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 XXII | DECEMBER 1961 | NUMBER 4

CONTENTS

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

THE ENHANCED-SCAN, POST-ACCELERATION KINESCOPE*

By

H. B. Law, L. Davne, and E. G. Ramberg

RCA Laboratories
Princeton, N J.

Summary—An experimental deflection system is described in which a spherically curved mesh is placed in a conventional black-and-white tube at the position where the neck joins the funnel of the tube. It is shown that by operating the mesh and the deflection region of the tube at low potential, the scan power required is much less than that required in conventional systems. Scan sensitivity, raster distortion, resolution, contrast, and brightness available with the experimental kinescope are discussed, and the data obtained is compared to that of a conventionally operated black-and-white tube.

Best resolution is obtained with magnetic focus. The effect on contrast of secondary electrons scattered by the mesh is minimized by applying a bias voltage which tends to draw the secondary electrons toward the gun. A small residual effect is substantially eliminated through the use of specially shaped apertures in the mesh. The brightness potential of the tube is reduced by the interception of beam electrons by the mesh.

Introduction

THE trend in kinescope development is toward shorter tubes, larger deflection angles and higher phosphor-screen potential. With conventional deflection means, increased deflection power is required to bring about these changes. Technical as well as cost problems are encountered. The actual power required becomes of importance, especially where battery-operated portable television sets are concerned, because of weight and playing-time consideration. There is need, therefore, for a deflection system that requires lower power.

Many arrangements have been proposed and some are in limited use at the present time. For example, oscilloscope tubes with the wall coating in the form of a spiral of resistive material are available commercially. A voltage difference established across the spiral provides an electric field that permits deflection of the electron beam to occur at a voltage lower than that on the phosphor screen. Recently

* Manuscript received October 28, 1960.

a spiral coating has been combined with a mesh to achieve improved sensitivity as well as other desirable performance characteristics.[1]

Another type of post-acceleration tube employs a low-potential wall coating with a mesh located close to the phosphor screen. The phosphor screen operates at high potential and produces an acceleration of the beam from the mesh to the screen.

A magnetic scan-enhancement method has been described[2] in which the divergent lens action of a quadrupole arrangement of permanent magnets produces scan boost directly in the horizontal direction. At the same time the converging lens action for vertical beam displacements is of such magnitude that a crossover is formed very close to the lens. Divergence beyond the crossover results in a large scan boost.

The same paper[2] describes a negative electrostatic gauze lens, with full rotational symmetry, that is formed by inserting a flat, fine-mesh gauze in the neck of the tube. The screen, wall, and neck coatings are operated at high potential. The fine-mesh gauze, operating at low potential, is immersed in the high-potential region to form a negative electrostatic lens that effectively produces a symmetrical scan boost.

The present paper describes another approach to scan power reduction. The tube described employs a fine mesh, but in an arrangement differing significantly from that referred to above.

General Description

The enhanced-scan post-acceleration kinescope is similar in some respects to the post-acceleration tube mentioned above in which the wall coating and a mesh mounted close to the phosphor screen are operated at a potential which is low compared to that of the screen. In the enhanced-scan post-acceleration tube, however, the mesh is much smaller and is not located adjacent to the screen but is at the point where the neck joins the funnel as shown in Figure 1. In this case the screen and funnel coating are at high potential while the mesh and a thin metal cylinder to which it is attached are at lower potentials.

The mesh and cylinder are small enough to fit through the neck of the tube. The metal cylinder is located under the yoke and made of nichrome 1 to 2 mils thick to avoid eddy-current loss of deflection

---

[1] K. Schlessinger, "Electron Trigonometry, a New Tool for Electron Optical Design," 1960 Electron Devices Meeting, Washington, D.C.

[2] D. W. Taylor, N. W. Parker and H. N. Frihart, "Application of Rotationally Non-Symmetrical Electron Lenses to TV Image Reproduction," *1959 IRE Convention Record*, Part 7, pp. 93-102.

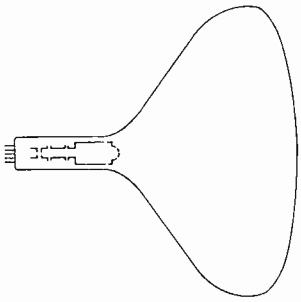


Fig. 1—Enhanced-scan post-acceleration kinescope.

power. Separate leads are brought out for the mesh and nichrome cylinder to permit the mesh to be operated at a lower potential than the cylinder. By this means most of the low-velocity secondary electrons that originate at the mesh can be drawn toward the gun rather than be accelerated to the screen, where they would lower contrast.

The mesh is spherically curved with the bulge toward the screen. A diverging electrostatic field is formed beyond the mesh that augments the magnetic deflection. An optical analog of the electron-lens system is shown in Figure 2.

Both magnetically and electrostatically focused guns were used. In the case of electrostatic focusing, unipotential as well as direct and inverse bipotential lenses were tested as described later.

DEFLECTION

Scan Sensitivity

In a conventional kinescope the yoke current required for full deflection is proportional to the square root of the screen potential.[3] Figure 3 shows the curve for a 14-inch 90° rectangular tube in which V_s/V_n is plotted against $(I_H/I_n)^2$. V_s is the screen potential and I_H

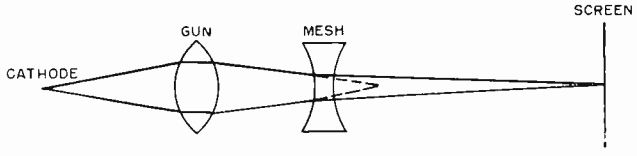


Fig. 2—Optical analogy of electron-lens system.

[3] V. K. Zworykin and G. A. Morton, *Television*, 2nd Edition, John Wiley and Sons, New York, N.Y., 1954, p. 554.

the horizontal yoke current required for full scan. V_n and I_n are potential and current values for normal black-and-white-tube operation, employed as standard of reference.

A similar curve may be obtained for the enhanced-scan post-acceleration kinescope. For comparison with operation of a standard black-and-white tube, it is convenient to maintain the screen at the

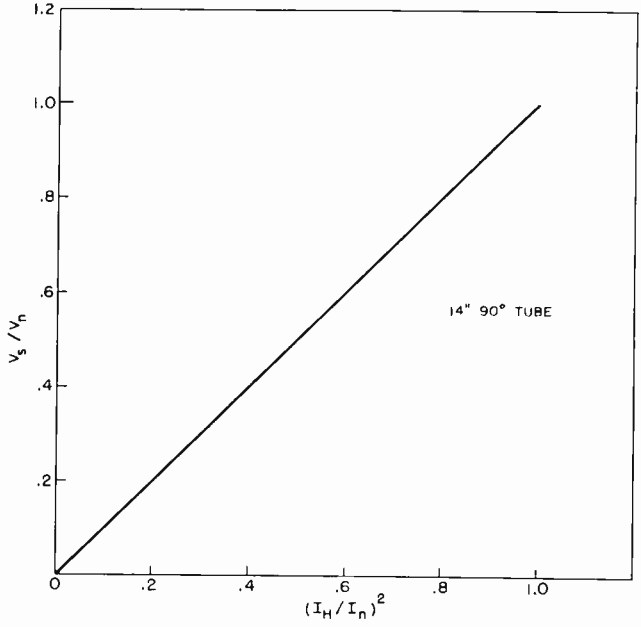


Fig. 3—Yoke current versus screen potential for conventional kinescope.

normal potential, V_n, and to plot V_m/V_n versus $(I_H/I_n)^2$, where V_m is the mesh potential. The curve for a typical enhanced-scan tube compared with that for the standard tube is shown in Figure 4.

It should be emphasized that, for the standard tube, the ordinate in Figure 4 represents a change in screen potential with a consequent change in light output; for the enhanced-scan tube, the screen remains at high potential so that high light output is maintained as the mesh is lowered to achieve increased scan sensitivity. Therefore, the two curves do not show a comparison of scan sensitivity of the standard and enhanced-scan tubes for the same light output, but rather under the condition where the screen potential in the standard tube is lowered in the same manner that the mesh potential is lowered in the enhanced-scan tube.

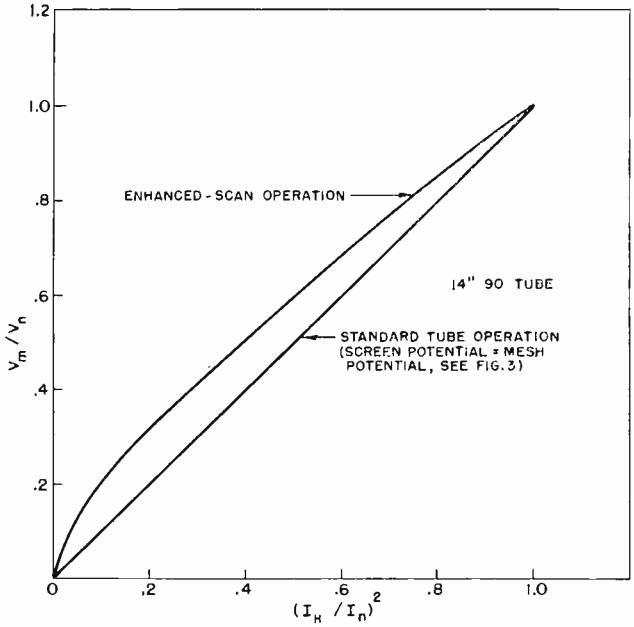


Fig. 4—Comparison of yoke current required for enhanced-scan and conventional operation.

The results obtained may be compared with those of calculations (described in greater detail in the Appendix) of electron paths in simplified electrode systems roughly approximating those of the enhanced-scan kinescope. Figure 5 shows paraxial paths for electrons

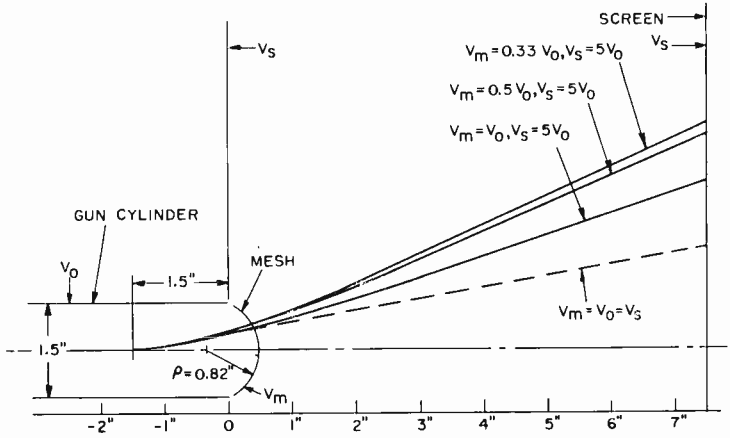


Fig. 5—Electron paths diverging from fixed deflection center in idealized enhanced-scan kinescope (calculated).

diverging with the same angle of elevation from a common center of deflection when the mesh potential V_m is equal to V_0, 0.5 V_0, or 0.333 V_0, where V_0 is the potential of the final anode cylinder of the gun. V_s is held at 5 V_0. It is seen that there is a scan enhancement produced by refraction of the rays by the equipotential surfaces equal to 1.66 for $V_m = V_0$, 2.11 for $V_m = 0.5\ V_0$, and 2.23 for $V_m = 0.333\ V_0$. When $V_m = V_0$, the required deflection current is reduced by a factor of 0.69 for $V_m/V_n = 0.2$ in Figure 4. If the deflection took place entirely at the potential V_0, the current-reduction factor would be 1/1.66, or 0.60 (see Figure 5). The somewhat larger value of the deflection current actually required is readily explained by the fact that the deflection takes place, in part, at a potential higher than V_0. For similar reasons, namely that part of the deflection takes place at potentials *lower* than V_0, the preceding figures for $V_m = 0.5\ V_0$ and 0.333 V_0 can be expected to *underestimate* the scan enhancement over and above that resulting from lowering the voltage in the deflection plane from V_s to V_0. Thus lowering of the mesh voltage alone should be much more effective in increasing the raster size for given deflection current than is indicated by the figures given.

To provide an indication of the effect of mesh position on scan sensitivity, measurements were made in a demountable 110° 21-inch rectangular tube using different mesh structures. In one case the mesh was a hemisphere with a diameter of 5¼ inches mounted against the bulb wall. Another arrangement employed a nearly hemispherical mesh ⅝ inch in diameter that would fit into the tube neck. With the smaller mesh, scan-sensitivity measurements were taken as a function of position of the mesh in the tube neck. This procedure also permitted the determination of the range of position for the smaller mesh that would result in full scan without shadowing of the beam by the mesh frame or bulb wall. The results are plotted in Figure 6.

The influence of mesh shape is indicated by comparing curves obtained with a hemispherical and a flat mesh. The results are plotted in Figure 7, which shows that the curved mesh results in a considerable increase in scan sensitivity.

The curves in Figure 7 were obtained with tubes having identical geometry, except that in one case the mesh was stretched flat across its mounting ring while in the other case the mesh was nearly hemispherical in shape. The beam travels a somewhat longer path at mesh potential in the hemispherical case than in the flat case and would, therefore, be deflected through a larger angle. Perhaps more important is the direction of acceleration of the beam as it leaves the mesh in the two cases. Experiments were performed in which it was ob-

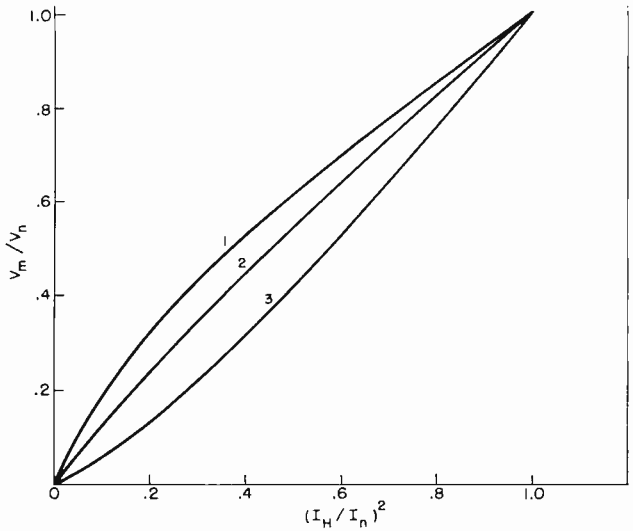


Fig. 6—Influence of mesh position on yoke current for enhanced-scan operation. Mesh-frame to reference line distance: curve 1 – .55 inch; curve 2 – .80 inch; curve 3 – 1.050 inch. Preferred distance is .67 inch (no beam shadowing); the curve for this case lying slightly below curve 1. (5¼-inch mesh gives same curve as 1.)

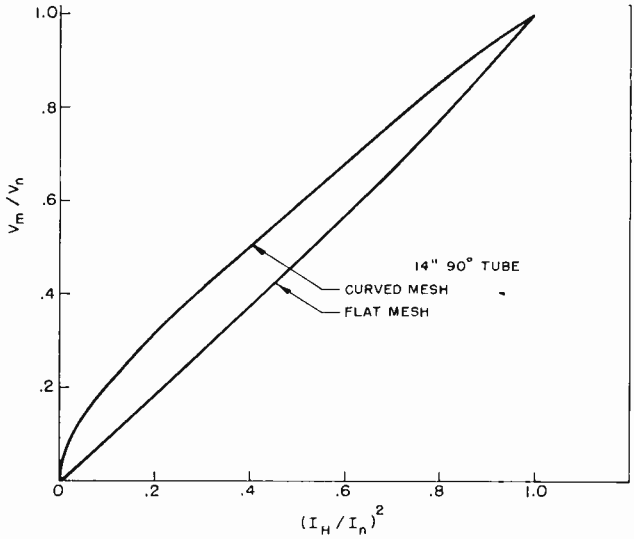


Fig. 7—Influence of mesh shape on yoke current for enhanced-scan operation.

served that the electrostatic field between the mesh and screen tends to decrease the picture size when a flat mesh is used and to increase the picture size for an hemispherical mesh. The data were taken by measuring the picture size as a function of screen potential with the mesh potential held constant. The results are given in Figure 8.

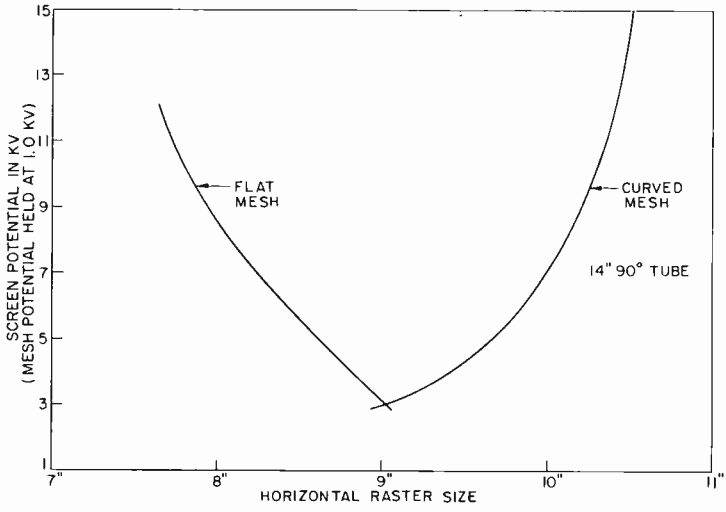


Fig. 8—Raster size versus screen potential for constant mesh potential and constant scan power.

Raster Distortion

Pincushion distortion has been observed in enhanced-scan tubes for both flat and hemispherically curved meshes. A study of the effect of mesh curvature on raster distortion was made by photographing a crossline pattern reproduced by tubes using meshes of various curvatures. As a result of these observations, a qualitative description of the effect of mesh curvature on raster distortion was arrived at.

Figure 9 illustrates a flat mesh with equipotential lines sketched in. Three beam paths are estimated as shown. Above the mesh, the paths are curved because of the magnetic deflection field. As the beam electrons penetrate the mesh, the paths are bent strongly toward the axis of the tube by the electric field. Since the force on the electrons is normal to the equipotential lines, the bending soon becomes radially outward instead of inward because of the changing slope of the equipotential lines.

The ultimate landing position of the electrons on the phosphor

screen is thus determined by the resultant action of an initial inward force and a subsequent outward force on the beam electrons. The inward force just beyond the mesh overrides the later outward force for a flat mesh as illustrated by the curve in Figure 8. That is, the picture size is reduced as the screen potential is raised above the mesh potential.

The predominance of the inward force becomes less with increasing radial distance from the tube axis as evidenced by the fact that the

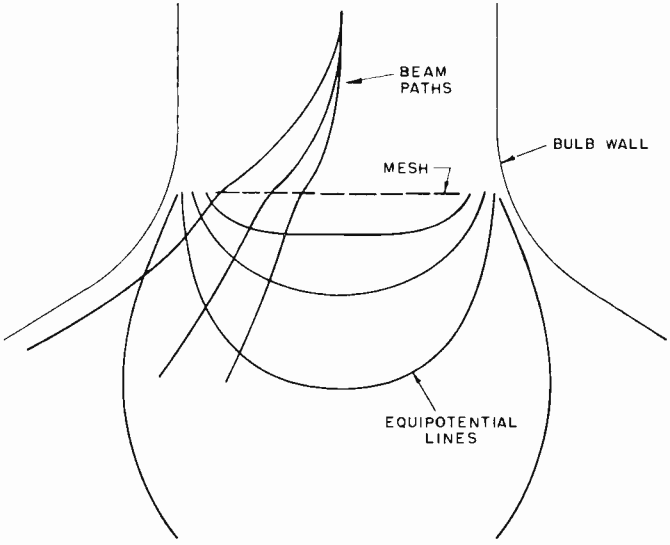


Fig. 9—Effect of flat mesh on raster distortion.

raster shows pincushion distortion. A plot of the pincushion radius of curvature near the edge of that raster as a function of the ratio of screen potential to mask potential is shown in Figure 10.

Figure 11 shows an estimate of the beam paths when the mesh is spherically curved. It is seen that the beam approaches the mesh almost normally, so that the initial radial force inward on the beam as it penetrates the mesh is quite small. Hence the outward radial force as the beam crosses equipotential lines predominates. This is shown in Figure 11 where it can be seen that for a hemispherical mesh, the picture size increases as the screen-to-mesh potential is increased. Some pincushion still exists, as can be seen in Figure 10, although it is less than for a flat mesh.

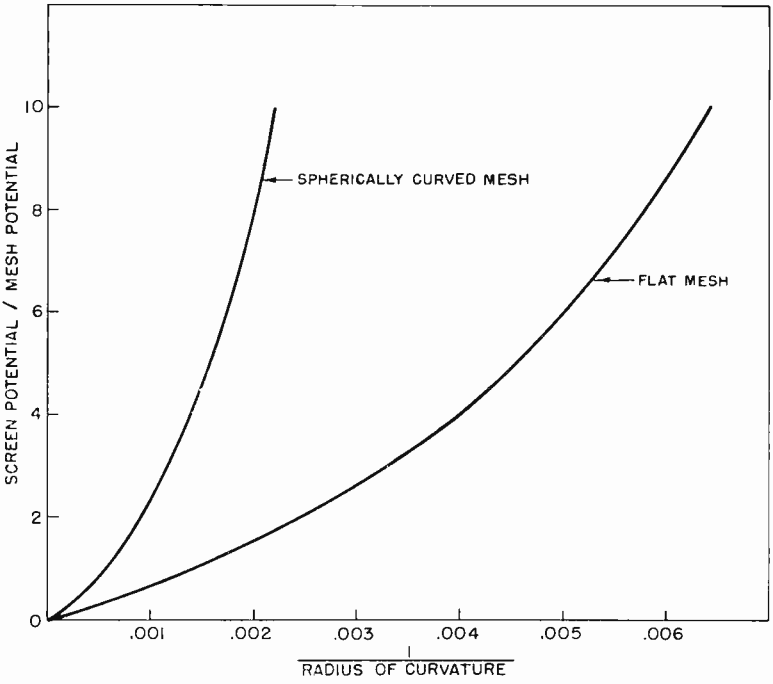


Fig. 10—Pincushion distortion versus screen potential/mesh potential.

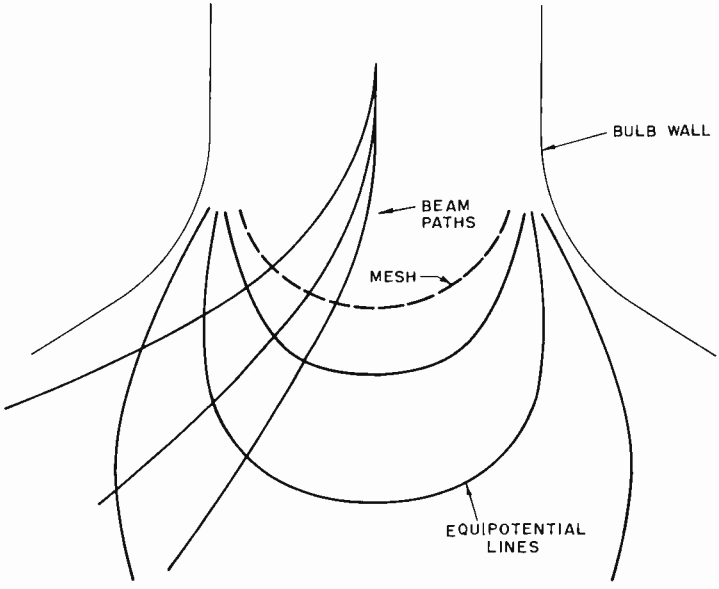


Fig. 11—Effect of curved mesh on raster distortion.

Resolution

For usual types of electron guns, the gun–anode potential available in the enhanced-scan tube is the mesh potential, which may be as low as ⅕ the screen potential. Tests with a standard black-and-white-tube gun operating at 4 kilovolts showed only a small deterioration in resolution at 100 microamperes beam current. Therefore, low-voltage gun operation is not a major factor affecting resolution in the enhanced-scan tube at beam currents of 100 microamperes or less.

A more important influence on resolution is the lens action of individual apertures in the mesh. On the gun side of the mesh, there is essentially a field-free space that extends to the region under the yoke. Between the mesh and screen there is an accelerating field that makes each aperture an electron lens.

Mesh aperture focus may be observed experimentally as a function of the ratio of screen voltage to mesh voltage, V_s/V_m, and thereby an estimate may be made of the effect of aperture focus on resolution. In one experiment an unfocused or spray beam from the cathode was used without scanning. At $V_s/V_m = 1$, no mesh pattern was resolved at the screen because the source of electrons was too large to produce a shadow of the fine-mesh wires. A mesh with 500 wires per inch was used. At a ratio of 2, the mesh pattern appeared. Best focus was obtained at a ratio of about 6, while at 8 or 9 the mesh pattern appeared to reverse polarity (white wire image) because of the overfocus condition. Lack of smoothness in the mesh made it difficult to determine a precise focus ratio because field strength at the mesh is dependent on surface irregularities such as slight dents. It was found that the ratio for focus was about 6 whether using the 5¼-inch-diameter or the smaller ⅝-inch-diameter mesh.

This can be readily understood by considering mesh focusing for a concentric mesh and screen, with radii of curvature R_m and R_s, respectively. The path equations for such an electrode configuration (Equation (13) of Reference 4) lead to the following focusing conditions:

$$\frac{V_s}{V_m} = \left(1 + 2\sqrt{\frac{R_m}{R_s}}\right)^2.$$

Thus, sharp mesh focusing would be expected at

[4] H. B. Law and E. G. Ramberg, "The Reflected-Beam Kinescope," *Proc. IRE*, Vol. 48, pp. 1409-1417, August, 1960.

$V_s/V_m = 9$, for $R_s/R_m = 1$; (flat mesh and screen)

$V_s/V_m = 6$, for $R_s/R_m = 1.9$;

$V_s/V_m = 5$, for $R_s/R_m = 2.6$;

$V_s/V_m = 4$, for $R_s/R_m = 4$;

$V_s/V_m = 3$, for $R_s/R_m = 7.5$.

The agreement with observation for the large-radius mesh is reasonably satisfactory. For the small-radius mesh, conditions clearly deviate too much from radial symmetry to make the analysis applicable.

The mesh pattern may be observed while the beam is scanning if the beam is focused in the deflection plane. Mesh-aperture focus occurs at nearly the same voltage ratio, although the image is not as sharp as the image obtained with an unscanned spray beam.

The fact that the focusing ratio is nearly the same with the beam focused in the deflection plane as for the spray beam suggests that the deflection imparted by the aperture fields is large compared to the angle subtended by the aperture at the source, even when the latter is in the deflection plane. The reduced sharpness, on the other hand, must be attributed to the fact that the angle subtended by the source at the apertures is much larger for focus in the deflection plane than for the unfocused spray beam.

It is probable that a voltage ratio of less than 6 would be used in an enhanced-scan kinescope, and therefore the aperture lenses would be underfocused. Experimental tubes were built in which the diameter of the ring on which the mesh was mounted was $\frac{5}{8}$ inch. The length of the intersection of the spherical surface with a plane through the axis was $\frac{3}{4}$ inch. Mesh with 500 wires per inch was used so that the vertical height scanned on the mesh was about 225 apertures. A limit of about 450 television lines of resolution is observed at a voltage ratio of 6. The use of a lower voltage ratio or finer mesh would raise the limit.

The resolution of an enhanced-scan 21-inch 110° rectangular tube was measured at various values of V_s/V_m. A $\frac{5}{8}$-inch-diameter spherical mesh with 500 wires per inch was mounted on an einzel type gun as shown in Figure 12. The data obtained are plotted in Figure 13.

As can be seen, the resolution falls off markedly as the ratio is increased. However, to maintain focus as the mesh potential is lowered, the einzel gun focus cylinder must be made increasingly more negative to strengthen the lens and overcome the increased diverging action of the field beyond the mesh.

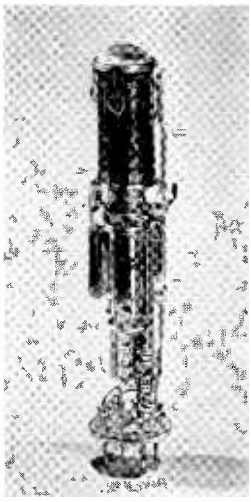


Fig. 12—Experimental gun and mesh assembly for 21-inch 110° kinescope.

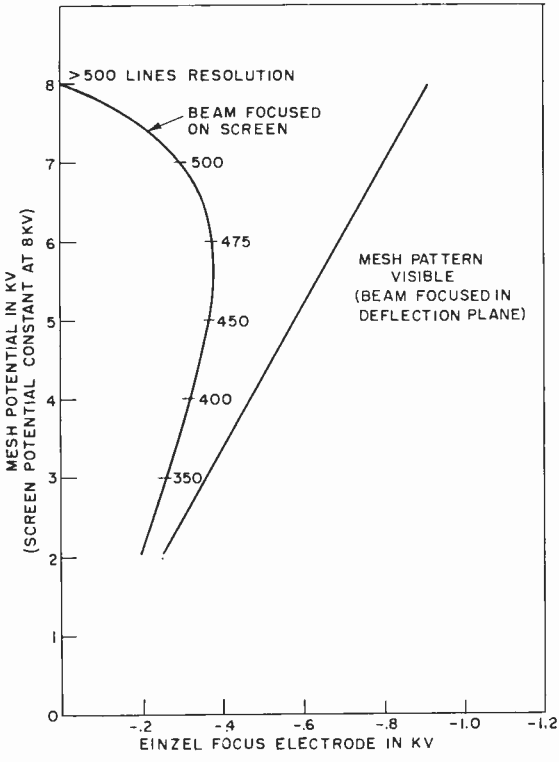


Fig. 13—Experimentally determined resolution in enhanced-scan kinescope.

To eliminate the einzel lens, the tube was magnetically focused. The focus electrode was tied to the anode cylinders and resolution read as a function of V_s/V_m. At 100 microamperes the resolution did not drop below 450 lines for a ratio of 5.

The tube was tested while the gun was operating in two different ways, electrostatic focus and magnetic focus. Although a restricted voltage range was necessary in operating the tube because of arcing difficulties, the following sets of conditions both resulted in 500-line resolution at 100 microamperes beam current.

Focus Type	1st Anode (kv)	2nd Anode (Mesh) (kv)	Screen (kv)
Electrostatic	.63	3	12
Magnetic	3	3	12

In neither of the tests was the mesh pattern detectable.

At currents above 100 microamperes, the resolution falls rapidly because of spot blooming. In either case, resolution would deteriorate rapidly for operation of the mesh below 3 kilovolts.

Contrast

Beam electrons that strike the mesh will produce secondary electrons. Some of the high-velocity secondaries may pass through the mesh and be accelerated to the screen, thus reducing contrast. However, the relative number of secondaries that have high velocity is not large, and only those high-velocity secondaries that originate from bombardment of the sides of the mesh wires will have a good chance of striking the screen.

Of more concern is the relatively large number of low-velocity secondaries that are liberated from the mesh. Some of these will go through the mesh. Since the accelerating field between the mesh and screen is divergent, these secondaries tend to form an enlarged image of the mesh area from which they originate.

Figure 14 shows photographs of a small raster, the first with normal exposure and the second overexposed. The screen potential was 15 kilovolts and the mesh potential 3 kilovolts. It appears that secondary electrons originate from the entire mesh even though only a small area of the mesh is scanned by the beam. It is likely that the strong curvature of the mesh permits low-velocity electrons emitted

in a back direction to reach all portions of the mesh with a good chance of being accelerated through the mesh to the screen.

