > H_0$,

$$B = \frac{1}{X}\int_0^X H dx = \sqrt{\frac{2}{3}}\left(\frac{\lambda}{X}\right)\left(\frac{H_0}{v_{00}}\right)\int_0^{1/2\,cv_{00}} dv_0$$

$$= 1.10973\, H_0\left(\frac{\lambda}{X}\right). \tag{85}$$

Knowing X as a function of H_a, we can now get $-4\pi M$ as a function of H_a. A plot is given in Figure 5. For H_a larger than about $2H_0$, it

is convenient to expand $-4\pi M$ as an inverse power series in (H_a/H_0),

$$-\left(\frac{4\pi M}{H_0}\right) = 0.302834 \left(\frac{H_0}{H_a}\right) + 0.088740 \left(\frac{H_0}{H_a}\right)^3 + \ldots \quad (86)$$

The reader will note that there is a peculiarity in these results as regards M as a function of H_a. There is a well-known theorem[11] which states that the *total* area under the $-M(H_a)$ curve is $H_0^2/(8\pi)$. This theorem is supposedly model-independent and based on thermodynamics alone. But we see immediately that the area under our $-M(H_a)$ curve is greater than $H_0^2/(8\pi)$; the area under that portion of the curve lying to the left of $H_a = H_0$ is $H_0^2/(8\pi)$. The resolution of this paradox lies in the fact that a hidden assumption is made in the standard derivation of the theorem — an assumption which is incorrect for the present situation. This is the assumption that the total free energy (i.e., the spatial integral of G) is a *continuous* function of H_a. Actually, there is a discontinuity in the free energy at every value of H_a for which there is a change in the total number n of superconducting layers. These critical values are denoted by H_{an}. For $n \ll W/\lambda$,

$$H_{an} \cong H_0 \left[1 + 2.46 \exp\left\{ -\frac{W}{n\lambda} \right\} \right], \quad (87)$$

where W is the size of the superconductor in the direction normal to the plane of the superconducting layers. This equation results from setting $2nX \cong 2nX_1 = W$ in Equation (74). For a macroscopic superconductor, W/λ will be a huge number ($\sim 10^5$). The fractional change in free energy at each discontinuity is of the order $1/n$. Thus the appreciable discontinuities occur at small values of n, where $H_{an} \cong H_0$ to an extremely good approximation. In calculating the magnetization, we have ignored that portion of the superconductor given over to the normal phase. This normal region, occupying a fractional volume of order X/W, results from the fact that W/X is not an integer for all values of H_a. Since (X/W) is an extremely small number for all $H_a > H_0$ except those H_a lying extremely close to H_0, we are justified in calculating M in this approximate fashion.

We have seen that for $H_a < H_0$ there is a complete Meissner effect. Let us examine in detail this case. Consider a semi-infinite superconductor having its surface at $x = 0$ ($x < 0$ corresponding to the region of the superconductor). As $x \to -\infty$, $H \to 0$ and $v_0 \to 0$. Thus Equation (67) has the form

$$\frac{1}{8\pi}(H^2 - H_0^2) = W_0(v_0). \tag{88}$$

We can combine this with Equation (38) to find the value of v_0 (designated v_{0s}) at the surface where $H = H_a$. We get

$$\frac{v_{0s}}{v_{00}} = \left(\frac{3}{2}\right)^{1/2} \left(\frac{H_a}{H_0}\right), \tag{89}$$

if $(H_a/H_0) \leqslant (2/3)^{1/2}$; whereas, if $(2/3)^{1/2} \leqslant (H_a/H_0) \leqslant 1$, v_{0s} is given by the implicit relation

$$\begin{aligned} 1 - \left(\frac{H_a}{H_0}\right)^2 &= \frac{e^{2q}}{3(1+q)^2}(1 - 3q^2 + 2q^3) \\ \frac{v_{0s}}{v_{00}} &= \frac{e^q}{1+q}. \end{aligned} \tag{90}$$

The mean distance of penetration of the magnetic field into the superconductor can now be calculated.

$$\begin{aligned} \lambda_H \equiv \frac{1}{H_a}\int_{-\infty}^{0} H\,dx &= \lambda\left(\frac{H_0}{H_a}\right)\left(\frac{2}{3}\right)^{1/2}\int_0^{v_{0s}} \frac{dv_0}{v_{00}} \\ &= \lambda\left(\frac{H_0}{H_a}\right)\left(\frac{2}{3}\right)^{1/2}\left(\frac{v_{0s}}{v_{00}}\right). \end{aligned} \tag{91}$$

A plot of λ_H/λ versus H_a/H_0 is given in Figure 6. It is seen that $\lambda_H = \lambda$ for $H_a/H_0 < (2/3)^{1/2}$, where the present theory becomes identical with the linear London theory.

The discussion of the last paragraph can be made to apply to the case of a macroscopic superconducting wire carrying current, in which case the external magnetic field H_a is due to the current in the wire. As with the London theory, a wire of circular cross section (radius R) will have a current

$$I = 2\pi R\int_{-\infty}^{0} J\,dx = \frac{1}{2}\,c\,R\,H_a \tag{92}$$

The critical current, defined as the maximum total current possible in the superconducting phase, will be

$$I_c = \frac{1}{2} c R H_0 . \tag{93}$$

On the other hand, if we have a wire with cross-sectional radius much smaller than λ, then the critical current is

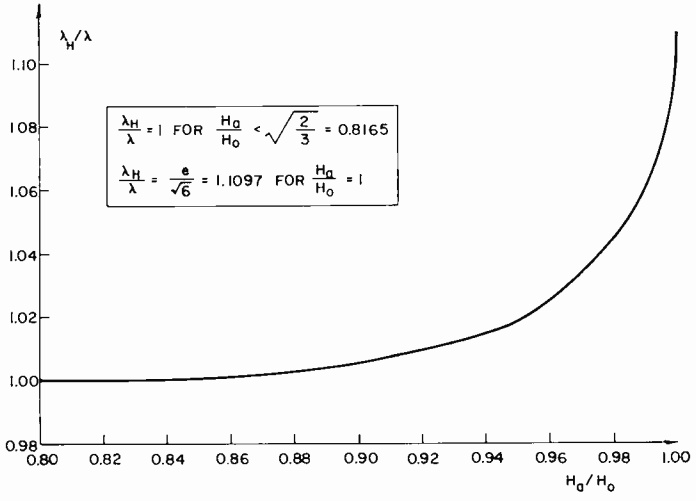

Fig. 6—Mean depth of penetration of magnetic field as a function of applied magnetic field.

$$I_c = 0.8256\, c \left(\frac{R^2}{4\lambda} \right) H_0 . \tag{94}$$

Here we have made use of the identity

$$\frac{c}{4\pi} \left(\frac{H_0}{\lambda} \right) = \left(\frac{3}{2} \right)^{1/2} n_0 e\, v_{00} , \tag{95}$$

and the fact that the *uniform maximum* current density is

$$J_{\max} = 1.0112\, n_0 e\, v_{00} . \tag{96}$$

Although Equation (93) agrees with the London theory of current

quenching,[18] Equation (94) does not. (In the London theory, the factor 0.8256 is missing.) This is a consequence of the assumption, in the London theory, that there is, under all conditions, a critical electron drift velocity of $(3/2)^{1/2}\, v_{00}$, corresponding to a critical drift kinetic energy density of $H_0^2/(8\pi)$. In the present theory, it was seen that there is a maximum drift velocity of 1.0112 v_{00}, but this maximum velocity can be described as a critical velocity only in the case of small specimens. It has been suggested[19] that current quenching occurs when the thermal gap $2(\epsilon_0 - p_F v_0)$ goes to zero, i.e., when $v_0 = v_{00}$. The results here show that this is not correct, although the error is only one per cent for the case of wire of small cross section.

We have seen in this section that superconductivity may be maintained, through the intermediate state, in magnetic fields much higher than the thermodynamic critical magnetic field H_0. In Equation (83), it was estimated that superconductivity would quench at fields of the order of $(\lambda/\xi)H_0$. It is tempting to associate such a field with the so-called Kunzler field H_K required for the destruction of superconductivity in certain superconducting alloys and compounds[20] (e.g., Nb_3Sn, Mo_3Re, V_3Ga, Nb + 25% Zr). Like $(\lambda/\xi)H_0$, H_K may be one order of magnitude or more larger than H_0. The difficulty with making such an association lies in the fact that as yet we have been unable to extend our analysis to the case of an intermediate state where a finite net current is flowing. Indeed, there is reason to believe that such an intermediate state is impossible in a *homogeneous* superconductor. Gorter[21] has argued that each individual current-carrying domain or filament will be impelled to move, with consequent energy dissipation, by the Lorentz force associated with the magnetic field of the other domains. Presumably, inhomogeneities or imperfections (e.g., dislocation lines) could pin down the current filaments and prevent their motion.

Extension to Finite Frequencies

Thus far, we have considered a superconductor to be without dissipation. Under the action of a finite-frequency electric field **E**, however, the normal carriers will get out of phase with the Cooper pairs so that,

---

[18] H. London, "Phase Equilibrium of Superconductors in a Magnetic Field," *Proc. Roy. Soc.*, Vol. A152, p. 650, November 15, 1935.

[19] See Reference (9).

[20] J. E. Kunzler, "Superconductivity in High Magnetic Fields at High Current Densities," *Rev. Mod. Phys.*, Vol. 33, p. 501, October, 1961.

[21] C. J. Gorter, "Note on the Superconductivity of Alloys," *Physics Letters*, Vol. 1, p. 69, May 1, 1962.

at any given instant, the former are not in the distribution f_k of minimum free energy in the laboratory coordinate system. We shall assume the new distribution function differs from that of Equation (6) only in that $-\mathbf{v}_0$ is replaced by $-\mathbf{v}_0 + \Delta\mathbf{v}_0$. The presence of $\Delta\mathbf{v}_0$ causes an increment in $\mathbf{v}_Q$ which is parallel to $\Delta\mathbf{v}_0$. The drift velocity $\mathbf{v}$ becomes

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_Q(\mathbf{v}_0 - \Delta\mathbf{v}_0), \tag{97}$$

i.e., $\mathbf{v}_Q$ is here a function of $\mathbf{v}_0 - \Delta\mathbf{v}_0$. We assume

$$\frac{d\mathbf{v}_0}{dt} = -\frac{e}{m}\mathbf{E}, \tag{98}$$

$$\frac{d\Delta\mathbf{v}_0}{dt} = -\frac{e}{m}\mathbf{E} - \frac{\Delta\mathbf{v}_0}{\tau}, \tag{99}$$

where τ is the lifetime associated with the normal-state d-c conductivity

$$\sigma_n = \frac{n_0 e^2 \tau}{m}. \tag{100}$$

Equation (98) is equivalent to Equation (47). Equation (99) is equivalent to a Boltzmann transport equation for the modified f_k. Two cases must be considered: (1) the a-c electric field and current are parallel to the d-c current; (2) the a-c electric field and current are perpendicular to the d-c current. We define

$$\begin{aligned} K(v_{0dc}) &= 1 + \frac{dv_Q}{dv_0}, \text{ case (1)} \\ &= 1 + \frac{v_Q}{v_0}, \text{ case (2).} \end{aligned} \tag{101}$$

Here the argument of K is the d-c value of v_0. Equation (101) is equivalent to Equation (60). Now $\mathbf{v}_0$ and $\mathbf{v}$ have both d-c and a-c components, but $\mathbf{E}$ and $\Delta\mathbf{v}_0$ have only a-c components. We can write

$$\mathbf{v}_{ac} = K\mathbf{v}_{0ac} + (1 - K)\,\Delta\mathbf{v}_{0ac}. \tag{102}$$

Taking all a-c quantities proportional to $e^{i\omega t}$, Equations (98), (99), and (102) can be solved for $\mathbf{v}_{ac}$ in terms of $\mathbf{E}_{ac}$. The resultant a-c conductivity,

$$\sigma = \sigma_1 - i\sigma_2 = -\frac{n_0 e v_{ac}}{E_{ac}}, \tag{103}$$

is

$$\frac{\sigma_1}{\sigma_n} = \frac{\pi K}{\tau}\,\delta(\omega) + \left(\frac{1-K}{1+\omega^2\tau^2}\right), \tag{104}$$

$$\frac{\sigma_2}{\sigma_n} = \frac{K}{\omega\tau} + \frac{(1-K)\,\omega\tau}{1+\omega^2\tau^2}. \tag{105}$$

We have inserted a term in σ_1 involving the zero-frequency delta function $\delta(\omega)$. This insures that $\sigma_1(\omega)$ and $\sigma_2(\omega)$ obey the Kramers–Kronig relations, so that causality is not violated.[22]

The above analysis is restricted to frequencies less than the optical energy gap frequency $\omega_g = 2\epsilon_0/\hbar$, since we have not allowed for the generation of normal-particle pairs by absorption of radiation. In practice, ω_g is usually less than the relaxation frequency τ^{-1}. This is especially true when there is appreciable scattering by boundaries or disorder. This means that Equations (104) and (105) are accurate only when $\omega\tau \ll 1$. Thus, for nonvanishing frequencies,

$$\begin{aligned} \sigma_1 &= (1-K)\,\sigma_n \\ \sigma_2 &= K\sigma_n/(\omega\tau). \end{aligned} \tag{106}$$

In the limit of vanishing d-c current, $K_\parallel = K_\perp = \gamma$, as defined in Equation (55), while Kn_0, $(1-K)n_0$ become the densities of superconducting and normal electrons, respectively, in the usual two-fluid model of superconductivity. Whenever the slope of the v versus v_0 curve goes negative, $K_\parallel$ is negative. This corresponds to a negative inductive reactance, and thus instability, for microscopic specimens carrying a d-c supercurrent. By measuring the a-c conductivity (real or imaginary part, parallel or perpendicular geometry), we can infer the d-c $v(v_0)$ curve.

Comparison with the Ginzburg–Landau Theory

In conclusion, we wish to point out that the Ginzburg–Landau (G–L) nonlinear phenonenological theory of superconductivity,[17] in the limit $\lambda/\xi \to \infty$ (in G–L notation, $\kappa \to \infty$), can be put in a form similar to the present theory. We redefine $v(v_0)$,

---

[22] See e.g., C. Kittel, *Elementary Statistical Physics*, John Wiley and Sons, Inc., New York, 1958.

$$v(v_0) = \gamma v_0 \left[1 - \frac{1}{2}\gamma \left(\frac{v_0}{v_c} \right)^2 \right] \tag{107}$$

where the temperature-dependent constants γ and v_c are defined, respectively, by Equation (55) and by

$$\frac{1}{2} n_0 m v_c^2 = \frac{H_c^2}{8\pi} . \tag{108}$$

H_c is the thermodynamic critical magnetic field, so that v_c is, in the London theory, the critical drift velocity. The effective wave function Ψ of the G–L theory is given by

$$|\Psi|^2 = \frac{v}{v_0} . \tag{109}$$

Subject to this redefinition of $v(v_0)$, Equations (46), (51), and (65) hold in the G–L theory. Following Gorkov,[23] if we associate Ψ with ϵ_0/ϵ_{00}, then

$$\frac{\epsilon_0}{\epsilon_{00}} = \gamma^{1/2} \left[1 - \frac{1}{2}\gamma \left(\frac{v_0}{v_c} \right)^2 \right]^{1/2} \tag{110}$$

At $T = 0$, where $\gamma = 1$, $v_c = (3/2)^{1/2} v_{00}$, Equations (107) and (110) become

$$v(v_0) = v_0 \left[1 - \frac{1}{3} \left(\frac{v_0}{v_{00}} \right)^2 \right] \tag{111}$$

$$\epsilon_0(v_0) = \epsilon_{00} \left[1 - \frac{1}{3} \left(\frac{v_0}{v_{00}} \right)^2 \right]^{1/2} \tag{112}$$

These curves differ appreciably from the curves of Figure 2. For example, Equation (111) has a maximum v of (2/3) v_{00} occurring at $v_0 = v_{00}$. Thus the G–L theory cannot be expected to be accurate at $T = 0$, even when $\lambda \gg \xi$.

ACKNOWLEDGMENT

The author is indebted to colleagues at RCA Laboratories for many helpful discussions.

[23] L. P. Gorkov, "Microscopic Derivation of the Ginzburg-Landau Equations in the Theory of Superconductivity," *Jour. Exptl. Theoret. Phys.* (USSR), Vol. 36, p. 1918, June, 1959; *Soviet Physics JETP*, Vol. 36 (9), p. 1364, December, 1959.

EFFECTS OF FADING ON QUADRATURE RECEPTION OF ORTHOGONAL SIGNALS

By

GILBERT LIEBERMAN

RCA Aerospace Communications and Controls Division,
Camden, N. J.

Summary—*An analysis of effects of fading, frequency instability, and noise on quadrature reception of orthogonal signals is given. The results are in terms of formulas and graphs of error probability for binary signaling as a function of the following parameters: (1) signal-energy-to-noise spectral density, (2) relative fading rate, (3) relative frequency instability, (4) ratio of the specular component of the received signal to the r-m-s value of the random component, (5) proportion of the additive noise which is fading. The analysis includes effects of the shape of the spectrum of the fading carrier. The model assumes nonselective fading consisting of a specular component superposed on a Gaussian-process random component. The measure of error probability performance used consists of quantiles of error probability as well as the mean probability of error.*

INTRODUCTION

THE COMMUNICATION SYSTEM considered here is designed to operate reliably in a narrow-band, low signal-to-noise ratio channel.[1] The signal from a power-limited transmitter must be of a duration such that the ratio of signal energy to noise-power density after detection is sufficiently large. However, the longer the signal duration, the more vulnerable the system is to channel variations. For example, in the case of frequency instability, the system performance can be completely degraded regardless of signal-to-noise ratio (see Figure 2).

The measure of frequency instability used is denoted by ΔfT, where T is the transmitted signal duration, and Δf is a frequency displacement between the receiver local oscillator and the incoming carrier. In addition to considering independent signal and noise fading, error probabilities are given for identically fading desired and undesired signals in the slow-fading case. The term fading refers here to signal envelope and phase fluctuations due to the transmission medium. A

[1] G. Lieberman, "Quantization in Coherent and Quadrature Reception of Orthogonal Signals," *RCA Review*, Vol. 22, No. 3, p. 461, Sept. 1961.

parameter λ is considered which is a measure of the proportion of the additive noise which is subject to fading. That is, two independent additive white noises are assumed, one of which is subject to fading. When $\lambda = 0$, only the signal is subject to fading; when $\lambda = 1$, both desired and undesired signals are subject to fading. For the case where $\lambda = 0$, formulas have been derived for the "quantiles" of error probability (Equation (42)) and for the mean error probability (Equation (54)). In these equations γ represents the ratio of the amplitude of the constant component to the standard deviation of the random component of the received signal. The measure of fading rate used is the product of the single-sided bandwidth B of the fading carrier and the signal duration T. Equations (54) and (58) for $\lambda = 0$ and $\lambda = 1$, respectively, apply to slow signal fading when $BT = 0$. Note that $BT = 0$ does *not* refer to no fading. The mathematical model involves a narrow bandwidth Gaussian process. No matter how small the bandwidth B, the process is still random, and therefore subject to envelope and phase fluctuations.

In order to estimate the performance of communication systems employing a particular modulation technique in an operational environment, it is necessary to analyze the effects of given environments on communication reliability. Before particular environments are considered, however, an expression for the received signal will be given in a general form that includes a large number of possible situations. The effects of the signal transmission medium on received signals is expressed here in terms of its response to a single frequency input $\cos \omega_0 t$. It is assumed that for any ω_0 within the transmission bandwidth, the received signal can be written in the form

$$\sigma(t) = A a_s(t) \cos [\omega_0 (t - \tau) + \phi_s(t) + \theta(t)]. \tag{1}$$

The parameter τ represents lack of time synchronization. In this analysis τ will be taken to be zero. The parameter A is proportional to the received signal r-m-s value. The functions $a_s(t)$ and $\phi_s(t)$ represent signal envelope and phase fluctuations due to the transmission process. $\theta(t)$ represents phase variations of the signal carrier relative to the receiver local oscillator. At this point, no assumption has been made as to the nature of the functions $a_s(t)$ and $\phi_s(t)$. For any given message transmission interval, particular but arbitrary realizations of $a_s(t)$ and $\phi_s(t)$ are observed. It is useful to define $\theta(t)$ as

$$\theta(t) = 2\pi \Delta f t + \theta_0 . \tag{2}$$

The parameters Δf and θ_0 represent frequency and phase instability.

Superposed on the above-defined received signal, there can be additive noise $\eta(t)$. This additive noise is divided here into two additive components $\eta_1(t)$ and $\eta_2(t)$. The first noise function corresponds to external noise (e.g., jamming) which is subject to envelope and phase fluctuations due to the medium, as in the case of the above signal. The second noise function corresponds to the aggregate of all noise present which is not subject to envelope and phase fluctuations. The first noise function can be written

$$\eta_1(t) = a_n(t)\ R_1(t)\ \cos\ [\omega_0 t + \phi_1(t) + \phi_n(t)]. \tag{3}$$

In the analysis which follows, $R_1(t)$ and $\phi_1(t)$ correspond to the envelope and phase of band-limited white Gaussian noise occupying the signal transmission band; $a_n(t)$ and $\phi_n(t)$ represent envelope and phase fluctuations of the noise caused by the medium.

The second noise function can be written

$$\eta_2(t) = R_2(t)\ \cos\ [\omega_0 t + \phi_2(t)]. \tag{4}$$

This function is a band-limited stationary white Gaussian noise process occupying the signal-transmission band. In the analysis, all distinct random processes will be defined to be mutually statistically independent.

In the case of binary orthogonal phase-shift-keyed (PSK) signals, the original transmitted signal is of the form $S(t)\ \cos\ \omega_0 t$, where $S(t)$ is the baseband mark signal. That is, we have suppressed-carrier amplitude modulation of a carrier of angular frequency ω_0 by a pseudo-random type of square wave $S(t)$, assuming the values $+1$ and -1. For such transmitted signals, the over-all receiver input signal and noise is

$$x(t) = S(t)\ \sigma(t) + \eta_1(t) + \eta_2(t). \tag{5}$$

The approach taken in the analysis which follows is to calculate the conditional probability of error for given medium envelope and phase realizations. This is obtained by averaging over the two additive white Gaussian noise functions defined above. This conditional probability of error function can be expressed in terms of A, $a_s(t)$, $a_n(t)$, $\phi_s(t)$, $\phi_n(t)$ and $\theta(t)$. The purpose is to show how these functions contribute toward the probability of error. When there is no signal fading, this function *is* the error probability function. When there is signal fading, it is necessary to study the statistical properties of the conditional error probability function. The usual mean error probability is the mean

value of the conditional error probability averaged over the fading distribution. Since the mean ignores the inherent variability of error probability due to signal fading, the distribution function of conditional error probability has been calculated in addition to the mean error probability.

Conditional Probability of Error

Figure 1 shows a block diagram of the quadrature detector. The input signal and noise $x(t)$ is split up into two signals in quadrature, $y(t)$ and $\hat{y}(t)$. $y(t)$ is obtained by low-pass filtering the product of

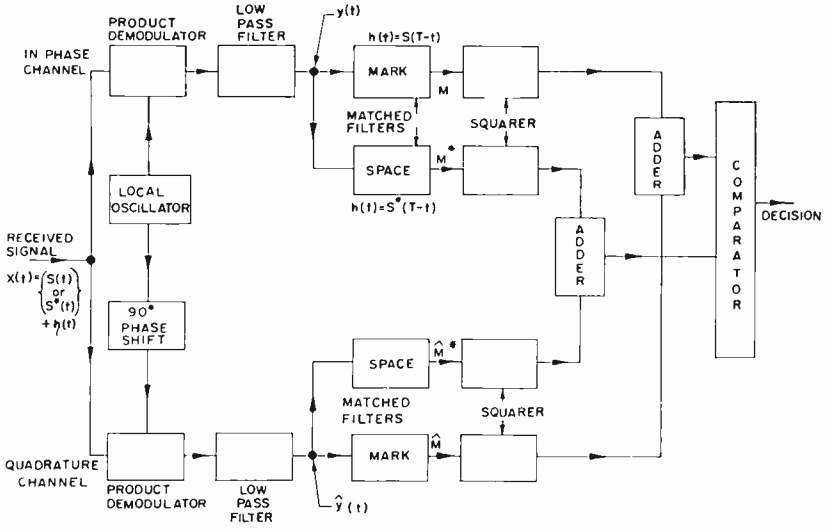

Fig. 1—Block diagram of an ideal quadrature detector.

$x(t)$ and $\cos \omega_0 t$; $\hat{y}(t)$ is obtained by low-pass filtering the product of $x(t)$ and $\sin \omega_0 t$. Applying the definition of $x(t)$ in Equation (5), the resulting functions can be written

$$y(t) = AS(t)a_s(t) \cos [\phi_s(t) + \theta(t)] + a_n(t)\, R_1(t) \cos [\phi_1(t) + \phi_n(t)] + R_2(t) \cos \phi_2(t)$$

$$-\hat{y}(t) = AS(t)a_s(t) \sin [\phi_s(t) + \theta(t)] +$$

$$+ a_n(t)\ R_1(t)\ \sin\left[\phi_1(t) + \phi_n(t)\right]$$
$$+ R_2(t)\ \sin \phi_2(t) \tag{6}$$

The next step is to cross correlate $y(t)$ and $\hat{y}(t)$ with local replicas of the possible orthogonal signals. In the binary case, there are 2 possible orthonormal signals $S(t)$ and $S^*(t)$, the baseband mark and space signals, respectively. There are therefore four product detector outputs. For the transmitted signal,

$$M = \int_0^T S(t)\ y(t)\ dt,$$
$$\hat{M} = \int_0^T S(t)\ \hat{y}(t)\ dt. \tag{7}$$

For the nontransmitted signal,

$$M^* = \int_0^T S^*(t)\ y(t)\ dt,$$
$$\hat{M}^* = \int_0^T S^*(t)\ \hat{y}(t)\ dt. \tag{8}$$

where M = inphase matched filter output with the correct signal,

$\hat{M}$ = quadrature matched filter output with the correct signal,

M^* = inphase matched filter output with the wrong signal,

$\hat{M}^*$ = quadrature matched filter output with the wrong signal.

The decision procedure shown in Figure 1 is equivalent to comparing a rooter output with the correct signal (Q) with a rooter output with the wrong signal (Q^*):

$$Q = \sqrt{M^2 + \hat{M}^2},$$
$$Q^* = \sqrt{(M^*)^2 + (\hat{M}^*)^2}. \tag{9}$$

In appendix A, it is shown that the conditional probability of error for given realizations of signal envelope and phase fluctuations is given by

$$P_e = \frac{\exp\left\{-\frac{\beta}{2+\delta}\right\}}{1+\frac{1}{1+\delta}} \tag{10}$$

where

$$\beta = \frac{E}{N_0}\left\{\frac{X^2+Y^2}{\lambda Z^2+(1-\lambda)}\right\} \qquad E = \frac{A^2T}{2}$$

$$\delta = \frac{E}{N_0}\left\{\frac{(X^*)^2+(Y^*)^2}{\lambda Z^2+(1-\lambda)}\right\} \tag{11}$$

$$X = \frac{1}{T}\int_0^T a_s(t)\cos\left[\phi_s(t)+\theta(t)\right]dt$$

$$Y = \frac{1}{T}\int_0^T a_s(t)\sin\left[\phi_s(t)+\theta(t)\right]dt \tag{12}$$

$$X^* = \frac{1}{T}\int_0^T S(t)\,S^*(t)\,a_s(t)\cos\left[\phi_s(t)+\theta(t)\right]dt$$

$$Y^* = \frac{1}{T}\int_0^T S(t)\,S^*(t)\,a_s(t)\sin\left[\phi_s(t)+\theta(t)\right]dt \tag{13}$$

$$Z^2 = \frac{1}{T}\int_0^T a_n^2(t)\,dt \tag{14}$$

where β = effective ratio of post-detection signal energy to noise power,

δ = an energy-to-noise density factor due to coherence between the reference space signal and the incoming fading mark signal,

X = time average over the information bit duration of the inphase component of the received signal,

Y = time average over the information bit duration of the quadrature component of the received signal,

X^* = time average over the information bit duration of the inphase component of the received signal multiplied by the product of the mark and space signal,

Y^* = time average over the information bit duration of the quadrature component of the received signal multiplied by the product of the mark and space signal,

Z^2 = time average over the information bit duration of the square of the undesired signal fading envelope.

λN_0 would be the spectral density of the noise function $\eta_1(t)$ if it did not contain envelope and phase fluctuations due to the medium. $(1 - \lambda)\ N_0$ is the spectral density of the nonfading noise $\eta_2(t)$. Thus λ is the proportion of the total noise power which is fading. Note that the phase fluctuations of the noise due to fading, $\phi_n(t)$ have dropped out in the analysis.

A basic assumption in deriving the conditional error probability in Equation (10) is that δ should be small compared to unity. It is shown in Appendix B that in an ensemble of possible long-duration signals generated from Bernoulli trials, the values of δ will be of the order of the input signal-to-noise power ratio with high probability. Therefore for relatively weak signals, which is the case of interest, δ will indeed be negligible. Consequently, the conditional probability of error is approximately

$$P_e = \frac{1}{2} \exp \left\{ \frac{-\beta}{2} \right\}. \tag{15}$$

When there is no fading and no frequency instability, a_s, $a_n = 1$. The phase $\phi_s = 0$, and θ is constant. It follows from Equation (11) that $\beta = E/N_0$. Thus Equation (15) is the familiar formula for binary noncoherent frequency-shift keying (FSK).[2]

When there is no fading, but there is frequency instability, it follows from the definition of $\theta(t)$ in Equation (2) that

[2] S. Reiger, "Error Probabilities of Binary Data Transmission Systems in the Presence of Random Noise," *I.R.E. Convention Record*, Pt. 8, p. 72, 1953.

$$\beta = \frac{E}{N_0}\left[\frac{\sin \pi \Delta f T}{\pi \Delta f T}\right]^2 \tag{16}$$

where $\Delta f T$ is a measure of relative frequency instability. Note that when $\Delta f T = 1$, $\beta = 0$ regardless of E/N_0. Thus increasing T for given Δf can result in substantial system degradation. Error probabilities can be calculated by applying Equation (16) to Equation (15). Figure 2 contains curves of error probability as a function of $10 \log_{10} E/N_0$ for $\Delta f T = 0$, 1/4, 1/2, and 1. These curves show that the quadrature detection system is not affected significantly by frequency instability as long as $\Delta f T < 1/4$; Figure 2 shows that there is a loss of less than 1 decibel. On the other hand, when $\Delta f T = 1$, the loss becomes infinite.

Distribution Functions of Conditional Probability of Error

The statistical properties of β used in conjunction with Equation (15) determine the statistical properties of conditional error probability performance. The distribution function of error probability is

$$\begin{aligned} F_{Pe}(x) &= P\,[P_e \leq x]\,;\ 0 \leq x \leq 1/2, \\ &= P\left[\frac{\exp\left\{-\frac{\beta}{2}\right\}}{2} \leq x\right] \\ &= P\left[\beta \geq 2 \ln\left(\frac{1}{2x}\right)\right] \\ &= 1 - F_\beta\left[2 \ln\left(\frac{1}{2x}\right)\right]. \end{aligned} \tag{17}$$

The function $F_\beta(z)$ is the distribution function of the random variable β.

The "quantiles" of error probability, x_p, are defined by

$$F_{Pe}(x_p) = p. \tag{18}$$

From Equation (17), it follows that

$$x_p = \frac{1}{2} \exp\left\{-\frac{z_{1-p}}{2}\right\} \tag{19}$$

where z_{1-p} is the solution of the equation

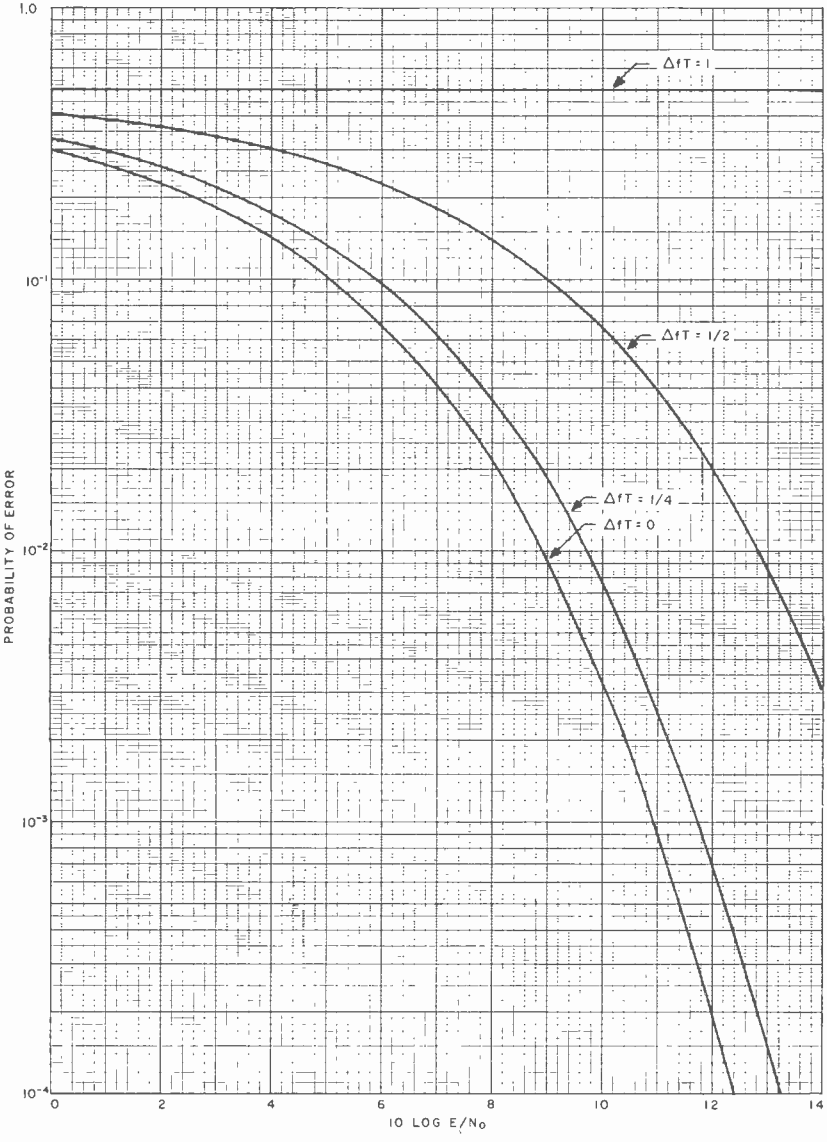

Fig. 2—Probability of error versus 10 $\log_{10}$ (E/N_0) for $\Delta fT = 0, 1/4, 1/2, 1$.

$$F_\beta(z_{1-p}) = 1 - p. \tag{20}$$

The quantity x_p is that value of conditional error probability which is exceeded in a proportion $1 - p$ of a large number of successive signal transmissions. Generally, for each signal transmission the realization of fading is different, and therefore the conditional probability of error will also be different. The above analysis considers the probability distribution of these error probability variations. In the slow-fading case, this information is of particular significance for systems which are designed to adapt to the channel variations.[3]

In defining the model to include a constant received signal component, it is assumed that the inphase component of the fading carrier is a Gaussian noise process with mean γ and unit variance, and the corresponding quadrature component is an independent Gaussian noise process with zero mean and unit variance. In general, there will be parameters γ_s and γ_n corresponding to the desired and undesired signals. The envelope and phase functions a_s, a_n, ϕ_s, and ϕ_n defined in Equations (1) and (3) correspond to the envelope and phase of sinusoids with respective amplitudes γ_s and γ_n, each superimposed on a Gaussian noise function.

In making comparisons, it is important to employ an appropriate measure of signal-to-noise ratio. From Equation (1), the instantaneous signal power can be defined as $A^2 a_s^2(t)/2$. Its mean value[4] is $A^2(1 + \gamma_s^2/2)$. Similarly from Equations (3), (4), and (5), the instantaneous noise spectral density can be defined as $N_0\ [\lambda a_n^2(t) + 1 - \lambda]$ by averaging over the additive noise but not over the multiplicative noise. Its mean value is therefore $N_0\ [\lambda(\gamma_n^2 + 2) + 1 - \lambda]$. The measure of signal-to-noise ratio used is the ratio of the mean signal energy to the mean noise spectral density. This quantity can be written

$$\frac{<\mathrm{E}>}{<N_0>} = \frac{E}{N_0}\left[\frac{2 + \gamma_s^2}{\lambda(2 + \gamma_n^2) + 1 - \lambda}\right]; \qquad E = \frac{A^2 T}{2}. \tag{21}$$

In the error probability expressions given below, E/N_0 is always expressed in terms of $<E>/<N_0>$ in accordance with Equation (21). For independent Rayleigh desired and undesired signal fading with $\lambda = 1$, the ratio of means in Equation (21) is equal to E/N_0. For

[3] G. Lieberman, "Adaptive Digital Communication for a Slowly Varying Channel," CP #61-1157, *AIEE Fall General Meeting*, Oct. 16, 1961. (To be published in *Trans. AIEE Communications and Electronics.*)

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

$\lambda = 0$, $<E>/N_0 = E/N_0(2 + \gamma_s^2)$. The removal of the averaging symbol on N_0 corresponds to the fact that the undesired signal is not fading.

Slow Fading and Frequency Instability

In the slow-fading case, the signal envelope and phase fluctuations a_s, a_n, and θ_s are slowly varying functions compared with the information rate $1/T$. The quantity β in Equation (11) reduces to

$$\beta = \frac{E}{N_0}\left(\frac{\sin \pi \Delta f T}{\pi \Delta f T}\right)^2 \left[\frac{a_s^2}{\lambda a_n^2 + 1 - \lambda}\right] \tag{22}$$

In Appendix B, the quantity δ in the error probability expression, Equation (10), is shown to be of the order of the input signal-to-noise ratio, and is therefore neglected here.

In the case of independent Rayleigh desired and undesired signal fading, $a_s^2/2$ and $a_n^2/2$ will be independent random variables with exponential probability densities. From Equations (22) and (17),

$$\begin{aligned}
F_{Pe}(x) &= P\left[\beta \geqq 2 \ln \frac{1}{2x}\right] \; ; \qquad 0 \leqq x \leqq \frac{1}{2} \\
&= P\left[\frac{a_s^2}{2} \geqq \lambda g(x)\, \frac{a_n^2}{2} + \frac{(1-\lambda)\, g(x)}{2}\right] \\
&= \int_0^\infty \int_{\lambda g u + \frac{1-\lambda}{2} g}^\infty e^{-(u+v)}\, dv du \\
&= \frac{\exp\left[-\frac{1}{2}(1-\lambda)\, g(x)\right]}{1 + \lambda g(x)} \\
g(x) &= \frac{N_0}{E}\left[\frac{\sin \pi \Delta f T}{\pi \Delta f T}\right]^{-2} 2 \ln\left(\frac{1}{2x}\right), \\
\frac{<E>}{<N_0>} &= \frac{2}{1+\lambda}\left(\frac{E}{N_0}\right).
\end{aligned} \tag{23}$$

Comparing Equations (23) and (17), it is seen that

$$F_\beta(z) = 1 - \frac{\exp\left[-\frac{(1-\lambda)}{2}\frac{N_0}{E}\left(\frac{\sin \pi \Delta f T}{\pi \Delta f T}\right)^{-2} z\right]}{1+\lambda\frac{N_0}{E}\left[\frac{\sin \pi \Delta f T}{\pi \Delta f T}\right]^{-2} z} \tag{24}$$

Figure 3 shows graphs of these distribution functions for $\lambda = 0$, corresponding to signal only fading. These curves show the increasing

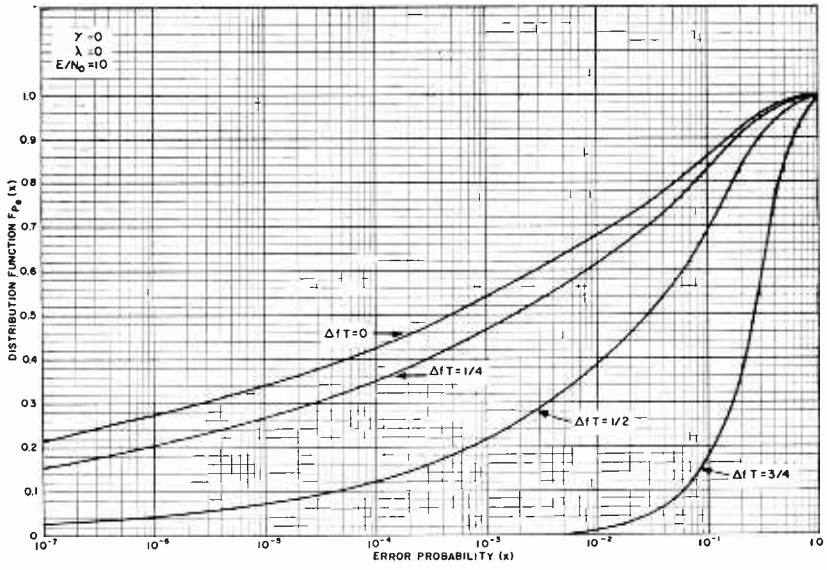

Fig. 3—Distribution functions of error probability for $\Delta fT = 0, 1/4, 1/2, 3/4$.

likelihood of occurrence of large error probabilities as ΔfT gets larger. Figure 4 shows the relative insensitivity of the quartiles of error probability to the parameter λ. The smallest median error probabilities occur for $\lambda = 0$. However, the spread of the distribution is larger when $\lambda = 1$. This is in agreement with results calculated previously[5] that the required fading margin with undesired signal fading is somewhat less than when there is only signal fading. Applying Equations (19) and (20) to Equation (24) leads to simple formulas for the

[5] A. B. Glenn and G. Lieberman, "Effect of Propagation Fading and Antenna Fluctuations on Communication Systems in a Jamming Environment," *Trans. I.R.E. PGCS*, Vol. CS-10, March 1962. See also *Proc. National Aeronautical Electronics Conf.*, Dayton, Ohio, May 1960.

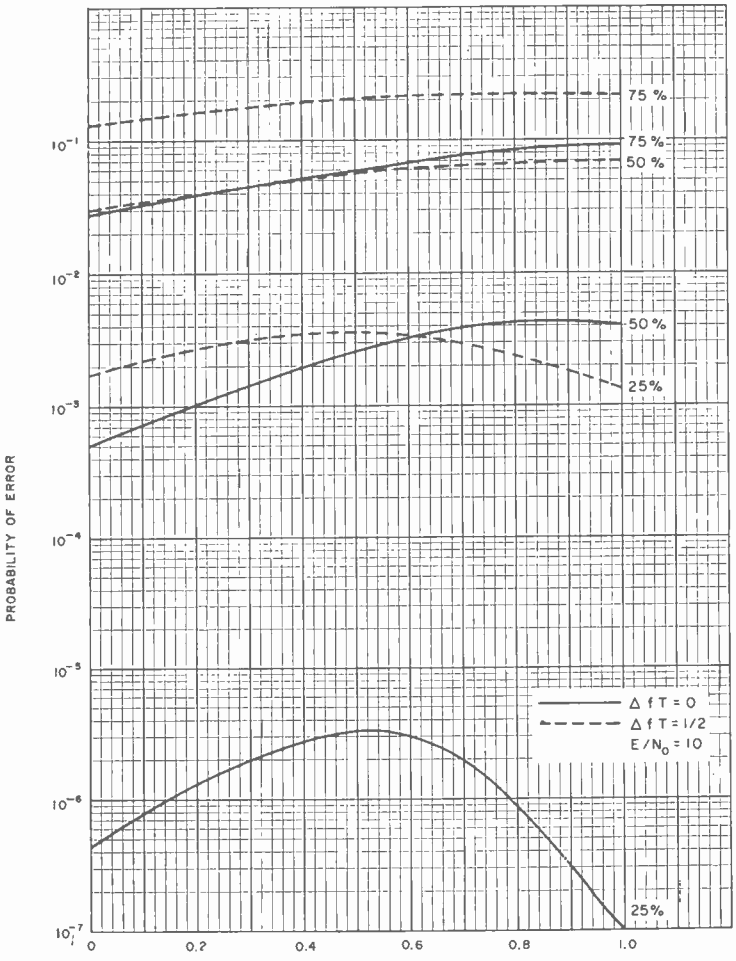

Fig. 4—Quartiles of error probability as a function of λ.

quantiles when $\lambda = 0, 1$. For $\lambda = 0$ (signal only fading),

$$x_p = \frac{1}{2}\,(p)^{(<E>/2N_0)(\sin\pi\Delta fT/\pi\Delta fT)-2}. \tag{25}$$

For $\lambda = 1$ (signal and noise fading),

$$x_p = \frac{1}{2}\exp\left\{\left[-\frac{<E>}{2<N_0>}\right]\left(\frac{\sin\pi\Delta fT}{\pi\Delta fT}\right)^2\left(\frac{1}{p}-1\right)\right\} \tag{26}$$

In the case of identical desired and undesired signal fading, $a_s = a_n = a$,

$$F_{Pe}(x) = P\left[\beta \geqq 2 \ln \frac{1}{2x}\right]$$

$$= P\left[\frac{\frac{a^2}{2}}{\frac{\lambda a^2}{2} + \frac{1-\lambda}{2}} \geqq g(x)\right]; \quad \frac{\langle E \rangle}{\langle N_0 \rangle} = \frac{2}{1+\lambda}\left(\frac{E}{N_0}\right) \tag{27}$$

$$= 1 - F_{a^2/2}\left[\frac{\left(\frac{1-\lambda}{2}\right) g(x)}{1-\lambda g(x)}\right]$$

when $\quad \frac{1}{2}\exp\left[-\frac{E}{2\lambda N_0}\left(\frac{\sin \pi \Delta f T}{\pi \Delta f T}\right)^2\right] < x < \frac{1}{2}$;

$= 0$ otherwise.

The function $F_{a^2/2}(z)$ is the distribution function of $a^2/2$. The reason for the truncation in Equation (27) is that β is largest when a is infinite. However, when a is infinite, this results in the minimum error probability given in Equation (27). The function $g(x)$ is defined in Equation (23).

For a fading model consisting of a constant component superposed on a Gaussian process random component,[6,7] the distribution function of $a^2/2$ has been tabulated.[7-9] This function is

$$F_{a^2/2,\,\gamma}(y) = \int_0^y \exp\left\{-\left(x + \frac{\gamma^2}{2}\right)\right\} I_0(\gamma\sqrt{2x})\, dx. \tag{28}$$

---

[6] G. L. Turin, "Error Probabilities for Binary Symmetric Ideal Reception Through Nonselective Slow Fading and Noise," *Proc. I.R.E.*, Vol. 46, p. 1603, Sept., 1958.

[7] K. A. Norton, L. E. Vogler, W. V. Mansfield, and P. J. Short, "The Probability Distribution of a Constant Vector Plus a Rayleigh-Distributed Vector," *Proc. I.R.E.*, p. 1354, Oct., 1955.

[8] *Table of Circular Normal Probabilities*, Bell Aircraft Corp., Report No. 02-949-106, June, 1956 (AD #139515).

[9] J. I. Marcum and P. Swerling, "Studies of Target Detection by Pulsed Radar," *Trans. I.R.E. PGIT*, Vol. IT-6, April, 1960.

Figure 5 contains graphs of this function for different values of γ. The data for these graphs were obtained from Reference (7). The parameter γ is the ratio of the amplitude of the constant signal component to the r-m-s value of the Gaussian random signal component. The quantiles of error probability x_p can be computed from

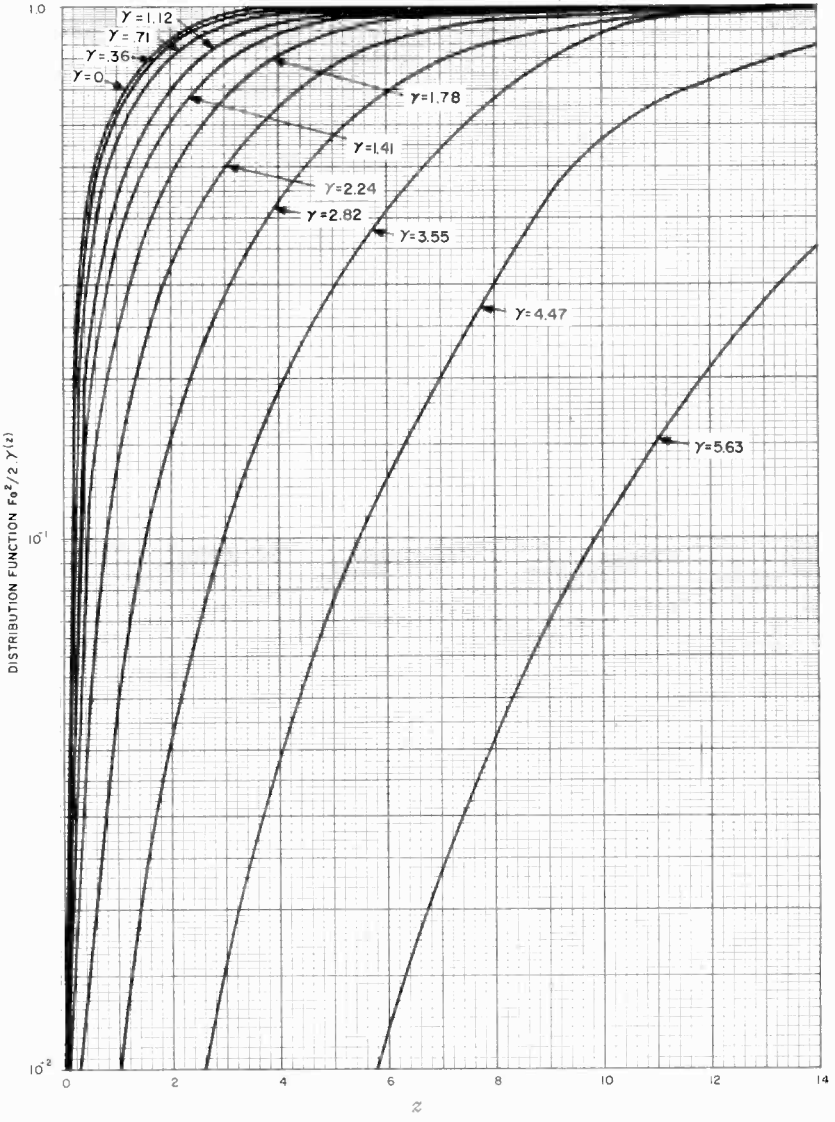

Fig. 5—Distribution function $F_{a^2/2,\gamma}(z)$ for different values of γ.

$$x_p = \frac{1}{2}\exp\left[-\frac{\dfrac{E}{2N_0}\left(\dfrac{\sin \pi \Delta f T}{\pi \Delta f T}\right)^2 y_{1-p}}{\left(\dfrac{1-\lambda}{2}\right)+\lambda y_{1-p}}\right]; \quad \frac{<E>}{<N_0>} = \frac{2}{1+\lambda}\left(\frac{E}{N_0}\right) \tag{29}$$

where y_{1-p} satisfies the equation

$$F_{a^2/2,\,\gamma}(y_{1-p}) = 1-p, \tag{30}$$

and x_p must exceed the truncation point defined in Equation (27). Note the similarity of Equation (29) to Equation (26). When $\lambda = 0$ (signal-only fading), Equation (29) reduces to Equation (25) when $\gamma = 0$, since in this case

$$y_{1-p} = \ln \frac{1}{p}. \tag{31}$$

When $\lambda = 1$ (signal and noise fading), Equation (29) shows that the error probability performance is the same as that for no fading (apply Equation (16) to (15)) since the distribution function in Equation (27) becomes a step function. Figure 6 shows the variation of the quartiles of error probability as a function of λ for slow identical Rayleigh fading. When $\lambda = 1$, the envelope fluctuations of the desired and undesired signals cancel each other out, thus resulting in a constant error probability. Figure 6 shows the approach to this constant value with increasing λ. Figure 7 shows curves of quartiles of error probability as a function of 10 $\log_{10} E/N_0$ for slow Rayleigh fading with $\lambda = 1/2$. Figure 8 shows the dependence of the quartiles of error probability on the fading parameter γ when there is no frequency instability for $\lambda = 0$. The curves for signal-only fading differ very little from those for signal and noise fading.

Fading With Arbitrary Fading Rate

In this case a_s, a_n and ϕ_s are random time functions with arbitrary spectral densities. The systematic phase instability $\theta(t)$ is assumed to be constant. The random variable β which determines the statistical properties of error probability is defined by Equations (11) and (12) where

$$\beta = \frac{E}{N_0}\left\{\frac{X^2+Y^2}{\lambda Z^2+1-\lambda}\right\},$$

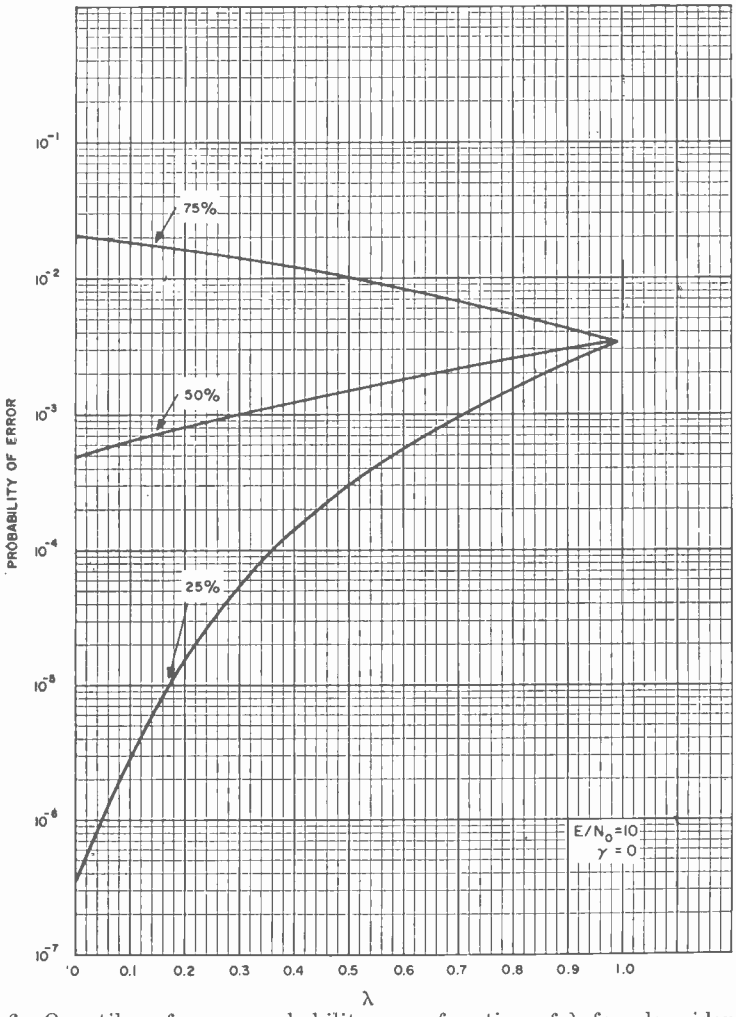

Fig. 6—Quartiles of error probability as a function of λ for slow identical Rayleigh fading; $E/N_0 = 10$.

$$X = \frac{1}{T}\int_0^T I_s(t)\,dt,$$

$$Y = \frac{1}{T}\int_0^T J_s(t)\,dt,$$

$$Z^2 = \frac{1}{T}\int_0^T a_n^2(t)\,dt. \tag{32}$$

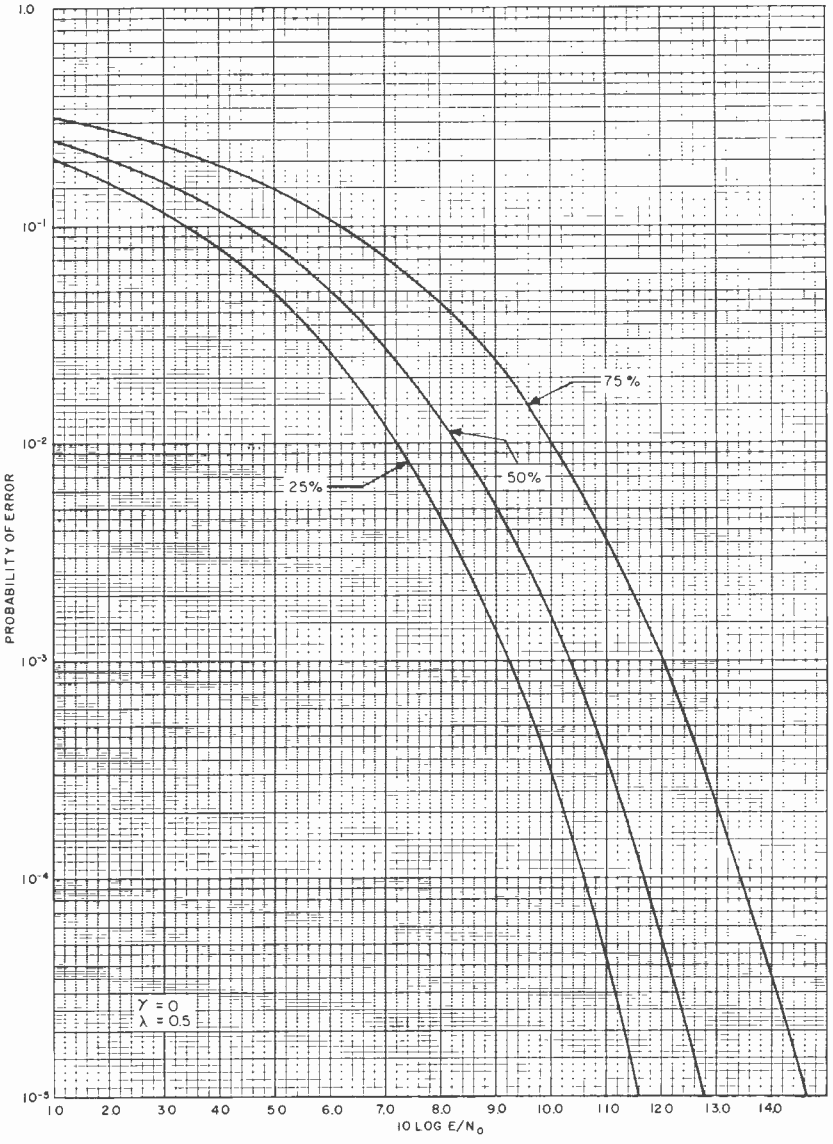

Fig. 7—Quartiles of error probability as a function of $10 \log_{10} E/N_0$ for slow identical Rayleigh fading; $\lambda = 0.5$.

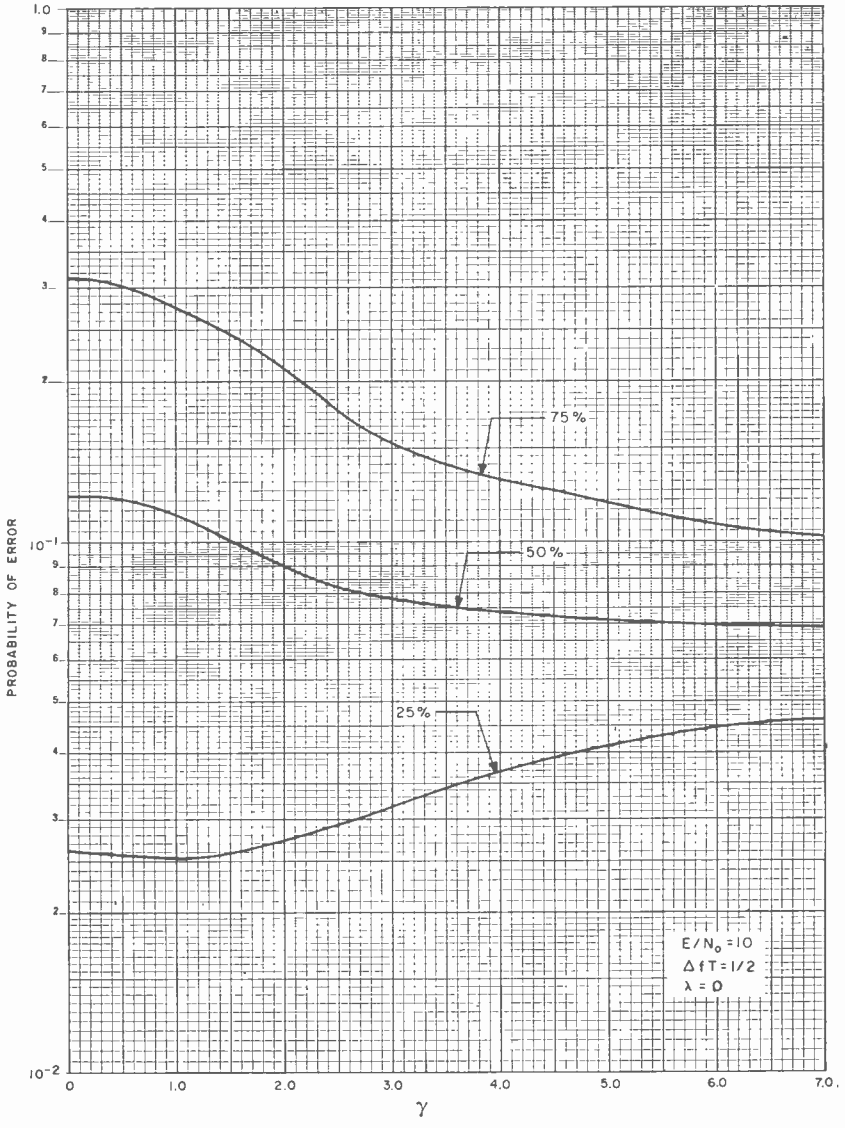

Fig. 8—Quartiles of error probability for slow signal fading as a function of γ; $\Delta fT = 1/2$, $\lambda = 0$, $E/N_0 = 10(1 + \gamma^2/2)$.