One method of improving contrast that has proven successful is the acceleration of the low-velocity electrons toward the gun by application of a potential, positive with respect to the mesh by 45 to 300 volts, to the cylinder facing the mesh on the gun side. A large

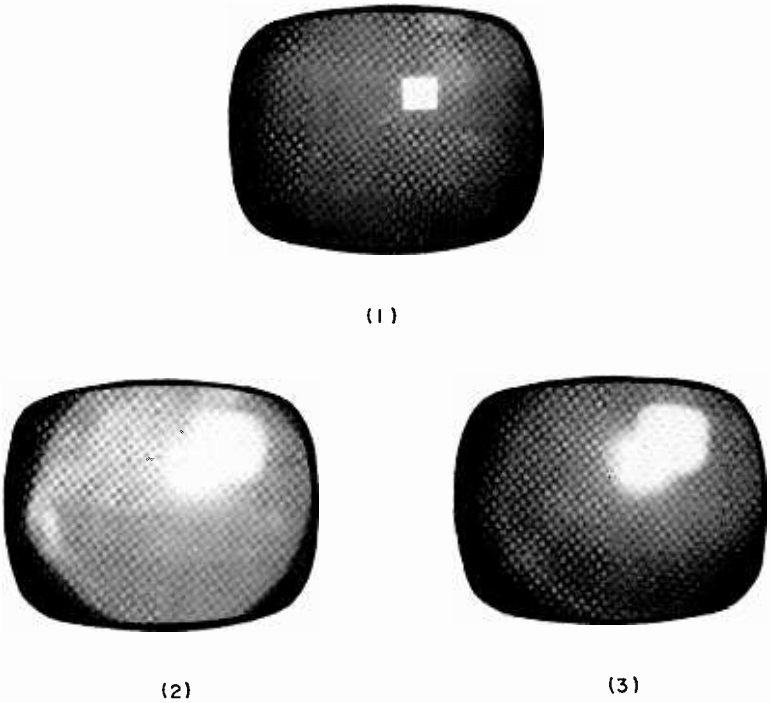


Fig. 14—Small scan raster on 90° 14-inch enhanced-scan tube; phosphor screen 15 kilovolts, mesh 3 kilovolts: (1) normal exposure, no suppression voltage; (2) same as (1) but overexposed to show patterns caused by secondary electrons; (3) same as (2) but 300-volt suppression voltage applied.

improvement in contrast is obtained as indicated in the third photograph in Figure 14, where a bias voltage of 300 volts was applied. The exposure times for photographs 2 and 3 were the same.

In the operation of sealed-off tubes in which low-velocity secondaries were accelerated back to the gun, it was noted that not all the secondary electron effects on contrast are eliminated by applying even 500 volts to the nichrome cylinder. This can be seen in the third picture in Figure 14 which shows a diffuse raster displaced radially

outward from the true raster. Overexposure has greatly accentuated the diffuse raster intensity. In the first photograph (which was made with normal exposure) neither the diffuse raster nor the low-velocity scatter is visible, even though no suppression voltage was used.

Since the field at the mesh is in a direction to accelerate electrons away from the mesh on both sides, there will be areas on the sides of the mesh wires subject to beam bombardment where low-velocity secondaries may still be accelerated to the screen. Low-velocity electrons from the mesh must be the cause of the spurious pattern in the third picture; this is evident from the fact that the pattern disappears whenever the difference in potential between the screen and mesh is less than the stopping power of the aluminum film on the phosphor screen, regardless of the screen-voltage-to-mesh-voltage ratio or actual values of potential.

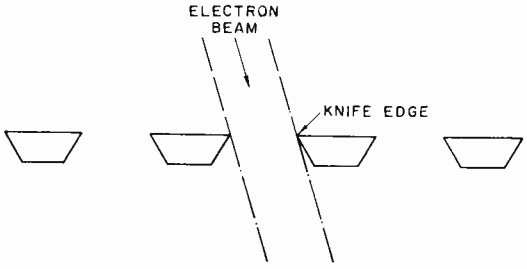


Fig. 15—Mesh with knife-edge apertures.

The beam electrons may be prevented from striking the sides of the apertures by tapering the aperture holes as shown in Figure 15. The knife-edge side of the mesh is toward the gun, and the taper is made sufficient to prevent the beam from striking the aperture sides.

A mesh was made with tapered knife-edge apertures by etching holes from one side in a thin piece of supernickel and then forming the thin metal mesh into a hemisphere with a diameter of 5/8-inch. The holes were 6 mils apart and arranged in hexagonal fashion; that is, each hole surrounded by six equally spaced holes.

Contrast was measured using the knife-edge aperture mesh in a 21-inch 110° tube. For $V_s/V_m = 3$, the contrast was measured to be 300 to 1. A 2 x 2-inch raster displaced to the edge of the screen was used and measurements were taken on and just off the raster where the secondary pattern appears. For comparison, similar measurements were taken using electrolytic-type mesh with 500 wires per inch. A contrast of 70 to 1 was measured.

Further contrast measurements were made with the tube having the knife edge apertures by using a 6 x 6-inch raster in the center of the tube and measuring 2 inches inside the raster and 2 inches off the raster. Readings of contrast as a function of V_s/V_m were made for three conditions — with no bias on the nichrome cylinder, with the cylinder 300 volts positive with respect to the mesh, and with the cylinder 300 volts negative with respect to the mesh. The results are plotted in Figure 16.

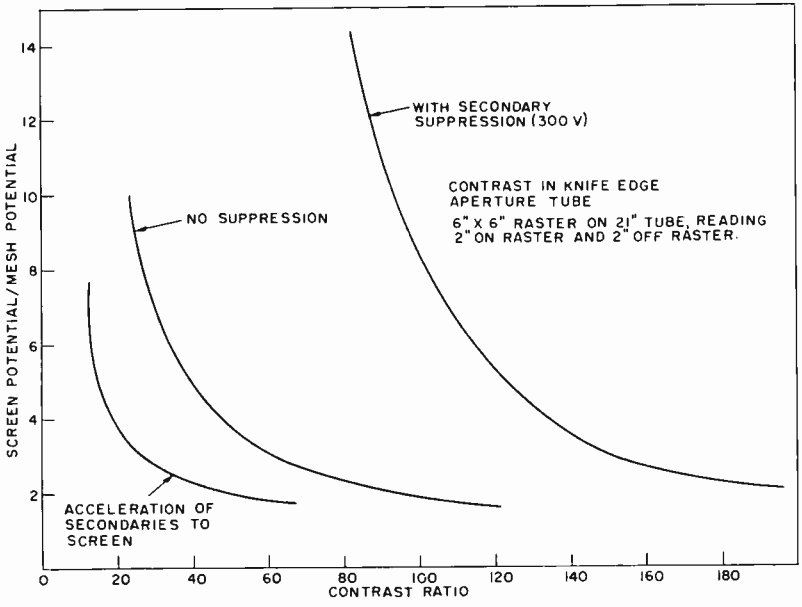


Fig. 16—Contrast in knife-edge-aperture tube.

BRIGHTNESS

The brightness capability of the enhanced-scan tube is lower than that of a conventional tube because the beam must pass through the fine mesh located in the tube. Experimental tubes were made in most cases by using 500 wires per inch electrolytic type mesh with about 50 per cent transmission. Metallized Corning Pyroceram* mesh or etched mesh might be used, but it is unlikely that a transmission appreciably greater than 50 per cent could be realized in any case. The brightness would then be limited to one half that of a conventional tube operating at the same voltage and current.

* Registered Trademark

An experiment was performed to determine whether the power-dissipation capability of the mesh would limit the brightness of the tube. A 21-inch enhanced-scan tube using 500 mesh of the electrolytic type was scanned with a .14 x .14-inch raster on the mesh. The mesh and screen were operated at 10 kilovolts. It was found that at 160 microamperes beam current the mesh underwent a slight deterioration after 10 minutes bombardment. If a value of 100 microamperes were taken as a safe current, then ½ watt would be dissipated, assuming the transmission of the mesh to be 50 per cent. Under the conditions of the test, about 1/12 of the screen area was scanned, so that roughly 1 milliampere beam current could be used safely with the mesh at 10 kilovolts.

In practice the mesh would likely be operated at 5 kilovolts or less. The beam current could then reach 2 milliamperes or more without exceeding the permissible dissipation value. It is concluded, hence, that power dissipation in the mesh would not limit tube brightness.

Fabrication of the Mesh

Electrolytic type mesh was used in the enhanced-scan tube. This mesh is made by a plating technique[5] that was developed for producing very fine mesh with high transmission. The mesh is very fragile and has insufficient body to permit severe drawing operations. It was necessary, therefore, to devise a means other than drawing to give the initially flat mesh a hemispherical form.

A successful technique for forming the mesh to the desired shape was developed by utilizing the shrinking action on the mesh[6] of surface tension forces at high temperature as well as the yielding of the mesh when tension is applied. Successful application of these two processes for forming a nearly hemispherical mesh depends on mounting the mesh between two rings with the correct amount of looseness or excess mesh. The excess mesh is evenly distributed around the mounting ring in the form of very fine overlaps.

Smoothing out of the mesh to a curved shape is completed as follows. The assembly of mesh sandwiched between two rings is placed on an oxidized, chrome-plated ball. The rings are weighted slightly to cause some tension on the mesh and then the assembly is heated to 950°C for 25 minutes in a hydrogen atmosphere. This treat-

---

[5] H. B. Law, "A Technique for the Making and Mounting of Fine Mesh Screens," *Rev. Sci. Instr.*, Vol. 19, pp. 879-881, December, 1948.

[6] H. B. Law, U.S. Patent No. 2,654,940, October 13, 1953.

ment causes the combination of stretching and shrinking to occur simultaneously. The resulting spherical shape obtained is uniformly smooth provided the initial amount of material within the rings was neither too much nor too little. Where an excess of material is provided, the retaining ring will seat on the ball but the mesh will not attain the spherical shape. If insufficient material is provided, disruption usually occurs unless weighting by the rings is reduced.

Conclusions

It was demonstrated that the scan power required for a black-and-white kinescope may be greatly reduced by placing a spherically curved mesh in a conventional tube at the position where the neck joins the funnel; the mesh and the deflection region of the tube are then operated at low potential. Best resolution was obtained when using magnetic focus.

The effect on contrast of secondary electrons scattered by the mesh was minimized by applying a bias voltage which tended to draw the secondary electrons toward the gun. A small residual effect could be substantially eliminated by the use of specially shaped apertures in the mesh.

The brightness of the tube is reduced by the interception of beam electrons by the mesh.

Acknowledgments

The authors wish to thank John Mirsch for making many of the measurements here reported and Paul Messineo for much help in tube fabrication as well as testing. The direction and encouragement of Harwick Johnson, under whose leadership the work was done, is greatly appreciated.

Appendix—Computation of Paraxial Rays in Electrode Configuration Approximating That Realized in the Enhanced-Scan Kinescope

The assumed electrode configuration is shown in Figure 5. It is treated analytically by assuming that the potential distribution to the left of the mesh is that within two coaxial equidiameter cylinders of diameter D, with the left cylinder ($z < 0$) at the potential V_0 and the right cylinder ($z > 0$) at a potential V' such that the equipotential with radius of curvature ρ has the prescribed potential, V_m. If distances are measured in units of D/2.63, the potential distribution along the axis is given by

$$\phi = \frac{V' + V_0}{2} + \frac{V' - V_0}{2} \tanh z, \qquad z < z_m,$$

and the radius of curvature by

$$\rho = -\frac{2\phi'}{\phi''} = \coth z.$$

Hence

$$V' = \frac{2V_m\rho - (1-\rho)V_0}{\rho + 1}.$$

The field to the right of the mesh is assumed to be that in the neighborhood of a circular aperature of diameter D, with zero electric field at large distances to the right of the aperture and the field E at large distances to the left; E is established such that the equipotential with the curvature $1/\rho$ of the mesh has the potential of the mesh, V_m. In the same units as before, the potential distribution along the axis is given by

$$\phi = V_s + \frac{Ez}{2} - \frac{Ez}{\pi} \arctan \frac{z}{1.315} - \frac{1.315\,E}{\pi}, \qquad z > z_m,$$

and

$$\rho = \frac{2.63\pi}{4}\left[1 + \left(\frac{z}{1.315}\right)^2\right]\left\{\left(1 + \left(\frac{z}{1.315}\right)^2\right)\left(1 - \frac{2}{\pi}\arctan\frac{z}{1.315}\right) - \frac{2z}{1.315\pi}\right\}$$

The value of z must be found for which $\rho = \coth z$. This is fulfilled for $z_m = 0.87$, $\rho = 1.43$. The condition that $\phi(z_m) = V_m$, finally, leads to $E = -6.88\ (V_m - V_s)$. Ray paths can now be integrated by the well-known equation

$$R'' = -\frac{3}{16}\left(\frac{\phi'}{\phi}\right)^2 R, \qquad R = r\phi^{1/4}$$

taking due account of the continuity of r' at $z = z_m$.

THERMAL EMISSIVITY AND CONDUCTIVITY OF ALUMINA HEATER COATINGS*

BY

N. RUDNICK† AND J. J. CARRONA‡

Summary—The experimental determination of the thermal properties of alumina heater coatings is described. Inconsistencies between observed thermal behavior in electron-tube heaters and the behavior predicted from a consideration of published conductivity and emissivity values are eliminated by the establishment of new values of spectral and total emissivity. An analysis of the transfer of heat from the electron-tube heater to its cathode shows that the predominant mechanism is thermal radiation from the alumina.

An analysis is made using the new low-emissivity values which indicates that significant improvements in the rate of heat transfer may be expected if the emissivity of the alumina coating is increased. Experimental confirmation of such improvements is described. With graphite coatings on alumina (emissivities four times that of alumina), the heat transfer is increased by a factor of four.

INTRODUCTION

IN most electron tubes, the cathode must be operated at a temperature of about 1000°K to provide the electron emission required for proper functioning of the device. The power required to maintain this temperature is supplied by passing an electrical current through a coil or series of folds of tungsten wire enclosed by the cathode and electrically insulated from it. The insulation is generally provided by a coating of sintered aluminum-oxide particles on the wire. Because heat transfer is quite poor, the heater wire must operate at temperatures of 1500 to 1800°K to keep the cathode temperature at the required level.

These high operating temperatures limit the useful life of the heater. High diffusion rates permit alteration of the tungsten-wire structure which weakens and deforms it. Cyclic operation produces thermally induced stresses between the coating and the wire which exceed the strength of the weakened tungsten and cause failure. It is

* Manuscript received 6 October 1961.

† Formerly RCA Electron Tube Division; now with Gulton Industries.

‡ RCA Electron Tube Division, Harrison, N. J.

known that heater life is exponentially dependent on temperature and that a small decrease in heater temperature can extend life many times. A study of the mechanism of heat transfer in the heater–cathode system was undertaken with the expectation that means could be developed to provide reduced heater temperatures and enhanced heater reliability.

During this study, a number of uncertainties were uncovered concerning the thermal properties of the alumina heater coating. This paper describes the measurements performed to resolve these uncertainties.

Theoretical Considerations

Heat is transferred from the tungsten heater wire to the surrounding cathode sleeve almost totally by radiation. The power P in watts (9.48×10^{-4} BTU/sec) which is transferred is given by

$$P = \frac{5.67\, A_H\, (T_H^4 - T_K^4)}{\frac{1}{\epsilon_H} + \frac{A_H}{A_K}\left(\frac{1}{\epsilon_K} - 1\right)} \times 10^{-12} \qquad (1)$$

where ϵ_H and ϵ_K are the total emissivities of the radiating surface of the heater and of the inner cathode surface, respectively; A_H and A_K are the effective radiating surface area of the heater and the inner surface area of the cathode in square centimeters; and T_H and T_K are the temperatures of the radiating heater surface and of the cathode in degrees Kelvin.

Thus, the dominant heater characteristics affecting power flow are temperature, emissivity, and surface area. All three of these characteristics, however, are based on measurements of doubtful accuracy. Some of the factors questioned in this survey are discussed below.

Transparency of Alumina Coating

Because sapphire, a pure aluminum oxide, is sometimes used as a window for infrared radiation, it is often assumed that these aluminum-oxide heater coatings are also largely transparent to heat radiation. If this assumption were true, the heater–cathode heat flow would consist of radiation directly from the heater wire rather than from the coating surface. The amount of heat radiated would be determined by the wire temperature, surface area, and emissivity, and T_H, A_H, and ϵ_H in Equation (1) would refer to the wire rather than the coating. Improved heat transfer could then be achieved by altering the

wire characteristics. The coating would serve only as electrical insulation and participate in the heat flow mainly by furnishing small conduction paths at the scattered points of contact with the cathode.

The alumina heater coating is, however, a sintered ceramic material rather than a clear crystal-like sapphire. Even if individual crystal grains were transparent, the multiple reflections and refractions within the ceramic might result in effective opacity. In this event, heat would flow by conduction from the wire through the coating, and then be radiated from the coating surface. Radiation would be determined by the emissivity, temperature, and area of the coating surface. An increase in heat output at a given temperature, or the ability to radiate the required heat at a lower temperature, would then be attainable only through an increase in the emissivity or effective surface area of the coating.

Thus, the transparency of the alumina coating must be established before Equation (1) can be used.

Total Emissivity

The total emissivity of a surface is its relative capacity to radiate energy at a given temperature as compared with a hypothetical "black body." The black body emits the maximum possible energy per unit area at a given temperature. Total emissivities of both tungsten and aluminum oxide are given in published tables, but these values refer to specially prepared specimens. Because smoothness, color, cleanliness, and other factors exert a strong influence on emissivity, published figures are not necessarily applicable to a practical situation.

For example, commercial tungsten wire retains marks from the dies through which it has been drawn, an oxide film of variable degree depending on the history of the wire, and other contamination from various sources. Published aluminum-oxide emissivities are usually for the oxide formed on smooth metal rather than for a ceramic. Even values for the ceramic may not be valid for a porous heater coating. In fact, as discussed in the previous section, the coating might prove to be transparent.

It is important to know how much the heater differs from an ideal black body to estimate the extent to which temperatures may be reduced by "darkening" the emitting surfaces.

Spectral Emissivity

The spectral emissivity of a surface is also a comparison between the amount of radiation from that surface and the radiation from a black body, but consideration is restricted to energy lying within a

narrow band close to a particular wavelength. Because it may vary with wavelength, spectral emissivity must be specified in association with a specific wavelength.

Spectral emissivity is important in the present work because of the use of optical pyrometry. Because all work in this field deals with heater temperatures at which visible radiation is involved, optical pyrometry is a conventional method used for temperature measurements. However, an optical pyrometer reading must be adjusted in accordance with the spectral emissivity of the surface being tested; therefore, the spectral emissivities of heater materials must be known. Large errors in temperature data may result from inaccurate emissivity values.

The same shortcomings of published values mentioned above for total emissivities also apply to spectral emissivities.

Temperatures of Heaters

The operating temperature of heaters ranges from about 1500 to 1800°K. Because of the difficulties in using thermocouples within an evacuated glass bulb on bodies of such small mass, only two methods are feasible for determining temperatures in this range: (1) optical pyrometry, and (2) resistance thermometry in which the tungsten heater wire is itself the sensing resistor.

These methods have been used in the past with sometimes conflicting results. Thus, it was necessary to investigate the use of these methods and the extent of disagreement between their results.

Thermal Conductivity of Alumina Coatings

Temperature differences of as much as 200°C have been reported between heater wire and coating surface. The temperature of the coating surface depends on the conductivity of the particular alumina structure involved. Thermal conductivities of alumina ceramic in the temperature range from 1000 to 1500°K vary from 0.013 to 0.018 calorie per centimeter per degree Centigrade per second, as reported by Kingery *et al*[1] for zero porosity. For an average conductivity at zero porosity of 0.015 and an estimated actual porosity of 50 per cent, the coatings might be expected to have a conductivity of 0.0075 cal/cm/°C/sec or about 0.03 watt/cm/°C. The temperature drop ΔT (°C) across the coating is given by

---

[1] W. D. Kingery, J. Francl, R. L. Coble, and T. Vasilos, "Thermal Conductivity: X, Data for Several Pure Oxide Materials Corrected to Zero Porosity," *Jour. Amer. Ceram. Soc.*, Vol. 37, p. 107, February 1, 1954, Part II.

$$\Delta T = \frac{P}{k\,2\pi L} \ln \frac{D_0}{D_i}, \tag{2}$$

where P is the power flowing through the coating in watts, k is the thermal conductivity in watts/cm/°C, L is the length of the heater in centimeters, and D_0 and D_i are the outside and inside diameters of the coating, respectively.

In a relatively extreme example, the tube type 6DQ6A, heater power of 7.55 watts is delivered by 20.4 centimeters of folded wire. The outside and inside diameters of the coating are, respectively, about 17 and 5.8 mils. For a thermal conductivity of 0.03 watt/cm/°C, the drop across the coating would be approximately 2°C. Thus, the temperature difference between wire and coating should be negligible even at many times this power input.

The discrepancy between the expected and observed temperature drops could be caused by an effectively lower thermal conductivity than the published figure, due either to a difference in character between the porous coating and a conventional ceramic or to the existence of a thermal barrier at the wire–coating interface. Otherwise, either the temperature measurements or their interpretation must be in error.

Large temperature differences between wire and coating appeared to occur in two situations. In one case, the wire was exposed by a chip in the coating, and the temperatures of the exposed wire and the intact coating were measured by an optical pyrometer. In the second case, a bare and a coated wire having the same base-wire dimensions were heated to the same resistance (i.e., mean temperature), and the temperatures at the mid-points of both wires were measured optically. In both these cases, the apparent temperature drop depended on the accuracy of optical pyrometer measurements, which, in turn, were reliable only insofar as the assumed spectral emissivities of wire and coating were reliable.

Another possible area for misinterpretation lay in the geometry of the arrangements. The wire exposed at the chip lacked the cooling influence of the coating, which could dissipate more heat than the bare wire by radiation from a relatively larger surface. The chip thus tended to produce a higher temperature in the bare wire than if the coating were not broken away. Moreover, the slight rise in wire temperature produces a rise in local electrical resistivity which, for a given current, increases the power converted into heat in that region and further increases the temperature. The extent to which the temperature of the exposed wire may exceed the temperature of neighboring coated regions is moderated somewhat by the increasing ability to

lose heat by enhanced radiation (which rises as the fourth power of temperature) and by thermal conduction of heat from the hot spot along the wire to cooler portions.

It was shown by W. H. Fonger[2] that this wire temperature rise due to a chip in the coating could be high enough to account for part or most of the reported observation. He also pointed out that the measurement of an apparent temperature drop across the coating was actually a comparison of the wire temperature at the bare hot spot with the temperature of the coating over a cooler section of wire away from the exposed portion. That is, what was observed was, to a large extent, a lateral rather than a radial temperature drop.

In the second instance, involving one bare and one coated wire, the temperature distribution along the length of the coated wire was different from that of the bare wire. Thus, although the mean temperatures were the same, the temperatures at particular points might differ.

If the apparent temperature drop across the coating were a true one, it would impose a serious limit to the reduction of heater temperatures through improvement of heater–cathode heat transfer.

Transparency of Alumina Coating

For the investigation of the transparency of the aluminum-oxide heater coating, three types of heaters were mounted in dummy bulbs in the form of single straight lengths of about 3.5 centimeters. All of these heaters were from the same spool of tungsten wire, and were first cleaned in line hydrogen at 1650°C for five minutes. The three types used were as follows:

1. Bare wire, 3.7 mils diameter.
2. Wire cataphoretically coated with alumina and fired in line hydrogen at 1750°C for five minutes; outside diameter was 17.5 mils.
3. Bare wire threaded through a loose sleeve of alumina. The sleeve had an inside diameter of 11.6 mils and an outside diameter of 25 mils, and was made by cataphoretically coating molybdenum wire, firing as for type 2, dissolving out the base wire with acid, washing, and baking dry. The sleeve was fragile and was actually made of about five long pieces.

The cold resistances were tested at 20.5°C by measuring the voltage

[2] W. H. Fonger, RCA Laboratories, Princeton (Private Communication).

drop across each heater with 1.5 milliamperes flowing through it. D-C voltage was then applied to each heater from a battery source until the resistance increased to approximately 4.6 times the cold resistance. (According to the Jones and Langmuir table,[3] this is equivalent to about 1010°K.) The results of these measurements are given in Table I.

Table I—Effect of Wire-Coating Contact on Radiation from Single-Strand Heaters

Type of Heater	Length (mm)	Resistance (ohms) cold	hot	Voltage (volts)	Current (amperes)	Power (watts)
Bare wire	38.0	0.375	1.73	1.15	0.664	0.76
Alumina-coated wire	35.0	0.346	1.61	1.60	0.995	1.6
Bare wire within alumina sleeve	32.5	0.336	1.54	0.900	0.585	0.53

Mean Temperature (all cases): 1010°K
Base-Wire Diameter (all cases): 3.7 mils
Coating Diameter (coated wire): 17.5 mils
Sleeve Dimensions (sleeve wire): O.D. 25.0 mils; I.D. 11.6 mils

The cold resistances varied because of differences in heater length. Figure 1 shows the temperature distributions as measured by an optical pyrometer. The readings were not corrected for spectral emissivities, which were assumed unknown at that time. Despite the wide variation in brightness temperature along each heater, 1010°K represents the mean true temperature because of the almost linear relation between resistance and temperature for tungsten. There is no contradiction in this value even though the peak brightness temperatures are 1060, 1160, and 1350°K, and the true temperatures must then be even higher. Because the wires were short, the temperatures were not uniform, but ranged from the relatively cold temperatures at the heavy stem leads to the peaks at the centers. When temperatures vary so extensively, the maxima might be expected to exceed the means by a considerable amount.

Even without knowledge of the true temperatures, it is possible to draw valid conclusions. For example, the alumina is not transparent.

[3] Properties of Tungsten, *Handbook of Chemistry and Physics*, p. 2457, 33rd Edition, Chemical Rubber Publishing Co., Cleveland, Ohio (1951-1952).

If it were, the radiation from the wire in all three cases would emerge relatively unimpeded and there would be only small differences among the three filaments. The large differences observed demonstrate that the coating is an active participant in the thermal transfer.

The input power represents the rate at which heat must be generated electrically to replenish losses by radiation and end conduction

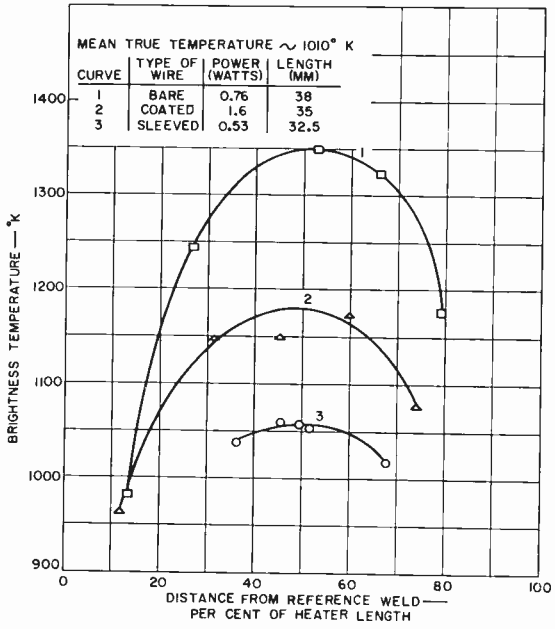


Fig. 1—Brightness temperatures; short wires variously covered.

and thus to maintain the wires at equilibrium temperatures. The bare wire required a power input of 0.76 watt. The coated wire, in which the tungsten was at the same mean temperature as the bare wire, required over twice this amount. When the coating was in the form of a loose sleeve, with conduction paths between wire and alumina almost completely absent, the power requirement fell sharply to 0.53 watt. Because the heat losses to the stem leads could not have differed drastically among the three cases, the differences in power levels must reflect, primarily, differences in radiated power. If these unknown, but large, stem-lead heat losses (large because of short wires and heavy leads) could be subtracted from the total powers, the remaining radiation losses would actually vary by much larger ratios than the gross power figures.

The fact that the radiation from the sleeved wire was much less than that from the bare wire indicates that the alumina sleeve is not transparent; the outflow of heat was diminished by its presence. The radiant energy emitted from the two tungsten surfaces was a function of their temperatures, which were approximately the same. (The effect of differences in temperature distribution is discussed below.) In the case of the bare wire, this energy was lost irretrievably. In the case of the sleeved wire, some of the energy was returned to the tungsten surface by reflection and by re-radiation back from the sleeve. The difference between the input power levels for the bare and sleeved wires must then denote the fraction of the primary radiation returned by the sleeve and absorbed back into the tungsten.

These results do not prove that the alumina is totally opaque; however, the large power difference between the sleeved and bare wires is inconsistent with any great degree of transmission.

Much greater radiation was emitted from the coated wire, in which there was an extensive conduction path between wire and coating. This increased radiation can result only from a greater emissivity, a larger emitting surface, a higher temperature, or a combination of these factors.

As regards emissivity, it is conceivable that the process of applying and firing the coating can so alter the tungsten surface as to raise its emissivity. However, such a hypothetical tungsten emissivity could not have affected the results because, as described above, the alumina sleeve, and therefore the alumina coating as well, blocked radiation directly from the metal. Thus, the governing characteristic was the coating emissivity. The emissivity of the coating may be higher or lower than that of tungsten, but it is improbable that it could be so high as to account for the observed radiation from the coated wire.

However, as a consequence of the opacity of the alumina, the emitting surface was increased to the area of the coating surface, almost five times that of the bare wire. This change could easily explain the larger amount of radiation.

Finally, with respect to temperature, the differences are such as to reduce radiation rather than increase it. First, the conduction of heat from the base wire to the coating surface must have been accompanied by a temperature drop. (Whether this drop was large or small is discussed later.) Therefore, the radiation originated at a surface which was at a lower temperature than that of the base tungsten. Second, although the wire under the coating was held at the same mean temperature as the bare wire, the temperature-distribution curve for the coated wire was flatter than that for the bare wire, as shown

in Figure 1. It can be demonstrated mathematically that a surface having such a narrow temperature spread radiates less energy per unit time than a second surface which has the same area, total emissivity, and mean temperature, but has some portions hotter and some cooler than the mean. The enhanced radiation from the hotter regions more than offsets the diminished radiation from the cooler regions. It can be concluded, then, that the coated wire radiated so much more heat than the bare wire mainly because of a greater emitting surface.

In Figure 1 there appears to be a large difference between the midpoint temperatures of the bare wire and of the coating on the coated wire. The temperatures plotted have not been corrected upward for spectral emissivity. Because the spectral emissivities of tungsten and alumina are not likely to be the same, the required corrections may be widely disparate. Part of the apparent difference might therefore disappear if the true temperatures could be determined. Part of the difference is also due to the dissimilarity in temperature distribution. Although both wires were at the same mean temperature, the coated wire had a smaller dispersion of temperatures and therefore a lower maximum temperature at the center than the bare wire. It would be premature to speculate about the magnitude of the temperature drop across the coating in the coated wire on the basis of such uncertain data.

The results of these tests show (1) that the alumina insulation is largely opaque to radiation at the wavelengths involved, and (2) that the heat transfer includes conduction from base wire to coating surface and radiation from the coating surface.

Resistivity, Emissivity, and Temperature of Tungsten

The results described thus far are essentially qualitative. Filaments were too short for accurate measurements, end losses were too high, and temperatures were too variable from point to point. Therefore, a more appropriate test filament was made by stretching a ten-inch-long tungsten wire taut along the axis of a special glass tube 12 inches long and 1⅛ inches in diameter, with an eight-pin button stem sealed into each end. The ends of the filament were welded to stem leads at the ends of the tube. Pairs of thermocouples were also welded to these same stem leads, one couple at each filament junction, and one added couple further down on each of the two stem leads. The remaining leads of the stems were used as auxiliary feedthrough conductors for the thermocouple wires. Because the 4-mil chromel–alumel thermocouple pairs were thin compared with the 30-mil nickel

stem leads, they had little effect on end heat losses. The tube was baked and evacuated, and a getter was flashed at each end.

The spacing between the thermocouples permitted measurement of the temperature gradients along the nickel filament supports, and subsequent calculation of the heat lost by end conduction. No thermocouple compensation or special reference junctions were used because only temperature differences were important. The couples at the filament–stem-lead joints were used to determine the temperatures of the weld points, but accuracy was not critical.

Table II—Electrical Measurements on Bare-Tungsten Long Filament (Length = 26.27 cm, Diameter = 2.83 mils = 7.19×10^{-3} cm)

Approx. Temp. (°K)	Voltage (volts)	Current (amps)	Resistance (ohms)	Power (watts)	End Loss (watts)
1200	4.37	0.2400	18.2	1.05	0.06
1400	8.00	0.3487	22.9	2.79	0.10
1600	11.70	0.4384	26.7	5.12	0.13

The wire was heated by direct current to three temperatures, in the neighborhood of 1200, 1400, and 1600°K. Currents, voltages, and thermocouple outputs were then recorded, and optical pyrometer readings were made along the length of the wire. A red filter was used which had characteristics such that the effective wavelength of the radiation seen by the pyrometer was approximately 0.65 micron. Dimensions of the wire and electrical measurements are shown in Table II. The end losses, P_E, in watts were calculated from the relation

$$P_E = 2k_N A_N \frac{\Delta T}{\Delta X}, \tag{3}$$

where k_N, the thermal conductivity of nickel in watts/cm/°C, is 0.59; $\Delta T/\Delta X$ is the temperature gradient along the stem lead in °C/cm, and A_N, the cross-sectional area of the stem lead in square centimeters, is 4.56×10^{-3}.