I_s and J_s are the inphase and quadrature components of the fading carrier. A Rayleigh fading model implies that I_s and J_s are independent zero-mean-value unit-variance Gaussian processes[4,10] with the same autocorrelation function $\psi(x)$. It follows that X and Y have zero mean and variance[10]

$$\sigma_s^2 = \frac{2}{T}\int_0^T \left(1 - \frac{x}{T}\right)\psi(x)\,dx$$

$$= \int_0^\infty G(f)\left[\frac{\sin \pi fT}{\pi fT}\right]^2 df,$$

where $G(f)$ is the spectral density corresponding to $\psi(x)$. When the fading carrier has a rectangular spectral density of twice its single-sided bandwidth, B, the spectral density $G(f)$ is

$$G(f) = \begin{cases} \dfrac{1}{B}; & 0 < f < B \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

It follows that

$$\sigma_s^2 = \frac{1}{B}\int_0^B \left[\frac{\sin \pi fT}{\pi fT}\right]^2 df$$

$$= \frac{\text{Si}\,(2\pi BT)}{\pi BT} - \frac{\sin^2 \pi BT}{(\pi BT)^2}$$

$$= h(BT). \tag{34}$$

For a "single-tuned" spectral density,[4]

$$G(f) = \frac{2B}{\pi(B^2 + f^2)}; \quad f \geq 0, \tag{35}$$

from which it follows that

---

[10] W. B. Davenport and W. L. Root, *Random Signals and Noise*, McGraw-Hill Inc., New York (1958).

$$\sigma_s^2 = h(BT) = \frac{1}{\pi BT}\left(1 - \frac{1 - \exp\{-2\pi BT\}}{2\pi BT}\right) \tag{36}$$

where B is the half-power-point bandwidth. For a Gaussian spectral density,[4]

$$G(f) = \frac{2}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{f^2}{2\sigma^2}\right\}; \quad f \geqq 0. \tag{37}$$

The standard deviation σ is

$$\sigma = \frac{B}{\sqrt{2 \ln 2}}. \tag{38}$$

The corresponding variance is

$$\sigma_s^2 = h(BT) = \frac{\sqrt{2 \ln 2}}{BT}\left[\frac{1}{\sqrt{2\pi}} H\left(\frac{2\pi BT}{\sqrt{2 \ln 2}}\right) - \frac{\sqrt{2 \ln 2}}{2\pi^2 BT}\left(1 - \exp\left\{-\frac{(\pi BT)^2}{\ln 2}\right\}\right)\right]$$

$$H(z) = \int_{-z}^{z} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt. \tag{39}$$

Figure 9 contains graphs of these three variance functions with BT as the independent variable. They are referred to as energy-loss functions since they are proportional to the mean coherent signal energy at the receiver. As the relative fading rate BT increases, the mean coherent signal energy approaches zero, since the effect of the signal fading is to destroy the coherence between the incoming signal and the corresponding stored signal at the receiver. For slow fading ($BT = 0$), there is no such loss of signal coherence.

Since I_s and J_s are independent normally distributed random variables with zero mean and variance $h = h(BT)$, it follows that

$(X^2 + Y^2)/h$ has a chi-square distribution[11,12] with two degrees of freedom. The corresponding distribution function is

$$F(z) = 1 - \exp\left\{\frac{-z}{2}\right\} \tag{40}$$

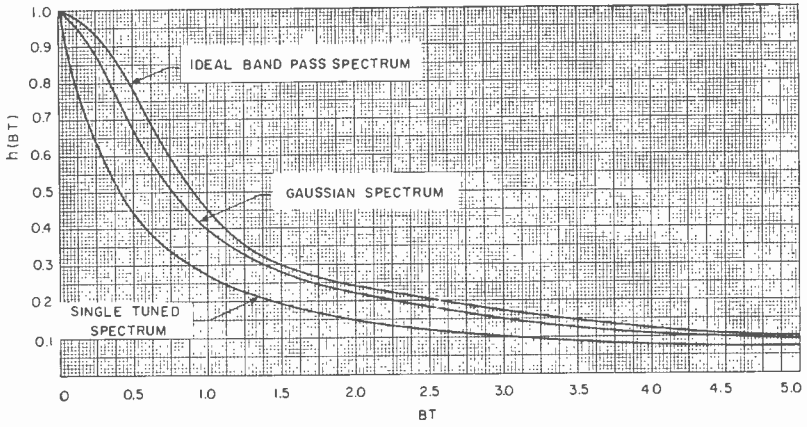

Fig. 9—Signal energy loss functions—quadrature detection system.

When $\lambda = 0$, it follows from Equations (17) and (32) that the distribution function of error probability is

$$F_{Pe}(x) = 1 - F_{n^2/2,\ \gamma/\sqrt{h}}\left[\frac{N_0}{hE}\ln\left(\frac{1}{2x}\right)\right] \tag{41}$$

The quantiles x_p can be computed from the formula

$$x_p = \frac{1}{2}\exp - \left\{\frac{hE}{N_0} z_{1-p}\right\}\ ; \quad \frac{E}{N_0} = \frac{<E>}{(2+\gamma^2)N_0}, \tag{42}$$

[11] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946.

[12] A. Hald, *Statistical Theory with Engineering Applications*, John Wiley and Sons, New York, 1952.

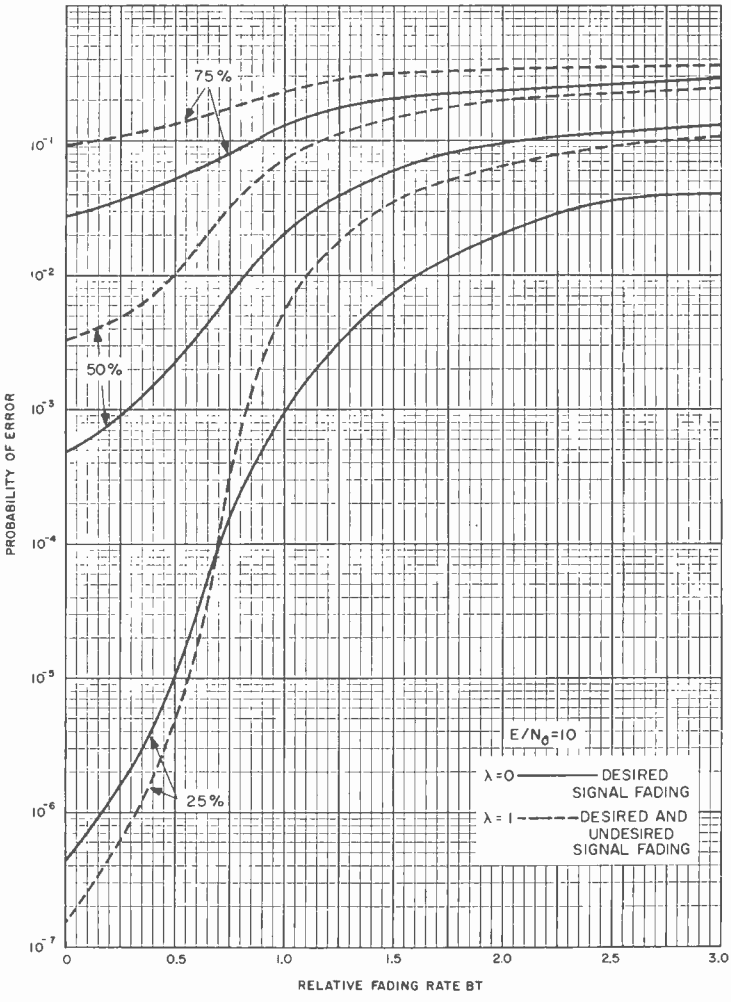

Fig. 10—Quartiles of probability of error for Rayleigh fading ($\gamma = 0$) as a function of relative fading rate BT; $E/N_0 = 10$.

where z_{1-p} represents the $1-p$ percentage point of the distribution function $F_{a^2/2,\ \gamma/\sqrt{h}}(z)$. Figure 10 shows the dependence of the quartiles of error probability on the relative fading rate BT for Rayleigh signal fading and $E/N_0 = 10$, assuming a rectangular fading spectrum. Note that the interquartile range, the difference between 75 and 25 per cent points, decreases with increasing BT. This is due to the decrease in the variance of β with increasing BT. The larger the relative fading rate BT, the more pronounced is this smoothing effect.

Figure 11 shows the dependence of the quartiles of error probability on the parameter γ for $BT = 1$, and $<E>/N_0 = 10$. As γ increases, the interquartile range approaches zero, and the quartiles approach

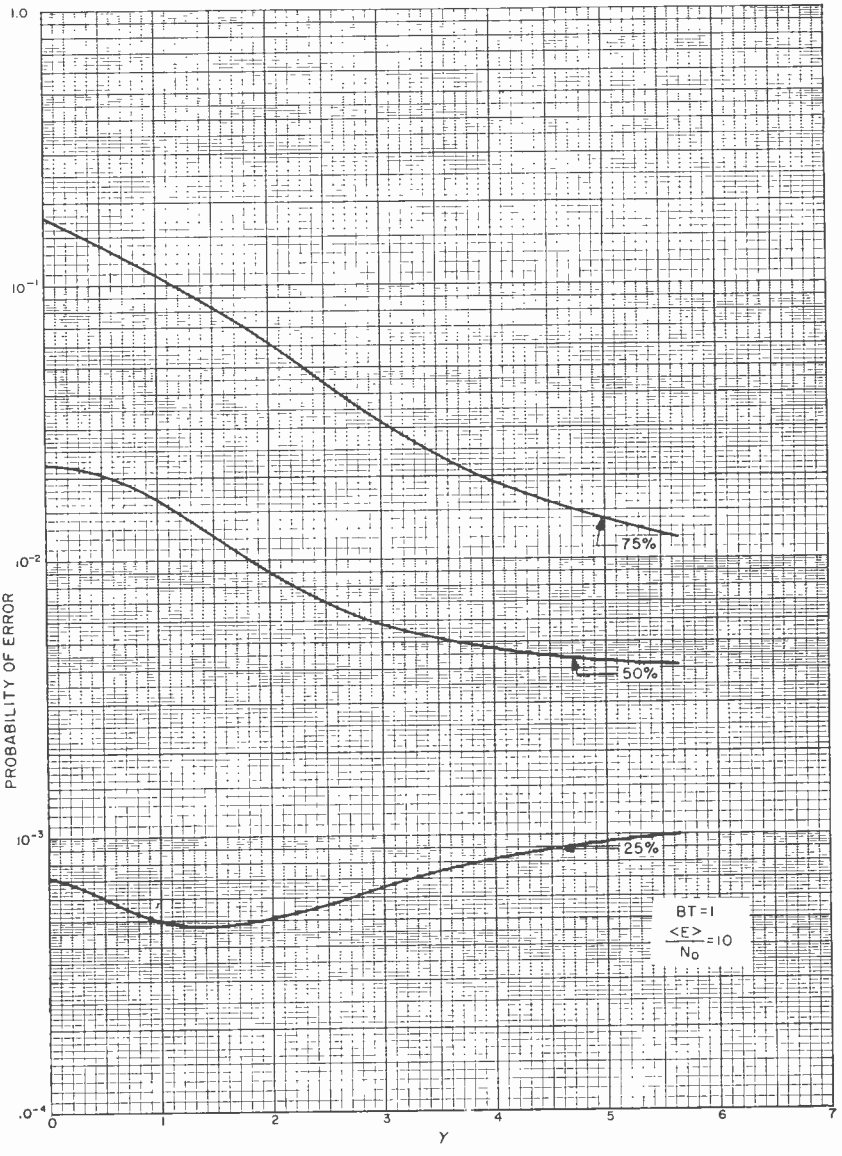

Fig. 11—Quartiles of probability of error as a function of γ for fading desired signals; $BT = 1$, $<E>/N_0 = 10$.

the error probability for no signal fading. The median probability of error is largest when $\gamma = 0$.

When $\lambda = 1$,

$$\beta = \frac{E}{N_0}\left[\frac{X^2 + Y^2}{Z^2}\right] \tag{43}$$

The random variable Z^2

$$Z^2 = \frac{1}{T}\int_0^T \left[I_n^2(t) + J_n^2(t)\right] dt, \tag{44}$$

where I_n and J_n are independent normalized Gaussian processes. The exact and approximate distributions of Z^2 have been studied in Reference (13). An approximate distribution can be based on an approximation of Z^2 for large BT by

$$Z^2 \cong \frac{1}{2BT}\sum_{k=1}^{2BT}\left[I_n^2\left(\frac{k}{2B}\right) + J_n^2\left(\frac{k}{2B}\right)\right]. \tag{45}$$

It follows that $2BZ^2$ has a chi-square distribution with $4BT$ degrees of freedom. It is shown in Reference (13) that this approximation is quite good for $2BT = 16/\pi$. However, in our problem this would only occur with relatively fast fading. For $2BT = 4/\pi$, Reference (13) shows good agreement at the tails, and in initial behavior, with disagreement in between. For the sake of obtaining some simple analytical results, this approximation will be retained. For $2BT \leq 1$, $2BZ^2$ will be approximated by a chi-square distribution with two degrees of freedom. (Note that when $BT = 0$, this is not an approximation.) Combining this approximation with the fact used earlier that $(X^2 + Y^2)/h$ has a chi-square distribution, implies that

[13] U. Grenander, O. Pollak, and D. Slepian, "The Distribution of Quadratic Forms in Normal Variates. A Small Sample Theory with Applications to Spectral Analysis," *Jour. Soc. Indust. Applied Math.*, Vol. 7, No. 4, Dec., 1959.

$$\beta = \frac{2BThE}{N_0}\left[\frac{\chi_2{}^2}{\chi^2{}_{4BT}}\right] \tag{46}$$

The notation $\chi^2{}_\nu$ refers to a random variable with a chi-square distribution with ν degrees of freedom; i.e., the distribution of a sum of squares of ν independent normally distributed variables each with zero mean and unit variance. The variable β can be expressed in terms of the random variable $F_{\nu 1,\,\nu 2}$, whose distribution function[11,12] has been extensively tabulated.[11] The definition of $F_{\nu 1,\,\nu 2}$ is

$$F_{\nu 1,\,\nu 2} = \frac{\dfrac{\chi^2{}_{\nu 1}}{\nu_1}}{\dfrac{\chi^2{}_{\nu 2}}{\nu_2}}, \tag{47}$$

the ratio of two independent chi-square variables which have been divided by their respective numbers of degrees of freedom. Since

$$F_{2,\,4BT} = 2BT\,\frac{\chi^2{}_2}{\chi^2{}_{4BT}}, \tag{48}$$

it follows that

$$\beta = \left(\frac{Eh}{N_0}\right) F_{2,4BT}. \tag{49}$$

Applying Equation (17), the distribution function of error probability is

$$F_{Pe}(x) = P\left[F_{2,4BT} \geqq \frac{2N_0}{hE}\ln\frac{1}{2x}\right]. \tag{50}$$

The desired error probability distribution can therefore be computed from the distribution function tables of $F_{\nu 1,\,\nu 2}$. The quantiles of error probability x_p, defined in Equation (18), can be calculated from the formula

$$x_p = \frac{1}{2} \exp \left\{ -\frac{hE}{2N_0} F(2,4BT,p) \right\} ; \quad \frac{E}{N_0} = \frac{<E>}{2N_0} . \tag{51}$$

The function $F(\nu_1, \nu_2, p)$ is tabulated in Reference (14). It refers to the "significance levels" of the distribution function of $F_{\nu 1, \nu 2}$ defined by

$$P[F_{\nu 1, \nu 2} \geq F(\nu_1, \nu_2, p)] = p. \tag{52}$$

The curves in Figure 10 for $\lambda = 1$ calculated from Equation (51) show how the quartiles of error probability depend upon the relative fading rate BT for $E/N_0 = 10$. Comparison of the curves for $\lambda = 0$ and $\lambda = 1$ shows that the error probabilities tend to be larger when both the signals and noise are fading ($\lambda = 1$) than when only the signal is fading ($\lambda = 0$). Equation (51) is in agreement with Equation (25) since it can easily be shown that

$$F(2, 2, p) = \frac{1}{p} - 1 ; \tag{53}$$

also, $h = 1$ for slow fading ($BT = 0$). This agreement is due to the fact that when $BT = 0$, no approximation is involved in applying the F distribution. Figure 11 illustrates the dependence of the quartiles on the parameter γ for $BT = 1$. As γ increases from zero, the spread of the distribution reaches a maximum, and then approaches zero as γ becomes infinite.

Mean Probability of Error

The mean probability of error function can be calculated by averaging the conditional probability of error function as defined by Equations (10) and (11) over the probability distributions of the fading.

Signal Only Fading ($\lambda = 0$)

Applying Equations (15) and (32) and the fact that $(X^2 + Y^2)/h$ has the distribution function $F_{a^2/2,\gamma}(z)$ defined in Equation (27), the mean error probability is

[14] L. E. Vogler and K. A. Norton, "Graphs and Tables of the Significance levels $F(\nu_1, \nu_2, p)$ for the Fisher-Snedecor Variance Ratio," National Bureau of Standards, Boulder Laboratories Report No. 5069.

$$\overline{P_e} = \frac{1}{2} \exp \overline{\left\{ -\frac{1}{2} \frac{hE}{N_0} \left(\frac{X^2 + Y^2}{h} \right) \right\}}$$

$$= \frac{1}{2} \int_0^\infty \exp \left\{ -\frac{1}{2} \frac{hEz}{N_0} - z - \frac{\gamma^2}{2h} \right\} I_0 \left(\frac{\gamma}{\sqrt{h}} \sqrt{2z} \right) dz$$

$$= \exp \left\{ -\left(\frac{\gamma^2}{2(2+\gamma^2)} \right) \frac{\frac{<E>}{N_0}}{2 + \frac{h<E>}{N_0(2+\gamma^2)}} \right\} \left[2 + \frac{h<E>}{N_0(2+\gamma^2)} \right]^{-1}$$

$$\frac{<E>}{N_0} = \frac{E}{N_0} (2 + \gamma^2). \tag{54}$$

The energy loss function $h = h(BT)$ approaches zero with increasing BT, in which case the mean error probability approaches

$$\overline{P_e} = \frac{1}{2} \exp - \left\{ \frac{\gamma^2}{4(2+\gamma^2)} \left(\frac{<E>}{N_0} \right) \right\}. \tag{55}$$

The presence of the constant received signal component can keep the error probability small. In the Rayleigh fading case, where $\gamma = 0$, the error probability is 1/2.

For slow fading ($BT = 0$), $h = 1$ in Equation (54), so that Equation (54) also applies to slow signal fading[6] with arbitrary γ. For slow Rayleigh fading, Equation (54) reduces to the formula given by Turin for noncoherent binary FSK[6] (see Equation (38) in Reference (6)).

Figure 12 shows graphs of the mean error probability as a function of 10 $\log_{10} E/N_0$ for $BT = 0, 1, 2$ for Rayleigh fading ($\gamma = 0$) as calculated from Equation (54). Similar curves can easily be plotted for particular values of γ.

Signal and Noise Rayleigh Fading ($\lambda = 1$)

Applying Equation (49), the mean error probability is approximately

$$\overline{P_e} = \frac{1}{2} \overline{\exp\left\{-\frac{Eh}{2N_0} F_{2,4BT}\right\}} \tag{56}$$

The probability density of $(2/\nu)\ F_{2,\nu}$, given in Equation (18.3.2) of Reference (11) is

$$g_{2,\nu}(x) = \frac{\nu}{2} (1 + x)^{-(\nu/2)-1};\ x > 0. \tag{57}$$

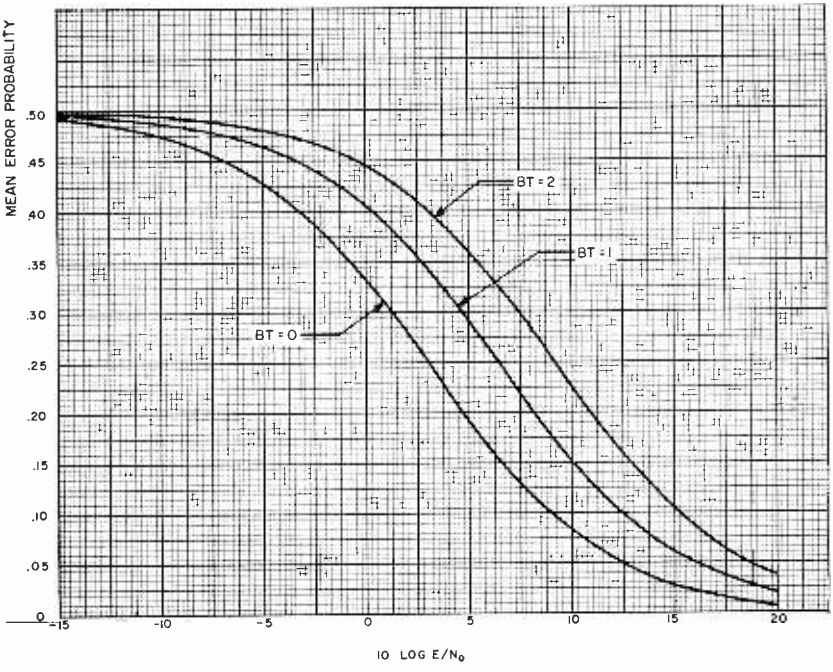

Fig. 12—Mean probability of error as a function of $10 \log_{10}(E/N_0)$ for Rayleigh signal fading.

Recalling the F distribution approximation, in computing $\overline{P_e}$, $\nu = 2$ for $BT \leq 1/2$ and $\nu = 4BT$ for $BT \geq 1/2$. With this definition, it follows that

$$\overline{P_e} = \frac{1}{2} \int_0^\infty \exp\left\{\frac{-Ehvx}{4N_0}\right\} g_{2,\nu}(x)\, dx$$

$$= \begin{cases} BT \exp\left\{\dfrac{hEBT}{N_0}\right\} E_{2BT+1}\left(\dfrac{hEBT}{N_0}\right); & BT \geqq \dfrac{1}{2}, \\ \dfrac{1}{2} \exp\left\{\dfrac{hE}{2N_0}\right\} E_2\left(\dfrac{hE}{2N_0}\right) & ; \; BT \leqq \dfrac{1}{2}, \end{cases} \tag{58}$$

where

$$E_n(x) = \int_1^\infty e^{-xt} \frac{dt}{t^n} \tag{59}$$

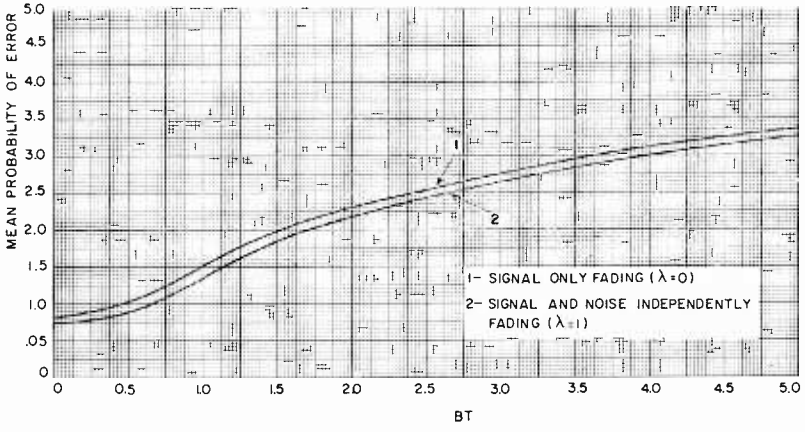

Fig. 13—Comparison of mean probability of error as a function of relative fading rate BT for Rayleigh signal and Rayleigh signal and noise fading, $E/N_0 = 10$.

is tabulated in References (15) and (16).

Figure 13 shows graphs of the mean error probability as a function of BT for $\lambda = 0, 1$, assuming Rayleigh fading and a mean signal-energy-to-noise density of 10. As was observed for slow fading, the error probabilities for $\lambda = 0$ and $\lambda = 1$ do not differ appreciably from each other. Furthermore, when the noise is also fading ($\lambda = 1$), the error probabilities are somewhat smaller than when only the signal is fading ($\lambda = 0$). Figure 14 compares the mean and median error probabilities

[15] *Tables of Functions and of Zeros of Functions*, National Bureau of Standards, Applied Mathematics Series No. 37 (1959).

[16] V. I. Pagurova, "Tables of the Exponential Integral Function," Computing Center, *U.S.S.R. Academy of Sciences M.* (1959).

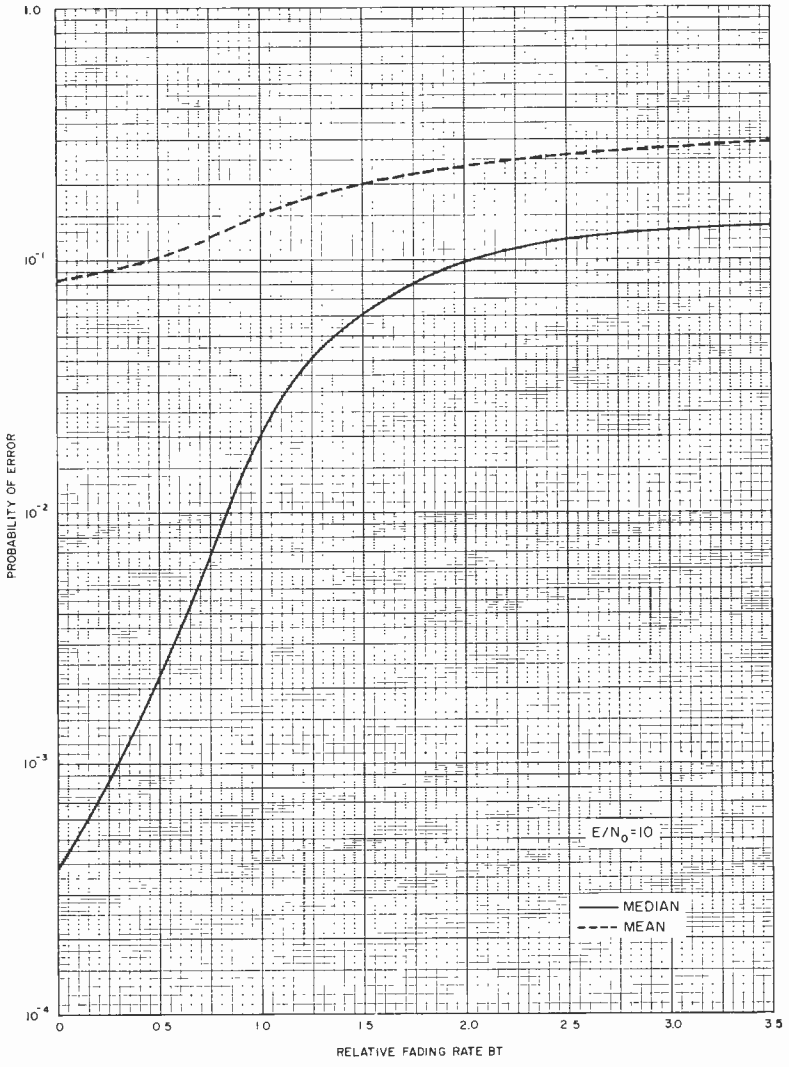

Fig. 14—Comparison of mean and median probabilities of error as a function of BT for $\lambda = 0$, $E/N_0 = 10$.

as a function of BT for $\lambda = 0$. For small BT, there is a substantial difference; the median is about two decades smaller. As BT increases, however, the difference becomes relatively small. Figure 15 compares the mean and the median error probabilities as a function of signal energy to noise density in decibels for $BT = 0$, 1. The separation between the mean and the median increases with increasing signal-to-

noise ratio, since the mean decreases inversely with increasing signal-to-noise ratio whereas the median decreases exponentially. Figure 16 compares the mean probabilities of error as a function of BT for the ideal bandpass, Gaussian, and single-tuned fading spectra (Figure 9 shows graphs of the energy loss functions for these spectra). As the fading rate increases, there is a constant displacement between the curves for the three spectra. The poorest performance occurs for the single-tuned spectrum, the best for the ideal bandpass spectrum, with the Gaussian spectrum in between. Since all the other plotted curves which take into account effects of fading rate assumed the rectangular spectrum, these curves are optimistic compared with expected curves for the other spectra. This conclusion is evident from Figure 9, if one notes the ordering of the energy-loss functions for the three spectra.

A comparison of the mean as a function of BT for $\lambda = 0$ and $\lambda = 1$ in Figure 13 shows the mean to be larger for $\lambda = 0$ than for $\lambda = 1$. This is opposite from the result obtained in comparing the medians in Figure 10. Evidently the increased spread of error probability for $\lambda = 1$ compared with $\lambda = 0$, has different effects on the mean and the median. Of course, the existence of this greater spread is not evident from the mean.

Acknowledgment

Appreciation is expressed to Mrs. Ewaugh Fields, who did the numerical and graphical work.

Appendix A—Derivation of the Conditional Error Probability Function

Applying Equation (6), and the assumption that $S(t)$ and $S^*(t)$ can only assume the values ± 1, the correlator outputs can be written

$$\begin{aligned} M &= m + \epsilon, \\ -\hat{M} &= \hat{m} + \hat{\epsilon}, \end{aligned} \tag{60}$$

$$\begin{aligned} M^* &= m^* + \epsilon^*, \\ -\hat{M}^* &= \hat{m}^* + \hat{\epsilon}^*, \end{aligned} \tag{61}$$

where

$$m = A \int_0^T a_s(t) \cos \left[\phi_s(t) + \theta(t)\right] dt,$$

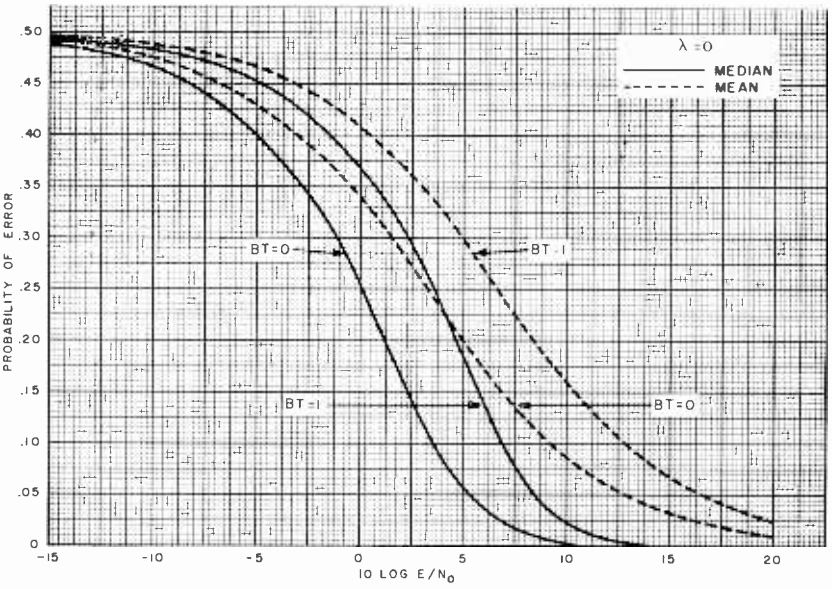

Fig. 15—Comparison of mean and median probabilities of error as a function of $10 \log_{10}(E/N_0)$ for $\lambda = 0$, $BT = 0$, 1.