The factor 2 in Equation (3) represents the contribution of both ends of the filament. Temperature gradients at the two ends were in close agreement.

Brightness temperatures for these filaments are shown in Figure 2. Temperatures at the ends are actually true temperatures, as measured

by a thermocouple, but would be close enough to brightness temperatures in this low region to make little difference in the curve.

A first approximation of the true temperatures was made by using the published value of spectral emissivity for tungsten at 0.65 micron between 1200 and 1600°K, which ranges from 0.44 to 0.45. In transmission through the glass bulb, eight per cent of the radiation is lost

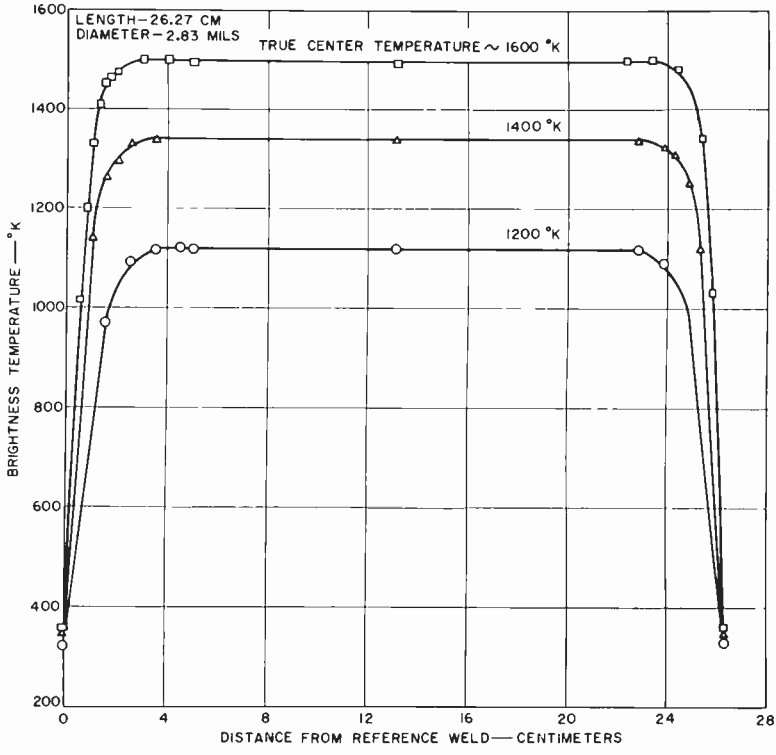


Fig. 2—Brightness temperatures of long wires; bare tungsten.

by reflections from the inner and outer glass surfaces, and the apparent emissivity is reduced to 0.92 × 0.45, or to between 0.41 and 0.42. Corrections for an effective emissivity of 0.42 yielded new temperature distributions with plateau temperatures of 1172, 1419, and 1599°K true, as compared with 1120, 1341, and 1501°K brightness, respectively.

The wire was then considered as divided into segments with associated temperatures equal to the mean of the temperatures at the segment endpoints. Each large central region of constant tempera-

ture was considered as a single segment, and the regions of rapidly changing temperatures near the stem leads were divided into 0.5- and 1-centimeter segments. The electrical resistivity for each temperature was determined from the Jones and Langmuir table. The expected total resistance R in ohms was then calculated from the relation

$$R = \frac{1}{A} \sum \rho_T \Delta l, \tag{4}$$

where A is the cross-sectional area in square centimeters, ρ_T is the

Table III—Representative Data and Resistance Calculation for Bare-Tungsten Long Filament

Measured Data			Calculations				
Position (cm)	T_B* (°K)	T_T** (°K)	Segment (cm)	Δl (cm)	T_{mean} (°K)	ρ_T (microhm-cm)	$\rho_T \Delta l$ (microhm-cm²)
0	——	341	0-1.0	1.0	769	18.08	18.08
1.00	1141	1197	1.0-1.5	0.5	1264	32.93	16.46
1.50	1265	1332	1.5-2.0	0.5	1350	35.60	17.80
2.00	1297	1368	2.0-2.5	0.5	1388	36.82	18.41
2.50	1334	1409	2.5-3.5	1.0	1414	37.64	37.64
3.50	1341	1419	3.5-22.77	19.27	1419	37.80	728.40
13.13	1341	1419	22.77-23.77	1.0	1410	37.50	37.50
22.77	1341	1419	23.77-24.27	0.5	1395	37.03	18.52
23.77	1326	1401	24.27-24.77	0.5	1353	35.70	17.85
24.27	1316	1389	24.77-25.27	0.5	1248	32.47	16.24
24.77	1252	1317	25.27-26.27	1.0	762	17.85	17.85
25.27	1125	1179					
26.27	——	344					$\Sigma \rho_T \Delta l = 944.7$

$$R = \frac{944.7 \times 10^{-6}}{4.07 \times 10^{-5}} = 23.2 \text{ ohms}$$

* Brightness temperature.

** True temperature based on effective emissivity of 0.42.

resistivity at the temperature of the segment in ohm-centimeters, and Δl is the length of the segment in centimeters.

Table III shows a representative calculation based on this method, and Table IV compares the calculated resistances with the measured resistances of Table II. The agreement is very good between resistances based on the Jones and Langmuir data and published spectral emissivities for tungsten and resistances based on a simple measurement of current and voltage. Although this agreement does not constitute rigorous, absolute proof, it makes reasonable the assumption

that the book values of resistivity and spectral emissivity accurately describe the characteristics of clean tungsten wire. Therefore, if the clean, bare wire is visible, its true temperature may be measured with confidence by optical pyrometer. If it is not visible, its temperature is represented by its resistivity. However, unless the temperature of the wire is uniform, or nonuniformity can be accounted for, the resistance method may not be practical.

Table IV—Comparison of Calculated and Measured Resistances of Bare-Tungsten Long Filament

Approx. Temp. (°K)	R calc. (ohms)	R meas. (ohms)	Deviation (%)
1200	18.2	18.2	0
1400	23.2	22.9	1.3
1600	27.0	26.7	1.1

For example, if the temperature distribution of the bare wire under discussion were not known and the wire were assumed to be at a single, uniform temperature, the resistivity ρ in ohm-centimeters for a temperature of 1400°K (center temperature 1419°K) would be calculated as follows:

$$\rho = \frac{RA}{L} = \frac{22.9 \times 4.07 \times 10^{-5}}{26.27} = 35.5 \times 10^{-6}, \tag{5}$$

where R is the resistance in ohms (volts/amperes), A is the cross-sectional area of wire in square centimeters, and L is the length of wire in centimeters.

In the resistivity–temperature table, this resistivity corresponds to a temperature of 1347°K, which is 72°K less than the true plateau temperature. This difference reflects the influence of the cooler ends of the wire on the average resistivity.

When the true temperatures are known, the total emissivity may be calculated. The most direct way is to consider the center of the wire. Because the wire is at thermal equilibrium, the power generated per unit length equals the power lost per unit length. Also, because the temperature gradient along the wire is zero on the central plateau, no heat is conducted along the wire from the center. Thus the power generated per unit length equals the power radiated per unit length

at the center, as shown by the following equation:

$$\frac{I^2\rho}{A_c} = 5.67\ \epsilon_t\ A_s\ T^4 \times 10^{-12}\ \text{watts/cm}, \tag{6}$$

where I is the current in amperes, ρ is the resistivity at center temperature in ohm-centimeters, A_c is the cross-sectional area of the wire in square centimeters, ϵ_t is the total emissivity of the wire, A_s is the surface area per unit length in square centimeters per centimeter, and T is the temperature at the center in °K.

This equation may be rewritten

$$\epsilon_t = \frac{I^2\rho}{5.67\ A_c A_s T^4} \times 10^{12}. \tag{7}$$

Table V—Total Emissivities of Bare-Tungsten Long Filament

True Temp. (°K)	Current (amps)	Resistivity (microhm-cm)	Total Emissivity calc.	published*
1172	0.2400	30.18	0.176	0.139
1419	0.3487	37.80	0.218	0.178
1599	0.4384	43.53	0.246	0.207

* American Institute of Physics Handbook 1957.

Then, the terms A_c and A_s can be combined;

$$A_c A_s = \frac{\pi D^2}{4} \times \pi D = \frac{\pi^2 D^3}{4}\ \text{cm}^3 \tag{8}$$

where D, the diameter of the wire in centimeters, is 7.19×10^{-3}. Equation (7) for total emissivity ϵ_t may then be rewritten

$$\epsilon_t = \frac{1.92\ I^2\rho}{T^4} \times 10^{17}.$$

Table V compares the calculated results with published values of total emissivity for tungsten. Although there is fair agreement between experimental and book values, the consistently higher experimental values suggest that the surface of factory wire is not as clean

and/or as polished as the specimen used in obtaining the published emissivities. A test wire was oxidized almost to blackness by heating to a red glow in air and its emissivity was then measured near 1200°K by the technique described above. The emissivity was found to have risen to 0.68. Apparent values of 0.3 and 0.4 have been recently reported by other workers. These high values were probably also due to surface imperfections or contamination, or possibly to some gas in the tube used. The spectral emissivity appears to be relatively insensitive to minor surface variations.

These results demonstrate that the resistivity–temperature table is consistent with published values of spectral emissivity when both are applied to tungsten heater wire. The total emissivity values derived from temperatures based on these characteristics agree approximately with published values for clean, polished tungsten, and deviate only as might be expected. It is reasonable to assume, therefore, that the properties of the wire are well defined, and that wire temperatures can be arrived at with assurance.

Emissivities and Temperature of Alumina Coating

The spectral emissivity of the alumina coating can be found if both the true and the brightness temperatures are known. The brightness temperatures can be measured by optical pyrometer. The true temperatures must be deduced indirectly. Brightness temperature, T_B, and true temperature, T, are related by the following equation:

$$T = \frac{T_B}{1 + \dfrac{\lambda}{1.438 \times 10^4} T_B \ln \epsilon_\lambda}, \tag{9}$$

where λ is the wavelength in microns corresponding to the pyrometer filter, and ϵ_λ is the spectral emissivity at this wavelength. For wavelengths near 0.65 micron, as is usually the case, this equation becomes

$$T = \frac{T_B}{1 + \dfrac{T_B \ln \epsilon_\lambda}{22{,}120}}. \tag{10}$$

This equation is plotted as a family of curves for easy reference in Figure 3. Although a wavelength of 0.665 rather than 0.65 micron was

used in the preparation of these curves, the temperature difference rarely amounts to more than a few degrees.

True temperatures are also needed for the calculation of total emissivity from the dimensions of the coating and the radiated power by means of Equation (6). True temperatures of the alumina coating were obtained by an indirect approach. From the same spool of 2.83-

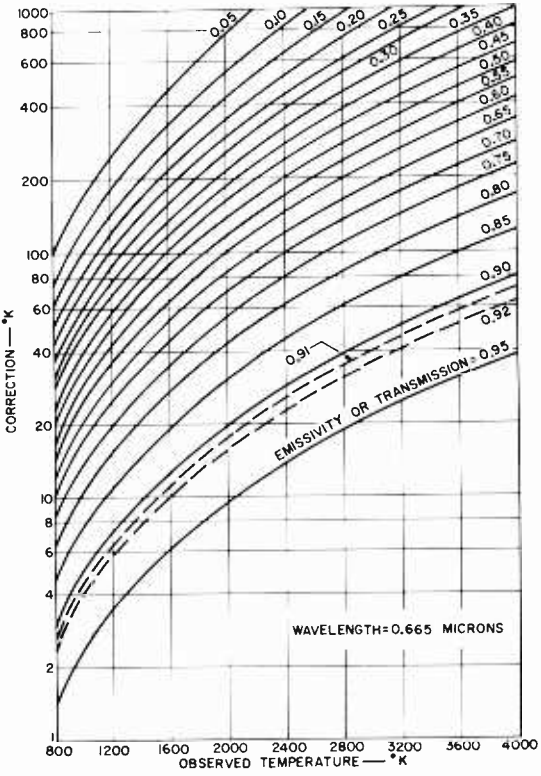


Fig. 3—Corrections to brightness-temperature determinations.

mil tungsten wire used previously, two more types of ten-inch-long filaments were made. First the wire was drag-coated with alumina and fired to a diameter of 9.2 mils. One ten-inch length of this coated wire was then mounted in a long, glass tube exactly as described before. The second piece of coated wire was dipped in a lampblack (dag) graphite dispersion diluted with methanol. This process deposited a very thin film of graphite on the alumina surface without appreciable change in the coating diameter and without penetration of the coating

which might have altered its electrical or thermal conductivity. The ten-inch length of black-coated wire was then mounted exactly like the others. The tubes were baked with the filaments glowing to drive out gas and volatilize any remaining vehicle and binder in the graphite.

The graphite provides a surface of high emissivity, both total and spectral. Surfaces of high emissivity are more likely to correspond to published radiation characteristics than a white or metallic surface. Also, the difference between true and brightness temperatures becomes very small for higher emissivities. At 0.76, the listed spectral emissivity of graphite, the difference is only 35 degrees at 1600°K. Thus, even a relatively large discrepancy between the actual behavior of the graphite film and the published emission characteristics introduces only small errors in the estimation of true temperatures.

Table VI—Electrical Measurements—White- and Black-Coated Long Filaments

Approx. Temp. (°K)	Voltage (volts)		Current (amps)		Resistance (ohms)		Power (watts)		End Loss (watts)	
	white	black	white	black	white	black	white	black	white	black
1200	8.70	15.9	0.4780	0.8844	18.2	18.0	4.16	14.1	0.09	0.22
1400	13.6	27.7	0.5800	1.176	23.4	23.6	7.89	32.6	0.14	0.36
1600	17.4	36.7	0.6540	1.374	26.6	26.7	11.4	50.4	0.19	0.50

White: Length = 25.54 cm, Diameter = 9.2 mils.
Black: Length = 25.33 cm, Diameter = 9.2 mils.

The white and black heaters were then raised to temperatures close to 1200, 1400, and 1600°K, based on the resistances expected for those temperatures, as done previously with the bare wire. Results are shown in Table VI and Figures 4 and 5.

The results show that the black heater radiates three to four times as much heat as the white heater of the same dimensions when the enclosed wires are close to the same temperature. If the brightness temperatures of the black heater are corrected for a spectral emissivity of 0.76 (less eight per cent for loss by reflection from the glass bulb, for an apparent emissivity of 0.7), the true plateau temperatures become 1170, 1442, and 1600°K, as compared with brightness temperatures of 1148, 1407, and 1558°K, respectively. Several facts testify to the accuracy of these true temperatures. First, they agree very well with the temperatures previously calculated for the bare tungsten wire (1172, 1419, and 1599°K). Second, if the spectral emissivity

were higher (e.g., unity, as for a perfect black body), the error would be only 42°K at worst; it is certainly much less than this figure. The true temperatures for a black body would be the same as the brightness temperatures. Third, if the emissivity were much less than the assumed value of 0.76, the calculated true temperatures would appear to be high as compared with those of the bare wire; it is inconceivable that the coating temperatures would exceed those of the wire.

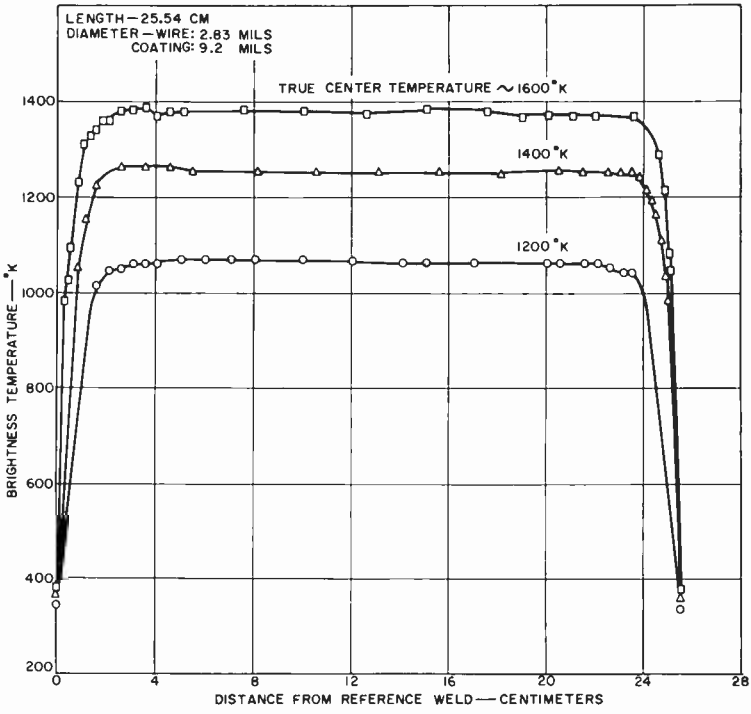


Fig. 4—Brightness temperatures of long wires; alumina coated.

This agreement between the coating-surface temperatures and those of the bare wire suggests that there is no large temperature drop across the coating. In testing this conclusion, it was assumed that the observed coating temperatures on the black heater were also those of the wire underneath. The resistances expected of the wire at these temperatures were then calculated, using the same procedure as for Table III. These calculated values agree very well with the actual measured resistances, as shown in Table VII. It may be concluded, therefore, that the assumed spectral emissivity is substantially correct

and that, even at these high power levels, there is very little drop in temperature across the coating thickness.

Equation (7) was used to calculate the total emissivities of the black coating at the centers of the plateaus of temperature. If the value $\pi D = 0.0735$ centimeter is substituted for A_s, the coating surface area per unit length, the total emissivity ϵ_t is given by

$$\epsilon_t = \frac{5.90\, I^2 \rho}{T^4} \times 10^{17}. \tag{11}$$

Total emissivity values calculated from this equation for the graphite-

Table VII—Comparison of Calculated and Measured Resistances for Long Filament with Graphite on Alumina Coating

Approx. Temp. (°K)	R calc. (ohms)	R meas. (ohms)	Deviation (%)
1200	18.3	18.0	1.6
1400	23.6	23.6	0
1600	26.8	26.7	0.4

covered alumina were 0.75, 0.72, and 0.74, respectively, for temperatures of 1200, 1400, and 1600°K. These values are very close to the spectral emissivity, as might be expected for relatively black surfaces.

Because the power flowing from the base wire to the white alumina surface in the case of the conventionally coated heater is less than one third that for the black heater, the temperature drop must be less in the same proportion. Therefore, the wire and the white coating are at essentially the same temperature. This temperature must be close to that of the black-coated wire because both have almost exactly the same length and hot resistances. In the black heater, however, the temperature plateaus are broader. Thus for a given hot resistance, or average temperature, the plateau temperature of the black heater may be expected to be slightly less than that of the white heater.

At a first glance, therefore, the spectral emissivity of the white alumina coating must convert brightness temperatures of about 1070, 1256, and 1385°K to true temperatures of about 1172, 1419, and 1599°K, respectively. A suitable value of ϵ_λ was calculated to be 0.15. However, when this value was used to calculate expected resistances for tentative true temperatures of tungsten, agreement between calculated and measured resistances was not as good as that found for

Table VIII—Comparison of Calculated and Measured Resistances for Long Filament with White Alumina Coating
(Assumed $\epsilon_\lambda = 0.13$)

Approx. Temp. (°K)	*R* calc. (ohms)	*R* meas. (ohms)	Deviation (%)
1200	18.2	18.2	0
1400	23.1	23.4	1.3
1600	26.4	26.6	0.75

the bare wire and the black-coated heater. A value of 0.12 was found to yield good agreement, as shown in Table VIII. This calculation corresponds to an actual value of 0.13 because an eight-per-cent difference is assumed for transmission loss in the glass bulb.

The true plateau temperatures of the white heater are then 1196, 1429, and 1595°K. Total emissivities for these values calculated from

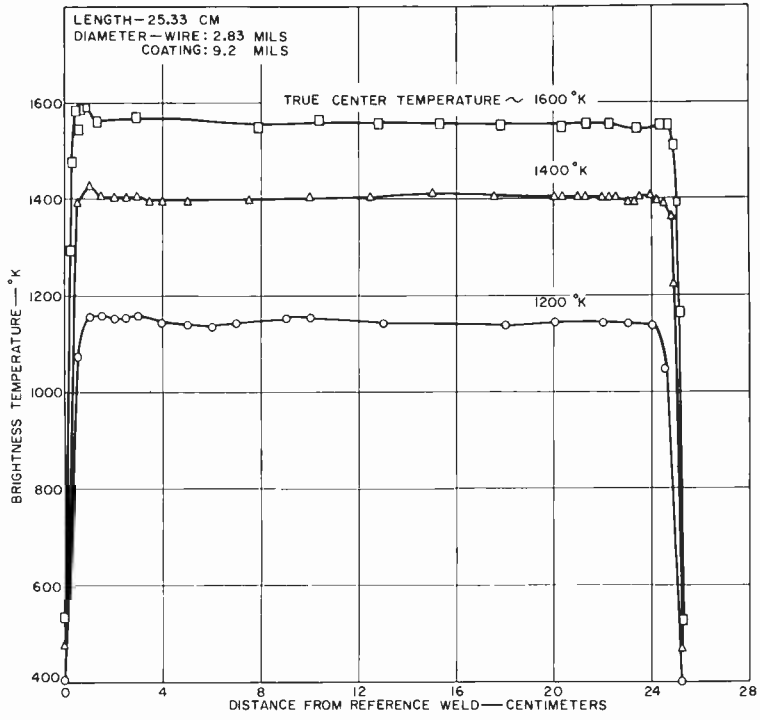


Fig. 5—Brightness temperatures of long wires; graphite over alumina coating.

Equation (11) (which is the same for the black and the white heaters because their dimensions are the same) were 0.20, 0.18, and 0.17, respectively. These emissivities are about one-fourth that of graphite. A graphite-coated heater thus radiates heat at a rate four times that of a white heater having the same dimensions and the same temperature.

Previous corrections of optical pyrometer sightings on alumina coatings were made for a spectral emissivity of 0.30, the number listed in the American Institute of Physics Handbook as the probable value for the oxide formed on the smooth metal. This value yields a true temperature of 1640°K for a brightness temperature of 1500°K. The calculated spectral emissivity of 0.13 yields a true temperature of 1760°K. The difference of 120°K emphasizes the importance of accuracy in spectral emissivity values.

A search of the literature had uncovered no information on the emissivity of heater coatings. A further search in other fields, however, yielded at least partial corroboration of the results of the present work. The diffuse reflecting power of aluminum oxide pigment is given in the Handbook of Chemistry and Physics as 84.1 per cent at 0.6 micron and 87.7 per cent at 0.95 micron. For opaque bodies, the sum of reflectivity and emissivity is unity. Thus, if the reflecting power is taken as a measure of reflectivity, the spectral emissivities for alumina pigment are 0.159 at 0.6 micron and 0.123 at 0.95 micron. The value of 0.13 determined in the present work for alumina coatings at 0.65 micron falls in this range. In a paper on alumina properties[4] there is reference to measurements of the emissivities of alumina ceramic bodies by various workers. The variation of total emissivities is shown in Table IX. The spectral emissivity was reported by one group as 0.15 at 0.655 micron from 1000°C to 1600°C. These figures are in reasonable agreement with the results in the present paper. The published emissivities, both spectral and total, are slightly higher.

Thermal Conductivity and the Temperature Drop Across the Coating

The data presented above show a very small temperature drop across the coating. Because the determination of the temperature drop was admittedly not direct, but was reasoned from sensible assumptions about the resistance of the wire and the emissivity of graphite, there is considerable uncertainty about the exact magnitude

[4] A. S. Russell, *et al.*, "Alumina Properties," *Technical Paper No. 10 (Revised)*, Aluminum Company of America, Pittsburgh (1956).

of the drop. The fundamental assumption was that the difference in temperature between the wire and the surface of the coating was too small to be detectable within the limits of the experimental technique. Because the agreement between measured wire resistances and those calculated by an assumption of zero temperature drop in the coating was better than 2 per cent, it might be supposed that 2 per cent is approximately the maximum temperature drop that might actually have existed. At 1600°K, this drop would be 32°K.

Table IX—Total Emissivity of Alumina Ceramic[4]

Temperature (°C)	Temperature (°K)	Emissivity
100	373	0.9
400	673	0.8
750	1023	0.5
1000	1273	0.3
1600	1873	0.18
2760	3033	0.12

Equation (2) may be rewritten as a formula for the thermal conductivity k, as follows:

$$k = \frac{P}{2\pi\Delta T L} \ln \frac{D_0}{D_i}. \tag{12}$$

For an outer coating diameter of 9.2 mils, a wire diameter of 2.83 mils, and a temperature drop of 32°C, Equation (12) becomes

$$k = 0.0059 \frac{P}{L}, \tag{13}$$

where P/L is the radial power flow per unit length. At the midsection of the filament, this factor is equal to the power generated per unit length of wire, $i^2\rho/A_c$. At 1600°K, $\rho = 43.55 \times 10^{-6}$ ohm-centimeter. A_c, the cross-sectional area of the wire, is 40.6×10^{-6} square centimeter. The current i was 1.374 amperes for the black coating. Because of its greater power flow, 2.03 watts per centimeter, this coating would be expected to exhibit the greater temperature drop. The calculated thermal conductivity for the graphite-covered coating was 0.012 watt per centimeter per degree centigrade, which is lower than the figure of 0.03 watt per centimeter per degree centigrade based on published

ceramic data and an assumption of 50 per cent porosity. However, the assumed temperature drop is an extreme estimate; the porosity may differ from 50 per cent, the contact between wire and coating may not be uniform, and the alumina coating might be expected to be different from the bulk ceramic in some respects. Therefore, it may be concluded that the conductivity values are in fair agreement. In terms of tube operation, where powers are usually less than 0.35 watt per centimeter, the difference is negligible, e.g., the temperature drop is typically under 2°C for a conductivity of 0.03, as compared with 8°C for 0.012.

Thus, the following conclusions can be drawn: (1) the published thermal conductivities for alumina ceramic are applicable to heater coating; (2) there is a negligible temperature difference between heater coating and heater wire (except where the coating is in direct contact with the cooler cathode); and (3) the apparently large temperature drops observed previously were not in fact large, but were misinterpretations based on inadequate emissivity data and misleading geometry.

Conclusions

The spectral emissivity of alumina heater coatings at 0.65 micron in the temperature range of heaters is close to 0.13. The total emissivity is between 0.17 and 0.20.

The spectral emissivity of bare tungsten heater wire is 0.45 at 0.65 micron at typical heater temperatures and is relatively insensitive to contamination of the wire surface. The total emissivity of bare, clean wire rises with temperature from about 0.18 at 1200°K to 0.25 at 1600°K, and is more dependent than spectral emissivity on the conditions of the surface, generally increasing with contamination or roughness.

There is very little difference in temperature between the heater wire and the heater-coating surface except at points where the coating is cooled by contact with the cathode. Thus, there is no serious obstacle to the reduction of heater temperatures by an increase in the coating emissivity. The emissivity of white alumina is so low that a great increase in radiation efficiency may be achieved by darkening the coating either in bulk or merely on the surface.

The alumina coatings are largely opaque to heat radiation. Heat is transferred almost solely by conduction from the base wire to the heater-coating surface and by radiation from there to the cathode.

Temperatures determined by optical pyrometry with proper emis-

sivity corrections agreed within two per cent with those determined from the electrical resistivity of the heater wire.

Since the completion of this work, the application of these principles has resulted in the improvement of the emissivity of heater coatings in presently produced electron tubes with a substantial increase in heater life.

ACKNOWLEDGMENT

The authors acknowledge the assistance and guidance obtained from the work of and discussions with O. Schade, Jr., of the Design Engineering Laboratory at Harrison, N. J., and W. Fonger and M. Kestigian of the RCA Laboratories, Princeton, N. J.

THE EFFECT OF A D-C MAGNETIC FIELD ON THE UHF PERMEABILITY AND LOSSES OF SOME HEXAGONAL MAGNETIC COMPOUNDS*†

By

R. L. Harvey, I. Gordon, and R. A. Braden

RCA Laboratories,
Princeton, N. J.

Summary—The properties of certain ferromagnetic materials having a hexagonal crystal structure are described with particular reference to their use as tuning elements at frequencies above 200 megacycles. Tuning is accomplished by varying the permeability by means of an adjustable biasing magnetic field. This paper discusses the nature of the materials, the method of evaluation, and the magnetic properties at frequencies from 250 to 1750 megacycles. In this frequency range the hexagonal crystal materials are found to be superior to the better known cubic ferrites.

Introduction

FERRITE MAGNETIC COMPOUNDS have relatively high permeability and low losses at high frequencies, and have made a major impact on the design of some electronic components and systems. However, ferrites have an inherent upper frequency limit of usefulness because of a magnetic resonance associated with the domain walls.[1]

In 1957 a group of magnetic compounds was described by workers at Philips Laboratories.[2] These compounds, although similar in many respects to better known cubic ferrites, have a different crystal structure. They crystallize in the form of hexagonal plates. Their chemical relation to better-known oxidic magnetic compounds is illustrated in Figure 1. The ferrites are composed of Fe_2O_3 and MeO. (The symbol Me represents any combination, in suitable proportions, of the divalent elements listed in the figure.) The permanent magnet compound (PM) contains only Fe_2O_3 and BaO. The hexagonal materials are

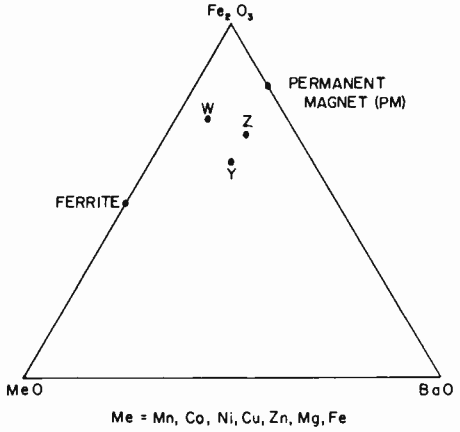
* Manuscript received 15 May 1961.

† This work was carried out under contract with the U. S. Army Signal Research and Development Laboratory.

[1] R. L. Harvey, "Ferrites and Their Properties at Radio Frequencies," *Proc. 1953 N.E.C.*, Vol. 9, p. 287 (1954).

[2] G. H. Jonker, H. P. J. Wijn, and P. B. Braun, "Ferroxplana, Hexagonal Ferromagnetic Iron-Oxide Compounds for Very High Frequencies," *Philips Tech. Rev.*, Vol. 18, p. 145 (1956/57).

specific compounds of Fe_2O_3, BaO, and MeO. They may be classified in three subgroups which, following the Philips Laboratories nomenclature, are known as W, Y, and Z compounds. In some of the compounds the hexagonal crystal structure causes the magnetic anisotropy to be very high. In the Y compounds and some of the W and Z compounds the basal plane of an individual crystallite is an easy plane of magnetization, and the crystal axis is difficult to magnetize. This characteristic makes it possible to orient the individual crystallites



FERRITE	CUBIC	$Me\ O \cdot Fe_2O_3$
PM	HEXAGONAL	$Ba\ O \cdot 6Fe_2O_3$
W	HEXAGONAL	$Ba\ O \cdot 2Me\ O \cdot 8Fe_2O_3$
Y	HEXAGONAL	$2Ba\ O \cdot 2Me\ O \cdot 6Fe_2O_3$
Z	HEXAGONAL	$3Ba\ O \cdot 2Me\ O \cdot 12\ Fe_2O_3$

Fig. 1—Compositional arrangement of the ferrites, permanent magnets, and W, Y, and Z hexagonal compounds.

magnetically while forming the core so that its nature approaches that of a single crystal. Some of the interesting properties of these compounds are enhanced by magnetic orientation.