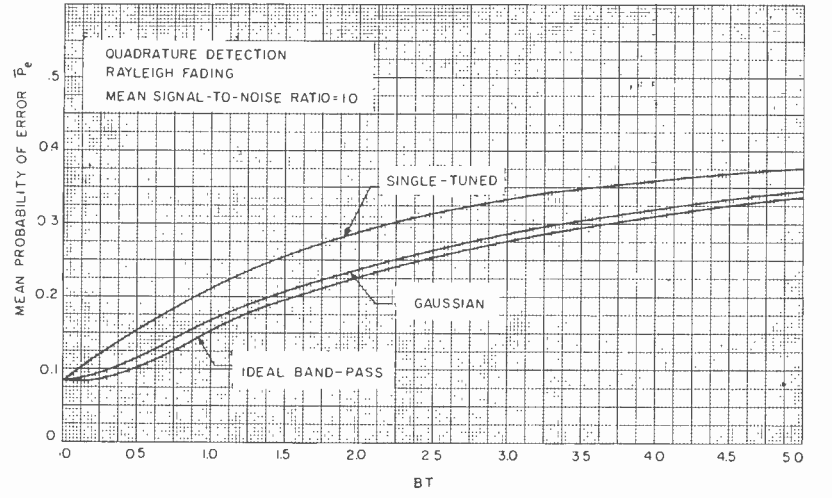

Fig. 16—Comparison of mean probabilities of error as a function of BT for three Rayleigh signal fading spectra; $\langle E \rangle / N_0 = 10$, $\lambda = 0$.

(62)

$$\hat{m} = A \int_0^T a_s(t) \sin [\phi_s(t) + \theta(t)]dt,$$

$$m^* = A \int_0^T S(t)S^*(t)\, a_s(t) \cos [\phi_s(t) + \theta(t)]dt,$$

(63)

$$\hat{m}^* = A \int_0^T S(t)S^*(t)\, a_s(t) \sin [\phi_s(t) + \theta(t)]dt,$$

and

$$\epsilon = \int_0^T a_n(t)\, R_1(t) \cos [\phi_1(t) + \phi_n(t)]S(t)\,dt + \int_0^T R_2(t) \cos [\phi_2(t)]S(t)\,dt,$$

(64)

$$\hat{\epsilon} = \int_0^T a_n(t)\, R_1(t) \sin [\phi_1(t) + \phi_n(t)]S(t)\,dt + \int_0^T R_2(t) \sin [\phi_2(t)]S(t)\,dt,$$

$$\epsilon^* = \int_0^T a_n(t)\, R_1(t) \cos [\phi_1(t) + \phi_n(t)]S^*(t)\,dt + \int_0^T R_2(t) \cos [\phi_2(t)]S^*(t)\,dt,$$

(65)

$$\hat{\epsilon}^* = \int_0^T a_n(t)\, R_1(t) \cos\left[\phi_1(t) + \phi_n(t)\right] S^*(t)\, dt + \int_0^T R_2(t) \sin\left[\phi_2(t)\right] S^*(t)\, dt.$$

As long as the additive noises $\eta_1(t)$ and $\eta_2(t)$ have no d-c term, the conditional mean values of ϵ, $\hat{\epsilon}$, ϵ^* and $\hat{\epsilon}^*$ are all zero. The conditional variances of the ϵ's for given a_n and ϕ_n are calculated here. Since the $2WT$ product has been assumed to be large ($2W$ is the r-f signal transmission bandwidth), each of the integrals in Equation (65) can be approximated by the sampling theorem. For example,

$$\epsilon \cong \frac{1}{2W} \sum_{k=1}^{2WT} \left\{ a_n\left(\frac{k}{2W}\right) R_1\left(\frac{k}{2W}\right) \cos\left[\phi_1\left(\frac{k}{2W}\right) + \phi_n\left(\frac{k}{2W}\right)\right] S\left(\frac{k}{2W}\right) + R_2\left(\frac{k}{2W}\right) \cos\left[\phi_2\left(\frac{k}{2W}\right)\right] S\left(\frac{k}{2W}\right)\right\} \quad (66)$$

Since the fading and nonfading noise functions have been assumed to be statistically independent, it follows that

$$\sigma^2(\epsilon) \cong \frac{1}{(2W)^2} \sum_{k=1}^{2WT} \left[a_n^2\left(\frac{k}{2W}\right) S^2\left(\frac{k}{2W}\right) 2W\lambda N_0 + 2W(1-\lambda) N_0 S^2\left(\frac{k}{2W}\right)\right], \quad (67)$$

since $R_1 \cos\phi_1$ and $R_1 \sin\phi_1$ are independent with variance $2WN_0$, whereas $R_2 \cos\phi_2$ and $R_2 \sin\phi_2$ are independent with variance $2W(1-\lambda)N_0$. Approximating the summation in Equation (67) by an integral gives

$$\sigma^2 \cong \sigma^2(\epsilon) = \lambda N_0 \int_0^T a_n^2(t)\, S^2(t)\, dt + (1-\lambda) N_0 T. \tag{68}$$

By applying the same process to the variance of $\hat{\epsilon}$, it is found that $\sigma^2(\hat{\epsilon}) = \sigma^2$. Correspondingly,

$$\sigma^2(\epsilon^*) = \sigma^2(\hat{\epsilon}^*) \cong \lambda N_0 \int_0^T a_n^2(t)\, S^{*2}(t)\, dt + (1-\lambda) N_0 T. \tag{69a}$$

If we make the additional assumption that $S(t)$ and $S^*(t)$ are binary signals assuming the values ± 1, it follows that the ϵ's have the common variance

$$\sigma^2 \cong \lambda N_0 \int_0^T a_n^2(t)\, dt + (1-\lambda) N_0 T, \tag{69b}$$

a quantity independent of the actual signal elements and of the phase fluctuations due to fading $\phi_n(t)$.

It will now be shown that the ϵ's are mutually uncorrelated. The covariance between ϵ and $\hat{\epsilon}$ reduces to the mean of the product of the first integrals in Equation (64). The other cross products average to zero since the two additive noise functions have been assumed to be zero. Applying the relations

$$\begin{aligned} R_1 \cos(\phi_1 + \phi_n) &= I_{c1} \cos\phi_n - I_{s1} \sin\phi_n \\ R_1 \sin(\phi_1 + \phi_n) &= I_{s1} \cos\phi_n + I_{c1} \sin\phi_n \end{aligned} \tag{70a}$$

where I_{c1} and I_{s1} are independent in-phase and quadrature components of the noise which is subject to fading, leads to the covariance

$$\sigma(\epsilon, \hat{\epsilon}) = \int_0^T dt \int_0^T dt' a_n(t) a_n(t')\, S(t) S(t') \psi(t-t') \sin[\phi_n(t') - \phi_n(t)] \tag{70b}$$

where $\psi(t-t')$ is the autocorrelation function of the noise inphase and quadrature components. This covariance is evidently small, since for small values of $t-t'$, the sine term goes to zero for relatively slow phase variations $\phi_n(t)$, whereas for large values of $t-t'$, the autocorrelation function tends to be small. Also the signal products tend to cancel each other out for values of $t-t'$ greater than T_0, the signal code keying interval. If we average this covariance over all signal codes, this average goes to zero since the mean of $S(t)S(t')$ is nonzero only for $|t-t'|$ less than T_0, and since $\phi_n(t)-\phi_n(t')$ is assumed to be constant over this range. In the same manner the variance of the covariance averaged over all signal codes can be shown to be negligible. The same type of reasoning can be used to show that the other covariances are negligible. For the covariances between the starred and unstarred ϵ's, we have the additional factor that the mark and space signals S and S^* are assumed to be orthogonal. A simpler approach, however, is to invoke the sampling theorem.[17] For example, in the case of the covariance of ϵ and $\hat{\epsilon}$,

$$\sigma(\epsilon,\hat{\epsilon}) \cong \frac{1}{(2W)^2}\left\{\sum_{k=1}^{2WT}\sum_{l=1}^{2WT} a_{nk}a_{nl}S_kS_l \sin(\phi_{nl}-\phi_{nk})\,\overline{I_{ck}I_{cl}}\right\} = 0, \tag{71}$$

since the noise samples I_{ck} and I_{cl} are mutually uncorrelated for $k \neq l$. For $k=l$, the sine term drops out. In the case of the covariance between ϵ and ϵ^*, we have

$$\begin{aligned}\sigma(\epsilon,\epsilon^*) &= \frac{1}{(2W)^2}\sum_{k=1}^{2WT}\sum_{l=1}^{2WT} S_kS_l^*\,\overline{I_{c2}\left(\frac{k}{2W}\right)I_{c2}\left(\frac{l}{2W}\right)} \\ &= \frac{(1-\lambda)N_0}{2W}\sum_{k=1}^{2WT} S_kS_k^* \\ &= 0,\end{aligned} \tag{72}$$

since the mark and space signals are orthogonal. The same result occurs with the covariances of the other ϵ's.

It follows from the above, and the definitions of the M's in Equations (60) and (61) that they are mutually independent normal random vari-

[17] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, McGraw-Hill Inc., N. Y., 1953.

ables. It follows from this and the definition of the mark and space quadrature detector outputs in Equation (9) that the probability densities of Q/σ and Q^*/σ are[4]

$$p(y) = y \exp\left\{-\frac{1}{2}(y^2 + 2\beta)\right\} I_0(y\sqrt{2\beta}),$$
$$p^*(y) = y \exp\left\{-\frac{1}{2}(y^2 + 2\delta)\right\} I_0(y\sqrt{2\delta}), \tag{73}$$

where

$$\beta = \frac{m^2 + \hat{m}^2}{2\sigma^2}.$$
$$\delta = \frac{m^{*2} + \hat{m}^{*2}}{2\sigma^2}. \tag{74}$$

These quantities have been defined in Equations (11) through (14).

In order to simplify the required error probability calculation, the probability density $p^*(y)$ can be approximated assuming small values of δ. That this approximation is reasonable can be seen by examining the definitions of m^* and $\hat{m}^*$ in Equation (63). If the envelope and phase tend to be constant during a time interval of duration T, the orthogonality of $S(t)$ and $S^*(t)$ will tend to make the integrals in Equation (63) close to zero. If the envelope and phase vary rapidly during a time interval of duration T, the integrals should also tend to be small due to the sine and cosine components which are being integrated. It is shown in Appendix B that δ can be assumed to be small for large bandwidth–time products and small receiver input signal-to-noise ratio.

For small δ, $p^*(y)$ can be approximated by the Rayleigh distribution

$$p^*(y) = \frac{y}{1+\delta} \exp\left\{-\frac{y^2}{2(1+\delta)}\right\}. \tag{75}$$

The probability of correct reception is

$$P_c = P[Q^* < Q]$$
$$= \int_0^\infty dy\, p(y) \int_0^y p^*(x)\, dx. \tag{76}$$

Applying Equation (75) to Equation (76), and integrating first with respect to x and then with respect to y,

$$P_e = \frac{\exp\left\{-\dfrac{\beta}{2+\delta}\right\}}{1+\dfrac{1}{1+\delta}}. \tag{77}$$

This is the result given in Equation (10).

Appendix B—Evaluation of the Parameter δ

It is shown below that for small input signal-to-noise power ratios, there is a high probability that the parameter δ will be small compared to unity.

From Equation (11), δ is proportional to the sum of squares of X^* and Y^* where

$$X^* = \frac{1}{T}\int_0^T S(t)S^*(t)a_s(t)\cos\left[\phi_s(t)+\theta(t)\right]dt,$$

$$Y^* = \frac{1}{T}\int_0^T S(t)S^*(t)a_s(t)\sin\left[\phi_s(t)+\theta(t)\right]dt. \tag{78}$$

Assume that the mark and space signals $S(t)$ and $S^*(t)$ are based on binary digital codes

$$S(t) = \sum_{k=1}^{n} a_k f[t-(k-1)T_0],$$

$$S^*(t) = \sum_{k=1}^{n} b_k f[t-(k-1)T_0], \tag{79}$$

where n is the number of code elements in time T, a_k and b_k are independent mark and space code elements equal to ± 1, T_0 is the signal keying interval ($T = nT_0$), and

$$f(x) = \begin{cases} 1; & 0 < x < T_0 \\ 0 & \text{otherwise.} \end{cases} \tag{80}$$

Under the signal orthogonality assumption

$$\sum_{k=1}^{n} a_k b_k = 0. \tag{81}$$

Applying Equation (79) to Equation (78),

$$X^* = \frac{1}{n} \sum_{k=1}^{n} a_k b_k \frac{1}{T_0} \int_{(k-1)T_0}^{kT_0} a_s(t) \cos(\phi_s(t) + \theta(t))\, dt$$

$$Y^* = \frac{1}{n} \sum_{k=1}^{n} a_k b_k \frac{1}{T_0} \int_{(k-1)T_0}^{kT_0} a_s(t) \sin(\phi_s(t) + \theta(t))\, dt. \tag{82}$$

In the slow-signal-fading case ($BT = 0$), assuming

$$\theta(t) = 2\pi\Delta f t + \theta_0 \tag{83}$$

implies that

$$X^* = \frac{a_s}{n} \sum_{k=1}^{n} a_k b_k \frac{1}{\pi\Delta f T_0} \sin \pi\Delta f T_0 \cos[\phi_s + \theta_0 + \pi f T_0 (2k-1)],$$

$$Y^* = \frac{a_s}{n} \sum_{k=1}^{n} a_k b_k \frac{1}{\pi\Delta f T_0} \sin \pi\Delta f T_0 \sin[\phi_s + \theta_0 + \pi\Delta f T_0 (2k-1)], \tag{84}$$

$$(X^*)^2 + (Y^*)^2 = \frac{a_s^2}{n^2} \left[\frac{\sin \pi\Delta f T_0}{\pi\Delta f T_0} \right]^2 \sum_{k=1}^{n} \sum_{l=1}^{n} a_k b_k a_l b_l \cos[2\pi\Delta f T_0 (k-l)]. \tag{85}$$

As long as $\Delta f T_0$ is very small, it follows from Equation (85) and the orthogonality assumption in Equation (81) that δ is approximately zero. However, if $\Delta f T_0$ is not zero and if random codes are used which are not exactly orthogonal, it is useful to estimate the magnitude of δ.

If a mean value of $X^{*2} + Y^{*2}$ is taken by averaging over an ensemble of codes generated by means of Bernoulli trials, we have

$$\overline{(X^*)^2 + (Y^*)^2} = \frac{a_s^2}{n}\left[\frac{\sin \pi \Delta f T_0}{\pi \Delta f T_0}\right]^2, \tag{86}$$

since when $k = l$ the code elements are mutually independent with zero means. It follows from Equation (11) and the above, that the mean of δ is

$$\bar{\delta} = \frac{E a_s^2(t)}{n N_0}\left[\frac{\sin \pi \Delta f T_0}{\pi \Delta f T_0}\right]^2 [\lambda a_n^2 + (1-\lambda)]^{-1} \tag{87}$$

for slow fading and frequency instability. Since

$$\frac{E}{nN_0} = \frac{PT}{nN_0} = \frac{PT_0}{N_0} = \frac{P}{N}, \tag{88}$$

where P is the average signal power, N is the average noise power,

$$N = 2WN_0 = \frac{N_0}{T_0}, \tag{89}$$

and $2W$ is the signal transmission bandwidth, it follows that $\bar{\delta}$ is a random variable proportional to the input signal-to-noise power ratio P/N. Therefore for small input signal-to-noise ratios, there is a high probability that δ will be small compared to unity. For example, when $\lambda = 1$,

$$\delta = \frac{P}{N}\left[\frac{\sin \pi \Delta f T_0}{\pi \Delta f T_0}\right]^2 F_{2,2}. \tag{90}$$

The random variable $F_{2,2}$ has been defined by Equation (57). It has probability density

$$g_{2,2}(x) = \frac{1}{(1+x)^2}; \quad x \geq 0. \tag{91}$$

It follows that

$$q = P[\bar{\delta} > \epsilon] = \left\{1 + \frac{\epsilon}{\frac{P}{N}\left[\frac{\sin \pi \Delta f T_0}{\pi \Delta f T_0}\right]^2}\right\}^{-1}. \tag{92}$$

Solving for P/N in this equation, and assuming $1/q$ to be large compared with unity,

$$\frac{P}{N} = \frac{q\epsilon}{\left[\dfrac{\sin \pi \Delta f T_0}{\pi \Delta f T_0}\right]^2}. \tag{93}$$

This equation can be used to determine the value of input signal-to-noise ratio for which $\bar{\delta}$ exceeds ϵ with probability q. For example, $q = 10^{-2}$, $\epsilon = 10^{-1}$, $\Delta f T_0 = 0$ implies $P/N = 10^{-3}$. For $q = 10^{-2}$, $\epsilon = 1$, $\Delta f T = 0$, $P/N = 10^{-2}$.

Performing a similar calculation for $\lambda = 0$,

$$q = P[\bar{\delta} > \epsilon] = \exp\left\{ - \frac{\epsilon}{\dfrac{2P}{N}\left[\dfrac{\sin \pi \Delta f T_0}{\pi \Delta f T_0}\right]^2} \right\} \tag{94}$$

Solving for P/N,

$$\frac{P}{N} = \frac{\epsilon}{2 \ln\left(\dfrac{1}{q}\right)\left[\dfrac{\sin \pi \Delta f T_0}{\pi \Delta f T_0}\right]^2}. \tag{95}$$

For example, $q = 10^{-2}$, $\epsilon = 10^{-1}$, $\Delta f T_0 = 0$ implies that $P/N \approx 0.02$. Evidently, the assumption of small δ is more easily fulfilled when $\lambda = 0$ than when $\lambda = 1$.

It is not enough to consider only the mean $\bar{\delta}$. It is also necessary to consider the standard deviation of δ, σ_δ. It follows from Equation (85) and the definition of δ in Equation (11) that, for slow fading,

$$\sigma_\delta = \frac{\dfrac{P}{N} a_s^2}{\lambda a_d^2 + (1-\lambda)} \left[\frac{\sin \pi \Delta f T_0}{\pi \Delta f T_0}\right]^2 \frac{1}{n^2} \left[\sum_{k=1}^{n} \sum_{l=1}^{n} \cos^2 \left[2\pi \Delta f T_0 (k-l)\right]\right]^{1/2}. \tag{96}$$

An upper bound for this standard deviation can be obtained by assuming that the cosine-square terms all equal unity, from which it can be shown that

$$\sigma_\delta < \bar{\delta}. \tag{97}$$

Therefore the above conditions for $\bar{\delta}$ to be small also apply for σ_δ to be small.

For the case of fading with arbitrary fading rate and no frequency instability,

$$(X^*)^2 + (Y^*)^2 = \frac{1}{T^2} \sum_{k=1}^{n} \sum_{l=1}^{n} a_k a_l b_k b_l \left[\int_{(k-1)T_0}^{kT_0} dt \int_{(l-1)T_0}^{lT_0} dt'\, a_s(t) a_s(t') \cos\left(\phi_s(t) - \phi_s(t')\right) \right] \tag{98}$$

Taking the mean value of δ by averaging over all signal codes, and assuming that the signal phase $\phi_s(t)$ does not change over the keying interval T_0, leads to the expression

$$\bar{\delta} \cong \frac{P}{N} \frac{\frac{1}{T} \int_0^T a_n^2(t)\, dt}{\left[\lambda \left(\frac{1}{T} \int_0^T a_s^2(t)\, dt \right) + 1 - \lambda \right]}. \tag{99}$$

An analysis similar to that done for slow fading could be done here. for arbitrary BT. However, an important point is that the actual values of δ will be less important as the BT product increases than for $BT = 0$ (slow fading). This is due to the fact that the variance of δ goes to zero with increasing BT since the time averages in Equation (99) become closer to their ensemble averages. Therefore the previous analysis for slow fading corresponds to the worst case insofar as the contributions of δ are concerned. This is not in agreement with the results for strictly orthogonal signals where, in the slow fading case, δ actually equals zero. This suggests the likelihood that the assumption of Bernoulli trial signals leads to larger values of δ, as compared with orthogonal signals. This occurs since the lack of orthogonality leads to terms which are products of the input signal-to-noise ratio and a random variable due to the fading, which can become large.

STABILITY OF TUNNEL-DIODE OSCILLATORS

By

F. Sterzer

RCA Electron Tube Division
Princeton, N. J.

Summary—*This paper discusses the factors which determine the dependence of the frequency of tunnel-diode oscillators on ambient temperature, d-c supply voltage, and r-f load. It is shown that fractional frequency shifts of less than* $3 \times 10^{-6}/°C$, $2 \times 10^{-6}/mv$, *and* 1×10^{-4} *for a change in load VSWR from 1 to 2 can be readily achieved. Frequency stabilization by means of fundamental and subharmonic locking signals is discussed. It is shown that the locking signals can be orders of magnitude smaller than the output of the oscillator, provided the frequency of the signals or their harmonics is close to the natural frequency of the oscillator.*

Introduction

TUNNEL-DIODE OSCILLATORS are compact r-f generators that have modest power-supply requirements and can be readily tuned by electrical or mechanical means.[1,2] The power output of these oscillators, which ranges from 100 milliwatts at 1.6 kmc to 4 milliwatts at 6 kmc and 2 microwatts at 90 kmc, is sufficient for many practical applications,[3-5] especially at frequencies up to 6 kmc.

Every application of tunnel-diode oscillators requires a certain degree of frequency stability; however, up to now little information has been published on this subject. The present paper discusses the dependence of tunnel-diode-oscillator frequency on ambient temperature, d-c supply voltage, and r-f load.

---

[1] H. S. Sommers, Jr., "Tunnel Diodes as High Frequency Devices," *Proc. I.R.E.*, Vol. 47, p. 1201, July 1959.

[2] F. Sterzer and D. E. Nelson, "Tunnel Diode Microwave Oscillators," *Proc. I.R.E.*, Vol. 49, p. 744, April 1961.

[3] A. Presser, RCA Electron Tube Division, Princeton, N. J., private communications. This oscillator utilizes several diodes in parallel.

[4] W. B. Hauer, "A 4 mw, 6 kMc Tunnel-Diode Oscillator," Digest of Technical Papers, *1962 International Solid State Circuits Conference*, Philadelphia, Pa., Feb. 1962.

[5] C. A. Burrus, "Millimeter Wave Esaki Diode Oscillator," *Proc. I.R.E.*, Vol. 48, p. 2024, Dec. 1960.

Equivalent Circuit of a Tunnel Diode

An approximate a-c equivalent circuit for an encapsulated tunnel diode consists of three elements connected in series: an inductance L_d, a resistance r_d, and a voltage-dependent resistance $R_d(v)$, shunted by a voltage-dependent capacitance, $C_d(v)$. The variation of R_d with voltage can be determined from the slope of the current–voltage characteristics of the diode,[2] and $C_d(v)$ may be approximated by

$$C_d(v) \approx K(\phi - v)^{-1/2}, \tag{1}$$

where K and ϕ are constants. The value of ϕ is 0.6 volt for germanium diodes and 1.1 volts for gallium arsenide diodes. Equation (1) is applicable only for voltages less than the valley voltage of the diode.

Changes in the frequency of tunnel-diode oscillators resulting from temperature variations of the parameters of the diode itself are caused primarily by variations in R_d, because it is the only one of the four tunnel-diode parameters which depends to an important extent on temperature. The variation of R_d with temperature is best described in terms of the temperature variation of the current–voltage characteristics of the diode. For oscillator applications, the region of the current–voltage characteristic of greatest interest is bounded by the peak and valley points. It has been shown[6-9] that (1) the peak voltage is substantially temperature independent, (2) the valley voltage generally decreases with increasing temperature, (3) the peak current may either increase or decrease with temperature, and (4) the valley current generally decreases with decreasing temperature. Little information has been published on the temperature dependence of the region between the peak and valley points, but it appears that the transition from peak to valley is smooth at all temperatures, and is generally of the shape shown in Figure 1.

Because the temperature dependence of the peak current is a function of carrier concentration, variations in peak current can be

---

[6] L. Esaki, "New Phenomenon in Narrow Ge P-N Junctions," *Phys. Rev.*, Vol. 109, p. 603, Jan. 1958.

[7] L. Esaki and Y. Miyahara, "A New Device Using the Tunneling Process in Narrow P-N Junctions," *Solid-State Electronics*, Vol. 1, p. 13, Jan. 1960.

[8] I. A. Lesk, *et al.*, "Germanium and Silicon Tunnel Diodes—Design, Operation and Application." *1959 I.R.E. WESCON Convention Record*, Pt. 3, p. 9.

[9] A. Blicher, R. Glicksman, and R. M. Minton, "Temperature Dependence of the Peak Current of Germanium Tunnel Diodes," *Proc. I.R.E.*, Vol. 49, p. 1428, Sept. 1961.

minimized by proper choice of carrier concentration. For example, Blicher *et al*[9] have measured variations in peak current of only 10 per cent over a temperature range from −40 to 140°C. These measurements were made on a germanium tunnel diode having a p-region carrier concentration of 8.4×10^{19} atoms/cm^3.

The temperature dependence of the valley current has been measured[10] for germanium tunnel diodes having p-region carrier concentrations between 1.0×10^{19} and 7.9×10^{19} atoms/cm^3. For such diodes, the valley current increased roughly by a factor of three for a temperature change from −50 to 150°C.

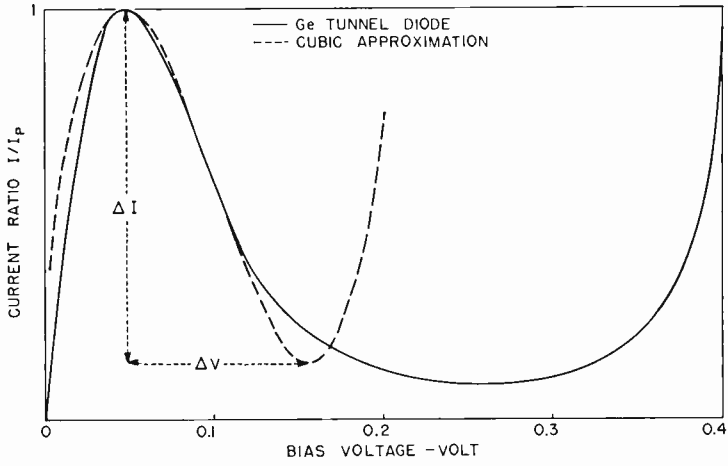

Fig. 1—Current–voltage characteristics of a germanium tunnel diode and cubic approximation used in the equivalent circuits of Figure 2.

Significant changes in valley voltage with temperature occur only if the ratio of peak current to valley current is less than approximately 4:1. Because many available diodes have peak-to-valley ratios exceeding 10:1 over most of their useful temperature range, variations in valley voltage are often negligible in practice.

Frequency of Tunnel-Diode Oscillators

Accurate calculations of the frequency of tunnel-diode oscillators can be carried out by graphical or numerical methods.[2,11] However,

[10] R. Gliksman, RCA Semiconductor Division, Somerville, N. J., private communication.

[11] W. A. Edson, *Vacuum-Tube Oscillators*, Chapter 4, John Wiley and Sons, Inc., New York, 1953.

because these methods are lengthy and do not lead to closed expressions for frequency, they are of limited usefulness for studying frequency stability.

Closed expressions for the frequency of simple tunnel-diode oscillators can be derived if the current–voltage characteristic of the tunnel diode is approximated by an antisymmetric cubic equation (see Figure 1) and if the variation of C_d with voltage is neglected.

Consider first the classic van der Pol oscillator shown in Figure 2(a). The steady-state frequency of this oscillator is given by[11,12]

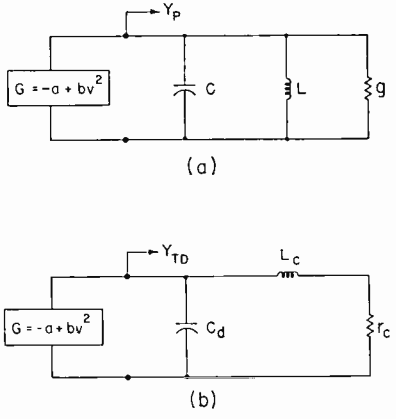

Fig. 2—Equivalent circuits of negative-resistance oscillators (a) circuit analyzed by van der Pol,[12] (b) tunnel-diode oscillator circuit. $L_c = L_d + L_o$, $r_c = r_d + r_o$, L_o = inductance of circuit, r_o = resistance of circuit.

$$\omega = \frac{1 - (a-g)^2 \dfrac{L}{16C}}{\sqrt{LC}}, \tag{2}$$

provided

$$(a-g)\sqrt{\frac{L}{C}} \ll 1. \tag{3}$$

Equation (2) is also applicable to the simple tunnel-diode oscillator

[12] B. van der Pol, "The Nonlinear Theory of Electrical Oscillations," *Proc. I.R.E.*, Vol. 22, p. 1051, Sept. 1934.

circuit of Figure 2(b) if the following transformations are made:*

$$
\begin{aligned}
g &\rightarrow \frac{r_c}{L_c} C_d \,, \\
C &\rightarrow C_d \left(1 - \frac{C_d}{L_c} r_c^2 \right) \,, \\
L &\rightarrow \frac{L_c}{\left(1 - \frac{C_d}{L_c} r_c^2 \right)^2} \,.
\end{aligned}
\tag{4}
$$

The frequency ω_{TD} of the tunnel-diode oscillator shown in Figure 2(b) can therefore be written as follows:

$$
\omega_{TD} = \frac{\left[1 - \left(\frac{3\Delta I}{2\Delta V} - \frac{r_c}{L_c} C_d \right)^2 \left(\frac{L_c}{16 C_d (1 - C_d r_c^2 / L_c)^3} \right) \right] \sqrt{1 - \frac{C_d r_c^2}{L_c}}}{\sqrt{L_c C_d}} \tag{5}
$$

where the coefficient a has been rewritten in terms of the differences between the peak and valley voltages and currents of the diode (see Figure 1).

Equation (5) is useful in estimating the effects on frequency stability of the various tunnel-diode and circuit parameters. However, the equation is, in general, not useful for the precise calculation of frequency of practical tunnel-diode oscillators.

Dependence of Frequency on Temperature, Bias Voltage, and R-F Load

Most tunnel-diode oscillators operating above a few hundred megacycles use transmission-line resonators. Distributed oscillators (and

---

* These transformations were derived from the conditions that

$$
\begin{aligned}
&\mathrm{Re}\,(Y_{TD})\,|_{\omega=\omega_0} = \mathrm{Re}\,(Y_p)\,|_{\omega=\omega_0} \\
&\mathrm{Im}\,(Y_{TD})\,|_{\omega=\omega_0} = \mathrm{Im}\,(Y_p)\,|_{\omega=\omega_0} = 0 \\
&\left.\frac{\partial\,\mathrm{Im}\,(Y_{TD})}{\partial\omega}\right|_{\omega=\omega_0} = \left.\frac{\partial\,\mathrm{Im}\,(Y_p)}{\partial\omega}\right|_{\omega=\omega_0}
\end{aligned}
$$

where Y_{TD} is the admittance of the tunnel-diode oscillator and Y_p is the admittance of the van der Pol oscillator.

even certain lumped oscillators) often exhibit frequency moding. In some cases, an oscillator may have simultaneous output power at two or more nonharmonic frequencies. In other cases, slight variations in bias voltage, r-f load, or temperature can cause the oscillator to jump modes, resulting usually in large changes in both frequency and output power. Frequency moding, which is difficult to treat analytically, can often be eliminated by careful circuit design and by the use of stabilizing resistors.[2,1,13] No attempt will therefore be made here to discuss the frequency stability of oscillators with moding problems.

Temperature dependence

The shape of the I–V characteristics of the diode used in an oscillator affects the frequency of oscillations, as shown, for example, in Equation (5), where the term $\Delta I/\Delta V$ appears in the expression for frequency. Any changes in the I–V characteristics of the diode resulting from temperature variations, therefore, cause changes in the oscillation frequency; for most tunnel-diode oscillators, this effect is the most important single cause of frequency changes resulting from variations in temperature. Variations in the I–V characteristics can be minimized by the use of (1) tunnel diodes with carrier concentrations chosen to minimize the variations of peak current with temperature and (2) tunnel diodes having ratios of peak current to valley current which are as high as possible. The second step results in minimum variations in valley voltage, and also minimizes the effect of the changes in the valley current.