The work on which the present paper is based is part of a continuing investigation of hexagonal magnetic compounds, the primary purpose of which is to discover compounds which may be useful in electrical circuits, especially for military applications. This investigation is very broad in nature, covering both a wide range of compositions and a wide frequency range (from d-c to microwave frequencies). One area on which attention has been focused is the ultra-high-frequency region (300-3000 megacycles). In this frequency range

there is need for high-permeability cores for inductance coils. Ferrite cores cannot be used above the lower part of this frequency band, because of the magnetic resonance which occurs in the material, as mentioned above. In the hexagonal compounds, however, the magnetic resonance occurs at a higher frequency, and useful values of permeability have been found over the entire UHF range and beyond. Obviously, the characteristics required for high-frequency inductance cores are high permeability and low loss. In some applications a circuit is tuned by applying a biasing magnetic field to the core so as to reduce its permeability. This adds the further requirement that there be a large change of permeability brought about by a relatively small d-c field. The evaluation of materials to determine their applicability to fixed and variable tuned circuits, and some of the results obtained, are discussed in this paper.

Experimental Apparatus and Techniques

The complex permeability, $\mu' - j\mu''$, is measured by a slotted-line method. The sample, in the form of a toroid, is placed against the shorted end of a coaxial line connected to a slotted line. The voltage-standing-wave ratio and the position of the voltage minimum are measured with the slotted line, first without the sample, and second with the sample in place. The real component of permeability is calculated from the shift of the voltage minimum position, and the imaginary component from the change in voltage-standing-wave ratio. Then, if (a) the thicknes of the sample is much smaller than a quarter wavelength in the material, (b) the sample fits the inner and outer conductors perfectly, and (c) the voltage-standing-wave ratio is not smaller than about 10, the following simple formulas[3] may be used.

$$\mu' = 1 + \frac{D}{T},$$

$$\mu'' = \left(\frac{1}{S_1} - \frac{1}{S_2} \right) \frac{\lambda_0}{2\pi T},$$

where

$D =$ displacement of voltage minimum when sample is inserted,

$T =$ thickness of sample,

[3] G. T. Rado, R. W. Wright, and W. H. Emerson, "Ferromagnetism at Very High Frequencies. III. Two Mechanisms of Dispersion in a Ferrite," *Phys. Rev.*, Vol. 80, No. 2, p. 273, 1950.

$S_1 =$ VSWR (with sample),
$S_2 =$ VSWR (no sample),
$\lambda_0 =$ free-space wavelength.

If the sample does not exactly fit the concentric line holder, the following relations may be used to obtain the true values of μ' and μ'':

$$\mu_t' - 1 = (\mu_m' - 1) \frac{\ln \frac{d_0}{d_i}}{\ln \frac{d_0'}{d_i'}}$$

$$\mu_t'' = (\mu_m'') \frac{\ln \frac{d_0}{d_i}}{\ln \frac{d_0'}{d_i'}}$$

where

$\mu_t =$ true permeability,
$\mu_m =$ measured permeability,
$d_o =$ diameter of outer conductor of sample holder,
$d_i =$ diameter of inner conductor of sample holder,
$d_0' =$ outside diameter of sample,
$d_i' =$ inside diameter of sample.

If the VSWR of the sample is low, or if the thickness approaches a quarter wavelength, a more exact set of expressions must be used.[4]

The Q of the sample is defined by the simple relation

$$Q = \frac{\mu'}{\mu''}.$$

For the measurement of μ' and μ'' in the presence of a d-c magnetic field, the sample holder shown in Figure 2 was designed. Direct cur-

[4] I. Bady and R. Franklin, "Measurement of Permeability and Q of Magnetic Materials over the Frequency Range of 50 to 500 Megacycles," *Engineering Report 1959, Signal Corps Engineering Laboratory,* Ft. Monmouth, N. J.; or *IRE National Convention Record,* Part 5, p. 137 (1957).

rent up to a maximum of 200 amperes flows along the inner conductor, as indicated by arrows in the figure. This produces a circumferential magnetic field in the core, the intensity of which can be easily calculated. The sample holder was designed for use at 250 megacycles. The shunt stub through which the current is introduced is resonant at this frequency so that it presents a high impedance to the slotted line. Measurement can also be made at odd multiples of 250 megacycles.

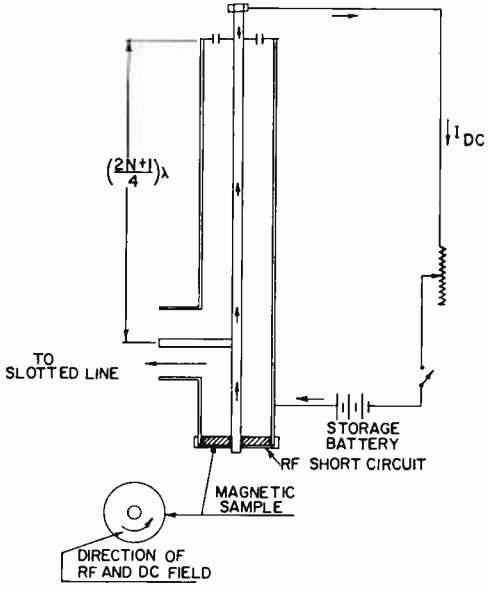


Fig. 2—Special sample holder for measuring μ' and μ'' in the presence of a d-c field.

Experimental Results

The data to be presented were obtained on toroidal cores which were magnetically oriented so that the plane of the toroid was the plane of easy magnetization. In the sample holder, both the r-f and the d-c magnetic fields were in the easy plane. The measurement procedure comprised four operations, as follows:

1. The VSWR and voltage minimum position were measured before the sample was inserted.
2. The sample, in a completely demagnetized* state, was inserted, and the above measurements were repeated.

* Demagnetization was effected by heating the sample to a temperature above the Curie Temperature.

3. A d-c current of 200 amperes was applied and then removed. This put the sample into the remanent† state. The VSWR and voltage minimum position measurements were made.
4. The VSWR and voltage minimum position measurements were repeated with various values of d-c current flowing in the sample holder.

During the last operation the higher values of magnetizing current were applied only during the actual measurement, so as to avoid heating the sample holder and thus changing the temperature of the sample. This was necessary because the permeability would be influenced by a temperature change.

Table I shows the experimental data on some of the compounds. F is a quality factor chosen as a means of comparing the various samples with respect to their performance as inductance coil cores. It is defined by the equation

$$F = \mu_0' Q_0 \frac{\mu_0'}{\mu_{40}'},$$

where

$\mu_0' = \mu'$ in the remanent state,
$Q_0 = Q$ in the remanent state,
$\mu_{40}' = \mu'$ with an applied magnetic field of 40 oersteds.

The data in the table are grouped to show several trends with respect to composition, orientation, firing temperature, etc.

(1) Cores 1-4 and 8-10 show how variations of the Me constituent influence the measured properties. In general, the values of μ_0', (μ_0'/μ_{40}'), and F can be optimized to maximum values by varying the composition.

(2) Cores 11-13 show how the firing temperature influences the properties. As the firing temperature increases (within limits), the values of μ', μ_0' and (μ_0'/μ_{40}') increase, but Q diminishes. The values of F and $(\mu_0' Q_0)$ pass through a maximum value.

(3) Cores 6 and 7, which are of the same composition, may be

† The remanent state has been found to be very reproducible. Irrespective of the value of the biasing field, the same values of μ' and μ'' are found when the field is removed.

Table I—Comparison of the Magnetic Properties of Some Hexagonal Magnetic Compounds

No.	Composition	T_f	μ'_d	Q_d	μ'_0	Q_0	μ'_0/μ'_{40}	$\mu'_0 Q_0$	F
1	$(Co)_2Z$	1220	13.9	29.0	11.6	41.3	2.30	478	1110
2	$(Co_{.85}\ Zn_{.15})_2Z$	1220	28.0	14.2	18.8	21.4	3.36	403	1350
3	$(Co_{.7}\ Zn_{.3})_2Z$	1220	30.9	13.0	20.3	18.2	4.10	370	1500
4	$(Co_{.55}\ Zn_{.45})_2Z$	1215	29.4	15.2	18.9	23.0	3.20	436	1390
5	$(Co_{.9}\ Cu_{.1})_2Z$	1215	19.4	13.7	14.3	33.4	2.64	480	1150
6	$(Co_{.75}\ Cu_{.25})_2Z$	1230	27.6	19.2	20.2	34.0	3.26	686	2230
7	$(Co_{.75}\ Cu_{.25})_2Z$	1230	12.5	18.4	10.5	21.0	1.55	220	342
8	$(Co_{.8}\ Cu_{.1}\ Zn_{.1})_2Z$	1215	22.3	24.2	16.2	34.5	2.91	560	1630
9	$(Co_{.7}\ Cu_{.1}\ Zn_{.2})_2Z$	1215	21.5	27.0	16.6	37.0	3.10	615	1700
10	$(Co_{.6}\ Cu_{.1}\ Zn_{.3})_2Z$	1215	22.0	24.6	17.3	29.3	2.93	508	1490
11	$(Co_{.8}\ Cu_{.1}\ Zn_{.1})_2Z$	1205	16.1	29.4	12.8	36.6	2.55	470	1200
12	$(Co_{.8}\ Cu_{.1}\ Zn_{.1})_2Z$	1215	22.3	24.3	16.2	34.5	2.91	560	1630
13	$(Co_{.8}\ Cu_{.1}\ Zn_{.1})_2Z$	1315	54.0	6.2	42.2	4.9	7.46	206	1540
14	$(Co_{.9}\ Ni_{.05}\ Cu_{.05})_2Z$	1220	20.7	26.3	15.3	35.5	2.79	545	1540
15	$(Co_{.7}\ Ni_{.15}\ Cu_{.15})_2Z$	1220	18.8	27.2	15.3	31.9	2.51	489	1225
16	$(Ni)_2Y$	1260	7.9	7.1	6.8	10.0	1.45	68	98
17	$(Ni_{.25}\ Zn_{.75})_2Y$	1260	29.9	2.5	24.5	2.5	3.36	61	206

$Z = Ba_3Fe_{24}O_{44}$ $Y = Ba_2Fe_{12}O_{22}$ Measurement frequency is 250 megacycles. All samples are oriented except No. 7, which is nonoriented. T_f is the firing temperature in °C. μ_d' and Q_d are demagnetized values. μ_0' and Q_0 are remanent state values. μ_{40}' is μ' with 40 oersteds d-c field. $F = \mu_0' Q_0 (\mu_0'/\mu_{40}')$.

compared to show that a substantial improvement results from orientation.

(4) Cores 8-15 represent compositions in which the Me component is a mixture of three of the permitted divalent elements. Obviously, a very large number of compositional variations are possible.

(5) When the magnetic material is changed from the demagnetized state to the remanent state, the permeability (μ') is always reduced. The Q increases at operating frequencies below the magnetic resonance frequency.

The values shown in Table I include some of the best achieved to date. Sample No. 6 is thought to have at least twice the F value at 250 megacycles that can be achieved using cubic ferrites. At higher frequencies the hexagonal compounds are more advantageous.

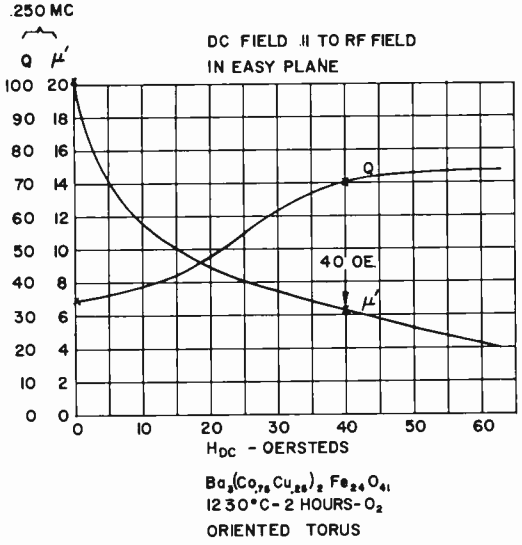


Fig. 3—μ' and Q as a function of d-c bias field.

In addition to the bias-field arrangement employed in the tests described above, two other configurations of the d-c magnetic field are possible. The bias field may be parallel to the axis of the torus, or perpendicular to it. Both of these arrangements were tested and the effects obtained were very similar to that illustrated in Figure 3; however, a large external field is required to effect the same change in permeability. The circumferential field arrangement is thought to

be more suitable for investigatory purposes, since it permits accurate calculation of the field intensity in the sample without the complication of demagnetizing factors.

Figure 4 shows μ' and μ'' as a function of magnetization state and frequency for the composition $(Co_{.45}Zn_{.55})_2Z$. The maximum value of μ'' occurs at the frequency of magnetic resonance (probably domain wall resonance). At frequencies sufficiently below the resonance frequency, μ' has a constant value, although some of the curves do not extend far enough to show this. As the resonance frequency is approached, the value of μ' falls, then rises again at still higher frequencies. Increasing the applied d-c magnetic field moves the resonance to a higher frequency. In the particular composition of Figure 4, a field of 18 oersteds placed the resonance at 1750 megacycles.

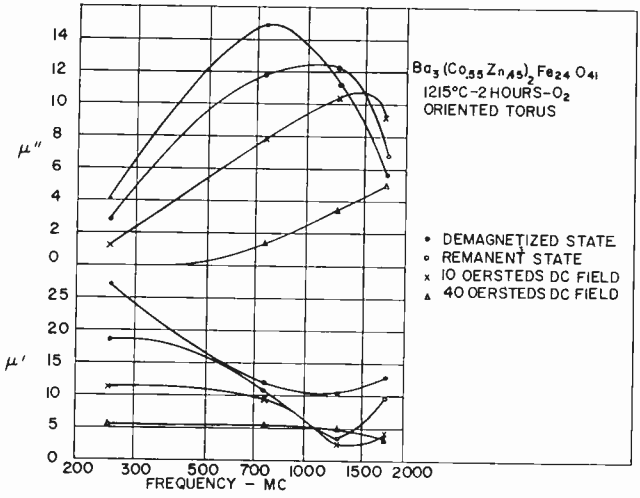


Fig. 4—μ' and μ'' as a function of frequency and magnetization state.

The compositions already discussed are Z-type materials and are the most suitable for tuning where the core is subjected to a varying d-c field. The Z-type compounds have a relatively poor temperature coefficient of permeability and may require temperature control in actual use. The temperature coefficient of permeability of the Z-type compounds is found to range from +0.1 to +0.5 per cent per degree centigrade.

Two Y-type compounds are shown in Table I as cores 16 and 17.

While the Y-type compounds are not as attractive as the Z-type compounds from a tuning viewpoint, they can be synthesized to have a good temperature coefficient of permeability (0.01 per cent per °C). In some applications, where a limited tuning range is acceptable, the Y-type compounds should be used.

The work to date has included only a small effort on the W-type compounds; however, preliminary indications are that these materials will have low values of permeability.

Conclusion

The oxidic compounds with a plane of easy magnetization possess improved properties for inductance coil cores at frequencies above 200 megacycles.

The Z-type compounds have been found to undergo a large change of permeability when the material is subjected to a biasing magnetic field. In the frequency range above 200 megacycles, this permeability change is greater than the change that can be obtained by using previously available materials if the condition of equal dissipation in the compared materials is imposed.

The large number of compositions which are possible, coupled with variation in processing parameters, give a wide range of selections with respect to tuning range and losses.

Acknowledgment

The authors acknowledge with thanks the helpful suggestions of E. Both, I. Bady, and W. Malinofsky of USASRDL.

TRANSFLUXOR FREQUENCY MEMORY*†

By

A. G. Samusenko

RCA Laboratories,
Princeton, N. J.

Summary—It is possible to change the frequency of an oscillator by changing one or more of the circuit parameters. This paper describes the use of a transfluxor as the frequency-control component of a magnetically coupled multivibrator. Several modes of operation are possible. One mode uses a base-band input to set the frequency of the multivibrator to a new value. Another mode uses a periodic input in which both the amplitude and the frequency of the input are instrumental in determining the new frequency of the multivibrator. Still another mode uses both the base-band and the periodic input to control the frequency of the multivibrator.

The analog storage characteristics of the transfluxor permit one to set the frequency of the multivibrator to any value within the frequency range. Improvement in the frequency-control characteristics of the multivibrator is realized by use of two transfluxors in the complementary mode of operation.

Introduction

In recent years cores made of square-loop magnetic materials have found wide application in various electronic circuits. A number of oscillators described in the literature make use of the properties of the square-loop magnetic cores to vary the frequency.[1-4] The frequency of the oscillators is varied either by changing the voltage across the oscillator or by varying the flux conditions in the core.[2,3] In order to maintain a given frequency of oscillation, the variable circuit parameter must be maintained at the proper level. The removal

* Manuscript received 28 August 1961.

† This material was presented at the AIEE Winter General Meeting, New York, N. Y., January 29-February 3, 1961.

[1] G. H. Royer, "A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage," *AIEE Trans., Part I, Communications and Electronics*, Vol. 74, p. 322, July, 1955.

[2] R. L. Van Allen, "A Variable Frequency Magnetic-Coupled Multivibrator," *AIEE Trans., Part I, Communications and Electronics*, Vol. 74, p. 356, July, 1955.

[3] A. J. Meyerhoff and R. M. Tillman, "A High-Speed Two-Winding Transistor-Core Oscillator," *I.R.E. Wescon Convention Record, Part 5, Instrumentation and Telemetry*, Vol. 1, p. 106, 1957.

[4] D. H. Schaefer, "Magnetic Core Event Counter for Earth Satellite Memory," *Elec. Eng.*, Vol. 77, p. 52, January, 1958.

of the input will cause the oscillator to assume its natural frequency. With the introduction of the transfluxor,[5] it has become possible to utilize partial switching in a ferrite core to store analog information nondestructively. This analog storage characteristic of the transfluxor has been utilized to build several variable-frequency oscillators, e.g., two-winding magnetically coupled multivibrators and sine-wave and blocking oscillators. A descriptive name—transfluxor frequency memory—has been adopted to denote this class of oscillators.

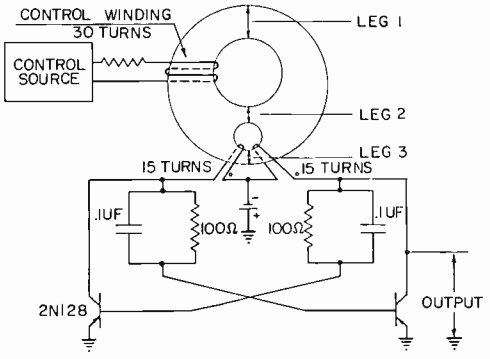


Fig. 1—Transfluxor frequency memory circuit.

The characteristics of a single circuit comprising a transfluxor and a two-winding magnetically coupled multivibrator (Figure 1) are discussed here. The nonlinear characteristics of the transfluxor and the transistors in the circuit make the experimental approach mandatory in the analysis of the circuit. The circuit is viewed as a four-terminal network, and the transfer characteristics between the input and the output terminals are studied.

Basic Circuit

A qualitative understanding of the circuit operation can be obtained from the approximate calculations of the frequency dependence of the multivibrator on the flux conditions around the small aperture of the transfluxor. The ON-OFF action of the multivibrator produces an alternating magnetomotive force (mmf) around the small aperture. For proper operation of the device, the magnitude of the alternating

[5] J. A. Rajchman and A. W. Lo, "The Transfluxor," *Proc. I.R.E.*, Vol. 44, p. 321, March, 1956.

mmf can assume any value between two limits. The lower limit is established when the alternating mmf cannot overcome the coercive magnetic intensity, H_c, of the ferrite. The upper limit is reached when the alternating mmf begins to switch flux around the large aperture.

The frequency of the multivibrator can be approximately determined by assuming that the counter electromotive force (emf) produced by the ON section of the multivibrator equals the power supply voltage, V, or

$$V = N\frac{d\phi}{dt}, \tag{1}$$

where N is the number of winding turns. Integrating over period T, the time the ON section of the multivibrator remains ON, and between two states of flux, $\pm\phi_r$,

$$\int_0^t V dt = \int_{-\phi_r}^{\phi_r} N d\phi. \tag{2}$$

Since V remains constant over the period T,

$$VT = 2N\phi_r, \tag{3}$$

where $2\phi_r$ is the flux switched by the ON section of the multivibrator. For the symmetrical output of the multivibrator, the frequency is

$$f = \frac{1}{2T}. \tag{4}$$

Substituting Equation (4) in (3),

$$f = \frac{V}{4N\phi_r}. \tag{5}$$

As shown in Equation (5), the frequency of the multivibrator can be varied either by changing the supply voltage, V, or by changing the flux, ϕ_r, switched by the multivibrator. In a sense, the multivibrator acts as a read-out circuit for the flux available for switching around the small aperture of the transfluxor. The frequency control is exercised by varying the flux available for switching around the small

aperture while V is kept constant. The flux around the small aperture of the transfluxor is a function of the flux condition around the large aperture. This flux condition can be changed by the application of current to the control winding. The two terminals of the control winding constitute the input to the circuit; the frequency of the multivibrator constitutes the output.

Direct-Current Transfer Characteristics

As a preliminary step, a large (blocking) pulse is applied to the control winding to orient the flux around the large aperture in one direction. The frequency of the multivibrator at this point of operation is given by

$$f = \frac{V}{4N\phi_b}, \tag{6}$$

where ϕ_b is the flux switched by the ON section of the multivibrator.

Due to the unidirectional flux orientation around the large aperture, the amount of flux switched by the multivibrator is low. In order to increase the flux available for switching, a direct current of opposite polarity with respect to the blocking pulse is applied to the control winding. The mmf produced by this direct current causes the flux around the innermost portion to reverse first. This flux reversal around the large aperture increases the amount of flux available for switching around the small aperture. If the increase in the available flux is denoted as $\Delta\phi_{dc}$, then Equation (6) becomes

$$f = \frac{V}{4N(\phi_b + \Delta\phi_{dc})}. \tag{7}$$

Upon increasing the direct current in the control winding, a point is reached when all the flux in leg 2 of the transfluxor is reversed. At this point the frequency of the multivibrator is a minimum. Further increase of direct-current input reverses the flux in leg 3 of the transfluxor, which, in turn, results in a decrease of available flux and, therefore, increases the frequency of the multivibrator.

In Figure 2 three different transfer characteristics for the transfluxor frequency memory are shown. The frequency of the multivibrator for the control characteristic shown by the curve was determined with direct-current input present in the control winding. The memory characteristic curve shows the frequency of the multivibrator after the direct-current input was removed. The frequency difference

between control and memory characteristics stems from slight "non-squareness" of the square-loop magnetic material of the transfluxor. By denoting the flux switched by the ON section of the multivibrator after direct current input was removed as $(\phi_b + \Delta\phi_m)$, one finds that the flux difference caused by removing the input is

$$\Delta\phi = \phi_b + \Delta\phi_{dc} - \phi_b - \Delta\phi_m = \Delta\phi_{dc} - \Delta\phi_m. \tag{8}$$

In order to obtain the setting characteristic curve of Figure 2, the multivibrator was disconnected from the power supply. The frequency

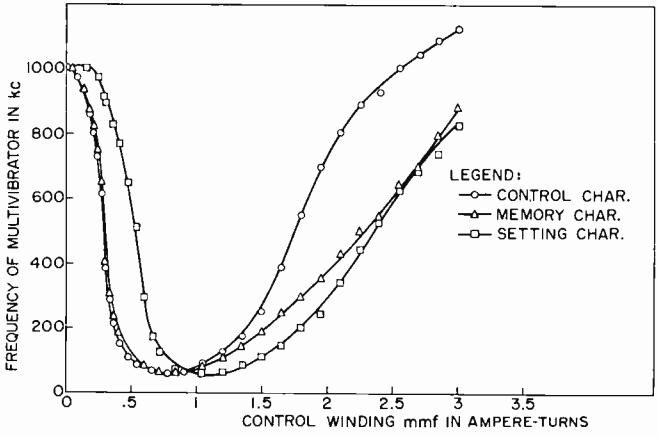


Fig. 2—D-C control, memory, and setting characteristics.

of the multivibrator was measured after a given direct-current input was applied to the control winding. The absence of the alternating mmf produced by the multivibrator during the setting of the transfluxor increases the reluctance around the large aperture of the transfluxor. For the control characteristics, the presence of the alternating mmf around the small aperture makes it possible for a given control-winding mmf to reverse more flux. This difference in required control-winding mmf between control and setting characteristics is due to the interaction that exists between control-winding mmf and alternating mmf produced by the multivibrator.

It should be observed in Figure 2 that the frequency control for the region of the control winding mmf between 0 and 0.7 ampere-turn is unidirectional. That is, if the input current is removed, the circuit "remembers" the frequency of oscillation, and if the amplitude is increased, the frequency can only be decreased. In order to increase the frequency of the multivibrator and thus be able to exercise bidirec-

tional control, the polarity of the input signal must be reversed; furthermore, the amplitude of the input signal must be increased to about the previously held value before the increase in the frequency will take place. (This effect is due to the square-loop hysterisis of the ferrite material of the transfluxor). Practical application of this form of the bidirectional frequency control imposes some difficulties.

Another mode of the bilateral frequency control can be achieved if, before the application of any new input, the circuit is reset by the blocking pulse. In a sense, this form of frequency control satisfies the basic characteristics of the term "memory". "Memory" implies the ability to "erase" and to "write in" information at will. In the transfluxor frequency memory, the blocking pulse acts as an "erase" signal. The input applied to the control winding "writes in" desired information. In order to write in a certain frequency with a small error, the amplitude of the base-band input must be precisely controlled (in order not to overshoot), or some kind of feedback mechanism is required. In the subsequent discussion, a novel means to write in a given frequency is discussed, wherein the amplitude and the frequency of the input is instrumental in determining the new frequency of the multivibrator.

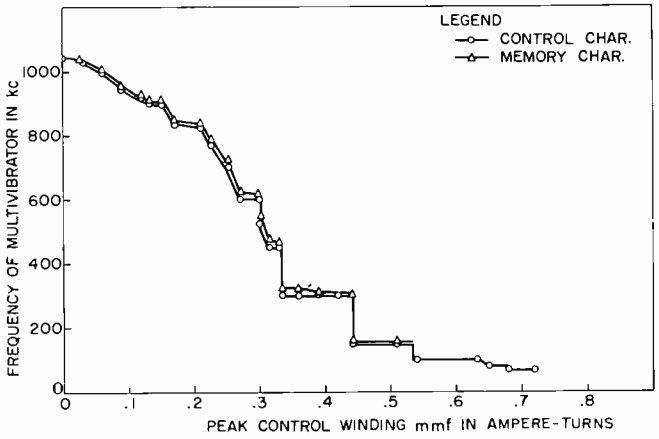


Fig. 3—Unipolar square-wave control and memory characteristics.

Unipolar Square-Wave Transfer Characteristic

If a unipolar square-wave input is applied to the control winding instead of direct current, completely different transfer characteristics are obtained. In Figure 3, control and memory characteristics for a 300-kilocycle unipolar square-wave are shown. The control winding mmf was limited in Figure 3 to 0.8 ampere-turn to obtain better

resolution of the characteristics. As the amplitude of the square wave is increased above 0.1 ampere-turn, the frequency of the multivibrator assumes different states of synchronization with respect to 300-kilocycle input frequency. The transition between states of synchronization is abrupt. For different frequencies of the unipolar square wave applied to the control winding, similar transfer characteristics are obtained. For any input frequency in the range of the frequency variation of the multivibrator, it is possible to synchronize the output frequency to the input frequency. If the input frequency falls outside the range of frequency variation of the multivibrator, the output frequency assumes either a harmonic or a subharmonic state of synchronization.

Memory characteristics follow the control characteristics closely. The frequency difference between control and memory characteristics during synchronization to the input frequency is a function of the amplitude of the unipolar square wave. At the beginning of synchronization the frequency difference is rather large, but as the amplitude is increased, the frequency difference diminishes. The minimum frequency difference occurs just before a jump to a subharmonic frequency takes place.

Preservation of constant multivibrator frequency over large variations of control-winding mmf in the synchronization regions suggests that the flux switched around the small aperture by the multivibrator remains constant. The mechanism of constant-flux preservation is of interest. It should be observed that the current in the control winding for the unipolar square-wave input is turned ON and OFF every cycle. Due to some nonsquareness in the characteristics of the magnetic materials of the transfluxor, the flux available for switching around the small aperture changes every cycle as shown in Equation (8).

$$\Delta\phi = \Delta\phi_{dc} - \Delta\phi_{m}.$$

In this case, however, $\Delta\phi$ is a function of input frequency, or $\Delta\phi_f$. If $\Delta\phi_f$ is inserted into Equation (7), the new relation becomes

$$f = \frac{V}{4N(\phi_b + \Delta\phi_{dc} \pm \Delta\phi_f)}, \tag{9}$$

where ϕ_b = flux switched by ON section of the multivibrator after a blocking pulse has been applied to the transfluxor,

$\Delta\phi_{dc}$ = additional flux switched by the ON section of multivibrator which is a function of the amplitude of the unipolar square wave,

$\Delta\phi_f$ = additional flux switched by ON section of the multivibrator caused by periodic input current.

The existence of a pulsating flux around the large aperture of the transfluxor causes the free-running multivibrator to be synchronized to the input frequency. The constant frequency is preserved by variation of $\Delta\phi_f$ as the amplitude of the input is increased. At the beginning of the synchronization, a large value of $\Delta\phi_f$ together with $(\phi_b + \Delta\phi_{dc})$ establishes the given frequency of oscillation. As $\Delta\phi_{dc}$ is increased, $\Delta\phi_f$ decreases to preserve the constant flux switched by multivibrator. Two experimental facts seem to support this contention:

1. The frequency difference between control and memory characteristics is large at the beginning of the synchronization;

2. The time coincidence of input and output waveforms changes as a function of input amplitude. At the beginning of the synchronization to the fundamental, the two waveforms are in phase. As the amplitude of the input is increased, the two waveforms move out of phase. A jump to a subharmonic occurs when two waveforms are more than 180° out of phase.

In comparison with the direct-current control characteristics, it is easier to write in a desired frequency with the unipolar square-wave input. Inasmuch as the frequency of the multivibrator remains constant in the region of synchronization, the adjustment of the unipolar square-wave amplitude does not have to be as precise as the magnitude of the direct current. An application of a short-duration burst of a given frequency and amplitude will lock the frequency of the multivibrator to the input frequency during the duration of the burst. The removal of the input signal will cause the output frequency to shift slightly.

Sine-Wave Transfer Characteristics

The sine-wave transfer characteristics do not differ substantially from the unipolar square-wave transfer characteristics. In Figure 4, control and memory characteristics for a 300-kilocycle sine-wave input are shown. Similar transfer characteristics were also obtained for sine-wave inputs of different frequencies. The regions of synchronization to harmonic, fundamental, and subharmonic frequencies are much more pronounced for the sine-wave input. This is believed to be due to an increase in the value of $\Delta\phi_f$ switched by the sine wave. Inasmuch as the sine wave has both positive and negative portions, some flux around the large aperture is reversed every cycle by the negative portion of the sine wave. This increases the magnitude of control-winding mmf required to obtain a given frequency for sine-wave input.

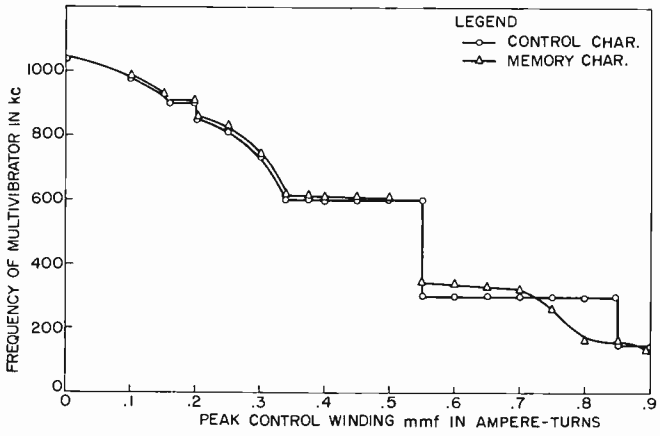


Fig. 4—Sine-wave control and memory characteristics.

The memory characteristics for the sine-wave input differ somewhat from the unipolar square-wave memory characteristics. Zero frequency difference between control and memory characteristics occurs in the middle of the synchronization range. This makes it possible to write in a given frequency without any error using sine-wave input; however, strict control of the input amplitude is required.