Equation (5) shows that frequency variations due to changes in $\Delta I/\Delta V$ for the circuit of Figure 2(a) approach zero if $\Delta I/\Delta V$ approaches $2r_cC_d/(3L_c)$. The physical interpretation of this condition is simply that as $\Delta I/\Delta V$ approaches $2r_cC_d/(3L_c)$, the amplitude of the oscillations approaches zero. As a result, changes in the peak or valley points no longer influence the shape of the small region of the I–V characteristics traversed during the oscillations. Similar considerations hold for tunnel-diode oscillators in general; if a high degree of temperature stability is desired, swinging of the oscillator past peak and valley points should be avoided.

Changes in the parameters of the passive-circuit portion of the oscillator also affect the frequency of oscillation (for example, see Equation (5)). The magnitude of these changes can be estimated for most oscillators from the requirement that the sum of the impedances

---

[13] C. S. Kim, "High-Frequency, High-Power Operation of Tunnel-Diodes," *Trans. I.R.E.*, Vol. PGCT-8, p. 416, Dec. 1961.

at any point in a steady-state oscillator must be zero.* Although this requirement is strictly true only for purely sinusoidal oscillators, its application also yields fairly accurate results for nearly sinusoidal oscillators.[2]

Figure 3 shows an experimental curve of frequency deviation as a function of temperature for an L-band tunnel-diode oscillator. This oscillator uses a GaAs tunnel diode in a straight strip-transmission-line cavity, similar to that shown in Figure 12 of Reference (2). The

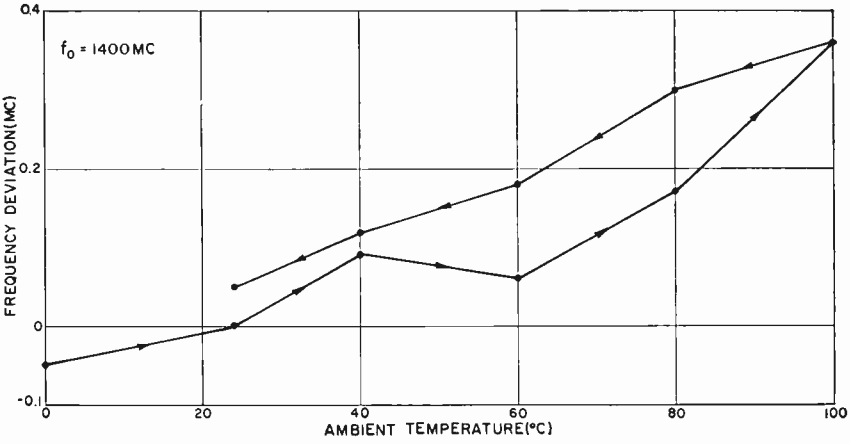

Fig. 3—Frequency deviation as a function of temperature for a GaAs tunnel-diode oscillator.

strip-line cavity was made from commercial copper-clad Teflon†-impregnated glass-fiberboard; no attempt was made to temperature stabilize the cavity.

The peak current of the diode was 47 ma at 25°C and 44.5 ma at 100°C; the corresponding values of valley current were 2 ma and 3 ma, respectively. Figure 3 shows that the average value of the fractional

* This requirement, while very general, yields little information on the stability of tunnel-diode oscillators. Consider for example, the variations in frequency due to variations in $R_d(v)$: Let $Z_T = R_T + jX_T$ be the total impedance across $R_d(v)$ for steady-state purely sinusoidal oscillations. The possible steady-state frequencies are determined solely by the condition that $X_T = 0$, and are not affected by $R_d(v)$. (The amplitude of the oscillations simply adjust to a value where $\overline{R}_d = -R_T$). Thus, for purely sinusoidal oscillations, the frequency is independent of the I–V characteristics of the diode. Actual tunnel-diode oscillators, like any other physically realizable oscillator, are, of course, never purely sinusoidal. Equations (2) and (5) take the nonsinusoidal nature of the oscillations of the circuits of Figure 2 into account.

† Trade Mark.

frequency shift for this oscillator is approximately 3×10^{-6} per degree centigrade.

The frequency stability of most tunnel-diode oscillators can be significantly improved by putting a stabilizing cavity at the output of the oscillator. In one test, for example, a 610-mc straight strip-transmission-line oscillator was stabilized by means of a cavity with 0.4 mc bandwidth. The frequency of this oscillator varied by less than 80 kc over the temperature range from 20-100°C.

Dependence on Bias Voltage

Variations in bias voltage affect the frequency of a tunnel-diode oscillator in two ways:

(1) The harmonic content of the output of the oscillator generally changes, because the oscillator swings over different portions of the I–V characteristic of the diode at different values of bias. These changes in harmonic content usually lead to changes in the fundamental frequency.*

(2) Changes in bias also cause changes in the effective value of C_d, because C_d is voltage-dependent (see Equation (1)). This change in C_d causes changes in frequency (for example, see Equation (5)).

The variations in frequency resulting from changes in harmonic content are caused by the nonlinearity of the current–voltage characteristics of the diode, and are therefore difficult to estimate.† However, variations in frequency resulting from changes in C_d can be readily estimated by assuming that the effective value of C_d is the same as the value of C_d at the bias point. The errors in C_d resulting from the use of this simplifying assumption will usually be only a few percent.

Tunnel-diode oscillators using low-Q circuits[14] can be voltage tuned over frequency ranges as high as 12 per cent. On the other hand, a curve of frequency as a function of bias voltage for a high-Q S-band ridge-waveguide oscillator (see Figure 4) shows that fractional fre-

---

* See Reference (11) for equations relating harmonic content and fundamental oscillation frequency for negative-resistance oscillators. These equations are useful for calculation of fundamental frequency only if the harmonic content of the output is known.

† Accurate calculations of frequency pushing have been made for a particular lumped-circuit tunnel-diode oscillator using a digital computer.[2] However, the results of these calculations are not directly applicable to most practical microwave tunnel-diode oscillators.

[14] J. K. Pulfer, "Voltage Tuning in Tunnel Diode Oscillators," *Proc. I.R.E.*, Vol. 48, p. 1155, June 1960.

quency shifts of less than 2×10^{-6}/mv can be obtained over a bias range of 60 mv.

Dependence on r-f Load

As previously discussed, it is possible to estimate the frequency of a tunnel-diode oscillator from the requirement that the susceptance across the negative resistance be zero in the steady state. Variations in frequency which result from changes in r-f load impedance can also be estimated by using this method.

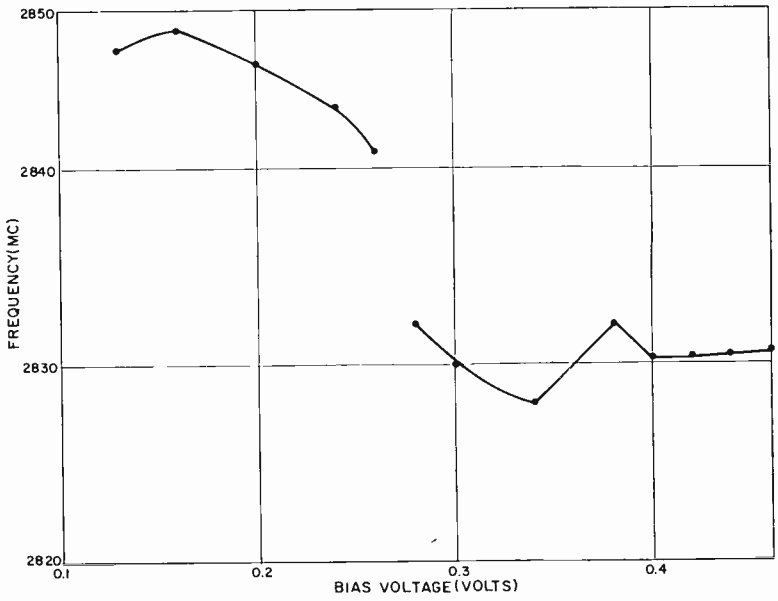

Fig. 4—Frequency as a function of bias voltage for a germanium tunnel-diode oscillator.

As with any oscillator, frequency pulling can be minimized in tunnel-diode oscillators by use of stabilizing cavities, or by decoupling the oscillator as much as possible from the load. Decoupling by means of nonreciprocal devices such as ferrite isolators is particularly useful, since in this case the loss in power delivered to the load due to decoupling is often negligible.

A complete Rieke diagram for a re-entrant strip transmission line oscillator is shown in Figure 16 of Reference (2). For the strip-line L-band oscillator discussed above, the maximum fractional frequency

deviation is about 1×10^{-4} for a change in load voltage-standing-wave ratio from about 1 to 2. This oscillator used no stabilizing cavity and no nonreciprocal elements, and the loss of available output power resulting from decoupling was approximately 60 per cent.

Frequency Locking

The frequency of a tunnel-diode oscillator can be locked to the frequency of an external signal.* Circuits for introducing the locking

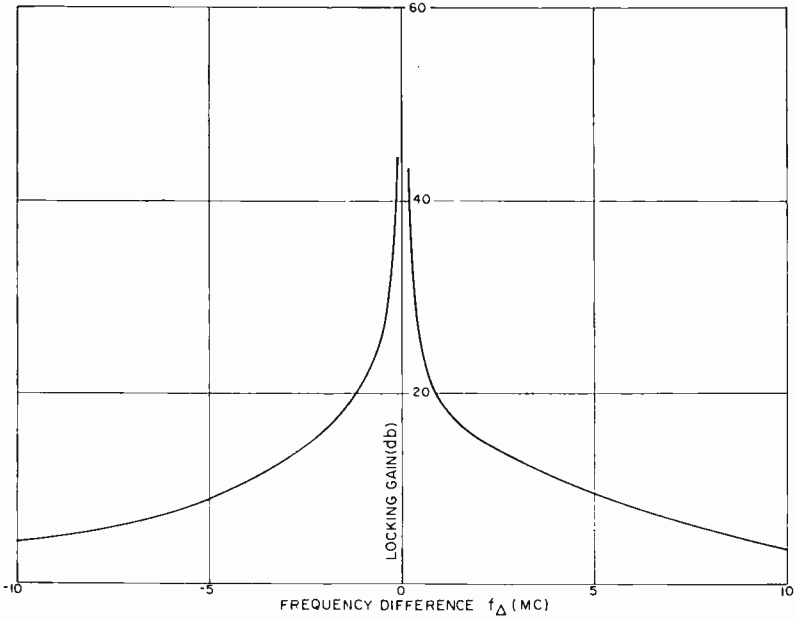

Fig. 5—Locking signal as a function of difference frequency f_Δ for a 1500 mc GaAs tunnel-diode oscillator. $f_\Delta = f_s - f_{LO}$, f_s = natural frequency of oscillator, f_{LO} = frequency of locking signal.

signal into the oscillator with minimum loss are discussed in Reference (2), where it is also shown that the locking signal can be orders of magnitude smaller than the power output of the oscillator, provided the locking frequency is close to the natural frequency of the oscillator. Figure 5 shows locking gain (the ratio of oscillator output power to the minimum locking power) as a function of frequency difference (the difference between the natural frequency of the oscillator and the frequency of the locking signal) for a strip-line oscillator. Simi-

* For a general discussion of the theory of locking or synchronization of oscillators see Chapter 13 of Reference (11).

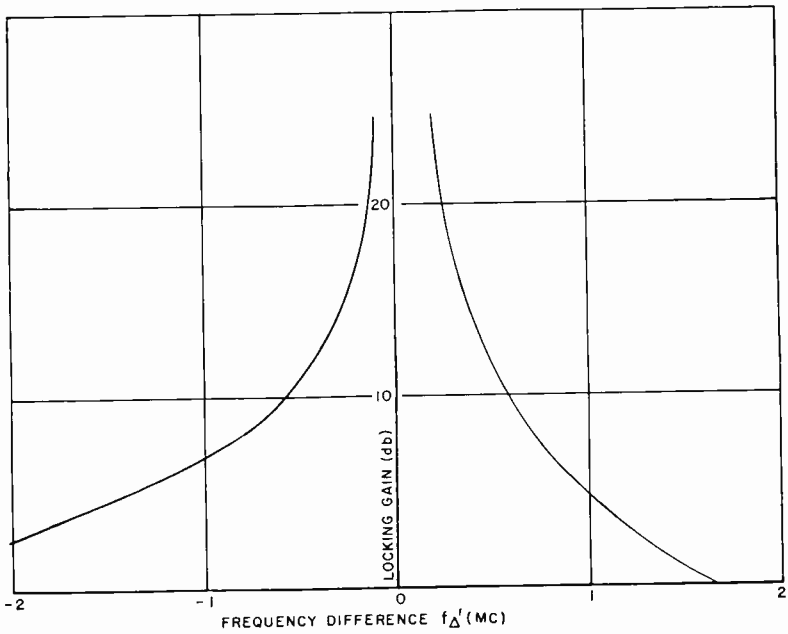

Fig. 6—Locking signal as a function of difference frequency f'_Δ for a 1500 mc GaAs tunnel-diode frequency. $f'_\Delta = f_s - 2f_{LO}$, f_s = natural frequency of oscillator, f_{LO} = frequency of locking signal.

larly, Figure 6 shows harmonic locking gain (the locking frequency is approximately one-half the output frequency of the oscillator).

The fractional frequency stability of the locked oscillator is equal to the stability of the locking signal. Thus, if the locking signal is derived from the harmonic of a low-frequency crystal-controlled oscillator, fractional frequency stabilities equal to that of the crystal oscillator are obtained.

ACKNOWLEDGMENT

The author extends his thanks to H. C. Johnson and J. L. Poirier for their assistance.

HETERODYNE RECEIVERS FOR RF-MODULATED LIGHT BEAMS*

By

D. J. Blattner and F. Sterzer

RCA Electron Tube Division
Princeton, N. J.

Summary—*The frequency of amplitude or phase modulation on a light beam can be shifted to a new value by passing the beam through an electro-optic modulator. Expressions for conversion loss are derived for a modulator using a crystal exhibiting the linear electro-optic effect. Good agreement with experiments is obtained. A suggested application is heterodyning high-frequency light modulation to the response ranges of available light demodulators.*

Introduction

THE HETERODYNE PRINCIPLE, widely used in r-f receivers, can also be used in receivers for r-f-modulated light beams. The modulation frequency f_s on the light beams can be heterodyned to a new value $|f_s \pm mf_{LO}|$ (where m is an integer) by passing the beams through light modulators driven at frequency f_{LO}.† This method can be used in detectors of light beams that are either intensity or polarization modulated.**

Figure 1 is a schematic diagram of a heterodyne demodulator for an intensity-modulated light beam. The heterodyning takes place in a mixer that modulates the intensity of the incident light beam at frequency f_{LO}. For the particular case of a mixer consisting of a crystal that exhibits the linear electro-optic effect when placed between two crossed polarizers, the conversion efficiency can be evaluated as follows:

The intensity of the light transmitted by the mixer is given by[1]

* Sponsored by Electronic Technology Laboratory, ASD, AFSC, Wright-Patterson AFB, Ohio, under contract AF33(616)-8199.

† A scheme using zero difference frequency has been proposed by Macek *et al*, "Microwave Modulation of Light," to be published in *I.R.E. Convention Record*, Part III, September, 1962.

** It should be noted that in conventional r-f receivers it is the carrier frequency that is shifted; for the heterodyne receivers discussed here, the modulation frequency is shifted.

[1] Readers unfamiliar with the theory of electro-optic modulators using crystals exhibiting the linear electro-optic effect may refer to R. H. Blumenthal, "Design of a Microwave Frequency Light Modulator," *Proc. I.R.E.*, Vol. 50, p. 452, April 1962. A more exhaustive treatment is given by B. H. Billings, "The Electro-Optic Effect in Uniaxial Crystals of the Type XH_2PO_4," *Jour. Opt. Soc. Amer.*, Vol. 39, p. 797, Oct. 1949.

$$I_T = I \sin^2 \left(\frac{\pi}{2} \frac{V}{V_{\lambda/2}} \right) \tag{1}$$

where I is the intensity of light incident on the crystal, V is the voltage across the crystal, and $V_{\lambda/2}$ is the voltage across the crystal necessary to produce a 180° differential phase shift.

If the incoming light is linearly intensity modulated at frequency f_s, then the light intensity I is given by

$$I = \frac{I_0}{2} (1 + M \cos \omega_s t). \tag{2}$$

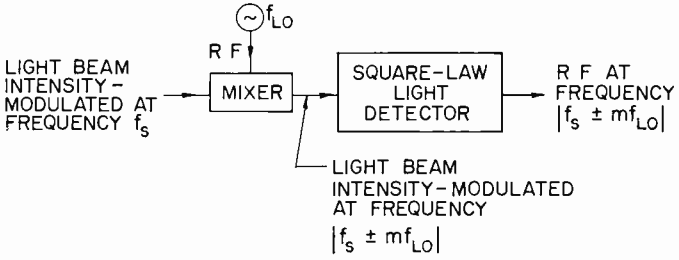

Fig. 1—Heterodyne light detector for intensity-modulated light. The mixer modulates the intensity of the light transmitted through it at frequency f_{LO}.

The voltage V across the mixer crystal can be expressed as:

$$V = V_B + V_{LO} \cos \omega_{LO} t, \tag{3}$$

where V_B is the d-c bias voltage across the crystal, V_{LO} is the amplitude of the local-oscillator voltage, and $\omega_{LO} = 2\pi f_{LO}$.

Substitution of Equations (2) and (3) into Equation (1) gives the following relation:

$$I_T = \frac{I_0}{2} (1 + M \cos \omega_s t) \sin^2 \left[\frac{\pi}{2V_{\lambda/2}} (V_B + V_{LO} \cos \omega_{LO} t) \right], \tag{4}$$

and spectral analysis of Equation (4) yields

$$I_T = \frac{I_0}{4} \left\{ 1 - \left[\cos \frac{\pi V_B}{V_{\lambda/2}} \right] J_0 \left(\frac{\pi V_{LO}}{V_{\lambda/2}} \right) + \right.$$

$$+ M \left[1 - \left(\cos \frac{\pi V_B}{V_{\lambda/2}} \right) J_0 \left(\frac{\pi V_{LO}}{V_{\lambda/2}} \right) \right] \cos \omega_s t$$

$$- 2 \left(\cos \frac{\pi V_B}{V_{\lambda/2}} \right) \sum_{n=1}^{\infty} (-1)^n J_{2n} \left(\frac{\pi V_{LO}}{V_{\lambda/2}} \right) \cos 2n\omega_{LO} t$$

$$+ 2 \left(\sin \frac{\pi V_B}{V_{\lambda/2}} \right) \sum_{n=0}^{\infty} (-1)^n J_{2n+1} \left(\frac{\pi V_{LO}}{V_{\lambda/2}} \right) \cos (2n+1)\omega_{LO} t$$

$$- M \left(\cos \frac{\pi V_B}{V_{\lambda/2}} \right) \sum_{n=1}^{\infty} (-1)^n J_{2n} \left(\frac{\pi V_{LO}}{V_{\lambda/2}} \right) \left[\cos (\omega_s + 2n\omega_{LO}) t + \cos (\omega_s - 2n\omega_{LO}) t \right]$$

$$+ M \left(\sin \frac{\pi V_B}{V_{\lambda/2}} \right) \sum_{n=0}^{\infty} (-1)^n J_{2n+1} \left(\frac{\pi V_{LO}}{V_{\lambda/2}} \right) \left[\cos (\omega_s + [2n+1]\omega_{LO}) t + \cos (\omega_s - [2n+1]\omega_{LO}) t \right] \Bigg\}. \tag{5}$$

If it is assumed that the incoming light is linearly polarized, so that all of it reaches the crystal, then the conversion gain G_c is given by

$$G_c = \frac{1}{2} \left[\cos \frac{\pi V_B}{V_{\lambda/2}} \right] J_{2n} \left(\frac{\pi V_{LO}}{V_{\lambda/2}} \right), \text{ for } \omega_{IF} = |\omega_s \pm 2n\omega_{LO}|,$$

$$G_c = \frac{1}{2} \left[\sin \frac{\pi V_B}{V_{\lambda/2}} \right] J_{2n+1} \left(\frac{\pi V_{LO}}{V_{\lambda/2}} \right), \text{ for } \omega_{IF} = |\omega_s \pm (2n+1)\omega_{LO}|. \tag{6}$$

For the components at frequency $|\omega_s \pm \omega_{LO}|$, the maximum value of G_c is 0.291. The minimum conversion loss is therefore $1/0.291 = 3.44$, which corresponds to approximately 5.4 db.

The light beam modulated at the heterodyned frequencies is demodulated in a square-law detector, i.e., a detector whose output current is proportional to the intensity of the incident light. Detectors of this type include phototubes and semiconductor photoelectric cells.

Figure 2 is a schematic diagram of a heterodyne demodulator for a light beam whose intensity is constant, but whose polarization is modulated at frequency f_s. In this demodulator, the heterodyning

takes place in a mixer that modulates the polarization of the incident light beam. The mixer is followed by a Nicol prism that converts the polarization modulation into intensity modulation, so that a conventional square-law detector can be used to demodulate the light beam.

Analysis similar to that described above shows that the conversion gain for the case of a mixer consisting of a crystal exhibiting the linear electro-optic effect is given by

$$G'_c = \left[\cos\frac{\pi V_B}{V_{\lambda/2}}\right] J_{2n}\left(\frac{\pi V_{LO}}{V_{\lambda/2}}\right), \text{ for } \omega_{IF} = |\omega_s \pm 2n\omega_{LO}|,$$
$$G'_c = \left[\sin\frac{\pi V_B}{V_{\lambda/2}}\right] J_{2n+1}\left(\frac{\pi V_{LO}}{V_{\lambda/2}}\right), \text{ for } \omega_{IF} = |\omega_s \pm (2n+1)\omega_{LO}|. \quad (7)$$

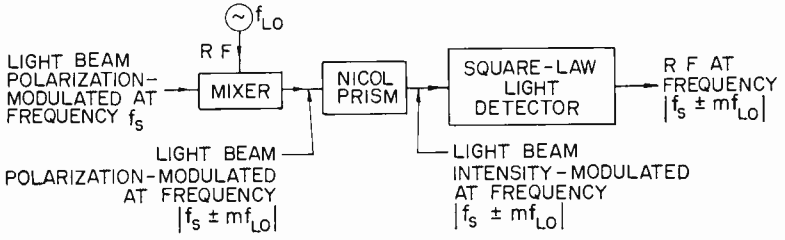

Fig. 2—Heterodyne light detector for polarization-modulated light. The mixer modulates the polarization of the light transmitted through it at frequency f_{LO}.

In these expressions, G'_c is defined as the ratio of the component of light that is intensity modulated at frequency $|f_s \pm mf_{LO}|$ and transmitted by the Nicol prism to the maximum value of intensity modulation at frequency f_s that can be produced by inserting wave plates and polarizers with the mixer removed from the incident light path. For the components at frequency $|\omega_s \pm \omega_{LO}|$, the minimum conversion loss is approximately 2.4 decibels.

Figure 3 is a plot of minimum conversion loss for $\omega_{IF} = |\omega_s \pm \omega_{LO}|$ as a function of $V_{LO}/V_{\lambda/2}$ calculated from Equation (7). Experimental points for an output frequency of $|\omega_s - \omega_{LO}|$ are also shown. The curve shows excellent agreement between measured and calculated values. The experiment was performed with two $NH_4H_2PO_4$ (ADP) modulators; one modulator was used to produce the polarization modulation, and the other served as the mixer. Data were measured at audio frequencies to permit precise determination of voltages.

Conversion loss was also measured as a function of signal voltage. The conversion loss remained constant even for signal voltages down to noise level.

The most important applications of heterodyne light detection involve heterodyning to the difference frequency $|f_s - f_{LO}|$. When f_s

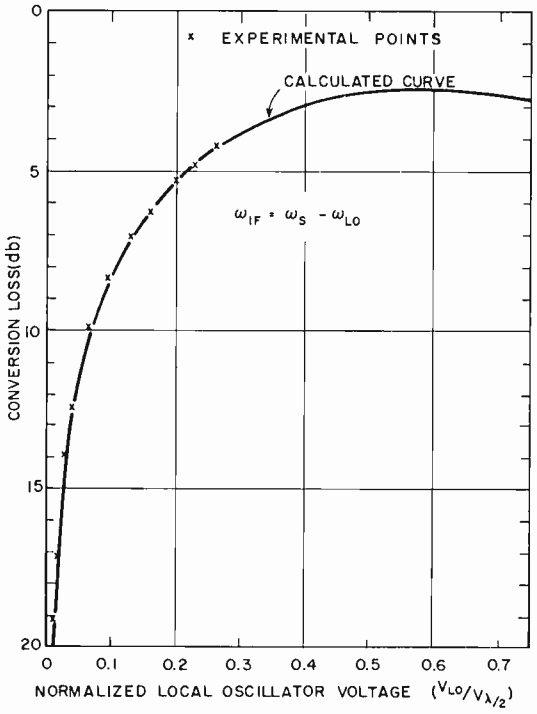

Fig. 3—Calculated and experimental values of conversion loss versus local oscillator voltage for the heterodyne detector of Figure 2. Experimental data were measured at audio frequencies using an $NH_4H_2PO_4$ crystal as mixer.

is at microwave frequencies, for example, if no heterodyning is used the square-law demodulator must respond to variations of light intensity at microwave rates. Although experimental light detectors capable of microwave response have been reported,[2] these devices are presently extremely inefficient; the same situation exists in the infrared range. The heterodyne system, by shifting the modulation down

[2] B. J. McMurtry and A. E. Siegman, "Photomixing Experiments with a Ruby Optical Maser and a Traveling-Wave Microwave Phototube," *Applied Optics*, Vol. 1, p. 51, Jan. 1962.

from the microwave range, permits use of conventional square-law detectors, such as phototubes and semiconductor photocells, which operate only up to VHF.

Many channels of information can be carried on a single light beam if multiplex techniques are used in conjunction with the heterodyne system. Each channel can then have a bandwidth equal to that of the detector.

ACKNOWLEDGMENT

The authors are grateful to H. R. Lewis, A. Miller, J. L. Poirier, F. E. Vaccaro, and L. M. Zappulla for discussions and assistance.

AN ORGANIC PHOTOCONDUCTIVE SYSTEM

By

H. G. Greig

RCA Laboratories
Princeton, N. J.

Summary—*An all-organic photoconductive system has been found which is useful in electrophotography. The two electrophotographic materials that are presently available commercially depend upon inorganic materials for their photoconductivity. The new uses introduced are based upon the physical and chemical differences between the organic and inorganic materials. For example, a transparent, photosensitive layer has been made which prints by conventional electrophotographic techniques. A thermoplastic photoconductive layer permits the development of an electrostatic latent image by heat-deformation to produce a master suitable for projection by schlieren optics. The durable, flexible, grainless, relatively low cost, organic photoconductive mixtures will increase the utility of the electrophotographic process.*

Introduction

Two types of photoconductive electrophotographic materials are presently in commercial use. One, a coherent layer of an inorganic photoconductor (selenium), was introduced in 1948.[1] The other, a particulate inorganic photoconductor (zinc oxide) dispersed in an electrically insulating organic binder, was introduced in 1954.[2]

Now an all-organic photoconductive material has been found which has sensitivity comparable with other electrophotographic systems and which prints by the conventional techniques. So far as is known, other organic photoconductors that have been considered for use in electrophotography have shown much lower printing sensitivity. Dyes with the highest reported photoconductivity have thus far been unsatisfactory for making a practical photosensitive layer. The use of an organic photoconductive system as the sensitive layer opens several new uses for electrophotography. For example, the photosensitive layer can be flexible, durable and transparent. Transparencies can now be produced directly. Figure 1 is a photograph of an electrophotographic transparency.

---

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

[2] C. J. Young and H. G. Greig, "Electrofax: Direct Electrophotographic Printing on Paper," *RCA Review*, Vol. XV, p. 469, Dec. 1954.

The organic photoconductors are mixtures of insulating materials, substituted polyarylmethanes and their derivatives. There are many possible combinations of these materials, so that desired electrical, physical and chemical properties can be tailored for a specific use. For example, a photosensitive layer has been made which is thermoplastic

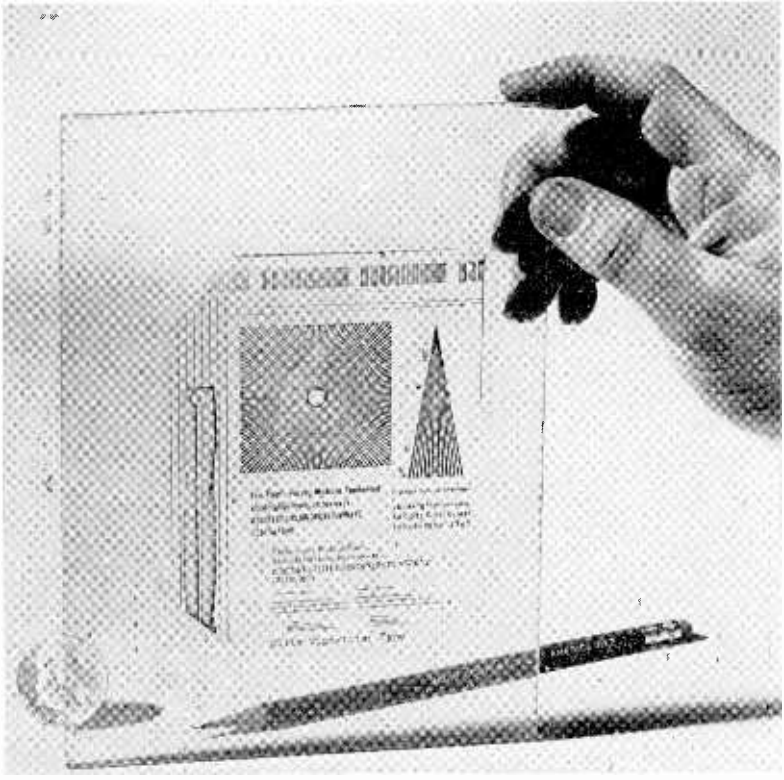

Fig. 1—Electrophotographic transparency.

and permits the development of the latent electrostatic image by heat-distortion and thus produces a master for image projection by schlieren optics.

The chemistry and physics of the system are complex and there is much work ahead in the measurements and interpretation of the inter-related electrical, thermal, and light effects which have already been observed. The material is a chemical mixture in equilibrium but subject to changes by heat, light, and chemical action. The sensitizers for example, are acid-base indicators and subject to change with *p*H

of the environment. The printing layers, on the other hand, are surprisingly durable and stable with use. Coated aluminum plates have shown no degradation with intermittent use and normal storage for well over a year.

The Organic Photoconductive System

There is increasing interest in the conductivity of organic materials from the purely academic standpoint, to try to gain a better understanding of the mechanisms involved. The search for new materials to fill a need must go concurrently with measurement and interpretation of the interactions between energy and materials. The complexity of organic systems, especially in the solid state, has of necessity placed them second to inorganic materials for theoretical treatment and measurement. However, purity is now well controlled and structure well defined for many inorganic materials of the semiconductive and photoconductive type. Refined techniques and instrumentation are being used to look at organics.