Multi-Input Control

If, instead of one input, two or more inputs are applied to separate windings wound around leg 1 of the transfluxor, the transfer characteristics will be modified. Figure 5 shows measured control charac-

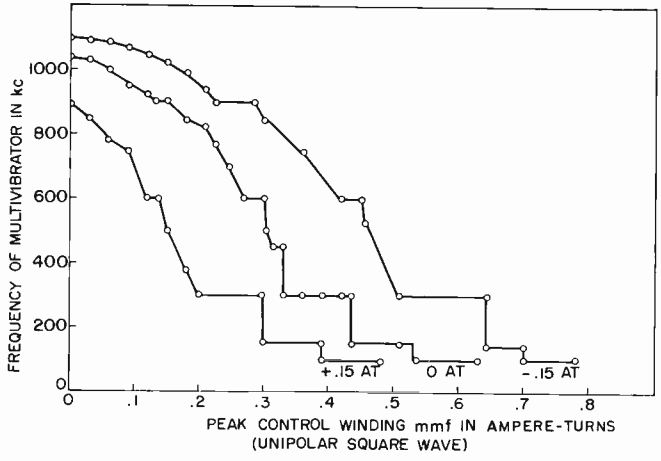


Fig. 5—Effect of bias mmf on the unipolar square-wave control characteristics.

teristics for a 300-kilocycle unipolar square-wave input and d-c bias mmf's of +0.15 ampere-turn, 0 ampere-turns, and —0.15 ampere-turn. The experimental data suggests that the mmf necessary to obtain a given frequency is constant. The required mmf equals the sum of the d-c bias mmf and the control-winding mmf. Similar transfer characteristics are obtained when a unipolar square wave is used as a bias mmf while the control function is established by direct current.

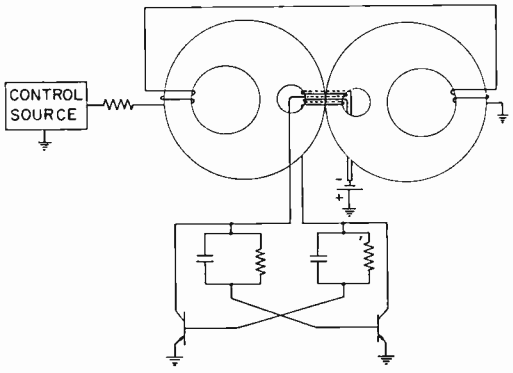


Fig. 6—Complementary operation of transfluxors.

There is also the possibility of obtaining a number of outputs for a single input to the transfluxor frequency memory. If a multiaperture transfluxor is used, with an identical multivibrator for each small aperture, and each multivibrator is energized with a different voltage, then for a periodic waveform input applied to the control winding, it is possible to obtain, simultaneously, harmonic, fundamental, and subharmonic frequency outputs for a given frequency input.

Complementary Operation of Transfluxors

A smaller frequency difference between control and memory characteristics for all transfer characteristics shown in Figure 2, 3, and 4 was observed when two transfluxors were used in the transfluxor frequency memory. The winding arrangement around two transfluxors for the complementary mode of operation is shown in Figure 6. After a blocking pulse has been applied to the control windings, the flux around the large apertures of two cores will be oriented in opposite directions with respect to each other. It is believed that due to the opposite orientation of flux in the two transfluxors a smaller flux change around the small aperture results as the input is removed.

For the single ON section of the multivibrator, the decrease in the available flux for switching around the small aperture in one core is countered by an increase in flux in another core. This complementary action reduces the frequency difference between control and memory characteristics.

Conclusions

An important property of the transfluxor frequency memory is its ability to adjust to and "remember" the input frequency. The flux switched by a multivibrator around the small aperture of a transfluxor assumes definite states which are determined by the frequency and the amplitude of input waveform. In contrast to voltage-controlled oscillators, the input signal can be removed without causing substantial frequency change. Unlike the closed-feedback-loop frequency-control systems, where the difference between the input frequencies and the oscillator frequency are used to change the frequency of the oscillator, the frequency adjustment in a transfluxor frequency memory is internal, and therefore requires no external circuitry. The versatility of the characteristics and the simplicity of the transfluxor frequency memory circuit offer new application possibilities.

Acknowledgment

The work on the transfluxor frequency memory was sponsored by RCA Laboratories under supervision of J. A. Rajchman and A. W. Lo. The author is indebted to members of the RCA Industrial Advance Development and Electronic Data Processing Division, Camden, New Jersey, and particularly to Dr. K. Li, for many interesting conversations.

HIGH-SPEED LOGIC CIRCUITS USING COMMON-BASE TRANSISTORS AND TUNNEL DIODES*

BY

J. AMODEI AND W. F. KOSONOCKY

RCA Laboratories,
Princeton, N. J.

Summary—*A complete set of logic circuits intended for operation at speeds of 100 megacycles or higher are described. Emphasis is placed on circuits that take maximum advantage of the logical features of both devices as a means of achieving simplicity even in the realization of more complex logical functions. The paper discusses the operation of the circuits, gives results of experimental tests and outlines analytical methods and models of evaluating their performance. The logic circuits included are basic gates, inverters, complementary gates, registers, and adders. The results of a tolerance analysis of the basic gates and the full binary adder are presented to demonstrate that these circuits are capable of operating under realistic conditions. Finally, the paper gives an evaluation of the circuits performance aspects as compared to conventional logic circuitry.*

INTRODUCTION

THE combination of transistors and tunnel diodes offers the possibility of utilizing the features of each device to greatest advantage in the realization of high-speed computer circuitry. When complementing each other, the devices are capable of achieving functions and performance features which could not be realized by either device alone without extreme complexity or prohibitive tolerance requirements.

In high-speed switching applications, the utilization of the transistor for current gain inescapably results in slower circuit performance regardless of the configuration in which the transistor is operated. The common-base configuration, which has no current-gain capabilities and, as a result, is the fastest choice available, is not practical since the lack of current gain is incompatible with the fan-out requirements of logic circuits. The choice of tunnel diodes for providing the current gain eliminates this difficulty and, further, results in circuit current-transfer parameters that show very small variation from one device to another.

In the circuits described in this paper, the transistor operates

* Manuscript received 11 July 1961.

between two definite OFF and ON states, thus standardizing the output-voltage levels and avoiding high dissipation and stabilization problems inherent in the operation of transistors in the active region. The flexible current threshold of the tunnel diode combined with the current-routing properties of the transistor, as a three-terminal device, is shown to permit the realization of very complex logical functions with exceptional simplicity. The decoder and adder circuits discussed

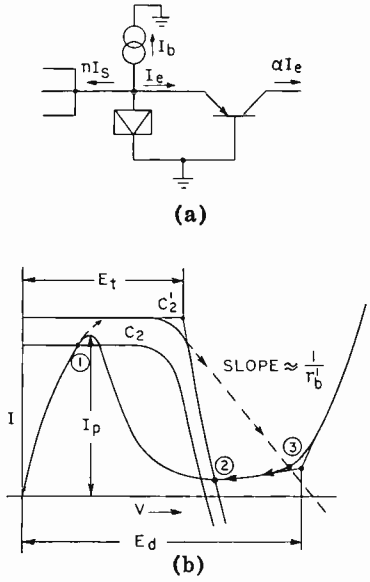


Fig. 1—Basic circuit.

in the text illustrate this. The voltage gain provided by the transistor permits the use of coupling resistances that impose negligible loading on the inputs of the driven stages, thus permitting the achievement of close to maximum theoretical gain from the tunnel diode, and complete isolation between stages.

Basic Circuit

Figure 1a shows the basic connection of transistor and tunnel diode with some of the parameters which are used. The static characteristic curves governing the operation of the circuit are shown in Figure 1b with the two positions of the load line C_2 and C_2' corresponding to the

OFF and the ON states, respectively.[1] C_1 represents the static volt–ampere curve for the tunnel diode which, together with its shunt capacitance, completely characterizes the device for this application. C_2 is the static emitter-input characteristic for the transistor inverted about the current axis and shifted vertically by the net current entering the node from the bias supply and input terminals. An approximate equivalent circuit to characterize the transistor input impedance under dynamic conditions is given in Figure 2. An ideal diode, D, is shown in the circuit, to account for the emitter-to-base conduction voltage.

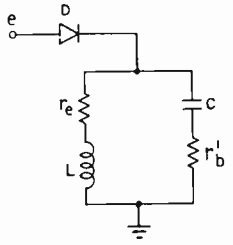


Fig. 2—Approximate equivalent for common-base emitter input impedance. (r_e = emitter resistance, r_b' = base resistance, C = diffusion capacitance, and $L = r_e r_b' C$.)

This circuit was derived from the small-signal equivalent circuit[2] by assuming that

$$1 - \alpha \ll \frac{r_e}{r_b'},$$

and proved very useful in predicting the switching performance of the combination. It also gives a good qualitative picture of the parameters that influence the speed and stability of the circuit. Using Figure 2, the dynamic path followed by the circuit during switching from point (1) to point (2) in Figure 1b can be traced as shown by the arrows in Figure 1b. It is assumed that the tunnel diode is capable of switching well within the $r_b'C$ time constant of the transistor, which is usually the case. It can be shown within the framework of the above assumptions that the performance to be expected from the transistor driven by a tunnel diode is the same as that to be obtained

[1] For a discussion on the use of nonlinear load lines for determining the logical properties of circuits, the reader is referred to M. H. Lewin, *Ph.D. Thesis*, Princeton University, University Microfilms, Inc., Ann Arbor, Michigan.

[2] Shea, *Transistor Circuit Engineering*, p. 40, John Wiley and Sons, New York, N. Y., 1957.

when the transistor is driven by a current step provided that the following condition is satisfied:

$$E_d \geqq E_t + (I_t - I_v) r_b'$$

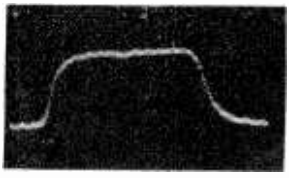
where

E_d = conduction knee voltage of the tunnel diode,

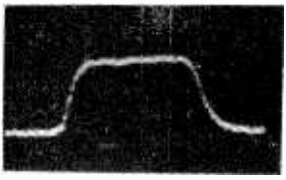
E_t = conduction voltage of the transistor,

I_t = total current entering the input node during switching,

I_v = tunnel-diode valley current.



(a) OUTPUT FOR CURRENT STEP



(b) OUTPUT WHEN DRIVEN BY TUNNEL DIODE

Fig. 3—Waveforms for current step and tunnel diode driven 2N769.

In other words, the intersection of the switching load line and the tunnel-diode characteristic, point (2) in Figure 1b, must occur within the valley of the tunnel diode. This is significant since a current source gives optimum switching time in this configuration and the above inequality is easily satisfied when operating at reasonably low current levels. Figures 3a and b show a comparison of switching times of a 2N769 transistor gate when driven by a tunnel diode and by a current step.

The basic AND and OR gates are obtained from the circuit shown in Figure 1a by properly choosing the bias current according to the function to be performed and the number of inputs to be handled. The

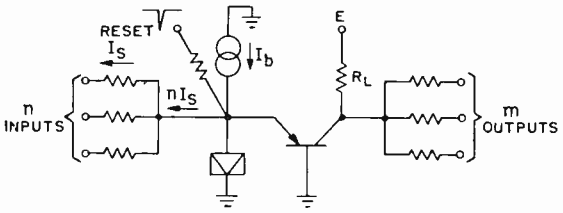


Fig. 4—Basic resistive gate.

general configuration is shown in Figure 4. The method of biasing the resistive OR gate is demonstrated in Figure 5. Here it is assumed that there are n inputs, and I_S flows out of the node through each when the gate is not energized. The effect of a pulse at any of the inputs is to increase the current input to the node by I_S, thus raising the load line to position C'; and the circuit proceeds to the high state (point (2)) in which all the input current, except for the small tunnel-diode valley current, flows into the emitter of the transistor. The output current when in this state will be α times the emitter current shown in Figure 5.

As illustrated by the above example, and following from the choice of the tunnel diode as the logic threshold element, current is the variable that carries the information in such a system. Therefore, by implication, the current outputs of each stage must be standardized and prevented from being a function of the number of inputs present, as would be the case in the above circuit. In order to achieve this, the gates should be designed to reach saturation at the smallest tolerable value of output current. The storage time resulting from saturated operation of these circuits does not have the usual serious consequences for two reasons:

(a) The magnitude of the storage time resulting under normal operating conditions is quite small, of the order of 4 to 5 nanoseconds when using 2N769 transistors. (This assumes a rather pessimistic

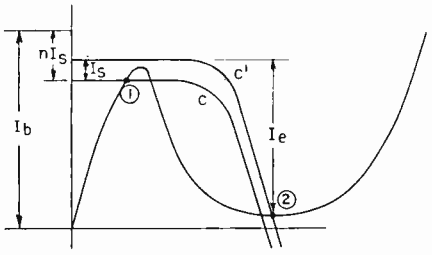


Fig. 5—OR-gate biasing.

case which would require a five-input OR gate with all inputs present at once.)

(b) In all the circuits except inverter and the complementary output gates, the storage time does not slow down the transfer of information. Indeed, under normal circumstances it is significant only if it becomes longer than the prescribed resetting interval, which is of the order of 5 nanoseconds in the circuits proposed. In the inverter and complementary gates described later, the degree of saturation does not depend on the number of inputs present, and therefore can be accurately kept to low values by proper biasing.

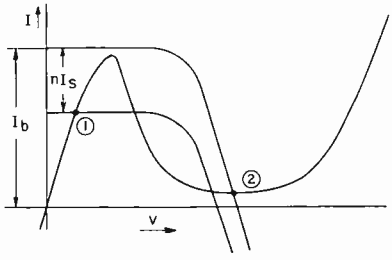


Fig. 6—AND-gate biasing.

The biasing method for AND gates is illustrated in Figure 6. Note that I_b is not dependent on the number of inputs to the gate, although n does determine the position of the quiescent point (1). Although an AND gate is illustrated, it should be apparent at this point that a general threshold device[3] where k out of n inputs are sufficient to activate the gate can be realized without difficulty.

Description of digital circuits would not be complete without reporting the tolerance requirements on the components. The uncertainty of the logical threshold of the basic gates depends on the variation of the tunnel-diode peak current as well as that of the input and bias currents. The ON and OFF output levels for the gate correspond respectively to the transistor in saturation and in cutoff. For an OR gate, the variation of the collector resistor and voltage supply of the driving stages must be included in the uncertainty of the input signal. In the case of AND gates, the maximum number of inputs is determined mainly by ratio of the variations of the collector saturation voltage, ΔV_s, to the collector voltage swing, V_c. The results of the

[3] R. O. Winder, "Single Stage Threshold Logic," *AIEE Conference Paper* 60-1261, October 1960.

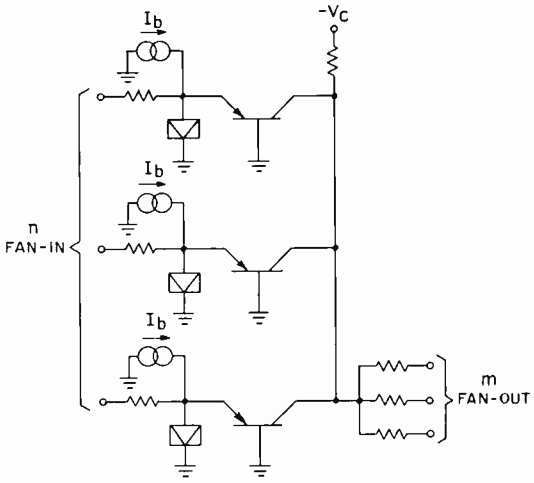


Fig. 7—Transistor-coupled OR gate.

worst-case analysis of the basic gates which are given include independent variations of all voltage supplies of individual circuits by ±2 per cent and variation of peak current of the tunnel diodes by ±5 per cent with a peak-to-valley current ratio of at least 10:1. Under these conditions, the OR gate of Figure 5, for fan-in = 2 and fan-out = 2, requires tolerance on the individual resistors of no more than ±3 per cent. The two alternate circuits for the OR gate that are shown in Figures 7 and 8 will relax the above requirements to ±6 per cent for the same fan-out but independently of the fan-in, which could be large. The tolerance requirement for the resistors of the AND gate for a fan-out = 2 is ±5 per cent assuming $\Delta V_s/V_c = .05$, and fan-in = 4.

The rise time of the output current is mainly dependent on the current step response of the transistor and the output capacitance between collector and ground; with voltage swings of the order of

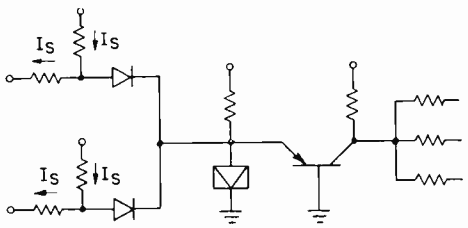


Fig. 8—Diode-coupled OR gate.

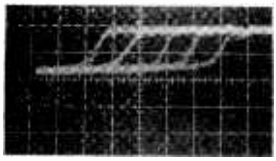


Fig. 9—Output of four cascaded OR gates.

2.5 volts, rise times of less than 2 nanoseconds were achieved. The total delay per stage is the sum of the tunnel diode and the transistor turn-on delay and the output rise time. Figure 9 shows the outputs of 5 consecutive OR gates with fan-in of two and fan-out of two, illustrating the total delay per stage.

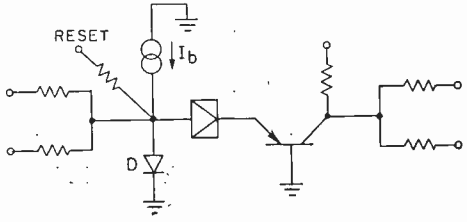


Fig. 10—Inverter.

Inverter and Complementary Gates

The circuit utilized to provide inversion is shown in Figure 10, and its operation will be explained by referring to the characteristic curves of Figure 11. The characteristic curve that describes the static behavior of the tunnel diode in series with the transistor is obtained by adding the two independent *I–V* curves along the voltage axis.[1] A silicon diode, *D*, serves as load line and, as shown in Figure 11, is nonconducting when the circuit is in state (1). When the current entering the input node is increased beyond the tunnel-diode peak

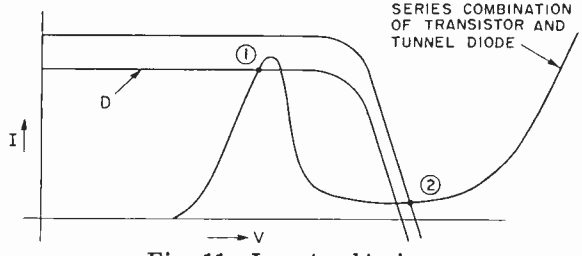


Fig. 11—Inverter biasing.

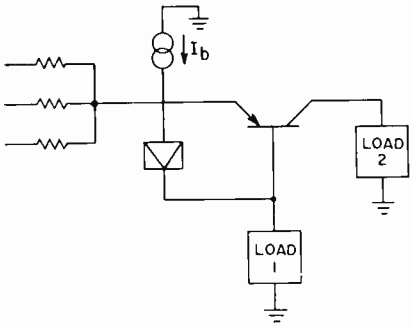


Fig. 12—Generalized complementary gate.

current, the circuit switches to its second stable point (2), and all but tunnel-diode valley current is steered into the diode. The circuit will remain at point (2) until reset by a negative pulse applied as shown in Figure 10 or by a positive pulse applied at the emitter of the transistor.

Transistor and tunnel-diode action can be combined very efficiently to provide the simple but highly useful circuit shown in its more basic form in Figure 12. In such a circuit the state of the tunnel diode determines which of the two loads will be energized. When the tunnel diode is in the low-voltage state, the transistor is not conducting and the bias current flows through the tunnel diode into "load 1." When the tunnel diode is in the high-voltage state, the transistor conducts and virtually all but valley current flows into the emitter and out the collector into "load 2." Under the conditions shown in Figure 12, there is a limitation to the size of the impedance that can be used for "load 1" without hampering the operation of the circuit. Figure 13 shows an alternate form of the circuit that provides two complementary outputs of the same impedance level.

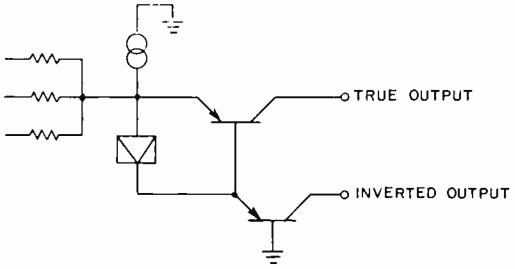


Fig. 13—Complementary gate.

The multiple threshold decoder shown in Figure 14 can be realized thanks to the unique current routing properties of the circuit just described. For proper operation of the decoder, the peak current of each successive tunnel diode must be higher than that of the previous one by the same amount, and the valley current of any of the tunnel diodes should not exceed the smallest peak current of the diodes in the string. If these conditions are met, each time that a diode switches to the high-voltage state the current will be routed into its corresponding output, and the tunnel diode immediately below will switch back to the low-voltage state. This insures that only one input will be energized at a time.

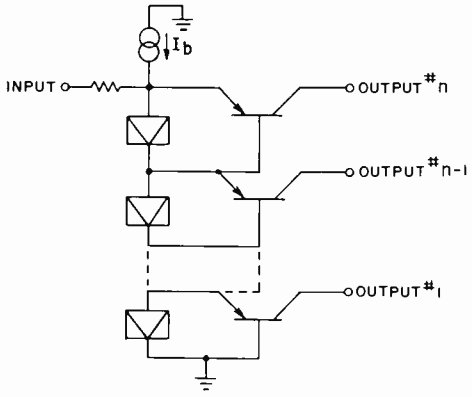


Fig. 14—Decoder.

REGISTERS

In present-day computers, most tasks requiring temporary storage are fulfilled by flip-flops in which memory is achieved by the use of regenerative feedback. The hybrid basic and complementary gates are natural registers that utilize the inherent memory of the tunnel diode, and, when operated as registers, they retain the ability to do their own gating thus permitting a substantial savings in circuitry. The hysteresis properties of the basic gate are illustrated by the static characteristic shown in Figure 15, and the emitter current versus input current plot derived from it is shown in Figure 16. This plot indicates that once the input current has been increased beyond I_{max}, the tunnel diode will not return to the low-voltage state, and thereby turn off the transistor, until the input current is decreased below I_{min}.

These plots assume that the transistor is fully characterized by its

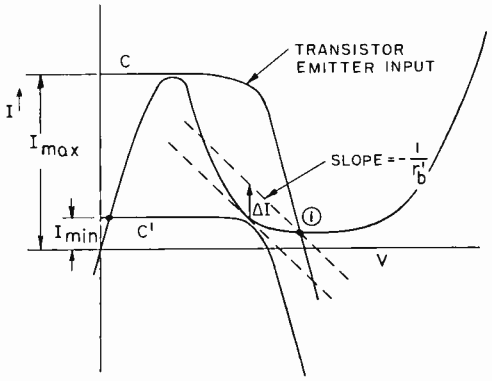


Fig. 15—Register biasing and dynamic performance.

static I–V curve and, therefore, only hold for very slowly varying input currents. If an abrupt decrease in input current should occur when the circuit is at point (1), Figure 2 shows that the transistor would be better characterized by a resistance of magnitude r_b' whose slope is shown by the dotted line in Figure 15. Under these conditions an abrupt decrease of ΔI in input current would be sufficient to reset the register. This difference between the static and dynamic reset current requirement should be taken into account to avoid designs that would result in a possibility for spurious reset. Although this difference did not present a practical limitation to the use of hybrid gates as registers, it could enter the picture if registers utilizing relatively slow transistors are to be driven by "set" pulses that have very steep fall times. The basic and the complementary gates used as registers can be set and reset in a total intervals of 5 nanoseconds.

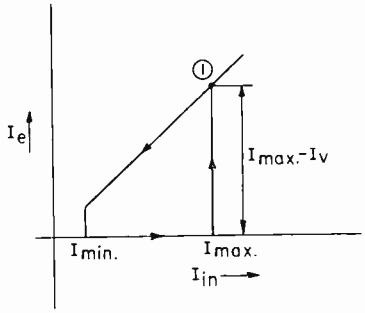


Fig. 16—Static hysteresis plot for common-base register.

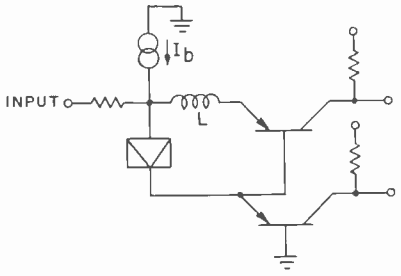


Fig. 17—Binary counter.

Counter Circuits

A close relative of the register, the counter, must be capable of being set and reset by the same polarity pulse and must produce an output that can drive a similar stage. One candidate for this application is the circuit shown in Figure 17. This circuit is essentially a complementary gate with an inductance added in series with the emitter of the first transistor as a means of achieving counter action. The circuit can be preset by a negative pulse so as to start with the tunnel diode in the low-voltage state and the bias current flowing into the bottom transistor. The first of a train of positive pulses slightly longer than the inductive time constant will cause the tunnel diode to switch to the high-voltage state diverting the current from the bottom to the top transistor. The next pulse will find the tunnel diode in the valley region, and during its duration it will increase the current flowing through the inductance; when the pulse is terminated, the current through the inductance must decrease rapidly to its original value and in so doing will generate a back voltage that triggers the tunnel diode back to its low-voltage state.*

Some preliminary results obtained with this circuit can be seen in Figure 18 which shows its performance when being set and reset

Fig. 18—Counter output during set and reset.

* A similar principle was first utilized by A. I. Pressman and H. Ditkovsky to design a hybrid gate with self-resetting properties.

by two closely spaced positive pulses. In accordance with these results the circuit should be able to operate at pulse repetition rates as high as 200 megacycles.

Adder Circuits

The flexibility of hybrid circuits permits binary addition to be performed by several alternate configurations with an improvement in speed and savings in components over conventional circuitry. The circuit shown in Figure 19 is a half adder that utilizes the tunnel diode threshold to detect the presence of two inputs and originate a

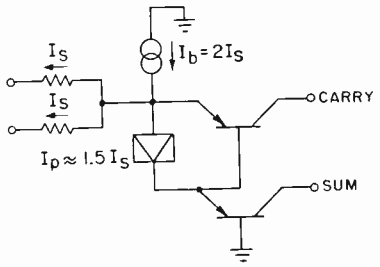


Fig. 19—Half adder.

carry output. The bias arrangement shown in the figure assumes that the circuit is being driven by resistive hybrid gates so that a positive bias current is provided which offsets the quiescent input current flow to provide zero net current when both inputs are zero. When either input becomes one, i.e., I_S drops to zero, a net current I_S flows into the node through the tunnel diode and out the collector of the sum transistor. When both inputs are high, the current into the node is $2I_S$, the tunnel diode peak is exceeded, and it switches to the high state, routing all but valley current into the carry transistor. Thus, the requirement of half binary addition is fulfilled with sufficient gain at the carry terminal to drive another stage. It can also be seen from the method of biasing used that these circuits return to their quiescent state when the inputs are removed, without the necessity of being reset.

With a very slight increase in the complexity of the circuit used for performing half addition, the much more useful function of full binary addition can be realized. The circuit of Figure 20 makes use of a second current threshold, the saturation current of the carry transistor, to detect the presence of a "one" in a third input and re-energize the sum output. The operation is similar to that of the

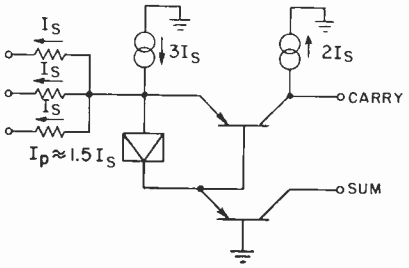


Fig. 20—Full adder.

half adder up to the point where two inputs are high and $2I_S$ flows into the carry transistor. When a third input goes high, requiring a carry and a sum output of "one," the collector current of the carry transistor, which was set to saturate at a value somewhat lower than $2I_S$, can increase no further. The third unit of input current is thereby routed through the base of the carry transistor and the tunnel diode and into the sum transistor, giving a sum of one as well as a carry. The storage time resulting from the saturated operation of the carry transistor does not slow down the carry ripple through time and therefore has little significance. The speed capabilities of the circuit are illustrated in the photographs of Figure 21 which show a total carry delay time of the order of 3 nanoseconds.

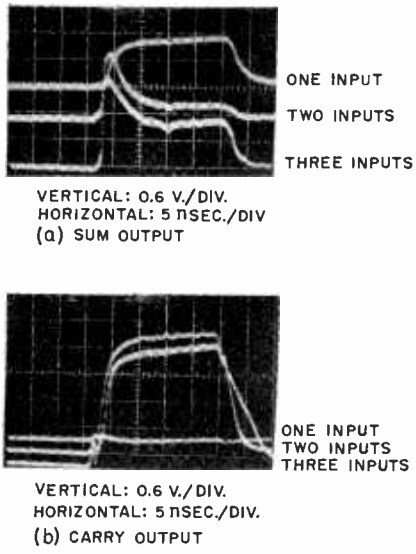


Fig. 21—Full-adder sum and carry outputs for three input combinations.

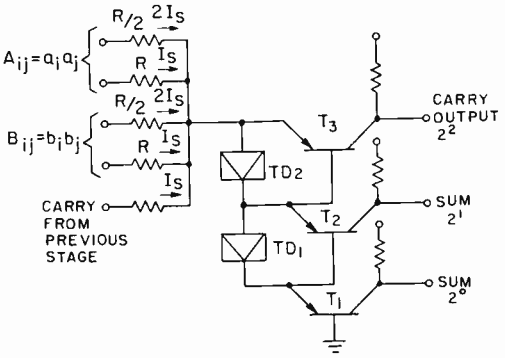


Fig. 22—Seven-bit full adder and decoder.

The principle of operation of the full binary adder can be extended to a multiple input adder capable of giving full sum and carry terms for two binary numbers of two or more bits and a carry. This can be achieved by weighting the inputs and using a plurality of stages which provide outputs in powers of two. Figure 22 shows how this can be realized for the case of two 2-bit numbers which could be two of n corresponding bits of numbers A and B. The peak current of TD_1 is chosen greater than one I_S unit and less than two. The saturation current of T_2 again is set at two I_S units. The peak current of TD_2 is set greater than $3I_S$ but less than four and, finally, the saturation current of T_3 is set to be equal to $4I_S$. The inputs are weighted in such a way that voltages representing bits of higher order produce twice the current as those of lower order. The circuit offers considerable advantage in speed and savings in components at the expense of tighter tolerance requirements than the three-input adder.

One of the circuits that was utilized for providing reset pulses is shown in Figure 23. The circuit consists essentially of a tunnel-diode

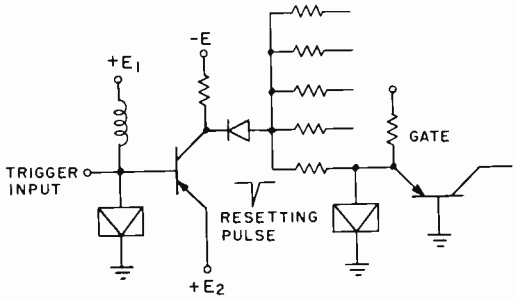


Fig. 23—Reset pulse generator.

blocked oscillator driving a transistor out of saturation for a short period of time, thus producing a negative spike of current which is used to reset five hybrid gates. Depending on the system organization, the circuit may be used to amplify a central clock pulse or to generate a locally controlled reset signal. The pulse widths obtained from the above circuit were of the order of 5 nanoseconds and amplitudes were sufficient to reset five hybrid gates in parallel.

Conclusions

It is felt that the significant improvements in speed and simplicity which are achieved by using the basic common-base hybrid gates described instead of conventional circuitry are well worth some of the added complexity introduced by their inherent memory. Furthermore, the simplicity and reliability with which the hybrid circuits could perform the more complex operations of registering, counting, and binary addition with no additional burden on peripheral circuitry cannot be matched by any other device or combination known at present.

AN ELECTRICAL ANALOG FOR ELECTROLUMINESCENT LAYERS*

BY

S. M. THOMSEN

RCA Laboratories,
Princeton, N. J.

Summary—An analog for EL (electroluminescent) layers can be devised using only capacitors and neon glow lamps. This suggests that a "breakdown" current, limited by capacitive impedance, flows in each light-emitting glow spot. This current is presumably delivered within the EL phosphor particle from internal electrodes formed during the phosphor synthesis. The deterioration of brightness during the operating life of an EL layer can be attributed to the electrical erosion of the internal electrodes.

INTRODUCTION

A NUMBER of equivalent circuits have been proposed for the correlation of the electrical properties of EL (electroluminescent) layers. One such circuit[1] employs only linear components (capacitors and resistors). Another[2] includes nonlinear components (diodes). However, the usefulness of an equivalent circuit of this type for EL layers is rather questionable.

Nonlinearity is not the only impediment to the derivation of a simple, conventional equivalent circuit. Complexity, or heterogeneity, of the EL layer is another impediment. Suppose that for a particular, sufficiently small area of EL layer a simple equivalent circuit could be derived having fixed component values. Suppose further, that for all other such small areas of the layer, similar equivalent circuits could be derived, identical in configuration, but with a wide range of component values. It does not then follow (indeed, it is unlikely) that any simple equivalent circuit would fit the large composite aggregation of these small areas, which constitutes an EL layer of moderate area.

The search for an equivalent circuit is facilitated by dividing the properties of an EL layer into two categories—intrinsic and extrinsic. The intrinsic properties (for example, the color of the emission), like

* Manuscript received 28 August 1961.

[1] A. N. Ince and C. W. Oatley, "The Electrical Properties of Electroluminescent Phosphors," *Philosophical Magazine*, Vol. 46, p. 1081, 1955.

[2] P. Zalm, G. Diemer, and H. A. Klasens, "Electroluminescent ZnS Phosphors," *Philips Research Reports*, Vol. 9, p. 81, April, 1954.

their counterparts in CL (cathodoluminescent) phosphors, are determined principally by the nature of the emission centers in the phosphor. The extrinsic properties (for example, the amount of power drawn), properties which have no counterparts in the CL phosphors, are principally determined by the EL layer structure and its inherent circuitry. The intrinsic properties result from the way the emission center converts electric energy into light; the extrinsic properties result from the way energy is taken from the electric field, processed, and delivered to the emission centers. This paper does not deal with intrinsic properties; it is assumed that the light emitted is proportional to the current that actually flows in the emitting phase (phosphor).

With due recognition of the complex or composite character of an EL layer, a useful *composite* analog circuit can be devised to simulate the extrinsic EL properties. It consists of an aggregation of *simple* analog circuits, each of which represents a functional EL unit. The term "analog" is used instead of "equivalent" because these circuits are somewhat unorthodox—they employ neon glow lamps. The more efficient an EL layer, the better the agreement of EL properties with those of the analogs.

Analog Circuits

The emission from an electroluminescing phosphor does not come uniformly from the whole phosphor particle, but only from small glow spots. Each spot is reported[3] to glow only once per cycle, while the composite EL layer glows twice per cycle. Therefore, presumably about half the spots glow on one half-cycle, and the remaining spots glow in the second half-cycle. These glow spots are more or less randomly distributed, possibly on, but probably within the phosphor particles. The phosphor particles, in turn, are more or less randomly distributed in an ambient dielectric (ceramic or plastic), between two electrodes.

The EL layer may be considered to consist of many small elements, situated side by side, all connected in parallel electrically, and each functioning substantially independently. The simplest such EL element is presumably one containing just two glow spots, the two glowing alternately, one each half cycle. The term "EL unit" will be applied to this structure of two glow spots and the associated capacitances.

The EL unit, with its two spots glowing alternately is suggestive of an a-c operated neon lamp. The possibility exists that the glow

[3] J. F. Waymouth and F. Bitter, "Experiments on Electroluminescence," *Phys. Rev.*, Vol. 95, p. 941, August 15, 1954.

spot undergoes impedance-limited breakdown as does the neon gas in a normally operated glow lamp. This suspicion is supported by a report[3] that the glow spots show a voltage threshold, and show a brightness proportional to the voltage increment above the threshold. Contrariwise, it has been reported[4] that a glow spot shows an emission-versus-voltage curve like that of the EL layer itself. Furthermore, the current–voltage characteristics of breakdown current in CdS single crystals are reported[5] to be similar to those of gas discharges. The term "threshold voltage" is used here for the lowest voltage that will

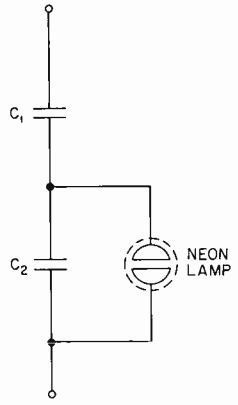


Fig. 1—Simple analog circuit of single "EL Unit."

maintain the breakdown current. The small difference, for neon lamps, between the firing voltage and the voltage as just defined, is ignored in this discussion. This difference does not generally exist for solid-state breakdowns, such as in Zener diodes.

A single EL unit will be represented by the *simple* analog circuit of Figure 1. This circuit consists of a voltage divider, C_1 and C_2, with a neon glow lamp connected across C_2. The EL glow spot pair is represented by the combination C_2 and the lamp (pure capacitance until the lamp ignites). C_1 represents dielectric material, both zinc sulfide and the ambient ceramic or plastic outside the glow spots. C_1 is both the driving capacitor and the limiting impedance. Substitution for the glow lamp of a pair of back-to-back Zener diodes gives an

[4] P. Zalm, G. Diemer, and H. A. Klasens, "Some Aspects of the Voltage and Frequency Dependence of Electroluminescent Zinc Sulphide," *Philips Research Reports*, Vol. 10, p. 205, June, 1955.

[5] G. Diemer, "Electric Breakdown and Light Emission in CdS Single Crystals," *Philips Research Reports*, Vol. 9, p. 109, April, 1954.

electrically equivalent circuit, but without the useful light-emitting feature.

The simple analog responds to d-c applied voltage, but only with a transient flash. To "sustain" the glow requires cyclical variation in the applied potential difference, and the glow then consists of two flashes per cycle, one from each electrode. For a particular lamp and specified values of C_1 and C_2, a certain minimum swing in potential difference is required. This will be termed the "peak-to-peak threshold voltage"; it is normally independent of the wave form of the applied voltage, because the maximum capacitor charge is determined by the peak voltage.

In the special case of sinusoidal a-c, it may be more useful to use the r-m-s threshold voltage. In either case, the threshold voltage is higher for the analog circuit as a whole than it is for the lamp itself, because the voltage applied to the analog divides between C_1 and C_2.

An EL layer is then represented by a *composite* analog, consisting of many parallel-connected simple analogs. The threshold voltages of these simple analogs extend over a wide range, having an unspecified distribution which might be either random or Gaussian.

Particle Structure

Before proceeding to compare the properties of EL layers with those of the analogs, the "internal electrode" hypothesis[6] of EL particle structure will be reviewed; it is consistent with the analogs, and is useful in correlating various EL properties.

By this hypothesis, zinc-sulfide EL phosphors contain a second material phase, composed of conducting copper sulfide. This conducting phase distinguishes EL phosphors from CL phosphors and is vital to field excitation (electroluminescence); the distributed conducting phase constitutes electrodes which supply current carriers, and produce high local fields.

In the preparation of EL phosphors, the product that has been fired, but not yet washed in a potassium-cyanide solution, is a rather dark gray. At this stage the product may have 10^4 to 10^6 times the conductivity that it has after the cyanide wash. The surfaces of the particles are covered with a thin layer of copper sulfide. Inclusions of copper sulfide within the particles must also be expected, perhaps in the form of thin sheets on the crystal interfaces of the polycrystalline particles.

---

[6] W. Lehmann, "Contact Electroluminescence," *Jour. Electrochem. Soc.*, Vol. 104, p. 45, January, 1957.

The cyanide wash, in addition to greatly reducing the conductivity, removes nearly all grayness. The finished, washed phosphor is brighter in emission and far more efficient than the dark gray product. Although the grayness disappears and the color lightens almost immediately as a cold 10 per cent cyanide solution wets the dark product, a more thorough treatment is required to develop the highest efficiency. There is no evidence that the cyanide treatment can be overdone; complete removal of the surface conductor appears to give the highest efficiency. The cyanide wash leaves only that small fraction of the original copper sulfide that is within the phosphor particle, to which the cyanide solution has no access. Any copper sulfide left on particle surfaces is probably harmful: it may shield the useful internal structures from the field; it may cause parasitic discharges through the weaker (lower breakdown strength) ambient dielectric; it may contribute to wasteful leakage conductance from electrode to electrode; and it may absorb some of the light emitted from the interior of the particle.

Thus, an efficient EL particle is free of conductor material on its surface, but has an internal electrode system to pick up energy from the electric field and deliver this energy to high-field regions within the particle. Such a particle can be considered a "device," in contrast with a particle of CL phosphor, which is merely a piece of homogeneous material.

The special particle structure required for high EL efficiency has not been achieved for oxygen-dominated phosphor host materials. What is presumably required for each such host is a conductor material that can, without poisoning luminescence, dissolve in the host during firing and segregate within the particle to form appropriate structures during cooling, and that can be removed from the surfaces of the fired particles.

D-C Transient Effects

For the simple analog, as a sufficiently high d-c voltage is applied, a flash appears at the first electrode of the lamp. After the voltage source has been disconnected, a flash at the second electrode can be obtained by shorting the circuit. Reversing the voltage also produces a flash at the second electrode.

For the composite analog, only a fraction of the lamps will flash; this fraction increases as the voltage is raised.

For an EL layer, a flash appears, as with the analogs, as a sufficiently high d-c voltage is applied. A second flash appears as the

voltage source is disconnected. Shorting is not required to produce the second flash, because the internal conductance of the EL layer is high enough to produce a rapid discharge.

Rectified A-C

For either analog, the interposing of a rectifier between the a-c supply and the analog reduces the light emission to a single flash. The capacitor system becomes charged to the peak voltage within a few cycles, and holds this charge in the absence of a discharge circuit. If a suitable resistor is connected across the analog, a sustained glow results. A suitable resistor is one which reduces the time constant for the circuit to the time of a half cycle or less.

For EL layers, using the same rectifier arrangement, the inherent conductance will usually provide a time constant low enough (e.g., 0.01 second) so the layer will glow at low frequencies. At high frequencies (e.g., 10 kilocycles) an efficient EL layer will not glow unless bridged by a suitable resistor. The more efficient the EL layer, the higher the time constant, and therefore the lower the frequency must be to produce a glow. This provides a simple and quick method of estimating the efficiency of an EL layer.

Brightness versus Voltage

For a neon lamp, the light output is proportional to the charge passing through the lamp. Consequently, the average emission intensity is proportional to the average current through the lamp.

For the simple analog the charge per unit time, or average current, passing through the lamp is proportional to the difference between the peak applied voltage and the circuit threshold voltage. The average emission intensity therefore increases linearly with peak applied voltage above the threshold.

For the composite analog, all lamps will glow at a sufficiently high applied voltage. In this case, the increment of light output, ΔL, will be, as with the simple analog, proportional to the increment of the applied voltage, ΔV; the value of n in the relation, $\Delta L = \Delta V^n$, will be unity. Gradual reduction of the voltage applied to the composite analog will cause a progressive extinction of the lamps. With each succeeding decrement of voltage, the percentage of lamps extinguished increases, and the value of n increases, ultimately approaching infinity.

For EL layers, the brightness, B, shows the same trend as does the average light emission of the composite analog; the value of n in the relation, $\Delta B = \Delta V^n$, rises as the applied voltage is reduced. How-

ever, the range of n values is more limited. Destructive breakdown intervenes at high voltages before all potential EL units achieve emission, and therefore n values below about 2 are not generally observed. At the other extreme, as the applied voltage is reduced, n values above about 8 are not generally reached before the brightness becomes too low to be measured conveniently.

Brightness versus Frequency

For both the simple analog and the composite analog, the light output per cycle remains constant as frequency is changed; the average light output is therefore proportional to frequency.

For EL layers, the light output per cycle is nearly independent of frequency below about 1 kilocycle. As the frequency is raised above 1 kilocycle, the light output per cycle falls off so that at 20 kilocycles it is only about one-tenth that at 0.2 kilocycle. This fall-off does not occur in the analog, and is therefore presumably attributable to changes within the glow spots, changes that progress as the current density increases.

Dissipation Factor versus Voltage

The dissipation factor, D, is equal to the ratio of the watts consumed to the reactive volt–amperes drawn. At low values, D is nearly equal to the power factor.

For the simple analog, as the voltage is raised, D remains at zero until the threshold voltage is reached and the lamp ignites. Then D rises rapidly as the conductance of the lamp rises, reaching a maximum as the lamp conductance equals the susceptance of capacitor C_2 in parallel with it.

For the composite analog, consequently, the value of D will rise with the voltage as long as the voltage increments cause more lamps to ignite.

For an EL layer, the value of D rises with voltage, as it does for the composite analog. The principal difference is that the EL layer has a residual value of D of a few per cent at very low voltages where no light emission is evident. The residual power loss presumably results from current flow in the copper sulfide structures within the EL-phosphor particle.

Dissipation Factor versus Frequency

For both the simple and the composite analogs, change of frequency produces no change in either the charge transfer per cycle, or the

voltages appearing across any of the components. Consequently, both the watts consumed, and the volt–amperes drawn, are proportional to frequency. The dissipation factor, D, is therefore independent of frequency. D remains unchanged, as frequency is changed, because the voltage drop across the lamp is constant in spite of the variation in the current.

For an efficient EL layer, D is very nearly constant over the frequency range from 0.2 to 20 kilocycles, as shown by the data in Table 1. This relationship apparently cannot be duplicated by any

Table I—Properties of an Efficient EL Layer (area 2.27 cm^2; thickness 50 microns; both electrodes transparent; phosphor ZnS:Cl:Cu, green at 0.2 kilocycles, blue at 20 kilocycles; dielectric, Aritemp* (epoxy)).

Frequency (kilocycles)	0.2			2.0			20		
Applied Voltage (Volts)	100	200	300	100	200	300	100	200	300
Brightness (Foot-Lamberts)	0.4	3.5	8	1.3	14	55	2.6	45	100
Current (ma)	.035	.08	.135	.37	.80	1.23	3.6	7.7	11.5
Capacitance ($\mu\mu$f)	326	354	384	296	319	347	286	308	336
Dissipation Factor	.06	.11	.13	.04	.10	.15	.05	.09	.13
Lumens/watt	10	10.5	7.4	4.0	4.3	5.0	0.77	1.6	1.15

equivalent circuit using linear elements for the power load. If a resistor and a capacitor are in series, D is proportional to frequency; if they are in parallel, D is inversely proportional to frequency. This frequency sensitivity of D persists in more complex networks; constancy of D can be achieved only over limited frequency ranges.

For EL layers of lower efficiency, the internal leakage conductance (nonlinear) provides a greater additional load extraneous to the emission process. The greater this internal leakage becomes, the more D becomes dependent on frequency, rising as frequency is diminished, as expected for a parallel-connected resistor–capacitor combination.

Capacitance versus Voltage

In the simple analog, the capacitance below the threshold voltage, is that of C_1 and C_2 in series. Substantially above the threshold volt-

* Registered trade mark.

age, C_2 is shorted out by the lamp conductance, and the capacitance for the circuit rises to a value nearly that of the driving capacitor, C_1, alone.

For the composite analog, the capacitance must rise progressively with the voltage, because additional lamps ignite, and conductance rises for those already glowing.

For EL layers, the capacitance rises with the applied voltage, as with the composite analog. However, EL layers show abnormal capacitance even at voltages so low that no emission is apparent. This abnormal capacitance is about double that of an EL-type layer, made with a non-EL phosphor. The greater capacitance can be attributed to the conductive copper sulfide structure within the EL-phosphor particles. These structures, though small in total volume, are presumably so arranged, either as thin sheets or thin needles, within the particle, that the average line of force travels only half of the layer thickness through dielectric; the other half of the way it follows a conductor.

Capacitance versus Frequency

For both the simple and the composite analogs, the capacitance is independent of the frequency. For the EL layer (Table I) the same is very nearly true. A small trend toward lower capacitance is observed as the frequency is raised. This is to be expected in a heterogeneous material in which a wide range of conductivities occur among the many microvolume regions. As the frequency is raised, the crossover frequencies are passed for the various regions, the susceptivity overtakes the various conductivities, and these regions are in turn converted from series conductors to series capacitors.

Emission Wave Form

For the simple analog, the duty cycle and, consequently, also the wave-form for the emitted light will depend on the magnitude as well as on the wave-form of the applied voltage. For simplicity, this discussion is limited to sinusoidal excitation. For an applied voltage just above the threshold, the duty cycle is low, and the light from each electrode is only a small pip emitted just as the voltage approaches its peak. As the voltage is raised, the duty cycle increases. Beyond voltages several times the threshold value, the emission wave-form approaches that of a complete half-cycle of sine wave, with its peak 90° ahead of the voltage peak. Thus, both the shape and the phase

relation change as the ratio of applied voltage to threshold voltage is changed.

In the composite analog, the emission wave form will be the sum for all the lamps. At any one applied voltage, a wide range of applied-to-threshold voltage ratios will exist, and duty cycles will range from near zero to near one half. The composite wave-form is not easily predicted, although its peaks should lead the applied voltage peaks somewhat, generaly by about 45°.

For the EL layer, two principal features of the emission wave forms are the same as those of the composite analog. There are two peaks of light per cycle, and these peaks do lead the applied voltage peaks. A third feature, the rise in the troughs as the frequency is raised, is attributable to the afterglow of the phosphor, and has no counterpart in the analog. A fourth feature, observed occasionally, is the secondary peak. Such peaks become less prominent as the EL layers become more efficient, and the simple analog suggests no mechanism for them. These secondary peaks presumably arise by the interaction of a nonemitting process competing with the EL emitting process. Except for the elevated troughs, and perhaps the secondary peaks, it may well be that every published emission waveform comprises a summation of many different duty cycles of assorted intensities. If so, such emission wave forms are not promising data for deducing the details of the emission process within the glow spot.

Emission Color versus Frequency

Although the analogs do not undergo appreciable change in emission color, the current-density changes that occur within the lamps are useful for correlating the emission color changes that do occur for some EL phosphors.

In the simple analog, increase of either the applied voltage or frequency increases the current through the lamp without appreciably changing the voltage drop across it. In spite of this similarity between effects of voltage and frequency changes, a difference becomes apparent in the composite analog. At a given voltage, a certain fraction of lamps will glow; increasing the frequency does not change this fraction, but does increase the current density within those lamps which are glowing. At a given frequency, however, an increase in voltage not only increases the current density of lamps already glowing, but also ignites additional lamps. The newly ignited lamps will be operating at low current density, because the applied voltage is not far above threshold for each. Voltage increase therefore probably

does not greatly increase the current density averaged for all glowing lamps.

In an EL layer whose phosphor has two emission bands, color shift towards the higher energy band can be produced by raising the frequency, but not by raising the voltage. If the color shift arises from increase in current density within the phosphor, as it does with cathode-ray excitation, there are two reasons why frequency change is more effective than voltage change in producing the color shift: (1) voltage can be increased only by a ratio of about 4 from threshold brightness to maximum brightness before breakdown, while the frequency-range ratio can be at least 1000 (e.g., from .02 to 20 kilocycles), and (2) voltage increments, while increasing current density in some glow spots, also ignite additional glow spots, which will initially be low in current density.

Brightness Maintenance

The properties of an EL layer change with time during its operating life; the most conspicuous change is the deterioration of brightness. The fraction of original brightness remaining after a specified duration of operation (brightness maintenance) depends[7] more or less upon such diverse factors as the procedure used in preparing the phosphor, structure of the layer, and operating conditions. This brightness deterioration can be attributed to damage of the vital internal electrodes as indicated by the following evidence.

High ambient humidity often results in abnormal deterioration. If, for example, the EL layer is contained between two sheets of glass, it is well protected except at its edges. In this case the abnormal deterioration starts at the edges, and the dark edge progressively widens. The internal copper-sulfide electrodes are probably being electrolytically destroyed, and traces of water-soluble electrolytes are probably present to facilitate the process. This abnormal deterioration can be avoided by proper protection of the layer.

A much slower normal deterioration takes place in layers well protected against humidity. During this deterioration process, the response to ultraviolet excitation does not suffer appreciably.[7] This indicates that the phosphor phase is not being damaged, and therefore probably the internal electrodes are being damaged. Also, as the brightness falls the power consumption falls nearly proportionately.

---

[7] W. A. Thornton, "Electroluminescence Maintenance," *Jour. Electrochem. Soc.*, Vol. 107, p. 895, November, 1960.

This, too, indicates that the electrodes are failing in their function of transmitting power from the field to glow spots.

The copper-sulfide electrode system constitutes only about one part per thousand of an EL phosphor. If it consisted of randomly distributed spheres, it could not be expected to have much electrical effect in the phosphor. It seems likely, therefore, that it takes the form of thin sheets or needles, having sharp edges or points. Furthermore, the power is consumed in small glow spots, all of which taken together constitute only a small fraction of the entire volume of an EL layer. Consequently, when such fragile electrode structures are called upon to deliver relatively enormous power densities to the glow spots, it is to be expected that the electrodes will be damaged or eroded. Raising the frequency shortens the life of an EL layer, not only in time duration, but in number of cycles.[7] This indicates that the internal electrode erosion rises even faster than the current density. But, in contrast, raising the applied voltage is almost without effect on the life of an EL layer.[7] This is in agreement with the earlier analysis of the analog, showing that voltage increases do not appreciably increase the average current density within the glow spots.

As the proportion of EL phosphor to plastic is increased in an EL layer, the life is shortened.[7] This again involves an increase in current density. At low phosphor proportions, the phosphor particles are in an ambient of low-dielectric-constant plastic. As the phosphor proportion is increased, the average dielectric constant rises. Thus, the field within the particles rises, and the current density rises in the glow spots. In terms of the analog, the driving capacitor becomes larger.

As the ambient temperature is raised, the life of an EL layer diminishes.[7] This suggests that high local temperatures in the glow spots are largely responsible for the erosion of the electrodes. Perhaps some of the erosion consists of redissolving some copper sulfide into the zinc sulfide because of these high temperatures.

Constant-current operation, instead of constant-voltage, achieved by connecting to a high-voltage source through a large external impedance lengthens the life of an EL layer.[7] Voltage rises automatically during operation as the impedance of the layer rises. In this case, a rise in the threshold voltage of an EL unit because of electrode damage does not necessarily result in its extinction.

Deterioration at a fixed voltage is selective; the deteriorated layer will operate at a somewhat higher voltage without much evidence of having deteriorated.[7] The higher voltage not only puts into service many other EL units that did not function at the original voltage, but

perhaps also puts back into service some whose threshold voltage had risen too much to permit them to continue to function at the lower voltage.

Within limits, the life of an EL layer improves with increasing copper proportion in the EL phosphor.[7] This could be the result of dividing the electrical load within the phosphor among more EL units since more electrodes result in more glow spots. Or alternatively, it may be that the electrodes are thicker and sturdier. The fall-off in brightness at the higher copper concentrations[7] may result from increased shielding or bypassing of an otherwise useful EL unit by a nearby conductor.

Higher phosphor firing temperatures improve EL layer life.[7] For higher firing temperatures, more copper sulfide presumably dissolves in the zinc sulfide and segregates as the phosphor cools. This could provide thicker, sturdier electrodes, and consequently slower erosion.

Larger particles would presumably have proportionately larger electrode structures. Because they are thicker, these electrodes would erode less rapidly, and deterioration of brightness would be slower for larger particles, as reported.[7] Furthermore, these electrodes would probably be more extensive in area or length and therefore do a better job of concentrating the field, leading to higher brightness for larger particles, particularly at the lower voltages, as reported.[8]

Practical EL phosphors are limited to zinc sulfide and selenide as host materials. Zinc-selenide EL phosphors have a longer life than the commoner zinc-sulfide EL phosphors.[7] This must be attributed to their having internal electrodes of copper selenide, which for some reason are more durable than those of copper sulfide.

Acknowledgment

Many helpful discussions with R. E. Shrader are gratefully acknowledged.

[8] W. Lehmann, "Particle Size and Efficiency of Electroluminescent Zinc Sulfide Phosphors," *Jour. Electrochem. Soc.*, Vol. 105, p. 585, October, 1958.

OPTIMUM BINARY FM RECEPTION USING DISCRIMINATOR DETECTION AND I-F SHAPING*†

By

Alan A. Meyerhoff# and William M. Mazer‡

Summary—The purpose of this study was to determine means of lowering the received carrier threshold in the detection of binary FM. Comparisons are made on the basis of received carrier power required to achieve a given error probability per pulse. The basis of evaluation of the detection under consideration is an ideal curve of error probability versus a normalized parameter, the ratio of energy in a signal element and noise power per cycle of bandwidth.

A large part of the paper is devoted to the derivation of the error probability relationships for discriminator reception with an i-f filter so narrow that the modulating pulses are shaped thereby. The results show that with a proper choice of parameters, this system requires only about 2 decibels more received carrier power than the ideal. For comparison, another detection method, the use of a narrow mark or space filter, is analyzed.

Introduction

THE problem examined is the achievement of better threshold performance for binary FM transmission than that obtained with present systems. The basis of comparison is the required received carrier power for a given independent bit-error rate.

The relationship of error rate and a normalized energy ratio for an ideal system has been presented in the literature. For example, Turin[1] and Lawton[2] give this relationship for a binary system involving unique mark and space wave forms having the same energy per

† Manuscript received 17 July 1961.

* This work was supported by the U. S. Army under Signal Corps Contract DA-36-039-sc-78045.

‡ Advanced Systems Center, U. S. Defense Group, International Telephone and Telegraph Corporation, Nutley, N. J., formerly with Radio Corporation of America.

RCA Surface Communications Division, New York, N. Y.

[1] G. L. Turin, "Error Probabilities for Binary Symmetric Ideal Reception through Nonselective Slow Fading and Noise," *Proc. IRE*, Vol. 46, p. 1603, September, 1958.

[2] J. G. Lawton, "Theoretical Error Rates of 'Differentially Coherent' Binary and 'Kineplex' Data Transmission Systems," *Proc. IRE*, Vol. 47, p. 333, February, 1959.

signal element (bit). The best results achievable with such a system are given by

$$P_e = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \sqrt{\frac{E}{n_0}}$$

where P_e is the probability of error per signal element, E is the energy in a signal element, and n_0 is the white noise power per cycle of bandwidth. This equation is given as curve A of Figure 1.

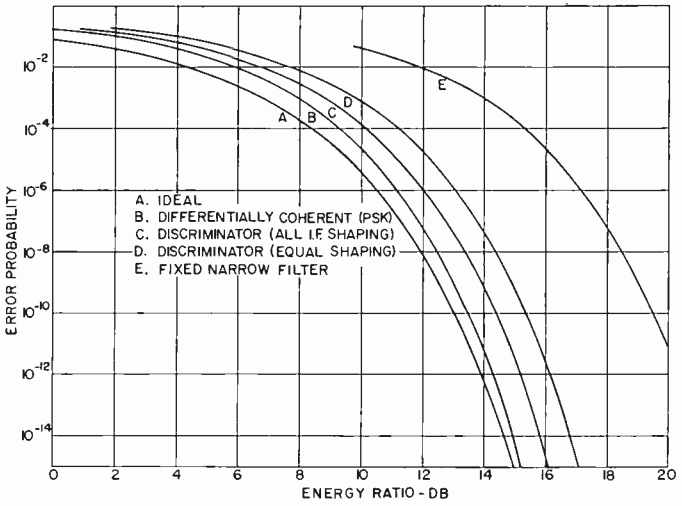


Fig. 1—Comparison of systems.

In the situation where continuous frequency modulation is allowed, although the energy in all signal elements is the same the actual r-f and i-f wave forms can vary from pulse to pulse. The modulating pulses are always one of two unique wave forms, but phase continuity of the carrier implies that the starting phase of each signal element can have almost any value. However, if the modulated carrier is regarded as a succession of equal-length signal elements, the actual wave form for any given signal element proceeds from that point in one of only two ways. Hence, the ideal relationship holds here too (as an upper bound on system performance).

The System

Consider a radio transmission system as represented schematically in Figure 2. The message supplied to point 1 is a random succession

of equal positive or negative impulses with regular spacing of h seconds. In passing through the low-pass network, the pulses become shaped at point 2 where, if the form of one pulse is $f(t)$, the signal is

$$s(t) = \sum_{n=-\infty}^{\infty} \pm f(t - nh). \tag{1}$$

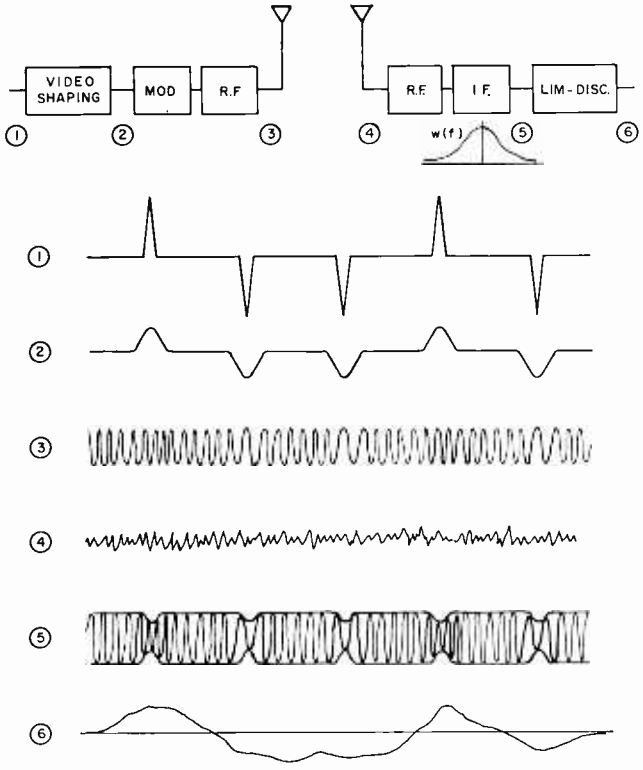


Fig. 2—Radio system and waveforms.

A carrier is frequency modulated by $s(t)$ to produce

$$g(t) = A \cos\left[\omega_0 t + \phi + 2\pi \int_{-\infty}^{t} s(x)\,dx\right]. \tag{2}$$

Flat Gaussian noise with density n_0 watts per cycle of bandwidth is added at the receiver. The carrier level and noise density are referred

to the receiver input, although we are interested in their relative effects at points in the receiver after amplification has taken place. Thus $n_0 = kTF$, where k = Boltzmann's constant, T = temperature in degrees Kelvin, F = noise figure. Let the received carrier amplitude be Q_0, and its power be P_0. The sum of the modulated carrier (3) and the noise (4) is passed through a band-pass filter, which produces the wave form (5) with reduced carrier amplitude, Q, at the centers of the modulating pulses. The instantaneous frequency deviation is also affected. The function of the limiter and discriminator is to produce the instantaneous frequency of the composite input signal.

The purpose of the receiver is to recover the transmitted sequence of marks and spaces* with a probability of error below an assigned limit, while keeping the required energy per signal element as small as possible. Only the ratio of carrier power to noise power is important, not their absolute values; thus a quantity proportional to $P_0/(b_0 n_0)$ will be a significant parameter, where b_0 is the equivalent noise bandwidth.

The energy in a signal element, E, is $P_0 h$, so that the parameter is

$$K \frac{E}{h b_0 n_0}.$$

However, hb_0 is constant for equivalent systems differing only in scale. By choosing $K = hb_0$, the parameter is

$$\frac{E}{n_0} = \frac{P_0 h}{n_0}.$$

Thus the same results can be achieved by increasing pulse power or spacing, or by decreasing noise density (noise figure) by a given factor. Studies of ideal systems in the literature[1,2] are based on this parameter.

In this analysis there is no receiver base-band shaping. That is, the mark or space decision is made directly at the output of the discriminator or after a flat base-band amplifier. It will be shown later that performance approaching the ideal is achievable without such shaping.

The decision is made by measuring the instantaneous discriminator output at the pulse centers. Perfect video synchronism is assumed.

* The designations "mark" and "space" are used for convenience even though the space is not literally that.

In practice, such synchronism is easily derived from the pulse train itself whenever the error rate is reasonably low. For example, a system which would be unusable for bit-error rates exceeding 1 in 10^6 would typically not show lack of synchronism until the received level is low enough to produce 1 error in 10^4.