There are at least two conditions that contribute to conduction of any magnitude in an organic material in the solid state. One is mobility of electrons within the molecules, as in a conjugate system of alternate single and multiple bonds. The other is an electron acceptor–donor (Lewis, acid–base) relationship for exchange between molecules. Examples to support this are the iodine-perylene complex (8 ohm-cm resistivity reported) and more recently the low-resistivity materials 7,7,8,8-tetracyanoquinodimethane (TCNQ) in complexes with various cations.[3] In order to be useful in electrophotography with the present processing techniques, an organic photoconductive system requires certain other characteristics: a dark resistivity high enough to meet the charge level and decay times for processing, sufficient light-induced drop in resistivity to meet the sensitivity and development requirements, and the desired physical properties for use as a printing surface. All of these are met by the present system.

The general chemistry of the materials will be found in literature relating to triphenylmethane dyes.[4,5] A description of one specific combination, however, will illustrate chemical relationships common to the

---

[3] "A New Series of Organics Has Low Resistivity," *Chemical and Engineering News*, Vol. 39, p. 42, Jan. 9, 1961.

[4] H. A. Lubs, *The Chemistry of Synthetic Dyes and Pigments*, ACS Monograph, Series No. 127, Reinhold Publishing Corp., New York, N. Y., 1955.

[5] Lyde S. Pratt, *The Chemistry and Physics of Organic Pigments*, John Wiley and Sons, Inc., New York, N. Y., 1947.

class. In this example the organic photoconductive system is a mixture of a vinyl chloride copolymer, 4,4′-benzylidenebis (N,N-dimethylaniline) and a trace of the dye malachite green. The trace of dye is formed *in situ* by reaction of the other two in the solid state by heat and or exposure to ultraviolet light. The dye can also be added to the system. The 4,4′-benzylidenebis (N,N-dimethylaniline) is the leuco base (colorless) of the dye malachite green. The dye is usually made from the base by oxidation with freshly prepared lead dioxide in acid solution. In a thin layer (0.20-0.5 mil thick) of the resin-base mixture applied to a metal plate from solvent solution, a trace of the dye is formed when the coating is dried by heating (about 1 minute at 150-180°C). This transparent green-tinted layer has an electrophotographic printing response to red light. The spectral characteristic of the photoconductive layer corresponds to the two major visible absorption peaks of malachite green (about 6300Å and 4300Å). In this example, only a trace of the dye is necessary to have this response to visible light.

The literature shows that degradation of halogenated polymers of the type used here, either by heat or ultraviolet light, involves the formation of free radicals, the splitting off of halogen atoms and the formation of halogen acids.[6] Photo-oxidation of the leuco base coupled with the degradation of the resin forms the trace cationic dye and provides halide anions for it in the solid system. A pure polystyrene or saturated hydrocarbon resin substituted for the vinylchloride copolymer apparently does not supply the anion necessary for dye formation by heat or photo oxidation. At best a yellow color develops under the same conditions, indicating oxidation of the leuco base to different products. Therefore, no appreciable printing response to red light has been observed, although this combination does have some printing response to ultraviolet or blue light.

It is too early to speculate on the various mechanisms involved in the formation and operation of this organic photoconductive system. One can, however, learn much by analogy from the publications of work on conductivity and photoconductivity in organic materials.[7-14] There

---

[6] H. F. Payne, *Organic Coating Technology*, Vol. I, p. 492, John Wiley and Sons, Inc., New York, N. Y., 1954.

[7] R. C. Nelson, "The Photoconductivity of Some Triphenylmethane Dyes," (Letters to the Editor), *Jour. Chem. Phys.*, Vol. 19, p. 798, June 1951.

[8] R. C. Nelson, "An Effect of the Anion on the Conductive Properties of Triphenylmethane Dye Salts," (Letters to the Editor), *Jour. Chem. Phys.*, Vol. 20, p. 1327, Aug. 1952.

[9] R. C. Nelson, "Organic Photoconductors. I. The Kinetics of Photoconductivity in Rhodamine B," *Jour. Chem. Phys.*, Vol. 22, p. 885, May 1954.

has been considerable work on the photoconductivity of dyes in thin layers on various substrates. Anthracene, naphthalene and the phthalocyanines are receiving much attention because the chemical and crystal structures are well defined. High-purity crystals of the first two materials can be grown for solid-state measurements. Anthracene, whose photoconductivity has been known since 1906,[15] was the first organic suggested for use in electrophotography.[16] Many other materials have shown up in more than thirty foreign patents relating to electrophotography.

For the system described here, a few of the structural formulas of the materials and their relationships are given to help illustrate the complexity of the system. Figure 2 shows the malachite green system of the example. The illustrated relationships are known to hold in an aqueous system. The acid–base indicators can be explained by the blocking or freeing (by the protons) of the auxochromic dimethylamino groups for participation in the conjugate system. The quinoid structure can be pictured in either substituted phenyl group since it is actually a limiting resonant structure contributing to the visible absorption peak at 6300Å. The third unsubstituted phenyl group can also participate in the resonance of the system resulting in a peak at about 4300Å for the dye.

There are close to 200 di- and tri-arylmethane dyes listed in the *Color Index*.[17] Many other combinations are possible in this general class all of which have different chemical structures for the corresponding leuco base. There are a large number of synthetic resins with various modifiers and stabilizers many of which can prove useful in

---

[10] John W. Weigl, "Spectroscopic Properties of Organic Photoconductors. I. Absorption Spectra of Cationic Dye Films," *Jour. Chem. Phys.*, Vol. 24, p. 364, Feb. 1956.

[11] John W. Weigl, "Spectroscopic Properties of Organic Photoconductors. II. Specular Reflection Spectra of Cationic Dye Films," *Jour. Chem. Phys.*, Vol. 24, p. 577, March 1956.

[12] John W. Weigl, "Spectroscopic Properties of Organic Photoconductors. III. Action Spectra for the Photoconduction of Cationic Dye Films," *Jour. Chem. Phys.*, Vol. 24, p. 883, April 1956.

[13] A. Vartanian, "Photoconductivity of Thin Layers of Dyes in the Solid State," *A.C.T.A. Physiochimica*, USSR, Vol. XXII, No. 2, 1947.

[14] J. W. Eastman, Gerrit Engelsma, and Melvin Calvin, "Free Radicals, Ions and Donor-acceptor Complexes in the Reaction: Chloranil + N, N-Dimethylaniline Crystal Violet," *Jour. Amer. Chem. Soc.*, Vol. 84, p. 1339, April 20, 1962.

[15] A. Pocchettino, *Acad. Lincei rend.*, Vol. 15, p. 355, 1906.

[16] Chester Carlson, U. S. Patent No. 2,297,691, Oct. 1942.

[17] *Color Index*, 2nd Edition, Chorley and Pickersgill, Ltd., Leeds, England.

formulating a photoconductive mixture of this type. The spectral response changes with the leuco base and dye chosen. For 4,4′,4″-methylidynetris (N,N-dimethylaniline) (the leuco base for crystal violet) the peak printing response is close to 6000Å. The thin printing element can still be transparent, but is usually more-highly colored. The triphenylmethane dyes are among the most brilliant known, and are efficient absorbers of light because of the well-balanced resonant systems.

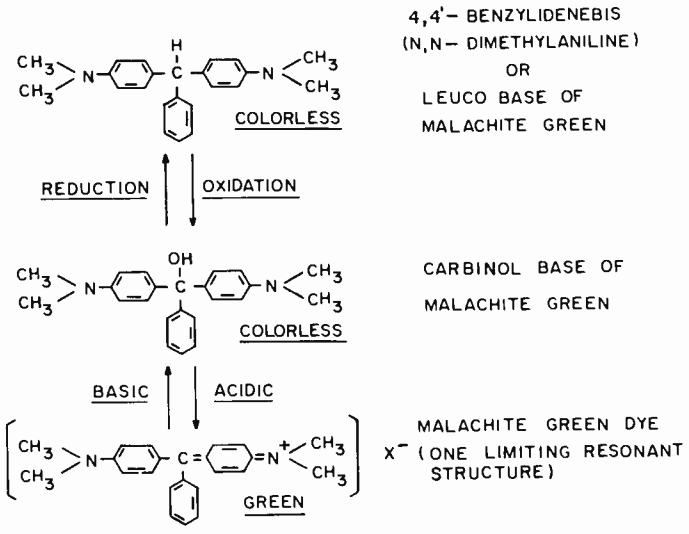

Fig. 2—Malachite green system.

Potential Uses for the Organic System

It is not anticipated that this new addition to the family of electrophotographic materials will replace the existing ones. The new properties offered by the organic system, however, will increase the utility of this branch of the graphic arts.

The direct printing transparent material of the example has been used successfully on conductive coated glass slides and on metallized film to produce masters for projection. By choosing colored developers which are transparent in the resin when fixed, multicolor prints have been made which allow color projection.

The printing elements will operate with either positive or negative corona charging. Prints have been made on self-supporting thin films with double corona charging, and on thin films which can be stripped after the image has been fixed.

The transparent printing element also lends itself to reflex printing techniques which opens several new possibilities.

N. E. Wolff, F. H. Nicoll, and H. G. Greig of these Laboratories have succeeded in producing thermoplastic photoconductive coatings on which the latent electrostatic image may be produced by standard electrophotographic techniques. This latent image can subsequently be heat-developed into the ripple pattern necessary for schlieren projection. A paper describing this work will be published in the future.

Toxicity

There are many conflicting assessments of the toxicity of dyes and dye intermediates reported in the literature. This is primarily because some of the physiological effects are little known and long range in effect. People differ in their sensitivities and reactions to different classes of chemicals. No hard and fast rules have been found to predict the toxicities of aromatic bases such as the tri-phenylmethane dyes and their intermediates. It seems reasonable to predict that many of these chemicals will be found to be relatively inert in the amounts used in photosensitive layers. There is, however, the possibility that in the bulk quantities some of the chemicals may be hazardous, and suitable precautions will be required.

Conclusion

An all-organic photoconductive printing element has been discovered which will result in several new uses for the electrophotographic process. A flexible, direct printing, transparent film has been demonstrated. Grainless multicolor transparencies have been made. The distortion of a thermoplastic photoconductive layer by the electrophotographic latent image when heated, has produced a master suitable for projection by schlieren optics.

A durable, flexible, grainless, relatively low cost electrophotographic printing element which can be processed by conventional electrophotographic procedures, has been produced by solvent coating. There are many possible combinations of the materials in this organic photoconductive system which can be modified to meet specific needs in a given use.

Acknowledgment

The author is grateful to C. J. Young for providing the environment, support, and encouragement that made this work possible.

LIMITING-CURRENT EFFECTS IN LOW-NOISE TRAVELING-WAVE-TUBE GUNS

By

A. L. Eichenbaum and J. M. Hammer

RCA Laboratories
Princeton, N. J.

Summary—*An experimental investigation of the radial variation of current density in the beams produced by ultra-low-noise traveling-wave-tube electron guns has been made. It was found that under certain operating conditions a virtual cathode is formed in the cathode–first-anode region. The studies also showed that the hysteresis which sometimes accompanies the forced formation of virtual cathodes between apertured grids remote from the actual cathode is strongly dependent on the trapping of positive ions.*

Introduction

THE NOISE PERFORMANCE of traveling-wave tubes (TWT) depends strongly on the d-c conditions in the electron gun. For this reason current–voltage characteristics and the radial distribution of current density in beams extracted from low-noise TWT guns with oxide cathodes were investigated. The experiments were mainly concerned with the formation of a virtual cathode.[1,2] Two regimes of operation were investigated. In the first regime the second- and third-anode voltages (V_2, V_3) were near to those used in ultra-low-noise TWT practice. In this case, the experiments revealed that for sufficiently positive beam-forming voltages (V_{BF}), a virtual cathode formed in the region between the actual cathode and first anode over a limited cross section of the beam.

A second regime of operation was studied to throw additional light on the nature of virtual-cathode formation. In this second regime, the second and third anodes were operated at voltages sufficiently low to cause virtual cathode formation in the region between second and third anodes. Both the limiting current and hysteresis that result from virtual cathode formation were found to be strongly dependent upon the trapping of positive ions.

The experiments were carried out on the tube shown in Figure 1.

---

[1] J. Berghammer, "Space-Charge Effects in Ultra-Low-Noise Electron Guns," *RCA Review*, Vol. 21, No. 3, p. 369, Sept. 1960.

[2] W. M. Mueller, "Reduction of Beam Noisiness by Means of a Potential Minimum Away from the Cathode," *Proc. I.R.E.*, Vol. 49, p. 642, March 1961.

Here an RCA ultra-low-noise S-band TWT gun is followed by a pinhole electrode and collector. The gun was immersed in a strong axial magnetic field and therefore no current was intercepted by the accelerating and beam-forming electrodes. The electron beam was scanned past the pinhole by accurately tilting the tube axis with respect to the magnetic field. The current density across the beam could thus be measured.

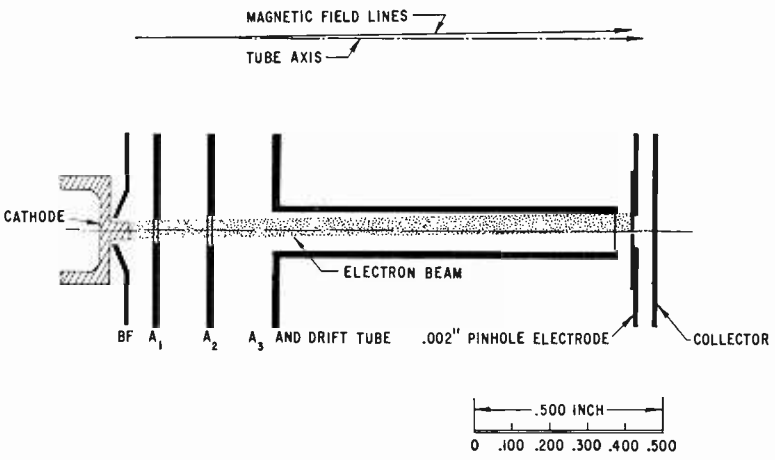

Fig. 1—Low-noise gun with pinhole analyzer. The beam is swept past the pinhole by accurately tilting the tube in the uniform magnetic field.

"Near" Virtual Cathode

In the first set of measurements the gun was operated with the voltages normally used in ultra-low-noise TWT practice. In this case V_2 and V_3 were sufficiently high to prevent the formation of a virtual cathode anywhere in the space between the first and third anodes. Relatively low voltages were present, however, in the cathode–first-anode region. This region is characterized by a potential distribution having a saddle point in the absence of space charge.[3] Injection of sufficient current density into this saddle-point region can cause the formation of a virtual cathode. This effect occurs between an entrance plane whose position is controlled by the value of the beam-forming

[3] M. R. Currie and D. C. Forster, "New Mechanism of Noise Reduction in Electron Beams," *Jour. Appl. Phys.*, Vol. 30, p. 94, Jan. 1959.

voltage (V_{BF}), and an exit plane which is approximately at the first anode. An understanding of the conditions in this region is complicated by the fact that the current density entering the region is controlled by the boundary voltages V_{BF} and V_1 (the voltage on the first anode).

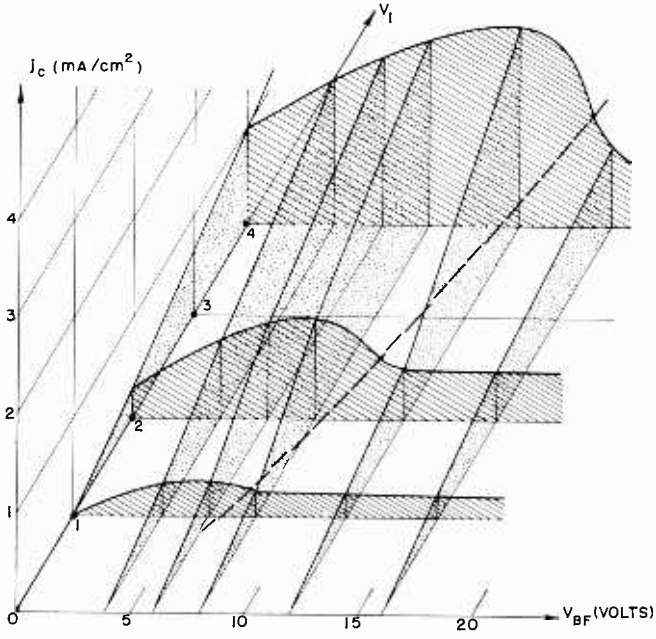

Fig. 2—Current density in the center of the beam versus V_1 and V_{BF} (with $V_2 = 30$ volts, $V_3 = 40$ volts). "Near" virtual cathode exists in plateau region to right of dashed line.

Figure 2 shows the measured transmitted current density on the beam axis (J_c) as a function of V_{BF} and V_1. For sufficiently high values of V_{BF}, the plateau region indicates the existence of a virtual cathode which limits the current. A similar set of curves (not shown) for the transmitted edge-current density shows no such limiting behavior. The edge current density always increases with either V_{BF} or V_1. This result shows that, in a beam of finite diameter, the virtual cathode does not extend to the beam edge. The voltage distribution across the beam is determined, however, by both the center current density and the edge current density.

The behavior of the virtual cathode can be conjectured from the evidence of Figure 2. For a fixed V_1, the center current density (J_c)

shows a very gradual decrease with increasing V_{BF} within the virtual-cathode plateau region. The action of increasing V_{BF}, insofar as V_{BF} establishes the input boundary voltage, is to move the virtual cathode away from the actual cathode, thus resulting in increased center current. At the same time, however, V_{BF} injects more current into the virtual-cathode region, thus tending to move the virtual cathode towards the actual cathode. The net result of these two tendencies

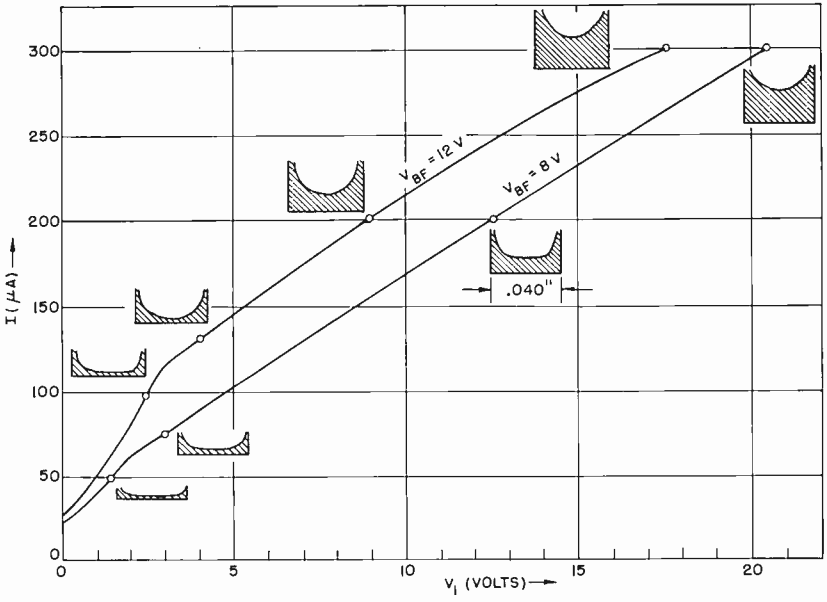

Fig. 3—Transmitted total current versus V_1 for $V_{BF} = 8$ and 12 volts (with $V_2 = 30$ volts, $V_3 = 40$ volts). "Near" virtual cathode exists to the left of the "knees". The shaded figures adjacent to various I–V points represent the measured current-density variation with radius across the .040-inch diameter beam.

appears to be a slow motion of the virtual cathode towards the actual cathode, since the center current decreases slowly with increased V_{BF}. The curves also show that, for a fixed V_{BF}, J_c increases rather slowly with V_1 within the plateau region. Because V_1 establishes the exit-plane voltage, one concludes that increased V_1 draws more core current from the virtual cathode and at the same time moves the virtual cathode toward the actual cathode. Since the distance between the virtual cathode and the exit plane increases with V_1, the rise of J_c is much slower than $V_1^{3/2}$.

As V_1 is further increased, at fixed V_{BF}, the rate of rise of J_c with V_1 first increases and then decreases. The virtual cathode gradually

vanishes in the region of steepest rise of J_c with V_1. The dashed line of Figure 2, which roughly delineates the virtual-cathode region, connects the "knee" points of Figure 3. This figure shows *total* transmitted current (I) versus V_1 for two values of V_{BF}. The virtual cathode exists on the low-V_1 side of the knee points. The change in slope of the I–V_1 curve at the knee reflects the rapid change in core current density J_c with V_1 along the dashed line of Figure 2.

Figure 3 also shows the current-density profiles across the transmitted beam for various points on the I–V_1 curves. For the present case of a "near" virtual cathode, the hollowness of the beam is due mainly to the fact that V_{BF} draws current from the edge of the cathode. The transmitted core current is only slightly reduced by reflection from the virtual cathode. On the other hand, for the "remote" virtual cathode described below, there is a well-defined drop in core current when the virtual cathode is formed.

The current-density profiles were also investigated visually by means of a fine-grained phosphor screen. The beam images agreed qualitatively with the results of the pinhole measurements.*

"Remote" Virtual Cathode

Figures 4 and 5 show the total transmitted current (I) versus V_1, for $V_{BF} = 8$ and 12 volts, respectively. These data were taken for sufficiently low values of V_2 and V_3 (14 volts) so that a virtual cathode would be formed between the second and third anodes upon the injection of sufficient current. The current injected into this region was not controlled by the boundary voltages V_2 and V_3, but rather by V_{BF} and V_1.

It can be observed from the I–V_1 curves that an abrupt transition into a virtual-cathode state and hysteresis occur provided V_{BF} is not too high. The abrupt transitions were found to occur at V_1 values which depended on the time rate at which V_1, and hence the injected current, was increased. The slower the rate of increase, the higher was the transition value of V_1.

This behavior is indicative of an ion effect. In the case of $V_{BF} = 8$ volts (Figure 4), the value of V_1 necessary to produce a virtual cathode is well above the ionization potential, and ions can be trapped in the V_2–V_3 region. Thus the electron current necessary to form a virtual cathode is increased (see dotted line of Figure 4).

As V_{BF} is made more positive, the value of V_1 necessary to draw

* These measurements were made by J. Berghammer.

limiting current is lower. If this value of V_1 becomes lower than the lowest ionization potential of the ambient gases, then for the given low values of V_2 and V_3 no ion trapping can occur and no hysteresis occurs. As further evidence of the absence of hysteresis in an ion-free finite-diameter beam, it was found that the hysteresis and the abrupt transition vanished when the collector was depressed to 5 volts

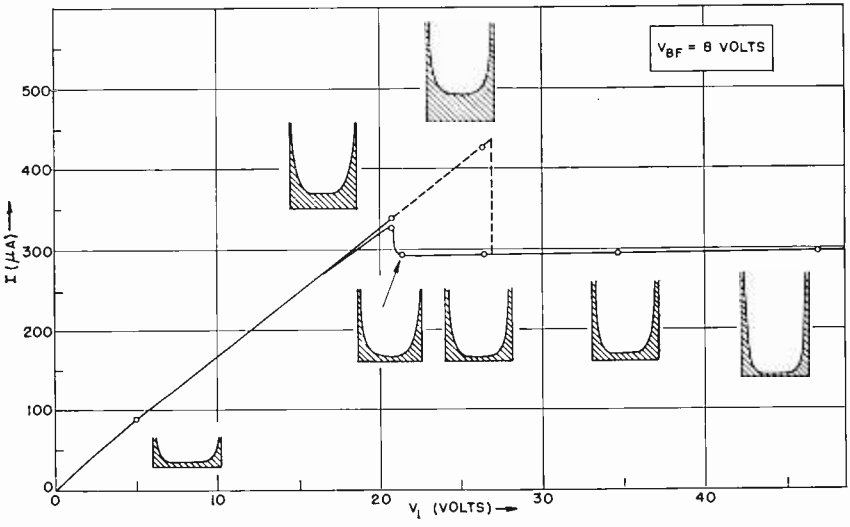

Fig. 4—Transmitted total current versus V_1 for $V_{BF} = 8$ volts (with $V_2 = V_3 = 14$ volts). "Remote" virtual cathode between A_2 and A_3 forms for values of V_1 above 21 volts. The dashed lines indicate virtual cathode formation when ion trapping occurs.

and ions were drained out of the virtual cathode trap. Also, measurements made on a similar gun maintained under high-vacuum conditions (pressure below 10^{-8} mm Hg) showed neither abrupt transitions nor hysteresis under any conditions. Since, in the "near" virtual cathode case discussed above, ions were drained to the cathode, it is clear why no hysteresis or sudden transitions were exhibited (see Figure 2). Thus it appears that for a *finite diameter* beam, hysteresis and the associated sharp transitions occur *only* if sufficient positive ions are trapped. This is in contrast to both the case of the ion-free one-dimensional beam and the case of complete space-charge neutralization. For the ion-free case, hysteresis and sharp transitions are necessitated by the axial boundary conditions of a one-dimensional beam. For the finite diameter beam, however, the potential minimum which forms when sufficient current density is injected has an additional degree

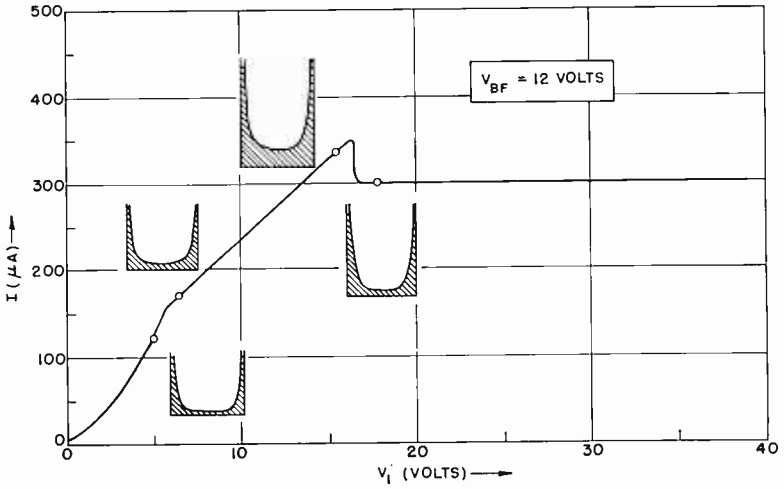

Fig. 5—Transmitted total current versus V_1 for $V_{BF} = 12$ volts (with $V_2 = V_3 = 14$ volts). A "remote" virtual cathode exists for values of V_1 above 17 volts; a "near" virtual cathode exists for values of V_1 below 6 volts. This "near" virtual cathode region is absent in Figure 4, since with $V_{BF} = 8$ volts insufficient current is injected into the cathode–first-anode region.

of freedom. Here, both the radial extent and the axial position of the virtual cathode can vary. Thus it is conjectured that the transverse boundary conditions remove the mathematical degeneracy and thereby remove the hysteresis and sharp transitions. For the case of *complete space-charge neutralization* it has been shown that virtual-cathode effects are eliminated.[4] Partial neutralization and ion trapping is an intermediate state and leads to hysteresis effects as described above.

ACKNOWLEDGMENT

The authors wish to thank S. Bloom and J. Berghammer for many discussions which materially aided the understanding of the phenomena described here. We are also grateful to E. E. Thomas for taking much of the data.

---

[4] K. G. Hernqvist, "Space Charge and Ion Trapping Effects in Tetrodes," *Proc. I.R.E.*, Vol. 39, p. 1541, Dec. 1951.

STUDIES OF THE ION EMITTER BETA-EUCRYPTITE*

By

Fred M. Johnson

RCA Laboratories
Princeton, N. J.

Summary—*The physical properties of β-eucryptite associated with the phenomenon of lithium-ion emission were investigated. Measurements were made of its work function and activation energy as a function of lithium content. The electronic conduction mechanism in β-eucryptite was explored and the application of this material for plasma synthesis (Li+ — e* plasma*) was demonstrated.*

Introduction

THE PRESENT INTEREST in ion emitters, in general, arises from such diverse fields as plasmas, thermionic energy converters, space propulsion schemes and space-charge neutralization of electron beams. For these applications a convenient and copious source of ions is of prime importance.

Beta-eucryptite is a ceramic material in which lithium is stored. When heated to temperatures in the range 1000-1300° C, lithium ions are emitted. This compound is one of the most copious ion emitters known.

The Structure of Beta-Eucryptite

Beta-eucryptite ($Li_2O \cdot Al_2O_3 \cdot 2SiO_2$) is one of the aluminosilicates. The crystal structure, shown in Figure 1, is similar to the high-temperature quartz structure (see Table I), yet differs from it in that half of the silicon atoms are replaced by aluminum atoms in alternate layers along the *c*-axis. The lithium ions are in the interstitial space on the level where the substitution takes place, surrounded by four oxygen atoms. This arrangement has a very satisfactory Pauling electrostatic bond structure. The oxygens surrounding the Li are shared by one Si and one Al, so that they receive electrostatic bonds

* The research reported was sponsored by the Advanced Research Project Agency, Department of Defense, under contract No. AF19(604)6175 and the U.S. Army Signal Corps, under contract No. DA36-039-sc-78151.

of strength 1¾ from this source. The tetrahedrally coordinated Li atoms contribute the remaining ¼.[1,2]

Beta-eucryptite is *extremely* anisotropic in thermal expansion, with coefficients of thermal expansion in the *a* and *c* crystallographic direc-

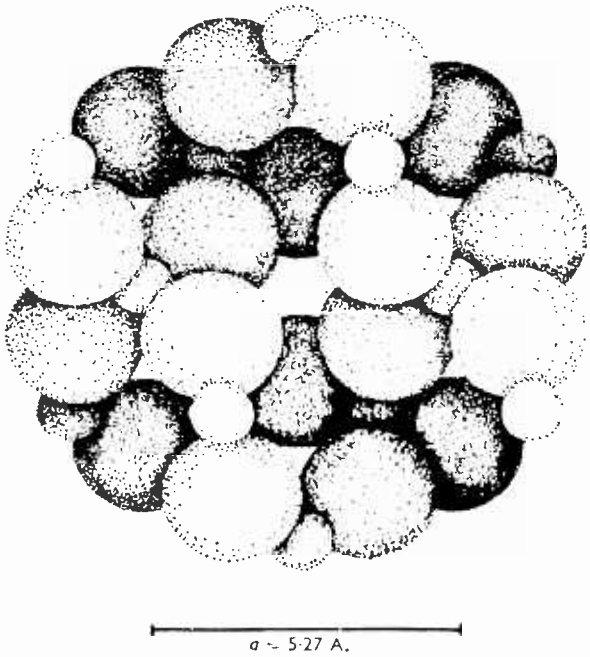

Fig. 1—Crystal structure of β-eucryptite viewed along *c* axis. Large spheres represent oxygen atoms, small spheres either Si or Al atoms. The lithium ions are situated in center opening. (From H. G. F. Winkler, *Acta Crystallographica*, see Ref. (1).)

tions of $+82 \times 10^{-7}$ and -176×10^{-7} per °C, respectively.[3] The net effect of this anisotropic expansion is to enlarge the centrally located openings of the crystal lattice in which the lithium ions are situated. This would tend to explain, in part, why lithium ions are readily

---

[1] H. G. F. Winkler, "Synthese und Kristallstruktur des Eukryptits, $LiAlSiO_4$," *Acta. Cryst.*, Vol. 1, p. 27 (1948); and, "Struktur und Polymorphie des Eukryptits (Tief-$LiAlSiO_4$) (Betrachtungen zur Polymorphie II.)," *Heidelberger Beiträge zur Mineralogie und Petrographie*, Vol. 4, p. 233 (1954).