The analysis here does not involve the ordinary FM concepts of high deviation, great bandwidth or restriction to operation above threshold. Instead, the FM treatment is kept as general as possible, and indeed the results indicate that best binary reception is obtained when some of the usual rules-of-thumb are violated. For example, the peak deviation limits can be well outside the relatively flat part of the i-f pass band.

Error Rate

Under the conditions stated, and with a symmetrical i-f bandpass characteristic the probability of error, as derived in Appendix I, is given by

$$F(\rho,G) = \frac{e^{-\rho}}{2} \int_0^\infty \frac{e^{\zeta}}{(1+\mu^2)^{3/2}} \left\{ I_0(\zeta) + (2\zeta + G^2\rho)\,[I_0(\zeta) + I_1(\zeta)] \right\} du \tag{3}$$

where $$\zeta \equiv \frac{\rho}{2}\left[1 - \frac{(\mu+G)^2}{1+\mu^2}\right]$$

$$\rho = \frac{Q^2}{2n_0b_0} = \frac{P}{n_0b_0} = \text{the signal-to-noise ratio at the detector,}$$

$$b_0 = \int_0^\infty w(f)\,df,$$

where $w(f)$ is the dimensionless bandpass power characteristic. G, the generalized deviation, is defined by

$$G = 2\pi S_M \sqrt{\frac{b_0}{b'_2}}, \tag{4}$$

where S_M is the peak instantaneous frequency deviation and

$$b'_2 = (2\pi)^2 \int_0^\infty w(f)\,(f - f_0)^2\,df, \tag{5}$$

where f_0 is the carrier frequency.

All of the parameters involved here appear at the input to the limiter–discriminator combination after the bandpass filtering has been accomplished. As plotted in Figure 3 the curves of $F(\rho, G)$ can be used to find the error rate if the various parameters at that point are known. If, as is usually the case, a minimum of carrier power is

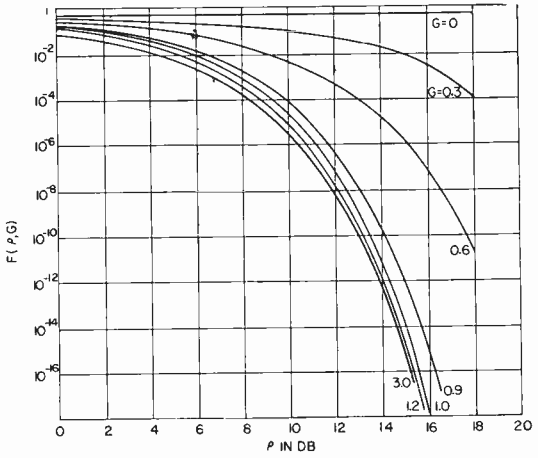


Fig. 3—Computed curves of $F(\rho,G)$.

required at the receiver input, the error probability must be referred to ρ_0, defined as

$$\rho_0 = \frac{Q_0^2}{2n_0 b_0}. \tag{6}$$

In the process, it will usually be advantageous to plot $F(\rho,G)$ versus ρ_0, taking into account that the amount of carrier reduction at the pulse centers depends on G, which is proportional to frequency deviation.

The procedure can be illustrated by considering a system in which the carrier is modulated by impulses, that is, the carrier phase is instantaneously shifted by $\pm\ b$ radians. The video shaping network of the transmitter is omitted in this case. Let the receiver bandpass

be Gaussian in shape,

$$w(f) = \exp\left\{-\frac{(f-f_0)^2}{\sigma_f^2}\right\}. \tag{7}$$

Find the value of b which requires least received carrier power for a given error rate.

Although the passage of FM through a frequency selective network has a nonlinear effect on the modulation, so that superposition does not hold, we assume here that the multipulse train behaves about the same as the sum of the individual pulses as far as peak deviation is concerned; that is, any given pulse peak is not appreciably altered by the presence of other pulses. It is also assumed that inter-symbol interference is no worse than it would be if the impulses were applied directly to one-half the bandpass network acting as a low-pass filter. The magnitude of the pulse peak itself, however, is computed on a true basis, taking the FM distortion into account. The parameters work out, as shown in Appendix II, to give

$$G = \frac{2}{\sqrt{\pi}} \tan\frac{b}{2}, \qquad b < \pi, \tag{8}$$

while the reduction in carrier amplitude at the pulse center is given by

$$\frac{Q}{Q_0} = \cos\frac{b}{2} = \frac{1}{\sqrt{1+\frac{\pi}{4}G^2}}. \tag{9}$$

Then

$$\frac{\rho}{\rho_0} = \frac{1}{1+\frac{\pi}{4}G^2}. \tag{10}$$

In order to replot the error curves with respect to ρ_0 it is only necessary to translate each constant-G curve to the right by the appropriate number of decibels, in accordance with Equation (10). The result of this process is given in Figure 4. To obtain the minimum more clearly, ρ_0 versus G is replotted for a given error probability in Figure 5; it is found that $G = 1$ is very close to the minimum value, and so will be used here. Then, by Equation (8),

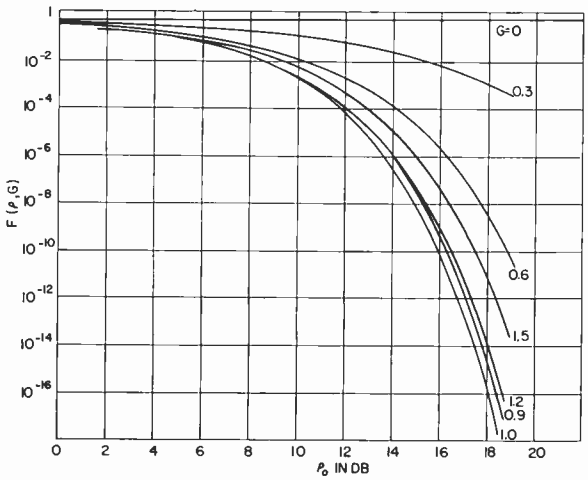


Fig. 4—Curves of $F(\rho,G)$ plotted against $\rho_0 = \left(1 + \frac{\pi}{4} G^2\right)\rho$.

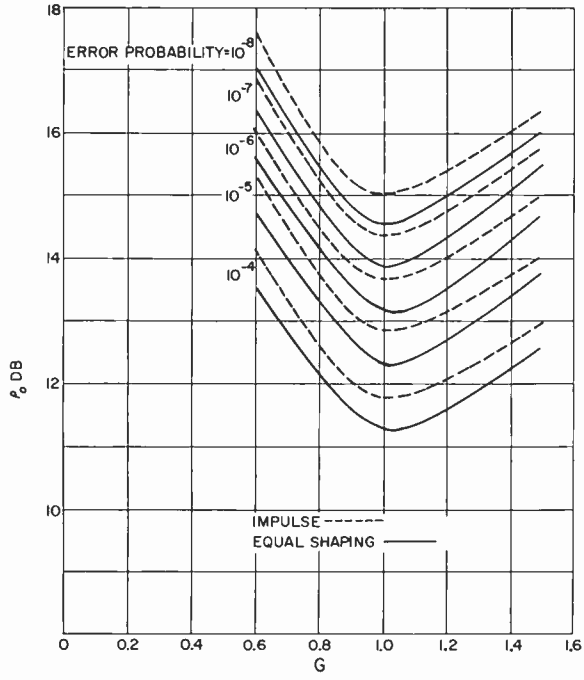


Fig. 5—Curves for choosing best G (dashed curves, impulse; solid curves, equal shaping).

$$b = 2 \tan^{-1} \frac{\sqrt{\pi}}{2} = 83.2°.$$

Also,

$$\rho_0 = \frac{\frac{Q_0^2}{2} h}{n_0} \frac{1}{b_0 h} = \frac{P_0 h}{n_0} \frac{1}{b_0 h}, \tag{11}$$

where h is the reciprocal of the pulse rate. The first factor of the right member of Equation (11) is the ratio of energy per signal element, $P_0 h$, to noise power per cycle, and is thus the normalized parameter used by Turin and others. The second factor, $b_0 h$, depends on the tolerable intersymbol interference. In the Gaussian bandpass case, $b_0 = \sigma_f \sqrt{\pi}$, as shown in Appendix II.

Table I—Representative Values of Intersymbol Interference

ϵ	$h\sigma_f$
.001	.591
.01	.482
.02	.446
.05	.390
.1	.342
.2	.286
.3	.247
.5	.187

Computations of the response of bandpass networks to FM by single pulses consistently show pulse outputs with greater peaks and smaller tails than the pulses out of an equivalent low-pass network. The greater peaks are taken into account as in Equation (8), where the equivalent low-pass network would give $G = b/\sqrt{\pi}$. The exact effect of the smaller tails would be difficult to compute, so the probably conservative assumption is made that the intersymbol interference is as given by the equivalent low-pass filter acting directly on the pulse train. Then the interference by one adjacent pulse, as a fraction of the peak, is $\epsilon = \exp(-2\pi^2 \sigma_f^2 h^2)$. Table I is a list of representative values.

A reasonable value of ϵ is 0.05, giving a total for the adjacent two pulses of 10 per cent of the tolerable interference; that is, 10 per cent

of the pulse peak itself. The corresponding reduction in allowable noise will not be taken into account now in computing the required carrier power, for it complicates the analysis and can be allowed for later by providing slightly more carrier power.

For $\epsilon = 0.05$, $h\sigma_f = 0.390$, so that for an error rate of 1 in 10^8, where the minimum $\rho_0 = 15$ decibels from Figure 5, the required energy ratio is $P_0 h/n_0 = \rho_0\sqrt{\pi}h\sigma_f$, or $15 + 2.5 - 4.1 = 13.4$ decibels. This result is quite general, being applicable to all pulse rates and noise figures. As an example, consider a pulse rate of 500,000 per second ($h = 2$ microseconds), and a noise figure of 10 decibels. The required carrier level is $(P_0)_{db} = -114 + 10 - 3 + 13.4 = -93.6$ dbm. This is the minimum value for the conditions stated and is about 3 decibels lower than the value required by previous design.

The transmitted spectrum for impulse modulation will be broader than it is when shaped pulses are used, but the corresponding narrower receiver bandwidth will tend to mitigate this effect as far as mutual interference of channels is concerned.

The case in which the transmitter low-pass and receiver bandpass networks have the same Gaussian parameter has also been worked out (Appendix III). In this case, the required ρ_0 is about 1/2 decibel less than for impulse transmission. However, for the same over-all intersymbol interference, the receiver bandwidth, measured by σ_f, is greater by the factor $\sqrt{2}$, so that the required carrier power is greater by $1.5 - 0.5 = 1$ decibel. Thus the impulse case is 1 decibel better. With equal Gaussian shaping, no simple expression has been found which relates b, the total phase deviation of one pulse, and G, the frequency deviation parameter after the bandpass. Therefore, the relation was established graphically after a point-by-point numerical integration. Figures 5, 6, and 7 pertain to the equal-shaping case.

Networks other than Gaussian can be treated in the same manner. In general, numerical integration will be required when the modulating pulses are not impulses. Perhaps the best procedure to obtain comparisons for non-Gaussian networks is to use impulse modulation, in which case the b, G relationship can be found easily.

An interesting point is that the curve relating ρ_0 and error rate for the optimum G in the best case thus far treated, namely modulation by impulses with Gaussian bandpass in the receiver, falls very nearly along the curve for differentially coherent reception as given by Lawton:[2]

$$P_e = \frac{1}{2}\exp\left(\frac{-E}{N_0}\right),$$

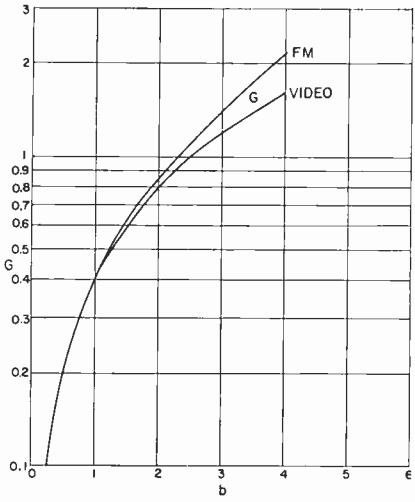


Fig. 6—G versus b for equal shaping.

where P_e is the probability of error, E the energy in a signal element, and N_0 = noise power per cycle. This relation is shown as curve B in Figure 1. Conveniently, the only G curve known in a simple closed form is the near-optimum $G = 1$. The expression is

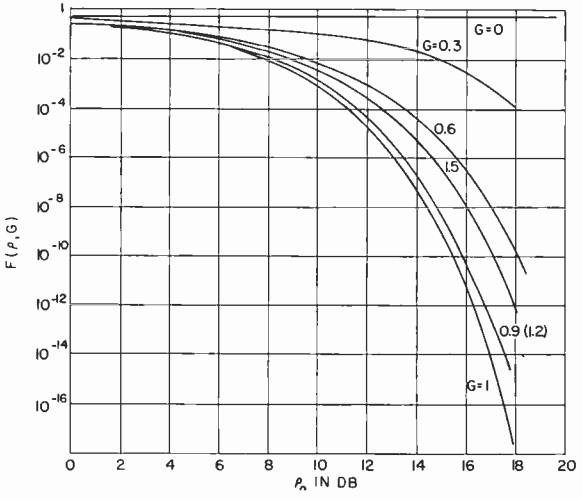


Fig. 7—$F(\rho,G)$ plotted against ρ_0 as given for equal shaping.

$$F(\rho,1) = \frac{e^{-\rho}}{2},$$

which, with Equation (10), becomes

$$F(\rho,1) = \frac{1}{2} \exp\left\{\frac{\rho_0}{1 + (\pi/4)}\right\}, \tag{12}$$

or, by Equation (11),

$$F(\rho,1) = \frac{1}{2} \exp\left\{-\frac{P_0 h}{n_0} \frac{1}{b_0 h} \frac{1}{1 + (\pi/4)}\right\}. \tag{13}$$

The term P_0h/n_0 in Equation (13) is identical with the term E/N_0 in Lawton's expression. Since

$$\frac{1}{b_0 h} \frac{1}{1 + \pi/4} = \frac{1}{\sqrt{\pi}(.390)} \frac{1}{1 + \pi/4} = 0.81,$$

it is seen that the carrier power for modulation by impulses with Gaussian bandpass in the receiver must be about 1 decibel higher than in the differentially coherent system. This result is shown in Curve C of Figure 1. An additional allowance of about 1 decibel should be taken to compensate for the reduction in noise margin due to intersymbol interference. A better choice of $h\sigma_f$ can probably be made to bring the results closer.

Furthermore, the differentially coherent result is, as pointed out by Lawton, only slightly poorer than in the ideal coherent situation, curve A, which is the true optimum. Thus, we are within about 2 decibels of the ideal.

FM Threshold

An interesting feature of the results shown by the curves of error probability versus carrier level is the lack of the sharp-breaking threshold usually associated with FM reception. The only threshold here is the carrier level required to yield the assigned maximum tolerable error probability.

The concept of an FM threshold as the carrier level below which a linear relation between carrier/noise and signal/noise no longer holds is not valid here. Attempts to apply program broadcasting and

direct voice channel techniques to the digital case can lead to erroneous results. The main differences are in the detected noise. In voice-channel transmission, the noise part of the output signal-to-noise ratio is taken with the signal absent. The noise statistics can then be handled easily, particularly above threshold. The effects of post-detection filtering are easily determined.

In digital-transmission, it was seen that the noise had to be evaluated at the sampling instants with signal present. In fact, the probability density function integrated in Equation (3) pertains to the combined output. If the noise-free signal value is subtracted from the composite output, a different amount of residual noise would result for each signal value on the pulse curve. The effects of low-pass filtering on such a varying kind of noise do not appear easy to determine, but it would be interesting to have such information.

Relationship to Phase-Shift Keying

Frequency modulation by impulses is equivalent to phase-shift keying. It is interesting that the optimum phase shift found here is very close to the value $\pm 90°$ used in PSK to make the two possible shifts as far apart as possible. Since quaternary PSK employs phase shifts of $\pm 45°$ and $\pm 135°$, it is probably best to use these values for quaternary transmission even with discriminator reception.

Practical Considerations and Typical Design

In the realization of a system employing FM by impulses, certain practical considerations should be taken into account. First, it must be decided how closely the modulating pulses should approximate impulses. It seems reasonable to require that the peak deviation be significantly greater than half the equivalent noise bandwidth of the receiver bandpass network. The pulse width would then be chosen to produce the required phase deviation through the entire pulse.

A second consideration is the sensitivity of the output to mistuning or drift in the receiver. A related problem is the inclusion of an order-wire signal, as required with most radio relay equipment. A voice-frequency signal added directly to the pulse trains would act for short periods of time like a mistuning. The subject of order-wire is complicated enough to require a separate investigation.

Although the use of impulse modulation has been demonstrated to be better than equal shaping from the standpoint of required received carrier power, some of the practical considerations above, together

with the fact that present designs are closer to equal shaping, lead to a design utilizing equal-shaping Gaussian filters.

It has been found that the best results occur at $G = 1$. From Figure (6), $b = 2.3$ radians or 132°. From Equations (58) and (64), of Appendix III, the peak deviation of the transmitted signal is

$$f_{DT}(0) = \pm \frac{b}{2\pi} \frac{1}{\sigma_{f_1}\sqrt{2\pi}} = \pm \frac{b\,\sigma_{f_1}}{\sqrt{2\pi}} = \pm 0.917\,\sigma_{f_1}. \qquad (14)$$

The intersymbol interference criterion given before, $\epsilon = 0.05$, is suitable here, so that $h\sigma_f = 0.390$, where σ_f is the over-all shaping parameter. With equal shaping, σ_{f_1} for one of the filters is $\sqrt{2}\,\sigma_f$. For example, with 12-channel 6-digit PCM, $h = 1/576{,}000$ second, so that $\sigma_f = 225$ kilocycles and $\sigma_{f_1} = 318$ kilocycles.

The over-all 3-decibel-down frequency is $f_{3db} = \sqrt{\ln 2}\,\sigma_f = 187$ kilocycles, while for each filter the 3-decibel-down frequency is 265 kilocycles. From Equation (14), the peak deviation is ± 290 kilocycles.

For an error rate of 1 in 10^6, ρ_0 must be 13.1 decibels, from Figure 7. The required received carrier power, $P_0 = \rho_0 n_0 b_0 = \rho_0 n_0 \sigma_{f_1}\sqrt{\pi}$, which is $13.1 - 144 + 10 + 25 + 2.5 = -93.4$ dbm, with a 10-decibel noise figure. An allowance of 1 decibel should be made for intersymbol interference.

For other values of h, the frequency terms and P_0 vary inversely with h. Thus, a 48-channel PCM system requires four times the bandwidth, deviation, and power.

For reasons of frequency-drift and order-wire allowance, one may want to increase the bandwidth. If all the relationships are maintained, the required power will vary directly as the bandwidth. The peak deviation must be increased appropriately. Incidentally, the intersymbol interference will be reduced.

Narrow Filter Tuned to Mark or Space Frequency

In binary frequency-shift transmission a mark is transmitted at frequency β and a space at frequency ω. Wide-band flat Gaussian noise of power n_0 watts per cycle is added. The receiver is essentially a very narrow filter tuned to β. These frequencies can be in an intermediate-frequency range, having been translated from the actual transmitted frequencies.

The incoming signal is fed into the filter from the beginning to the end of each signal element, with gating and discharging circuits

ensuring the mutual independence of all signal elements. At the end of each signal element, the envelope of the output is examined and a mark or space decision is made based on a threshold level. Simultaneous mark and space filters can be used in two receiver paths, in which case the outputs are subtracted and the decision threshold is zero.

Assume a filter as shown in Figure 8, with $R_L \gg R$. Using Laplace transforms,

$$\frac{E_0(s)}{E_i(s)} \cong g \frac{s + a_0}{(s+\alpha)^2 + \beta^2}, \tag{15}$$

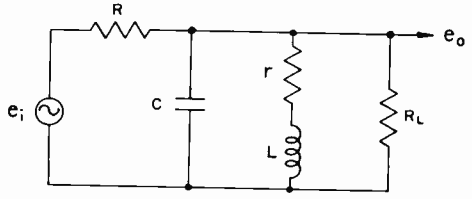


Fig. 8—Filter.

where $g = 1/RC$, $a_0 = r/L$, $\alpha = (g + a_0)/2$, and

$$\beta = \sqrt{\frac{R}{r} a_0 g - \frac{1}{4}(g - a_0)^2}.$$

For an input whose transform is

$$E_i(s) = \frac{\omega^2}{s^2 + \omega^2}, \tag{16}$$

the output envelope, E, becomes, after some manipulation,

$$E \cong \frac{g}{2} (x^2 + \alpha^2)^{-\frac{1}{2}} (1 + e^{-2\alpha t} - 2e^{-\alpha t} \cos xt)^{\frac{1}{2}}, \tag{17}$$

with $x = \omega - \beta$, and a_0, x, $\alpha \ll \beta$.

These results apply to a sinusoidal input of amplitude unity. The actual input will consist of a sinusoid of amplitude K, plus noise. The available carrier power, P_c, is $K^2/(8R)$. That part of the noise lying

in a band of width df can be represented by a sine wave of amplitude M, such that $M^2/(8R) = n_0 df$.

Then $K = \sqrt{8P_cR}$ and $M = \sqrt{8n_0Rdf}$. The total noise power out after time h is (suppressing the load resistance, R_L)

$$N_T \cong \int_{-\infty}^{\infty} 4n_0R\,[E(x)]^2 df = \frac{2n_0R}{\pi} \int_{-\infty}^{\infty} [E(x)]^2 dx$$

$$\cong \frac{g^2 n_0 R}{2\alpha}(1 - e^{-2\alpha h}), \tag{18}$$

which approaches $g^2 h n_0 R$ as $\alpha \to 0$.

The signal power at $t = h$ is, since $x = 0$,

$$P_s \cong \frac{K^2}{2} E^2(0) = \frac{P_c R g^2}{\alpha^2}[1 - e^{-\alpha h}]^2 \cong P_c R g^2 h^2. \tag{19}$$

The ratio of this signal power and the noise power at the end of the signal element is then $P_c h/n_0$, which is just the energy ratio.

If the space frequency is chosen so that x is a multiple of $2\pi/h$, the output space signal amplitude will approach zero as $\alpha \to 0$. Thus the FSK problem is reduced to an on–off keying problem where the mark frequency corresponds to "on" and space frequency corresponds to "off."

Montgomery[3] has treated the on–off keying problem; the probability of error for a space is

$$P_s = e^{-R_0/4}, \tag{20}$$

where R_0 is the carrier-to-noise ratio. For a mark, the probability of error is very nearly

$$P_M = \frac{1}{2}\left[\operatorname{erf}\frac{3\sqrt{R_0}}{2} - \operatorname{erf}\frac{\sqrt{R_0}}{2}\right] \operatorname{erf}\frac{\sqrt{R_0}}{2}. \tag{21}$$

The average probability of error, since equal mark and space probabilities are assumed, is

[3] G. F. Montgomery, "A Comparison of Amplitude and Angle Modulation for Narrow-band Communication of Binary Coded Modulation for Messages in Fluctuation Noise," *Proc. IRE*, Vol. 42, p. 447, February, 1954.

$$P_e = \frac{1}{2} e^{-R_0/4} + \frac{1}{4}\left[\operatorname{erf} \frac{3\sqrt{R_0}}{2} - \operatorname{erf} \frac{\sqrt{R_0}}{2} \right] \operatorname{erf} \frac{\sqrt{R_0}}{2}. \quad (22)$$

In the case under consideration, R_0 is the energy ratio $P_c h/n_0$. Furthermore, for any R_0 of reasonable size, the first term dominates, so that the approximate formula, plotted for comparison as curve E of Figure 1, is

$$P_e \cong \frac{1}{2} \exp\left\{ -\frac{P_c h}{4n_0} \right\}, \quad (23)$$

which is 6 decibels worse than differentially coherent detection.

The use of both mark and space filters will result in a 3-decibel improvement. Suppose that a mark is sent. Then, the probability density function of the envelope, M, of the sine-wave-plus-noise output of the mark filter is given by[4]

$$p(M) = \frac{M}{\psi_0} \exp\left[-\frac{M^2 + Q^2}{2\psi_0} \right] I_0\left(\frac{MQ}{\psi_0} \right), \quad (24)$$

where ψ_0 is the mean square value of the noise, and Q is the peak value of the sine wave. Similarly, the envelope, S, of the noise output of the space filter has a probability density function

$$q(S) = \frac{S}{\psi_0} \exp\left[-\frac{S^2}{2\psi_0} \right]. \quad (25)$$

In operation, the outputs of the two filters are subtracted, and a decision for mark is made when this difference is greater than zero. Thus, if a mark was sent, an error occurs whenever $M - S < 0$; the error probability is

$$P_e = \int_0^\infty \int_0^S p(M)\, q(S)\, dM dS = 1 - \int_0^\infty \int_0^M p(M)\, q(S)\, dS dM, \quad (26)$$

which reduces to

[4] S. O. Rice, "Statistical Properties of a Sine Wave Plus Random Noise," *Bell Syst. Tech. Jour.*, Vol. 27, p. 109, January, 1948.

$$P_e = 1 - e^{-R_0} \int_0^\infty (1 - e^{-x}) e^{-x} I_0(2\sqrt{R_0 x})\, dx, \tag{27}$$

where $x = M^2/(2\psi_0)$ and $R_0 = Q^2/(2\psi_0)$. Through use of a standard integral,[5] the result becomes

$$P_e = \frac{1}{2} e^{-R_0/2}, \tag{28}$$

which, by symmetry, holds for spaces as well as marks, and is thus the complete expression. As before, R_0 is simply the energy ratio, so that this case lies midway (in db) between curves B and E of Figure 1.

Conclusions

A study of detection of binary FM signals has been made on the basis of received carrier power required to achieve a given bit error rate threshold. The results show that, in the presence of white noise, discriminator reception with i-f pulse shaping can, with proper choice of parameters, be superior to other practical methods and that it approaches within about 2 decibels of the ideal represented by phase-coherent correlation detection. Curves and techniques presented in the course of the derivations are suitable for the design of systems to achieve the near-ideal performance.

Acknowledgment

The authors would like to thank J. Dutka and H. Zabronsky for helpful suggestions in connection with Appendix I.

Appendix I—Derivation of Expression for Probability of Error

Rice[4] derives a formula (No. (5.4)) for the probability density of θ', the derivative of the phase angle of a fixed sine wave plus random noise for any power spectrum of Gaussian noise. From this formula, an expression can be derived for the probability of error in discriminator detection of binary pulses frequency modulating a carrier, when the receiver bandpass characteristic is symmetrical about the carrier frequency.

---

[5] Gröbner and Hofreiter, *Integraltafel, Zweiter Teil, Bestimmte Integrale*, Springer-Verlag, 1950, p. 198, formula 5.

Rice's result is not immediately applicable to the problem considered here, since he assumes a carrier of fixed frequency and amplitude. To see the true relationship, it is first necessary to review some of the earlier development in Rice's paper (Section 3) on the definition and simple properties of R and θ, which are defined by

$$\begin{aligned} R\cos\theta &= Q + I_c, \\ R\sin\theta &= I_s \end{aligned} \tag{29}$$

where the sine wave is $Q\cos qt$ and the noise is

$$\begin{aligned} I_N &= I_c\cos qt - I_s\sin qt \\ I_c &= \sum_{n=1}^{M} c_n\cos\left[(\omega_n - q)\ t - \phi_n\right] \\ I_s &= \sum_{n=1}^{M} c_n\sin\left[(\omega_n - q)\ t - \phi_n\right]. \end{aligned} \tag{30}$$

The c_n are given by the spectrum of the noise, while the ϕ_n are random variables distributed uniformly over $(0, 2\pi)$. A continuation of Rice's development gives

$$\theta' = \frac{d}{dt}\tan^{-1}\frac{I_s}{Q + I_c} = \frac{(Q + I_c)\ I_s' - I_s I_c'}{(Q + I_c)^2 + I_s^2} \tag{31}$$

where Q and q are constant.

In relating the present case to Rice's work it is observed first that samples regularly spaced by h are of interest. It is assumed that the variations in the ϕ_n are such that their values at any sample are independent of their values at any other sample, while in a small neighborhood of any sample these values are fixed. Since the ϕ_n are chosen at random and uniformly from $(0, 2\pi)$, any other phase angles can be added without affecting the statistics. It will be convenient to add, in the neighborhood of each sampling point, t_k, the value which

$$\psi(t) \equiv 2\pi\int_{-\infty}^{t} s(x)\,dx - 2\pi s(t_k)\ t \tag{32}$$

takes at that point. Furthermore, the expansion for the noise will be

taken with respect to the angular frequency:

$$q = q_0 + 2\pi s(t_k),$$

where $s(t_k)$ is the same, S_M, for all marks.

With the assumption of variable Q and q, a new instantaneous phase-angle derivative may be written:

$$\theta' = \frac{(Q\cos A + I_c)\ (QA'\cos A + Q'\sin A + I_s')}{(Q\cos A + I_c)^2 + (Q\sin A + I_s)^2} - \frac{(Q\sin A + I_s)\ (-QA'\sin A + Q'\cos A + I_c')}{(Q\cos A + I_c)^2 + (Q\sin A + I_s)^2} \tag{33}$$

with

$$A = \psi(t) - \psi(t_k) \tag{34}$$

$$\psi_n \equiv \psi(t_k) + \phi_n \tag{35}$$

$$I_c = \sum_{n=1}^{M} c_n \cos\left[(\omega_n - q)t - \psi_n\right] \tag{36}$$

$$I_s = \sum_{n=1}^{M} c_n \sin\left[(\omega_n - q)t - \psi_n\right]. \tag{37}$$

At $t = t_k$, $A = A' = 0$, so that

$$\theta'_{t_k} = \frac{(Q + I_c)\ I_s' - (Q' + I_c')\ I_s}{(Q + I_c)^2 + I_s^2}. \tag{38}$$

Except for the Q' this is just the result that Rice uses. It is seen, then, that the form of $s(t)$ does not matter; only the value of $s(t)$ at the sampling time must be incorporated into the frequency, q. In the case of a symmetrical filter, Q' is zero at the peaks of the output pulses when the modulating pulses are symmetrical.

For all peaks of transmitted marks (or spaces), then, the distribution of θ' is the same as if there were a steady carrier at an angular frequency $q_0 + 2\pi\ S_M$, with additive Gaussian noise symmetrical about q_0, as illustrated in Figure 9.