[2] M. J. Buerger, "The Stuffed Derivatives of the Silica Structures," *Amer. Mineralog.*, Vol. 39, p. 600 (1954).

[3] F. H. Gillery and E. A. Bush, "Thermal Contraction of β-Eucryptite ($Li_2O{\cdot}Al_2O_3{\cdot}2SiO_2$) by X-Ray and Dilatometer Methods," *Jour. Amer. Ceramic Soc.*, Vol. 42, No. 4, p. 175, 1959.

emitted from this type of crystal structure if the material is sufficiently heated.

It should be pointed out that there are two polymorphs of eucryptite labeled α and β. The alpha form is a mineral that occurs in nature; it can be synthesized only by the hydrothermal growth method.[4] The beta form does not occur in nature but can be readily synthesized by heating, for example, stoichiometric compositions of lithium carbonate,

Table I

	Low-Temperature Quartz	High-Temperature Quartz	β-Eucryptite	
			Winkler[1]	Buerger[2]
a	4.913 Å	5.02 Å	5.27	2×5.275
c	5.404 Å	5.48 Å	2×5.625	2×5.61
c/a	1.10	1.09	2×1.067	1.064
cell constant	3 SiO_2	3 SiO_2	3 $LiAlSiO_4$	6 $LiAlSiO_4$
density	2.649	2.518 at 600°C	2.352 (actual)	2.305 (actual)
			2.305 (x-ray)	2.30 (x-ray)

silicon dioxide and aluminum oxide, which upon heating reduces the carbonate to the oxide;

$$Li_2CO_3 + 2SiO_2 + Al_2O_3 \xrightarrow{\text{heat}} Li_2O \cdot Al_2O \cdot 2SiO_2 + CO_2\uparrow. \qquad (1)$$

The $\alpha \rightarrow \beta$-eucryptite inversion was investigated by Isaacs and Roy[4] in the pressure range 4000 to 13,000 p.s.i. and temperatures ranging from 860 to 890°C. The $\alpha \rightarrow \beta$ inversion temperature extrapolated to zero pressure was found to be 848° ± 5°C. The inversion process, however, is only of academic interest since no hydrothermal growth techniques have thus far been employed in the work described in this paper.

Studies of phase equilibrium[5,6] in the system lithium metasilicate-β-eucryptite as well as studies of the compositional and stability rela-

[4] T. Isaacs and R. Roy, "The α-β Inversions in Eucryptite and Spodumene," *Geochimica et Cosmochimica Acta*, Vol. 15, p. 213, 1958.

[5] R. A. Hatch, "Phase Equilibrium in the System $Li_2O \cdot Al_2O_3 \cdot SiO_2$," *Amer. Miner.*, Vol. 28, p. 471 (1943).

[6] M. K. Murthy and F. A. Hummel, "Phase Equilibria in the System Lithium Metasilicate—β-Eucryptite," *Jour. Amer. Ceram. Soc.*, Vol. 37, No. 1, p. 14, Jan. 1954.

tionships[7,8] among the aluminosilicates have been made by a number of workers. The phase diagram for the ternary system $Li_2O \cdot Al_2O_3 \cdot SiO_2$ is shown in Figure 2, after Murthy and Hummel.[6]

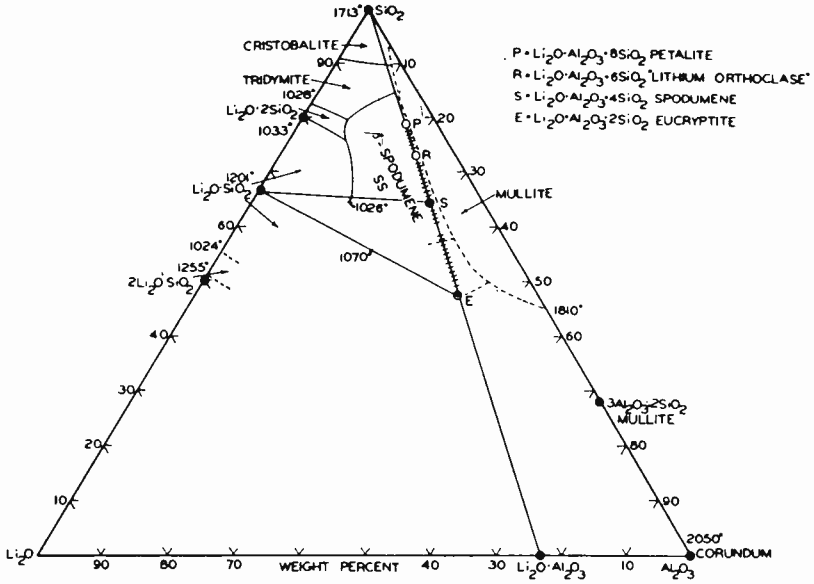

Fig. 2—The system $Li_2O \cdot Al_2O_3 \cdot SiO_2$ showing lithium metasilicate β-eucryptite join. (From M. K. Murthy and F. A. Hummel, *Jour. Amer. Ceramic Soc.*, see Ref. (6).)

Lithium Ion Emission

Blewett and Jones[9] were among the first to recognize the excellent ion-emission properties of β-eucryptite and in fact hypothesized that in the crystal lattice the lithium ions are situated in large "holes" (even though the detailed crystal structure was not then known). The alkali ion is therefore comparatively free to migrate within the lattice, and with the aid of small electrolytic potentials which are bound to be present, it migrates to the surface and is emitted easily as indi-

---

[7] R. Roy, D. M. Roy, and E. F. Osborn, "Compositional and Stability Relationships Among the Lithium Aluminosilicates: Eucryptite, Spodumene, and Petalite," *Jour. Amer. Ceram. Soc.*, Vol. 33, No. 5, p. 152, 1950.

[8] R. Roy and E. F. Osborn, "The System Lithium Metasilicate–Spodumene–Silica," *Jour. Amer. Chem. Soc.*, Vol. 71, No. 6, p. 2086 (1949).

[9] J. P. Blewett and E. J. Jones, "Filament Sources of Positive Ions," *Phys. Rev.*, Vol. 50, p. 464, Sept. 1, 1936.

cated by the low work function. The "work function" for ions determined at this laboratory[10] on a large number of samples of sintered β-eucryptite were all in the neighborhood of 2.0 volts in contrast to 2.7-4.0 volts as determined by Blewett and Jones[9] and 2.9 volts by Couchet.[11] The discrepancies between these values of the work function are presumably due in part to the fact that neither Blewett and Jones nor Couchet took the Schottky effect into account in the analysis of their data.

Studies of ion emission were undertaken by means of simple diode structures with an emitter collector spacing of about 1 millimeter. Diode characteristics of current versus voltage provide a great deal of important information. The ion emission data could be fitted to a Schottky-type equation:

$$I_s = AST^2 \exp\left(-\frac{e\phi}{kT} + \frac{e(eE)^{1/2}}{kT}\right) \text{ amperes}, \qquad (2)$$

where A = emission constant in amperes/cm^2,

S = emitter area in cm^2,

T = emitter temperature in °K,

ϕ = emitter work function for ions,

k = Boltzmann's constant,

e = electronic charge,

E = potential gradient = V/d for plane electrode, where d = spacing.

Equation (2) may be rewritten as

$$\log_{10} I_s = \log_{10} I_{s0} + \frac{1.9}{T_s}\left(\frac{V}{d}\right)^{1/2}, \qquad (3)$$

where V is measured in volts, d in centimeters, I_{s0} is the saturated ion emission current at zero applied voltage, and T_s is the apparent Schottky temperature. The second term in Equation (3) arises from the fact that an ion induces an image charge in the material as it leaves the surface and hence modifies the potential distribution immediately outside the material. (The formula, as given, actually was

[10] F. M. Johnson, "Electronic Phenomena in Alumina Silicate Ion Emitters," Meeting of the *American Physical Society*, New York, N. Y., Jan. 1960.

[11] G. Couchet, "Contribution a L'étude Des Sources, Solides Ioniques," *Ann. de Physique*, Vol. 9, p. 731, 1954.

derived for electron emission from a metal surface.[12] However, except for multiplicative constant in the second term, the formula appears to apply equally well for ion emission from a dielectric or semiconductor.) A plot of $\log I_s$ versus $V^{1/2}$ should therefore yield a straight line in the high-voltage region. This was indeed found to be the case. Samples were prepared with less than the required stoichiometric composition of Li_2O, in the ratios 30%, 50%, and 70% of the stoichio-

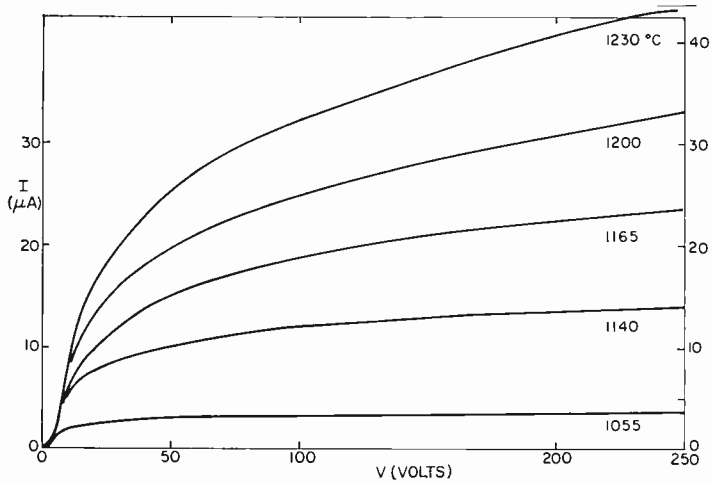

Fig. 3—Typical diode characteristics of current versus voltage.

metric value. Figure 3 shows the I–V diode characteristics for a sample which had the 50% composition. The corresponding Schottky plot of these data ($\log I_s$ versus $V^{1/2}$) is shown in Figure 4. From the intercept of the resulting straight lines on the ordinate axis, the saturation currents I_{s0} for various emitter temperatures were obtained. These saturation currents were then used to verify the applicability of the Richardson-Dushman equation, $I_{s0} = AT^2 \exp[-e\phi/kT]$, to the ion emitters. Figure 5 shows the necessary plot of I_{s0}/T^2 versus $1/T$. A value of the work function, ϕ, was obtained from the slope of the straight line. For each of the three compositions, the work function was found to be 2.0 volts. The results of an analysis of the 30% and 70% compositions are also shown in Figure 5. Since the slopes for each of these sintered samples are approximately equal, it appears that

[12] See, e.g., G. Hermann and S. Wagener, *The Oxide-Coated Cathode*, Chapman and Hall Ltd., London (1951).

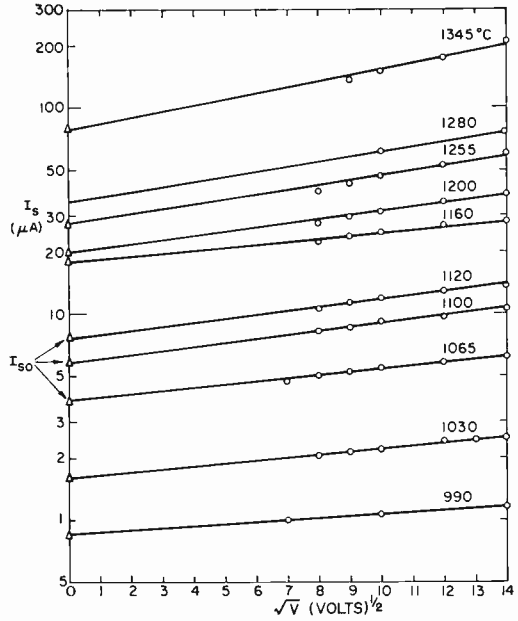

Fig. 4—Typical Schottky plot (50% lithium sample).

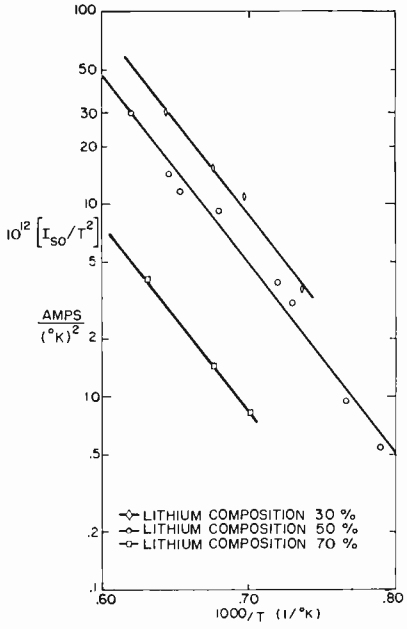

Fig. 5—Richardson–Dushman plot for three different lithium contents. Values of I_{so}/T^2 for the 70% sample have been divided by 10 to avoid overlap with the other lines.

the work function is independent of the lithium content in the temperature range* 1400-1600°K (for these concentration ranges).

In addition to yielding the field-free emission currents, the Schottky plots may also be used to obtain an apparent temperature from the slope of the straight lines in Figure 4. The values of this apparent Schottky temperature, T_s, range from 210 to 430°K depending on the emitter temperature (see Figure 6), whereas the actual temperature

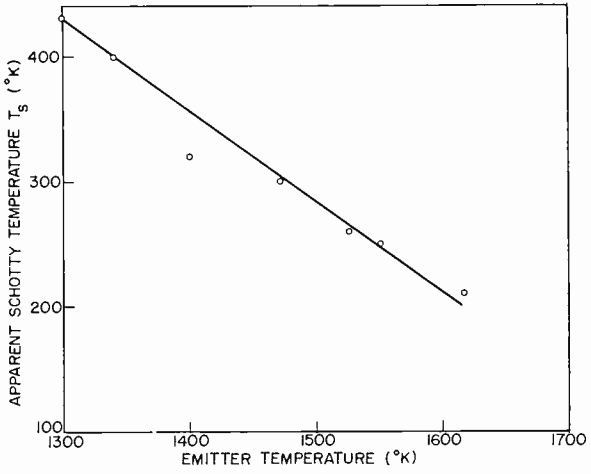

Fig. 6—Apparent Schottky temperature, T_s, versus emitter temperature.

of the ion emitter was in the 1300 to 1600°K range. This anomalous Schottky effect was previously observed by Nergaard[13] for the case of electron emission from oxide cathodes and was, in fact, of the same order of magnitude. The anomalous temperature dependence of the apparent Schottky temperature may be due to the temperature dependence of the dielectric constant of the semiconductor, a quantity heretofore neglected in the analysis.

The ion-emission data in the high-voltage region and 1300 to 1600°K temperature range may be fitted to the following equation, for the 50% lithium sample:

$$\log_{10} I_s = \log_{10} I_{s0} + \frac{1.9}{1375 - 0.726T} \left(\frac{V}{d} \right)^{1/2}.$$

[13] L. S. Nergaard, "Studies of the Oxide Cathode," *RCA Review*, Vol. XIII, No. 4, p. 464, Dec. 1952.

* The higher work functions reported by Blewett and Jones[9] (2.7-4.0 volts) were deduced from Richardson–Dushman plots at temperatures below about 1400°K.

In the low-voltage region, the lithium ion emission should be space-charge limited, which implies that the ion current is governed by the Child–Langmuir equation:

$$I = \frac{2.33 \times 10^{-6} V^{3/2} S}{d^2} \sqrt{\frac{m}{M}} \quad \text{amperes}$$

where m/M is the ratio of the electron mass to the ion mass, d is the

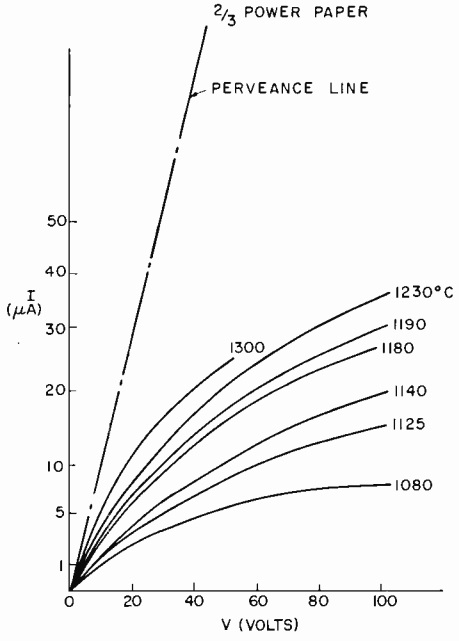

Fig. 7—Typical plot of (current)$^{2/3}$ versus voltage for a lithium ion emitter. The perveance line applies to the Child–Langmuir equation.

electrode spacing in centimeters and S is the emitting area in cm^2. A typical graph of I versus V, plotted on 2/3 power paper, is shown in Figure 7. In the space-charge-limited region, the experimental points are expected to lie on the theoretical curve, which is drawn as a dashed line in Figure 7 and marked "perveance" line. The deviations from the straight line are interpreted as being due to the resistance of the thermionic emitter. The ion current is then given by

$$I = P(V - IR)^{3/2},$$

where P is the perveance of the diode, i.e.,

$$P = \frac{2.33 \times 10^{-6} S}{d^2} \sqrt{\frac{m}{M}},$$

V is the plate voltage, and R is the internal resistance of the ion emitter. The values of this internal resistance may be inferred from

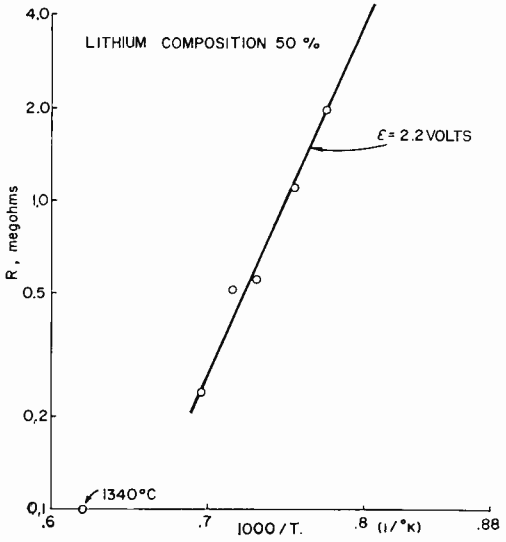

Fig. 8—Semi-log plot of the internal resistance, R, of the ion emitter versus reciprocal temperature.

Figure 7 for each emitter temperature by noting the voltage drop across the emitter for a given emission current. For example, at a temperature of 1300°C, the voltage drop across the sample is 4 volts at 5 microamperes and 8 volts at 10 microamperes. This indicates that for the low-voltage region, which is sufficiently removed from saturation, the internal resistance is ohmic. It should be noted that for higher voltages a fraction of the applied voltage appears across the emitter and not across the vacuum diode space. Having obtained the internal resistance as a function of temperature, one may then relate $R(T)$ to $\exp(-\epsilon/kT)$ where ϵ is an activation energy. Plots of log R versus $1/T$ gave straight lines for each of the samples tested. From the slope of these lines, activation energies were obtained which ranged in value from 2.0 to 2.4 volts depending on the composition of the sintered material. A logarithmic plot of R versus $1/T$ is shown for the 50% composition in Figure 8. Deviation from a straight line

occurs near the melting point of β-eucryptite (note the point at 1340°C), which implies a decrease in activation energy at these higher temperatures.

Plasma Synthesis Studies

For production of a plasma synthesized from electrons emitted from an electron emitter and lithium ions emitted from β-eucryptite, the following requirements must be satisfied: (1) the ion space-charge

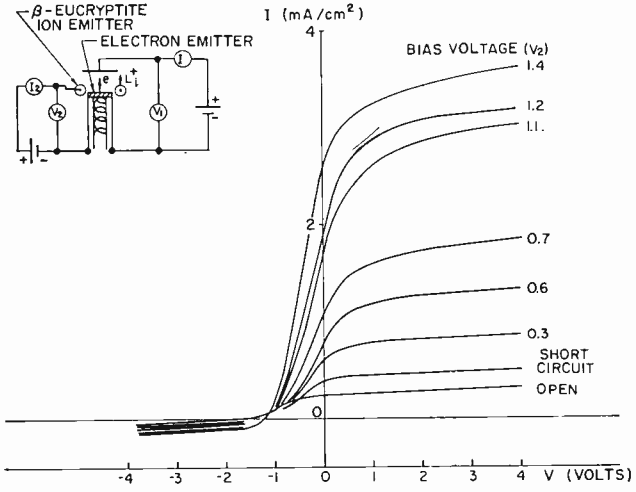

Fig. 9—Schematic diagram of the experimental arrangement of the plasma gun and its characteristic curves.

density must equal the electron space-charge density or, roughly, the ratio of ion to electron current density must be equal to the square root of the ratio of electron to ion mass; (2) the ions and electrons must enter the vacuum at the same potential level.[14] This latter condition requires a biasing voltage, V_2, for the ion emitter with respect to the electron emitter.

An experimental arrangement for plasma synthesis, first described in 1960,[10] is shown schematically in Figure 9. The electron emitter was a dispenser-type cathode which was in close contact with, and surrounded by, a ring-shaped β-eucryptite specimen. The structure was similar to the one shown in Figure 10 but without the center core. A heater was situated inside this specimen to allow independent heating of the lithium ion emitter. A bias voltage, V_2, was applied to the ion-

[14] K. G. Hernqvist, "Plasma Synthesis and Its Application to Thermionic Power Conversion," *RCA Review*, Vol. XXII, No. 1, p. 7, March 1961.

emitting ring. This so-called "plasma gun" yielded increased current output as the bias voltage was increased. Experimental results are shown in Figure 9. Current densities of the order of 3 ma/cm^2 were obtained. More recently, the author demonstrated plasma synthesis using lithium ions in an improved structure (Figure 10) together with a refined measuring technique. The plasma emitter was constructed as follows: the electrons were emitted from a ring-shaped cathode; the lithium ions originated from a β-eucryptite plug placed at the center of the cathode as well as from a ring-shaped β-eucryptite

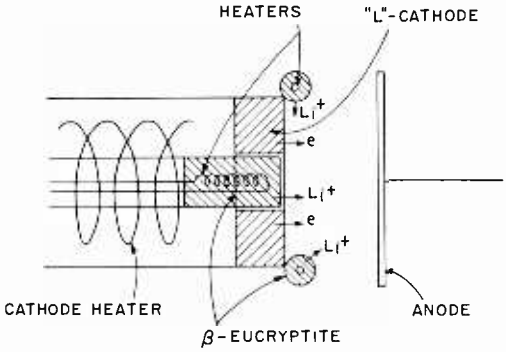

Fig. 10—Diagram of an "Electron-Lithium ion Plasma" synthesizer.

specimen placed adjacent to and outside the cathode. Both ion emitters could be heated independently.

In order to avoid spurious (60-cycle) signal pick up, electric heating of the ion emitters was performed by half-wave rectified current which was applied synchronously to all three heaters. All current and voltage measurements were then taken during the zero current interval of the heating cycle. The circuit diagram of the experimental arrangement is shown in Figure 11.

In order to obtain maximum ion current values, not limited by ion diffusion, all measurements were taken with pulses of 60 microseconds duration. A cutoff bias voltage of about —3 volts was applied to both ion emitters in series with a positive voltage pulse whose amplitude could be varied. No anode current was observed unless the pulse voltage on the ion emitters exceeded the cutoff bias voltage. Measurements were taken of anode current pulses, I_1, as a function of emitter pulse voltage, V_2, for a number of fixed values of anode potential, V_1. A typical oscillogram showing anode current and the bias pulse is shown in Figure 12.

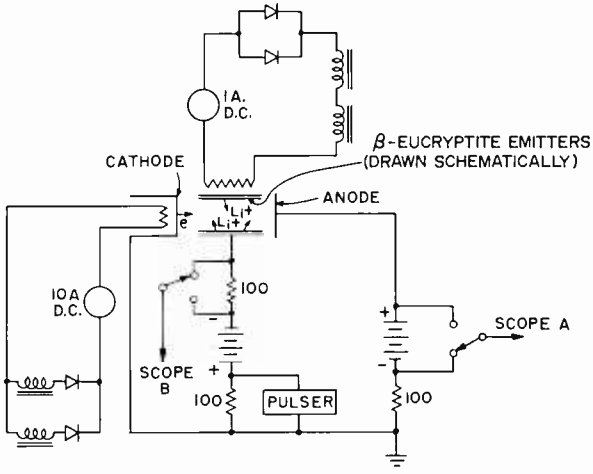

Fig. 11—Experimental arrangement for measuring "Electron-Li+" plasma.

Note the exponential rise and decay in anode current at the rising and trailing edges of the pulse bias voltage. These phenomena are presumably associated with lithium ion emission processes from β-eucryptite. The results of a number of such measurements are summarized in Figure 13. The important point to note in Figure 13 is that there is a fairly steep rise in anode current at the threshold bias voltage, V_2, which is independent to a first approximation of the anode voltage, V_1.

All data were taken while the tube was immersed in a magnetic field of 100 to 200 gauss perpendicular to the emitter surface, in order to minimize electron interception to the ion emitter. The one exception is so indicated. The enhanced anode current at threshold bias is here particularly pronounced. At the critical bias voltage, corresponding to peak anode current, a lithium-ion–electron plasma is presumably

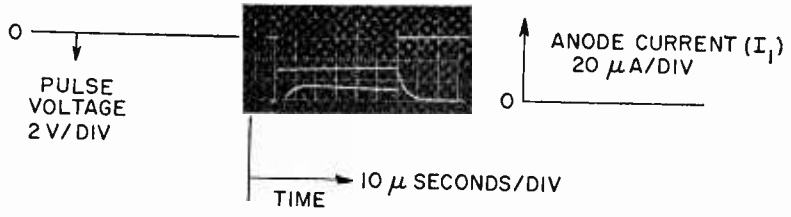

Fig. 12—Oscillogram of pulse voltage and anode current.

being synthesized. A similar behavior has been observed for synthesized cesium plasma.[14]

Energy balance requires that

$$V_{\text{bias}} = V_i - \phi_c + RI_p, \tag{7}$$

where V_i = ionization potential of lithium (5.4 volts), ϕ_c = electron

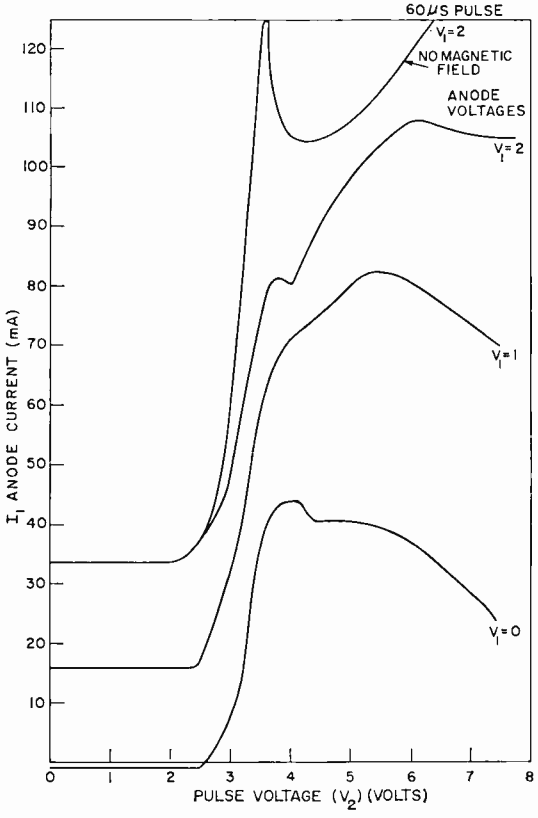

Fig. 13—Summary of data of anode current pulse height vs. pulsed bias voltage with anode voltage, V_1, as a parameter.

emitter work function, R = resistance of β-eucryptite at the operating temperature, and I_p = lithium ion current. It was not possible to perform an accurate check of this equation due to a lack of knowledge of the parameters ϕ_c, R, and I_p. However, on the basis of reasonable assumptions, a pulse bias voltage of about 4 volts is not unreasonable.

Mass-Spectrometer Studies

A knowledge of particle emission, involving the parameters of mass, charge, and the temperature at which this emission takes place, was thought essential to this type of investigation. Hence a mass-spectrometric analysis was performed on a sample of β-eucryptite weighing about 37 milligrams and consisting of 3 crystalline chips.

Both the neutral and the ion emission from the sample were observed at temperatures ranging up to about 1700°K, extending over a period of 433 hours of heating.

The results obtained may be summarized as follows:

(1) The β-eucryptite was observed to contain other alkali metals as impurities. These are tabulated below in order of concentration. Li has been arbitrarily taken as unity.

Li 1

Na 1.5×10^{-1}

K 9×10^{-3}

Rb 3×10^{-5}

Cs 5×10^{-8}.

No other impurities were observed which could be unambiguously ascribed to the sample.

(2) All of the impurities were observed before any Li was emitted. K was nearly exhausted by the time (and temperature) that Li was observed in an appreciable amount. Na persisted considerably longer. Rb lasted somewhat longer yet. Cs disappeared prior to the onset of Li emission.

(3) 1.3×10^{16} ions of Li were actually collected. Assuming a spectrometer efficiency of about 10^{-2}, approximately 10^{18} ions were emitted from the sample. This represents about 10% of the original sample atoms, and implies that the major fraction of its lithium content is emitted as ions.

(4) While the exact ratio of Li ions to Li neutral atoms emitted is not known, it appears that the ions predominate, probably by a large factor.

(5) A search was made for negative ions, particularly O^-. None were observed. Any neutral oxygen emitted could not have amounted to more than a few parts per million.

Fundamental Processes in the Beta-Eucryptite Ion Emitter

Since this material is one of the best ion emitters known, various approaches were explored in order to best utilize it for plasma generation. To accomplish this, however, it is essential to understand the conduction mechanism inside the material associated with the lithium ion emission. For every positive ion that is emitted, there must be a corresponding negative charge which leaves the β-eucryptite slab at the back surface to account for the overall charge balance. Since this material exhibits a large internal resistance, it is of prime importance to determine this conduction mechanism. There is a finite probability that negative oxygen ions are the charge carriers for the above mentioned "return" current. If this were the case, then it might be expected that the lithium ion emission is accompanied by the liberation of oxygen. The detection of such oxygen would consequently verify this hypothesis; however, its absence would not entirely rule out the hypothesis since it is conceivable, but not too likely, that a finite quantity of oxygen could be trapped in the crystal lattice, particularly since the oxygen ion has a fairly large atomic radius.

As a result of these mass spectrometer measurements, however, it appears unlikely that the oxygen ions are responsible for the negative ion conduction.

Another possible tool for determining the conduction mechanism is the measurement of the conductivity under pulsed conditions. The results of such measurements on a slab of sintered β-eucryptite are shown in Figures 14, 15 and 16. Figures 14 and 15 differ in the pulse width and applied voltage. The electrodes were tungsten and were directly applied to the slab under pressure (see Figure 17). All the measurements were taken with the sample at a temperature of about 900°C. Note the relatively large ion current that can be transmitted for 8 microseconds. The "ion-depletion effect" becomes quite noticeable as the pulse length is increased to 60 microseconds, and in fact at a pulse length of 1/60 of a second (see Figure 16) the polarization phenomenon at the conclusion of the applied pulse voltage is also quite apparent. This polarization effect is most noticeable if one expands the time scale, as shown in Figure 16. Note the change in polarity when the applied voltage is discontinued and the subsequent decay towards zero current. This curve is characteristic of the self-diffusion of lithium ions inside the slab and hence is a direct measure of the self-diffusion coefficient. The rather long time constant (of the order of 10 milliseconds) would seem to be consistent with the model that lithium ions play a dominant role in the conduction process.

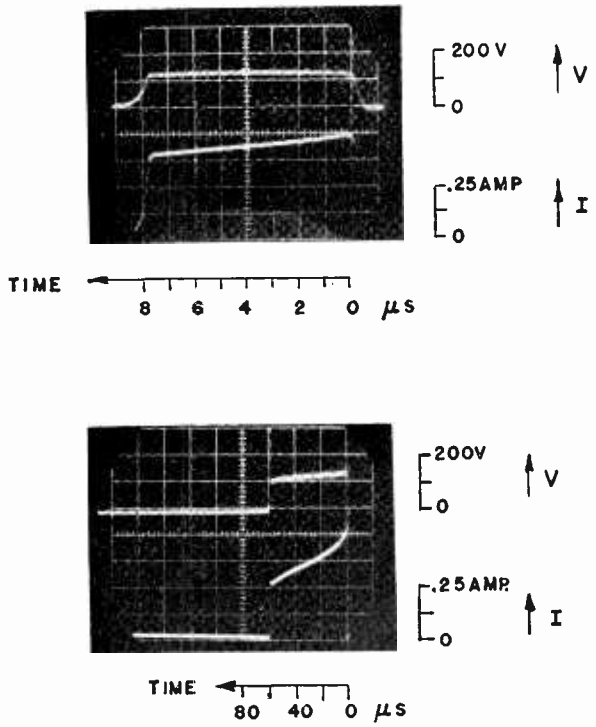

Fig. 14—Oscillograms of pulse current measurements on a Beta-eucryptite slab (short pulse duration, large current).

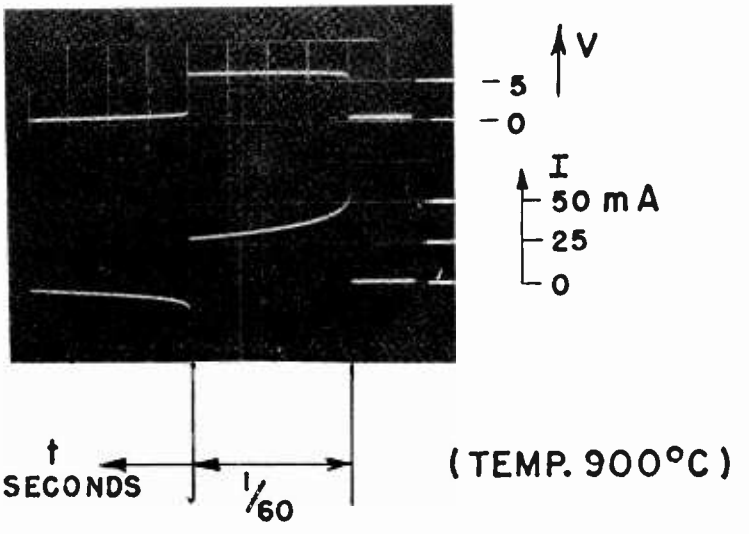

Fig. 15—Oscillogram of a pulse measurement on Beta-eucryptite slab (long pulse duration, low current).

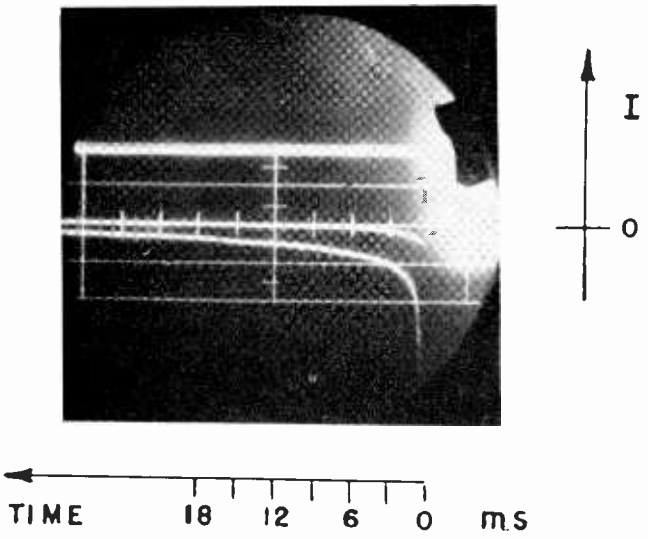

Fig. 16—Oscillogram showing characteristic current decay following cessation of several second long voltage pulse.

Although these studies are revealing, one ultimately must turn to single-crystal specimens in order to obtain definitive measurements on the conduction mechanism. Consequently, efforts were made to grow single crystals of β-eucryptite. These efforts have yielded crystal conglomerates which were essentially single crystals; i.e., x-ray analysis, using Laue patterns, showed that the symmetry axis was predominantly pointing in a unique direction throughout the crystal specimen. Slabs were cut from those specimens in such a way that the crystal symmetry axis was either in the plane of, or perpendicular to, the emitting surface. These slabs were then mounted in vacuum diode structures as shown in Figure 18. Potential probes as well as platinum–platinum-rhodium thermocouples were attached at the bottom and top surfaces of the crystal slab. The bottom of the crystal could be indirectly heated. Temperature measurements made with an optical pyrom-

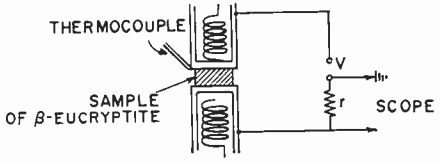

Fig. 17—Schematic diagram showing experimental arrangement for pulse measurements.

eter were in excellent agreement with those made with the thermocouple attached to the bottom surface. Since the top thermocouple was not actually in contact with the top of the crystal but slightly to one side and attached to the platinum sheet which pressed on the top of the crystal surface, there existed a thermal gradient between this top thermocouple and the crystal slab that prevented a direct reading of the surface temperature of the crystal. Nevertheless, it was possible to make qualitative measurements of the thermoelectric power (Seebeck voltage) of β-eucryptite by means of a high-impedance input voltmeter. The results of these measurements showed unambiguously

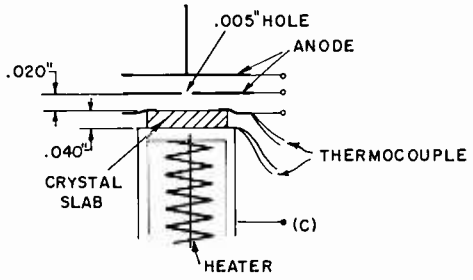

Fig. 18—Diode structure for studying single crystal of Beta-eucryptite.

that the crystal was *p-type* along the symmetry axis. This result is consistent with the physical model of having a concentration gradient of positive charge (lithium ions) established along the thickness of the slab.

If lithium ions are situated in channels which run parallel to the symmetry axis, then enhanced lithium ion emission is expected from single crystals whose emitting plane is *perpendicular* rather than parallel to the symmetry axis. This hypothesis has indeed been experimentally confirmed.

Life of Ion Emitter

At constant temperature and applied voltage, lithium ion emission slowly decreases as a function of time. On the assumption that most of available lithium inside β-eucryptite can be removed in the form of lithium ions, one cubic centimeter of the material, if it were emitting at a uniform rate of 1 ma/cm^2, would last about 450 hours. This finite life time may prove a limitation for some applications, but not necessarily an unsurmountable one. It may be possible to replenish the lithium in the β-eucryptite by diffusion from the back surface.

Conclusion

The use of β-eucryptite as an ion emitter source has been evaluated for possible application in devices requiring its special properties. Because lithium is condensable, localized plasmas can be produced in devices that otherwise require a high vacuum. Thus, it eliminates differential pumping. Ion current densities of 1-5 milliamperes per square centimeter can be drawn continuously, depending on source temperature. The finite life of the emitter due to exhaustion of lithium is a function primarily of the current drawn and should permit several hundred hours of operation if sufficient material can be used. Analogous ion-emitting compounds involving either potassium or sodium also exist.[11] These have not, however, been investigated here.

Acknowledgments

It is a pleasure to acknowledge a number of useful discussions with L. S. Nergaard and K. G. Hernqvist. J. R. Woolston kindly performed the mass spectrographic analysis.

RCA TECHNICAL PAPERS†

Second Quarter, 1962

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Bounds on Burst-Error-Correcting Codes," C. N. Campopiano, *Trans. I.R.E. PGIT* (April) (Correspondence) 1962
"Characteristics and Mode of Operation of Image Orthicons," R. G. Neuhauser, *Proc. NAB Convention* (April) 1962
"On the Control of Electroluminescent Cells by Unipolar Transistors," T. N. Chin, *Jour. Electronics and Control* (April). 1962
"Correction to 'Bandwidth of Hybrid-Coupled Tunnel-Diode Amplifier,'" R. M. Kurzrok, *Proc. I.R.E.* (April) (Correspondence) 1962
"Extreme-Value Theory Applied to False-Alarm Probabilities," T. L. Fine and M. J. Levin, *Trans. I.R.E. PGIT* (April) (Correspondence) 1962
"Ferroelectric Parametric Amplifier," E. Fatuzzo and H. Roetschi, *Proc. I.R.E.* (April) (Correspondence) 1962
"Landau Damping of Space-Charge Waves," J. Berghammer, *Jour. Appl. Phys.* (April) 1962
"Measurement of Photomagnetic Susceptibility by Means of a Vibrating Reed Balance," J. O. Kessler and A. R. Moore, *Rev. Sci. Instr.* (April) 1962
"Measurements of the Dielectric Constant of Rutile (TiO_2) at Microwave Frequencies between 4.2° and 300°K," E. S. Sabisky and H. J. Gerritsen, *Jour. Appl. Phys.* (April) 1962
"A New Approach to Testing High-Voltage Deflection Tubes," H. A. Wittlinger and J. A. Dean, *Trans. I.R.E. PGBTR* (April).. 1962
"Passive Radar Measurements at C-Band Using the Sun as a Noise Source," W. O. Mehuron, *Microwave Jour.* (April) 1962
"Some Characteristics of Tunnel-Diode UHF Mixers," J. Klapper, A. Newton, and B. Rabinovici, *Proc. I.R.E.* (April) (Correspondence) 1962
"Sources of Contamination in GaAs Crystal Growth," L. Ekstrom and L. R. Weisberg, *Jour. Electrochem. Soc.* (April) 1962
"A Transistor Vertical Deflection Circuit Employing Transformer Coupling," L. A. Freedman, *Trans. I.R.E. PGBTR* (April).. 1962
"Tunnel Diode Shift Register," B. Rabinovici, *Proc. I.R.E.* (April) (Correspondence) 1962
"A UHF Television Transmitter," M. L. Kaiser, *CQ* (April) 1962
"Frequency-Independent Voltage Dividers," C. L. Conner, *Electronics* (Reference Sheet) (April 6) 1962
"Microminiature Crystal Oscillator Using Wafer Modules," M. Lysobey, *Electronics* (April 13) 1962
"Effect of a High Electric Field on the Absorption Light by PbI_2 and HgI_2," R. Williams, *Phys. Rev.* (April 15) 1962
"Measurement of the Hall Effect in Metal-Free Phthalocyanine Crystals," S. E. Harrison and Coauthors, *Phys. Rev. Letters* (April 15) 1962

† Report all corrections to *RCA Review*, RCA Laboratories, Princeton, N. J.

"Goto Circuit TD's Give Rapid A/D Conversion," J. Asin, *Electronic Design* (April 26) 1962
"Air Bearing Headwheel Panels for RCA TV Tape Recorders," F. M. Johnson, *Broadcast News* (May) 1962
"Antennas and Transmission Lines," H. H. Beverage, *Proc. I.R.E.* (May) 1962
"Beam-Deflection and Photo Devices," E. G. Ramberg and Coauthor, *Proc. I.R.E.* (May) 1962
"Communication: A Responsibility and a Challenge," E. M. McElwee, *Proc. I.R.E.* (May) 1962
"Determination of Dislocation Densities in Silicon Crystals by an Optical Method," G. Eckhardt and S. R. Lederhandler, *Solid State Design* (May) 1962
"Direct Coupled Coaxial and Waveguide Band-Pass Filters," R. M. Kurzrok, *Trans. I.R.E. PGMTT* (Correspondence) (May).... 1962
"Early History of the Antennas and Propagation Field Until the End of World War I, Part I—Antennas," P. S. Carter and H. H. Beverage, *Proc. I.R.E.* (May) 1962
"Electronics and Health Care," V. K. Zworykin, *Proc. I.R.E.* (May) 1962
"Electronics Fifty Years Later," J. Hillier, *I.R.E. Student Quarterly* (May) 1962
"Energy Conversion Techniques," K. G. Hernqvist, *Proc. I.R.E.* (May) 1962
"Film Recording and Reproduction," M. C. Batsel and G. L. Dimmick, *Proc. I.R.E.* (May) 1962
"Investigation of the Electrochemical Characteristics of Organic Compounds—IX. Pyrimidine Compounds," R. Glicksman, *Jour. Electrochem. Soc.* (May) 1962
"Loudspeakers," H. F. Olson, *Proc. I.R.E.* (May) 1962
"Machines with Imagination," G. A. Morton, *Proc. I.R.E.* (May)... 1962
"Micromodule Reliability," D. T. Levy, *Proc. Electronic Components Conf.* (May) 1962
"Nuclear Radiation Detectors," G. A. Morton, *Proc. I.R.E.* (May).. 1962
"1-KW AM Transmitter with Space-Age Reliability," L. S. Lappin, *Broadcast News* (May) 1962
"Optimum HF Prediction," S. Krevsky, G. B. Bumiller, and Coauthor, *Trans. I.R.E. PGAP* (May) 1962
"On Percentage," E. A. Laport, *Microwave Jour.* (Letters to the Editor) (May) 1962
"Problems Peculiar to Very Large Systems," J. T. Mark and W. G. Henderson, *Vacuum* (May) 1962
"Processing of Sound," H. F. Olson, *Proc. I.R.E.* (May) 1962
"Proving Long Term Reliability for Alloy Transistors," J. R. Willey and Coauthor, *Electronic Industries* (May) 1962
"Stereophonic Phonograph Records, Phase Relations, and Stereo Broadcasting," H. E. Roys, *Broadcast News* (May) 1962
"The Technology of Television Program Production and Recording," J. W. Wentworth, *Proc. I.R.E.* (May) 1962
"TV Broadcasting from Satellites," N. I. Korman, *Signal* (May).. 1962
"Universal Pickup Cartridge for Broadcast Turntables," J. R. Sank, *Broadcast News* (May) 1962
"The Use of Electronic Computers in the Social Sciences," I. Wolff, *Proc. I.R.E.* (May) 1962
"Optical Absorption of Arsenic-Doped Degenerate Germanium," J. I. Pankove and Coauthor, *Phys. Rev.* (May 1) 1962
"Quantum Mechanical Effects in Stimulated Optical Emission," R. C. Williams, *Phys. Rev.* (May 1) 1962
"Measuring Temperature with Diodes and Transistors," L. E. Barton, *Electronics* (May 4) 1962
"More on Photochemical Retina," A. C. Schroeder (Comment), *Electronics* (May 4) 1962
"Generating a Worst-Case Noise Pattern in Coincident-Current Memories," Y. L. Yao, *Electronic Design* (May 10) 1962

"Instantaneous Measurement of Tape Flutter," A. Schulbach, *Electronics* (May 11) 1962
"Degenerate Germanium. I. Tunnel, Excess, and Thermal Current in Tunnel Diodes," D. Meyerhofer, G. A. Brown, and H. S. Sommers, Jr., *Phys. Rev.* (May 15) 1962
"Optical Spectra of Transition-Metal Ions in Corundum," D. S. McClure, *Jour. Chem. Phys.* (May 15) 1962
"Photoemission of Holes from Tin Into Gallium Arsenide," R. Williams, *Phys. Rev. Letters* (May 15) 1962
"Range of Excited Electrons in Metals," J. J. Quinn, *Phys. Rev.* (May 15) 1962
"Analog-To-Digital Converter Uses Transfluxors," N. Aron and C. Granger, *Electronics* (May 18) 1962
"Design Technique for Lumped-Circuit Hybrid Rings," R. M. Kurzrok, *Electronics* (May 18) 1962
"Acceleration of Receiving-Tube Life Using Reaction-Rate Principles," E. R. Schrader, *Proc. M.I.T. Conf. on Physical Electronics* (June) 1962
"Broadband Hybrid-Coupled Tunnel-Diode Amplifier in the UHF Region," H. Boyet, D. Fleri, and R. M. Kurzrok (Correspondence), *Proc. I.R.E.* (June) 1962
"The Calculation of Accurate Triode Characteristics Using a Modern High-Speed Computer," O. H. Schade, Sr., *RCA Review* (June) 1962
"Computer Memories—Possible Future Developments," J. A. Rajchman, *RCA Review* (June) 1962
"Dynamics of Cathode Sublimation in the Intermediate Pressure Range," V. Raag, *Proc. M.I.T. Conf. on Physical Electronics* (June) 1962
"Environmental Testing," R. H. Kirkland, Jr., *Encyclopedia of Electronics* (June) 1962
"Evaporation of Silicon and Germanium by rf Levitation," E. A. Roth, E. A. Margerum, and J. A. Amick (Notes), *Rev. Sci. Instr.* (June) 1962
"Flying-Spot Scanning," G. A. Robinson, *Encyclopedia of Electronics* (June) 1962
"High-Speed Logic Circuits Using Tunnel Diodes," R. H. Bergman, M. Cooperman, and H. Ur, *RCA Review* (June) 1962
"The Iconoscope," G. A. Robinson, *Encyclopedia of Electronics* (June) 1962
"The Monoscope," G. A. Robinson, *Encyclopedia of Electronics* (June) 1962
"A New Approach to Simplified Parametric Amplifier Tuning," J. A. Luksch and V. Stachejko (Correspondence), *Proc. I.R.E.* (June) 1962
"Optical Maser Action in $CaF_2:Tm^{2+}$," Z. J. Kiss and R. C. Duncan, Jr. (Correspondence), *Proc. I.R.E.* (June) 1962
"Optical Maser Action in $CaWO_4:Er^{3+}$," Z. J. Kiss and R. C. Duncan, Jr. (Correspondence), *Proc. I.R.E.* (June) 1962
"Project Relay Digital Command System," S. H. Roth and T. A. Janes, II, *Trans. I.R.E. PGANE* (June) 1962
"Properties of an As-S-Br Glass," A. G. Fischer and A. S. Mason, *Jour. Opt. Soc. Amer.* (Letters to the Editor) (June) 1962
"Pulsed and Continuous Optical Maser Action in $CaF_2:Dy^{2+}$," Z. J. Kiss and R. C. Duncan, Jr. (Correspondence), *Proc. I.R.E.* (June) 1962
"Retrieval of Ordered Lists from a Content-Addressed Memory," M. H. Lewin, *RCA Review* (June) 1962
"Sample Preparation Methods for X-Ray Fluorescence Emission Spectrometry," E. P. Bertin and R. J. Longobucco, *Norelco Reporter* (June) 1962
"Shadow-Mask Picture Tube," D. D. Van Ormer, *Encyclopedia of Electronics* (June) 1962
"Simple Method of Registering Evaporation Masks," G. W. Leck (Notes), *Rev. Sci. Instr.* (June) 1962

"Some Recent Developments in Photomultipliers for Scintillation Counting," R. M. Matheson, *Trans. I.R.E. PGNS* (June) 1962
"The Space-Charge-Neutralized Hollow Cathode," A. L. Eichenbaum, *RCA Review* (June) 1962
"Stability Criteria for Television Camera Tubes," K. Sadashige, *Jour. S.M.P.T.E.* (June) 1962
"The TFT—A New Thin-Film Transistor," P. K. Weimer, *Proc. I.R.E.* (June) 1962
"Tiros I, II and III—Design and Performance," A. Schnapf, *Aerospace Eng.* (June) 1962
"Transistor Mounting and Application Guide," J. Eimbinder, *Electronic Industries* (June) 1962
"Tunnel-Diode Balanced-Pair Switching Analysis," G. B. Herzog, *RCA Review* (June) 1962
"The Window Thickness of Diffused Junction Detectors," R. L. Williams and P. P. Webb, *Trans. I.R.E. PGNS* (June) 1962
"X-Ray Spectrometric Determination of Plate Metals on Plated Wires," E. P. Bertin and R. J. Longobucco, *Analytical Chemistry* (June) 1962
"Plate Dissipation Nomograph for Damper Diodes," H. A. Whittlinger and M. E. Foss (Electronics Reference Sheet), *Electronics* (June 8) 1962
"New Way to Measure Diode Recovery Time," D. R. Gipp, *Electronics* (June 13) 1962
"Solid-State Panels: Will They Bring Flat-Display TV?," B. Binggeli and E. Fatuzzo, *Electronics* (June 29) 1962
"Tolerances in Coaxial Low-Pass Filters," R. M. Kurzrok (Research and Development), *Electronics* (June 29) 1962
"An Accordion Communication System for Mobile Use," J. Klapper and B. Rabinovici, *Conf. Proc. 6th National Convention on Military Electronics*, June 25-27 1962
"Attitude Determination—a New Dimension for Radar," A. Reich, *National Convention on Military Electronics, Conf. Proc.*.... 1962
"Constant Channel Flow in an A.C. Generator," G. F. Miller, L. J. Kijewski, and J. B. Fanucci, *National Aerospace Electronics Conf. Proc.* 1962
"The DAMP Operational Use of a Satellite Navigation System," L. Farkas, *National Winter Convention on Military Electronics, Conf. Proc.* 1962
"On the Determination of the Position and Velocity Vector of a Navigator Using a Satellite as Reference," I. Kanter, *National Winter Convention on Military Electronics, Conf. Proc.* 1962
"Estimation of Component and System Failure Rates from Test Data," R. Mirsky, *Conf. Proc. 6th National Convention on Military Electronics*, June 25-27 1962
"Kinematic Factors in Satellite Detection," N. S. Potter, *Conf. Proc. 6th National Convention on Military Electronics*, June 25-27 1962
"Methodology for Performance Degradation Analysis," S. Weisman and B. Tiger, *Conf. Proc. 6th National Convention on Military Electronics*, June 25-27 1962
"Optical-Band Radio Communication," D. G. C. Luck, *Conf. Proc. 6th National Convention on Military Electronics*, June 25-27.... 1962
"Performance of Digital Communications Systems in an Arbitrary Fading Rate and Jamming Environments," A. B. Glenn and G. Lieberman, *Conf. Proc. 6th National Convention on Military Electronics*, June 25-27 1962
"System Design of Orbital Tracking Instrumentation," S. Shucker and R. Lieber, *Conf. Proc. 6th National Convention on Military Electronics*, June 25-27 1962

AUTHORS

DONALD J. BLATTNER received the B.S. (E.E.) and M.A. (Physics) from Columbia University, and has also taken courses at Harvard, MIT, and Rutgers. For four years he taught physics and did solid state research at Columbia. During World War II he served in the U. S. Navy. During the past nine years at RCA he has developed microwave tubes and circuits and components for laser systems, and has participated in the engineering of telecommunication systems. Mr. Blattner has taught courses in radio waves and circuits, and is the author of a number of technical papers and talks. He is a member of Sigma Xi and a senior member of the Institute of Radio Engineers.

C. O. CAULTON received the degree of B.S. in Physics and Mathematics from Juniata College in 1929. The same year he joined RCA as a development and design engineer in the loudspeaker and acoustical laboratories. During World War II, he was active in many areas of national defense research and acted as consultant to and member of committees and sections of the National Defense Research Committee of the Office of Scientific Research and Development. From 1946 to 1951, he was Commercial Product Development Manager responsible for television and radio receivers and phonographs and, subsequently, coordinator for television station and market expansion for RCA. Since 1954, Mr. Caulton has been active in planning and sales of commercial and industrial products. These products included industrial controls, scientific instruments, communications products and systems, highway vehicle detectors, and systems for electronic assistance and control of highway vehicles, among others. He has represented RCA at the Committee of Electronics of the American Association of State Highway Officials, the Highway Research Board, Columbia University's Institute of Highway Safety, and other national and local highway groups.

A. L. EICHENBAUM (See *RCA REVIEW*, Vol. XXIII, No. 2, June 1962, p. 288)

L. E. FLORY received the B.S. degree in Electrical Engineering, University of Kansas in 1930. He was with the research division of RCA Manufacturing Co., Camden, N. J. from 1930 to 1942 during which time he was engaged in research on television tubes and related electronic problems, particularly in the development of the iconoscope. In 1942 he was transferred to the RCA Laboratories Division, Princeton, N. J. From 1949 to 1954 he was in charge of work on storage tubes, and since 1949 he has been in charge of work on industrial television. More recently he has supervised the work of the General Research Laboratory on Medical Electronics, Astronomical Television and Highway Vehicle Control. Mr. Flory is a member of Sigma Xi and a Fellow of the Institute of Radio Engineers.

GEORGE W. GRAY attended Princeton University as a civilian and in the Navy V-12 Program until 1943 when he was assigned to active duty in the Navy. After release to inactive duty in 1946, he returned to Princeton University and received the B.A. degree in Physics. In March of 1947 he joined the technical staff of RCA Laboratories at Princeton, N. J., working on the development of industrial television systems. More recently, he has been associated with the development of the Electronic Highway System. Mr. Gray is a member of Sigma Xi.

H. G. GREIG received the B.S. degree in Chemical Engineering from Drexel Institute of Technology in 1931. He was with the Research Department of the National Aniline Division of Allied Chemical and Dye Corporation until 1943. Since 1943 he has been with RCA Laboratories in Princeton, N. J. Mr. Greig is a member of the American Chemical Society, the American Association for the Advancement of Science, and Sigma Xi.

J. M. HAMMER received his B.S. in Engineering Physics from New York University in 1950. In 1951 he received the M.S. in Physics from the University of Illinois. He received the Ph.D. in Physics from New York University in 1956. He taught physics at New York University from 1953 to 1956 and did research in the field of low-energy electron scattering from various atomic systems. From 1956 to 1959, he worked at the Bell Telephone Laboratories on microwave noise problems. In 1959, he joined RCA Laboratories as a member of the technical staff and has been working on low-noise microwave research and atomic interactions. Dr. Hammer is a senior member of the Institute of Radio Engineers and a member of the American Physical Society and Sigma Xi.

FRED M. JOHNSON received the B.S. degree from the City College of New York in 1949. He received the M.A. in Physics in 1951 and the Ph.D. in physics in 1958 from Columbia University. He was a teaching assistant at Columbia University from 1951 to 1954 and held a research assistantship at the Columbia University Radiation Laboratory from 1954 to 1958. In 1958 he joined the technical staff of RCA Laboratories where he has been engaged in fundamental research on ion emitters and plasma synthesis. Since 1959 he has been engaged in studies of thermionic energy conversion, with particular emphasis on fundamental processes in the cesium plasma. In 1961 he was appointed adjunct assistant professor at Columbia University where he is teaching a graduate course on Direct Conversion of Heat to Electrical Energy in the Electrical Engineering Department. Dr. Johnson is a member of Phi Beta Kappa, Sigma Xi, the American Physical Society and the American Astronomical Society.

GILBERT LIEBERMAN received a B.A. degree in Mathematics at New York University in 1948, and an M.A. degree in Mathematical Statistics at Columbia University in 1949. Since then he has done graduate work in Mathematics and Electrical Engineering at University of Maryland, American University, and the University of Pennsylvania. From 1950 to 1955, he was with the U.S. Naval Research Laboratory, Sound Division, Electronics Branch, where he worked on sonar signal processing techniques, analysis of acoustic signal and noise propagation data. From 1955 to 1959, he was with the U.S. Naval Ordnance Laboratory, where he worked on acoustic detection, navigation and guidance systems. During this period, he taught courses in probability and statistics at the University of Maryland. He joined RCA in March of 1959 as a Senior Engineer in the Airborne Systems, Communications Systems Group. His work has involved analysis of digital communication systems, studies of effects of antenna fluctuations and fading on communication system performance, and research on adaptive communication techniques. Mr. Lieberman is a member of the Institute of Radio Engineers, the Institute of Mathematical Statistics, the American Statistical Association, and the Mathematical Association of America.

ROBERT E. MOREY received his training in the United States Navy where he served as an Aviation Radioman from 1944 to 1946. He worked as Chief Analyzer for the Colonial Radio Corp., Buffalo, N. Y., until 1948. Mr. Morey joined the RCA Service Co. in Washington, D. C. in 1948 where he worked on black-and-white and color television receivers. In 1954 he joined the RCA Laboratories and was assigned to the Acoustic and Electromechanical Laboratory where he assisted in the development of Video Tape Recording Systems. Since 1958, Mr. Morey has worked on systems for the electronic control of Highway Vehicles in the General Research Laboratory.

R. H. PARMENTER received the B.S. degree in Engineering Physics from the University of Maine in 1947 and the Ph.D. degree in Physics from MIT in 1952. He was a staff member in the Solid State and Molecular Theory Group at MIT from 1951 to 1954, a Guest Scientist at Brookhaven National Laboratory in 1951 and 1952, and a staff member of Lincoln Laboratory from 1952 to 1954. He joined RCA Laboratories in Princeton in 1954, and spent six months in 1958 at Laboratories RCA Ltd. in Zürich, Switzerland. He was a Visiting Lecturer at Princeton University in 1960 and 1961. At present, he is acting head of the General Solid State Research Group at RCA Laboratories. Dr. Parmenter is a Fellow of the American Physical Society and a member of Sigma Xi and Tau Beta Pi.

FRITZ PASCHKE received the degree of Diplom-Ingenieur in 1953 and the Dr. techn. sc. in 1955 from the Technical University, Vienna. Between 1953 and 1955 he was a research assistant at the Institute of High-Frequency Techniques in Vienna. Dr. Paschke became a member of the technical staff at RCA Laboratories, Princeton, N. J., in 1956. In August 1961, he joined Siemens and Halske in Germany.

WINTHROP SEELEY PIKE received the B.A. degree in Physics in 1941 from Williams College. He served with U. S. Army Signal Corps during World War II as radar officer and later as project officer in charge of the Signal Corps Moon Radar project. In 1946 he joined the research staff of RCA Laboratories Division at Princeton, N. J., where he has worked on sensory aids for the blind, storage tube applications, color television, industrial television, and Highway Vehicle Control. Mr. Pike is a Member of Sigma Xi and the American Guild of Organists.

FRED STERZER received the B.S. degree in physics from the College of the City of New York in 1951, and the M.S. and Ph.D. degrees from New York University in 1952 and 1955, respectively. From 1952 to 1953 he was employed by the Allied Control Corporation, New York, N. Y. During 1953 and 1954 he was an instructor in physics at the Newark College of Engineering, Newark, N. J., and a research assistant at New York University. He joined the RCA Tube Division in Harrison, N. J., in 1956. He is now group leader in microwave physics. His work has been in the fields of microwave spectroscopy, traveling-wave tubes, backward-wave oscillators, solid-state microwave amplifiers, oscillators and converters, microwave computing circuits and r-f light modulators. Dr. Sterzer is a member of Phi Beta Kappa, Sigma Xi, and the American Physical Society.