Rice's formula for the probability density function of θ' is

$$p(\theta') = \frac{1}{8a}\left(\frac{2}{a\, b_0\, B}\right)^{1/2}\left[(\gamma+1)\, I_0\left(\frac{\gamma-c}{2}\right) + \gamma\, I_1\left(\frac{\gamma-c}{2}\right)\right]\exp\left(\frac{c+\gamma}{2} - \frac{b_0\, b_2\, \rho}{B}\right) \tag{39}$$

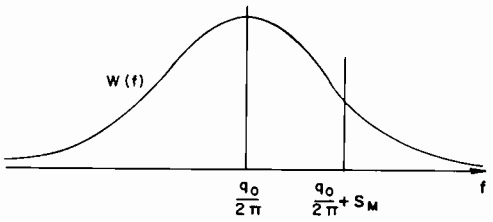


Fig. 9

where

$$b_0 = \int_0^\infty w(f)\, df$$

$$b_1 = 2\pi \int_0^\infty w(f)\left(f - \frac{q}{2\pi}\right) df \tag{40}$$

$$b_2 = (2\pi)^2 \int_0^\infty w(f)\left(f - \frac{q}{2\pi}\right)^2 df$$

$$a = \frac{(b_2 - 2b_1\,\theta' + b_0\,\theta'^2)}{2B}$$

$$B = b_0\, b_2 - b_1^2$$

$$\gamma = \frac{b^2}{a} \tag{41}$$

$$b = \frac{Q\,(b_2 - b_1\,\theta')}{2B}$$

$$c = \frac{b_1^2 \rho}{B}$$

$I_0(x)$ and $I_1(x)$ are modified Bessel functions of the first kind. Everything here is referred to the carrier angular frequency, q, which is taken to be $q_1 + 2\pi S_M$, in accordance with the above discussion. In order to refer the results to the unmodulated carrier frequency, q_0, at the center of the network, the b_i may be written in terms of the corresponding quantities, b_i', referred to q_0. Thus

$$\begin{aligned} b_0 &= b_0', \\ b_1 &= 2\pi \int_0^\infty w(f) \left(f - \frac{q_0 + 2\pi S_M}{2\pi} \right) df \\ &= b_1' - 2\pi S_M b_0, \\ b_2 &= (2\pi)^2 \int_0^\infty w(f) \left(f - \frac{q_0 + 2\pi S_M}{2\pi} \right)^2 df \\ &= b_2' - 4\pi S_M b_1' + 4\pi^2 S_M^2 b_0. \end{aligned} \tag{42}$$

For a symmetrical network, $b_1' = 0$. Therefore $b_1 = -2\pi S_M b_0$, and $b_2 = b_2' + (2\pi S_M)^2 b_0$. An error occurs for a transmitted mark whenever $\theta' < -2\pi S_M$. Thus the probability of error when a mark is sent is

$$P_e = \int_{-\infty}^{-2\pi S_M} p(\theta')\, d\theta'. \tag{43}$$

Performing the various substitutions, with

$$\zeta \equiv \frac{\rho}{2} \frac{1 - G(G + 2u)}{1 + u^2},$$

$$G \equiv \sqrt{\frac{b_0}{b_2'}}\, 2\pi S_M,$$

$$u \equiv -\left(G + \sqrt{\frac{b_0}{b_2'}}\, \theta' \right),$$

results in

$$\frac{1}{8a}\left(\frac{2}{a\,b_0\,B}\right)^{1/2} d\theta' = -\frac{du}{2\,(1+u^2)^{3/2}},$$
$$\gamma = 2\zeta + G^2\rho,$$
$$\frac{\gamma - c}{2} = \zeta, \tag{44}$$
$$\frac{c+\gamma}{2} - \frac{b_2\,b_0\,\rho}{B} = \zeta - \rho.$$

The integral then becomes

$$P_e = F(\rho, G) = \frac{e^{-\rho}}{2}\int_0^\infty \{I_0(\zeta) + (2\zeta + G^2\rho)\,[I_0(\zeta) + I_1(\zeta)]\}\,\frac{e^{\zeta}\,du}{(1+u^2)^{3/2}} \tag{45}$$

By symmetry, the same result holds for spaces as well as marks.

Evaluation of $F(\rho, G)$ for ranges of values of ρ and G has been done by machine computation. Some properties of F are:

$$\begin{aligned} F(0, G) &= F(\rho, 0) = \frac{1}{2}, \\ F(\rho, -G) + F(\rho, G) &= 1, \\ F(\infty, G) &= 0;\ G > 0, \\ F(\rho, \infty) &= 0, \\ F(\rho, 1) &= \frac{1}{2}\,e^{-\rho}. \end{aligned} \tag{46}$$

Appendix II — FM by Impulses through Bandpass Networks

The result of passing a carrier, $\exp(j\omega_0 t)$, on at all times, frequency modulated at $t = 0$ by an impulse, through a bandpass network is given by Gumowski.[6] For a symmetrical network, the result can be written

$$v(t) = e^{j\omega_0 t}\,[A(0) + (e^{jb} - 1)\,h(t)], \tag{47}$$

[6] I. Gumowski, "Transient Response in FM," *Proc. IRE*, Vol. 42, p. 819, May, 1954.

where b is the phase shift corresponding to the impulse in frequency and $h(t)$ is the step-function response of the low-frequency translation, with transfer function $A(\omega)$, of the bandpass network. The transfer function is normalized so that $A(0)$ is unity, and it is observed that for a symmetrical filter, $h(0) = 1/2$. Also,

$$\begin{aligned} h(-\infty) &= 0, \\ h(\infty) &= 1. \end{aligned}$$

Then

$$\begin{aligned} v(t) \Big|_{t\to\infty} &= e^{j\omega_0 t} \\ v(t) \Big|_{t\to\infty} &= e^{j(\omega_0 t + b)} \\ v(0) &= e^{jb/2} \cos b/2. \end{aligned} \tag{48}$$

Thus the carrier amplitude at the center of the pulse is reduced from its undisturbed value of Q_0 by the factor $\cos(b/2)$. The instantaneous angular frequency of $v(t)$ is found by taking the derivative of the phase angle;

$$\begin{aligned} v_d(t) &= \frac{d}{dt} \tan^{-1} \frac{h(t) \sin b}{1 + (\cos b - 1) h(t)} \\ &= \frac{h'(t) \sin b}{[h(t)]^2 + [1 - h(t)]^2 + 2h(t)[1 - h(t)] \cos b}. \end{aligned} \tag{49}$$

At $t = 0$, the center of the pulse,

$$v_d(0) = \frac{h'(0) \sin b}{\frac{1}{2} + \frac{1}{2} \cos b} = \frac{2h'(0) \sin \frac{b}{2} \cos \frac{b}{2}}{\cos^2 \frac{b}{2}} \tag{50}$$

or

$$2\pi S_M = h'(0)\, 2 \tan \frac{b}{2}. \tag{51}$$

This shows that the output increases with b at a greater than linear rate. For a Gaussian filter,

$$h'(0) = \sigma_f\sqrt{2\pi},$$

where
$$A(\omega) = \exp\left(-\frac{\omega^2}{8\pi^2\sigma_f^2}\right). \tag{52}$$

From Equation (7)

$$b_0 = \int_0^\infty w(f)\,df = \int_0^\infty \exp\left\{-\frac{(f-f_0)^2}{\sigma_f^2}\right\} df. \tag{53}$$

Since $f_0 \gg \sigma_f$ this is effectively unchanged by extending the lower limit to $-\infty$. Then.

$$b_0 \simeq \sigma_f \int_{-\infty}^{\infty} e^{-x^2}\,dx = \sigma_f\sqrt{\pi}. \tag{54}$$

Similarly,

$$b'_2 = (2\pi)^2 \int_0^\infty \exp\left\{-\frac{(f-f_0)^2}{\sigma_f^2}\right\} (f-f_0)^2 df$$

$$\simeq \sigma_f^3(2\pi)^2 \int_{-\infty}^{\infty} x^2 e^{-x^2}\,dx = \frac{\sigma_f^3\sqrt{\pi}(2\pi)^2}{2}. \tag{55}$$

With the above results, Equation (4) yields

$$G = \frac{2}{\sqrt{\pi}} \tan\frac{b}{2}. \tag{56}$$

Appendix III—Equal Shaping

Impulses of content $\pm b$ are applied at point (1) in Figure 2, so that at (2) there is the superposition of time functions of the form $\pm b a_1(t - nh)$, where $a_1(\tau)$ is the response of the transmitter shaping network to a unit impulse. One pulse, $\pm b a_1(t)$ centered at $t = 0$ will be examined going through the FM path to detection.

The transmitter is frequency modulated by $\pm b a_1(t)$, so that its output is proportional to

$$f_T(t) = \exp\{j\omega_0 t\} \exp\left\{ \pm jb \int_{-\infty}^{t} a_1(x)\,dx \right\}. \tag{57}$$

The actual instantaneous frequency deviation in cycles per second is

$$f_{DT} = \frac{1}{2\pi}\frac{d}{dt}\left[\pm b \int_{-\infty}^{t} a_1(x)\,dx \right] = \pm \frac{b}{2\pi} a_1(t). \tag{58}$$

In the receiver, all filtering is assumed to be accomplished in the i-f where the symmetrical bandpass network, given by $w(f) = [W(\omega)]^2$, acts on the received modulated wave. Let $a(\tau)$ be the impulse response of this network. Then the output can be found from the superposition integral

$$e_0(t) = \int_{-\infty}^{\infty} f_T(\tau)\, a(t-\tau)\,d\tau. \tag{59}$$

The problem can be referred to the associated low-pass network by setting $W(\omega) = X(\omega - \omega_0)$. Then

$$a(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W(\omega)\, e^{j\tau\omega} d\omega = \frac{e^{j\tau\omega_0}}{2\pi}\int_{-\infty}^{\infty} X(\omega')\, e^{j\tau\omega'} d\omega' = e^{j\tau\omega_0} a_2(\tau), \tag{60}$$

where $a_2(\tau)$ is the impulse response of $X(\omega)$. Then, substituting Equations (57) and (60) into Equation (59),

$$e_0(t) = \int_{-\infty}^{\infty} \exp\{j\omega_0 t\} \exp\left\{ \pm jb \int_{\tau}^{-\infty} a_1(x)\,dx \right\} \exp\{j(t-\tau)\omega_0\}\, a_2(t-\tau)\,d\tau$$

$$= \exp\{j\omega_0 t\} \int_{-\infty}^{\infty} a_2(t-\tau) \exp\left\{ \pm jb \int_{-\infty}^{\tau} a_1(x)\,dx \right\} d\tau. \tag{61}$$

This expression consists of a carrier term multiplied by a complex quantity which contains information as to the instantaneous phase.

The function of the discriminator is to extract the derivative of this phase angle;

$$e_d(t) = \pm \frac{d}{dt} \tan^{-1} \frac{\int_{-\infty}^{\infty} a_2(t-\tau) \sin\left[b \int_{-\infty}^{\tau} a_1(x)\,dx \right] d\tau}{\int_{-\infty}^{\infty} a_2(t-\tau) \cos\left[b \int_{-\infty}^{\tau} a_1(x)\,dx \right] d\tau} \tag{62}$$

Another important quantity from Equation (61) is the change in carrier power with respect to the unmodulated case. That is

$$\frac{\rho(t)}{\rho_0} = \left(\left\{ \int_{-\infty}^{\infty} a_2(t-\tau) \sin\left[b \int_{-\infty}^{\tau} a_1(x)\,dx \right] d\tau \right\}^2 + \left\{ \int_{-\infty}^{\infty} a_2(t-\tau) \cos\left[b \int_{-\infty}^{\tau} a_1(x)\,dx \right] d\tau \right\}^2 \right) \left\{ \int_{-\infty}^{\infty} a_1(t-\tau)\,d\tau \right\}^{-2} \tag{63}$$

Since the detection sampling occurs at the pulse center, we are interested in the quantity at $t = 0$, and we also must see that its derivative at this point is zero in order for the results of Appendix I to be applicable. Taking the derivative of Equation (63), setting $t = 0$, and assuming $a_1(\tau)$ is an even function, $\rho'(0)/\rho_0 = 0$.

Equation (62), in general, has to be evaluated by numerical integration. This was carried through for the case of equal Gaussian shaping, for several values of b. Equal Gaussian shaping implies that

$$a_1(\tau) = a_2(\tau) = \frac{1}{\sigma_{t_1}\sqrt{2\pi}} \exp\left\{ -\frac{\tau^2}{2\sigma_{t_1}^2} \right\}. \tag{64}$$

The evaluation is simplified by performing the inner integrations from 0 to τ instead of from $-\infty$ to τ. That this procedure does not change $e_d(t)$ is seen by adding $-b/2$ to the inverse tangent;

$$e_d(t) = \pm \frac{d}{dt}\{\tan^{-1} X\} = \pm \frac{d}{dt}\left\{-\frac{b}{2} + \tan^{-1} X\right\} \quad (65)$$

where X is the argument of the arctan in Equation (62). To show this, let

$$\left.\begin{aligned} Y &= -\frac{b}{2} + \tan^{-1} X; \\ \tan Y &= \frac{X - \tan\dfrac{b}{2}}{1 + X\tan\dfrac{b}{2}} \end{aligned}\right\} \quad (66)$$

Again, for $a_1(\tau)$ an even function,

$$X = \frac{\sin\dfrac{b}{2}\displaystyle\int a_2(t-\tau)\cos Z d\tau + \cos\frac{b}{2}\int a_2(t-\tau)\sin Z d\tau}{\cos\dfrac{b}{2}\displaystyle\int a_2(t-\tau)\cos Z d\tau - \sin\frac{b}{2}\int a_2(t-\tau)\sin Z d\tau},$$

where $Z = b\displaystyle\int_0^{\tau} a_1(x)\,(dx)$. When this is substituted into Equation (66),

$$\tan Y = \frac{\displaystyle\int_{-\infty}^{\infty} a_2(t-\tau)\sin\left[\, b\int_0^{\tau} a_1(x)\,dx\,\right] d\tau}{\displaystyle\int_{-\infty}^{\infty} a_2(t-\tau)\cos\left[\, b\int_0^{\tau} a_1(x)\,dx\,\right] d\tau}$$

Thus, $e_d(t)$ is unaffected by the change of limits.

For convenience in performing the numerical integration this expression is written, by change of variable,

$$\tan Y = \frac{\displaystyle\int_{-\infty}^{\infty} \exp\left\{-\frac{\tau^2}{2\sigma_{t_1}{}^2}\right\} \sin\left[\frac{b}{\sigma_{t_1}\sqrt{2\pi}} \int_0^{t-\tau} \exp\left\{-\frac{x^2}{2\sigma_{t_1}{}^2}\right\} dx\right] d\tau}{\displaystyle\int_{-\infty}^{\infty} \exp\left\{-\frac{\tau^2}{2\sigma_{t_1}{}^2}\right\} \cos\left[\frac{b}{\sigma_{t_1}\sqrt{2\pi}} \int_0^{t-\tau} \exp\left\{-\frac{x^2}{2\sigma_{t_1}{}^2}\right\} dx\right] d\tau}$$

Furthermore, let $z = t/(\sqrt{2}\,\sigma_{t_1})$ and $\zeta = \tau/(\sqrt{2}\,\sigma_{t_1})$.

$$\tan Y = \frac{\displaystyle\int_{-\infty}^{\infty} e^{-\zeta^2} \sin\left[\frac{b}{2}\frac{2}{\sqrt{\pi}} \int_0^{z-\zeta} e^{-x^2} dx\right] d\zeta}{\displaystyle\int_{-\infty}^{\infty} e^{-\zeta^2} \cos\left[\frac{b}{2}\frac{2}{\sqrt{\pi}} \int_0^{z-\zeta} e^{-x^2} dx\right] d\zeta}.$$

With the normalized variable z, a single set of curves holds for all pulse widths. A typical result of the evaluation is given as the dashed curve in Figure 10, where Y in radians is plotted as $e_1(t)$ against z for $b = 4$. The solid curve $e_2(t)$ is the result of a video approximation; that is, the integral of the impulse response of two identical low-pass filters in cascade.

The slope at the center, $z = 0$, is a measure of the peak frequency deviation out of the filter; that is,

$$f_{DR} = S_M = \frac{1}{2\pi}\frac{dY}{dt} = \frac{1}{2\pi}\frac{dY}{dz}\frac{dz}{dt} = \frac{1}{2\sqrt{2}\pi\,\sigma_{t_1}}\frac{dY}{dz}\bigg|_{z=0} \tag{67}$$

Equation (4), together with the computations of Appendix II, shows that

$$G = \sqrt{\frac{b_0}{b_2'}}\,2\pi\,S_M = \frac{\sqrt{2}}{2\pi\,\sigma_{f_1}}\frac{1}{\sqrt{2}\sigma_{t_1}}\frac{dY}{dz} = \frac{dY}{dz}\bigg|_{z=0}$$

The values of G are therefore obtained by finding the slopes at $z = 0$

of the Y curves. This was done by dividing the first computed value of Y by the corresponding value of z. The resulting curve of G versus b is given in Figure 6. Now, if a relation between $\rho(0)/\rho_0$ and b can be found, it will be possible to relate carrier reduction and G, and then plot $F(\rho, G)$ versus ρ_0, as desired. To that end, for equal

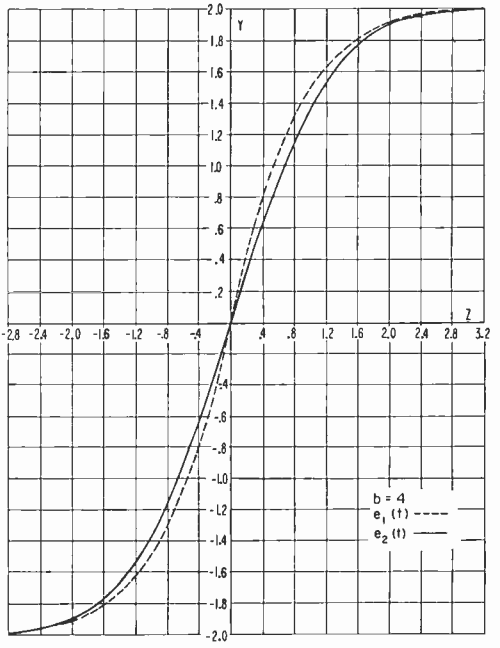


Fig. 10—Phase angle

symmetrical shaping, Equation (63) can be written, recognizing that the denominator of the right member is unity,

$$\frac{\rho(0)}{\rho_0} = \left\{ \int_{-\infty}^{\infty} a_2(\tau) \sin \left[b \int_{-\infty}^{\tau} a_2(x)\, dx \right] d\tau \right\}^2 + \left\{ \int_{-\infty}^{\infty} a_2(\tau) \cos \left[b \int_{-\infty}^{\tau} a_2(x)\, dx \right] d\tau \right\}^2$$

The main integrands are the derivatives of

$$-\frac{1}{b}\cos\left[\, b\int_{-\infty}^{\tau} a_2(x)\,dx \,\right]$$

and

$$\frac{1}{b}\sin\left[\, b\int_{-\infty}^{\tau} a_2(x)\,dx \,\right],$$

so that

$$\frac{\rho(0)}{\rho_0}=\left\{-\frac{1}{b}\left[\cos b-1\right]\right\}^2+\left\{\frac{1}{b}\sin b\right\}^2=\frac{1}{b^2}\left(2-2\cos b\right)$$

Thus the carrier power is reduced, at the time of the pulse center by,

$$\frac{\rho(0)}{\rho_0}=\frac{\sin^2 b/2}{(b/2)^2}. \tag{68}$$

Corresponding to each value of G in Figure 3, the value of b is found from the FM curve of Figure 6, and the value of $\rho(0)/\rho_0$ is computed from Equation (68). Expressed in decibels, this is numerically the amount each curve of Figure 3 was shifted to the right to produce Figure 7.

It is seen that, as in the case of all shaping in the receiver, the best value of G is again unity, and a confirmation of this reading is obtained from the solid curves of Figure 5.

It is interesting to observe that while b must always be less than π radians for impulse modulation, for equal shaping the results are continuous to 2π radians.

AN AUTOMATIC-FREQUENCY-CONTROLLED PHASE-SHIFT-KEYED DEMODULATOR*

By

W. Hannan and T. Olson

RCA Defense Electronic Products,
Camden, N. J.

Summary—*The principles of binary phase-shift-keyed (psk) demodulators are reviewed with regard to their use in airborne or satellite communication systems where large frequency errors due to doppler shift and oscillator drift are unavoidable and short lockup time is essential. A comparison of the possible demodulation techniques indicates that a phase-multilock demodulator in which all filtering is accomplished by a two-mode phase-locked-oscillator circuit, known as a quadricorrelator, offers the best performance. In this circuit, noise bandwidth is sacrificed during the lockup period so that wide pull-in range and short lockup time can be realized. After lockup the noise bandwidth is automatically narrowed. Design equations are given which indicate the trade-offs between pull-in range, noise bandwidth, and lockup time. Experimental results are given which show that: (1) probability of error versus signal-to-noise ratio within 1 decibel of the theoretically ideal PSK demodulator can be realized, and (2) the use of a quadricorrelator enables the phase-multilock demodulator to tolerate frequency errors as large as 20 kilocycles at an intermediate frequency of 200 kilocycles while maintaining lockup times of less than 1 millisecond.*

Introduction

ONE OF THE major problems in designing a phase-shift-keyed (PSK) demodulator which will tolerate the carrier frequency errors inherent in airborne and satellite data links is to maintain lockup times which are a small fraction of the message length and a narrow noise bandwidth. In a typical airborne data link operating in the VHF band, for example, frequency errors may be as large as ±10 kilocycles and message length may be as short as 15 milliseconds.

Conventional PSK demodulators can neither tolerate frequency errors as large as this nor provide short enough lockup times without a severe sacrifice of noise bandwidth. The automatic-frequency-controlled PSK demodulator described in this paper overcomes these deficiencies by utilizing a two-mode phase-locked-oscillator circuit (a quadricorrelator) as a filter. The phase-locked oscillator is more advantageous than passive filters because it enables the demodulator

* Manuscript received 1 August 1961.

to tolerate much larger frequency errors for a given noise bandwidth. However, noise bandwidth, lockup time, and pull-in range of a phase-locked oscillator are related such that the noise bandwidth is increased if either the lockup time is shortened or the pull-in range is widened. Once lockup is achieved and pull-in range and lockup time no longer restrict the circuit parameters it is possible to change the circuit parameters so that noise bandwidth is reduced. Thus, the two-mode phase-locked-oscillator circuit improves system performance.

Phase Multilock

In the phase-multilock circuit[1] shown in Figure 1 the phase reference is derived from the received signals by doubling the frequency

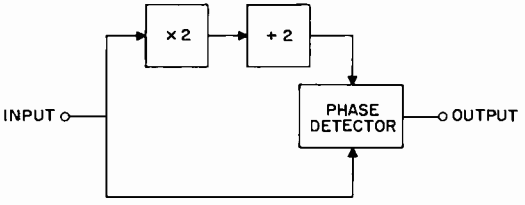


Fig. 1—Phase-multilock demodulator.

of the received signals. This eliminates the binary phase modulation. The multiplied signal is then divided by 2 to produce the phase reference. Ideally, a phase reference generated in this manner will be either in-phase or 180° out of phase with the received signals. (To eliminate the initial phase ambiguity, it is possible either to use differential PSK or to transmit a phase reference prior to each message.)

The effect of frequency errors must be considered. Because multiplication and division circuits employ analog circuits for smoothing noise and harmonics, frequency errors cause phase errors in the generated phase reference. If filtering is accomplished by a single tuned circuit, for example, phase error and frequency error are related by the Q of the filter. The universal resonance curve, Figure 2, illustrates this relationship. For example, if the phase error is not to exceed 25°, then the phase-versus-normalized-frequency curve indicates that the bandwidth, B, required to restrict the maximum phase error to 25° is

[1] C. Crafts, "Phase Multilock Communication," *Conf. Proc. 2nd National Convention on Military Electronics*, p. 262, 1958.

$$B \geqq 4\Delta f. \tag{1}$$

Even if an "ideal" filter (i.e., a filter having constant gain and zero phase shift over the passband and infinite attenuation outside the passband) were realizable, the bandwidth of the filter would still have to be at least as large as the expected frequency error.

With a passive filter, therefore, the maximum tolerable frequency error is directly proportional to filter bandwidth and the bandwidth

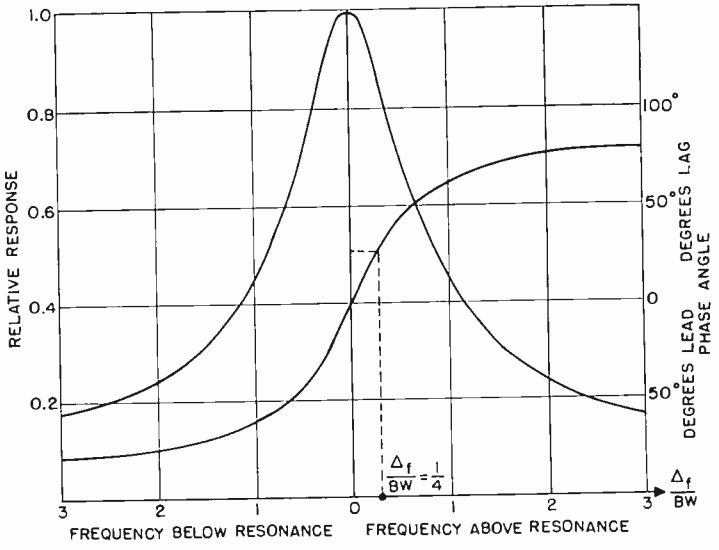


Fig. 2—Universal resonance curves.

of a single tuned filter must be on the order of 4 or more times the expected frequency error. Unfortunately, this leads to a conflicting situation. On one hand, the filter bandwidth should be as large as possible to make the phase-reference circuit insensitive to frequency errors, while on the other hand the bandwidth should be as narrow as possible to make the phase-reference circuit immune to noise. In other words, the price which must be paid to enable this PSK demodulator to tolerate larger frequency errors is a reduction in noise immunity.

Automatic-Frequency-Controlled Phase-Shift-Keyed Demodulator

The automatic-frequency-controlled PSK demodulator shown in

Figure 3 is the same as the phase-multilock demodulator previously described except that a phase-locked oscillator is used in place of passive tuned circuits to filter the $2f$ component from the frequency multiplier. This filtering technique enables the demodulator to tolerate large frequency errors without sacrificing noise immunity.

In this case, the price which must be paid for improved performance is increased lockup time. That is, for a given pull-in range, noise bandwidth cannot be reduced without sacrificing lockup time. In order to specify the tradeoffs between these parameters, the basic theory of phase-locked oscillators is reviewed briefly.

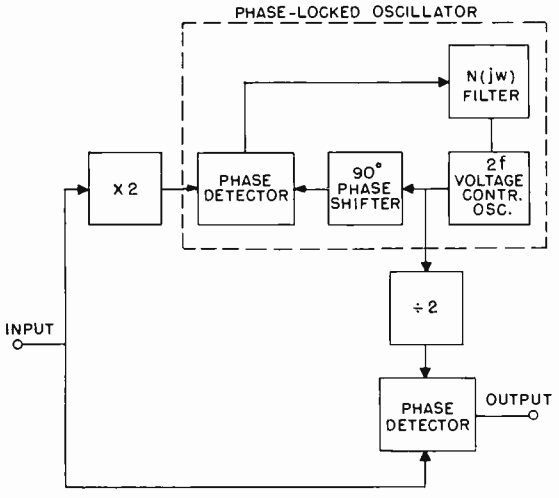


Fig. 3—Automatic-frequency-controlled phase-shift-keyed demodulator.

An oscillator can be phase locked to an input frequency through the use of a phase detector which generates a voltage proportional to the phase difference between the oscillator and the input frequency; this voltage is then used to force the phase of the oscillator to coincide with the phase of input frequency. McAleer[2] and Sonnenfeldt[3] have shown that an oscillator can be phase-locked even though the oscillator frequency, f, differs from the input frequency, f_0, by many cycles. When a frequency difference exists, the error voltage from the phase detector is a distorted sine wave at the difference frequency, Δf. This

[2] H. McAleer, "A New Look at the Phase-Locked Oscillator," *Proc. I.R.E.*, Vol. 47, p. 1136, June, 1959.

[3] R. Sonnenfeldt, "Automatic Phase and Frequency Control," *RCA Lecture Notes*, Lesson No. 44, January, 1957.

error-voltage frequency modulates the local oscillator. Among the sideband components of the FM spectrum are $f - \Delta f$ and $f + \Delta f$, one of which must equal the input frequency, f_0. The phase detector then generates a d-c voltage proportional to the phase difference between the two frequencies which, in turn, shifts the local-oscillator frequency toward f_0 until lockup occurs. The d-c voltage from the phase detector, however, indicates that a phase error exists. This "static" phase error is inherent whenever the local oscillator is locked to a reference frequency which is different from its unlocked frequency.

Gruen[4] has shown that the relationship between static phase error ϕ_e, initial frequency error Δf, and loop gain f_c is

$$\sin \phi_e = \frac{\Delta f}{f_c}. \tag{2}$$

Since the maximum value of $f_c \sin \phi_e$ is f_c, the maximum frequency error, Δf_{max}, that the automatic-phase-control (a-p-c) loop can tolerate is f_c. Further, since ϕ_e must be small ($< 25°$) to be of practical interest, the static phase error for given deviation Δf is

$$\phi_e \simeq \frac{\Delta f}{f_c}. \tag{3}$$

In the design of an a-p-c loop the maximum frequency error that the circuit must tolerate and the corresponding maximum static phase error are usually specified. Subject to these restrictions, the loop gain f_c must be at least large enough to satisfy the conditions imposed by Equation (3).

Richman[5,6] has shown how the ability of an a-p-c loop to follow phase error depends on $N(j\omega)$ in Figure 3. The anti-hunt (or phase-lag) filter shown in Figure 4 is often employed for N. A good compromise in this circuit is to design for critical damping. If the ratio of ac-to-dc loop gain, which from Figure 4 can be expressed as $m = x/(1 + x)$, satisfies the condition

---

[4] W. Gruen, "Theory of AFC Synchronization," *Proc. I.R.E.*, Vol. 41, p. 1043, August, 1953.

[5] D. Richman, "Color-Carrier Reference Phase Synchronization Accuracy in NTSC Color Television," *Proc. I.R.E.*, Vol. 42, p. 106, January, 1954.

[6] D. Richman, "The DC Quadricorrelator: A Two-Mode Synchronization System," *Proc. I.R.E.*, Vol. 42, p. 288, January, 1954.

$$m = \frac{4}{\sqrt{2\pi f_c T}}, \tag{4}$$

where $T = RC$, then the system will be critically damped.

If

$$f_c \gg \frac{1}{2\pi T(x+1)} \tag{5}$$

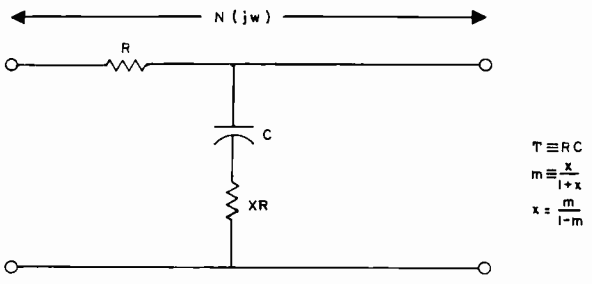


Fig. 4—Anti-hunt filter.

and the a-p-c loop is critically damped, the noise bandwidth f_{nn} is given by

$$f_{nn} = 1.25 \sqrt{\frac{2\pi f_c}{(1+x)T}}. \tag{6}$$

Also, if $m \ll 1$ the noise bandwidth is approximately

$$f_{nn} \simeq 2mf_c. \tag{7}$$

Richman[5] has shown that the lockup time is virtually instantaneous if

$$\Delta f < mf_c, \tag{8}$$

and that if the frequency difference falls within the range

$$mf_c < \Delta f < f_c\sqrt{2m}, \tag{9}$$

the lockup time, T_f, is given by

$$T_f = xT \frac{\left(\frac{\Delta f}{mf_c}\right)^2}{1 - \frac{1}{2m}\left(\frac{\Delta f}{f_c}\right)^2}. \tag{10}$$

The maximum pull-in range is the frequency deviation Δf_{max} at which the lockup time is infinite. It follows from Equation (10) that

$$\Delta f_{max} = f_c \sqrt{2m}. \tag{11}$$

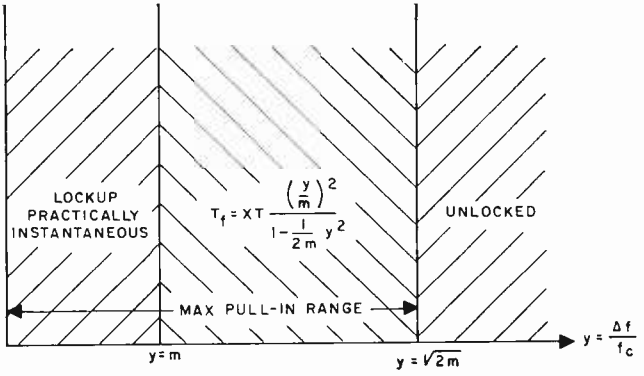


Fig. 5—Pull-in range of a phase-locked oscillator.

If the frequency difference is well within the pull-in range, i.e.,

$$|\Delta f| \ll f_c \sqrt{2m}, \tag{12}$$

then the lockup time can be expressed as

$$T_f \simeq xT \left(\frac{\Delta f}{mf_c}\right)^2. \tag{13}$$

These relationships are illustrated in Figure 5.

QUADRICORRELATOR

Because there are so many variables in an a-p-c loop, each design must be tailored to suit the needs of the system in which it is used. Compromises must be made in order to optimize the design. Usually, the four most important objectives are lockup time, pull-in range, noise bandwidth, and static-phase error. These are given by Equa-

tion (13), (11), (7), and (3), respectively. In a conventional, or single-mode, a-p-c loop, all four objectives are related to m and T, such that it is not possible to improve one parameter without degrading one or more of the others. For example, both pull-in range and lockup time can be improved by increasing loop gain; however, the price that must be paid is an increase in noise bandwidth.

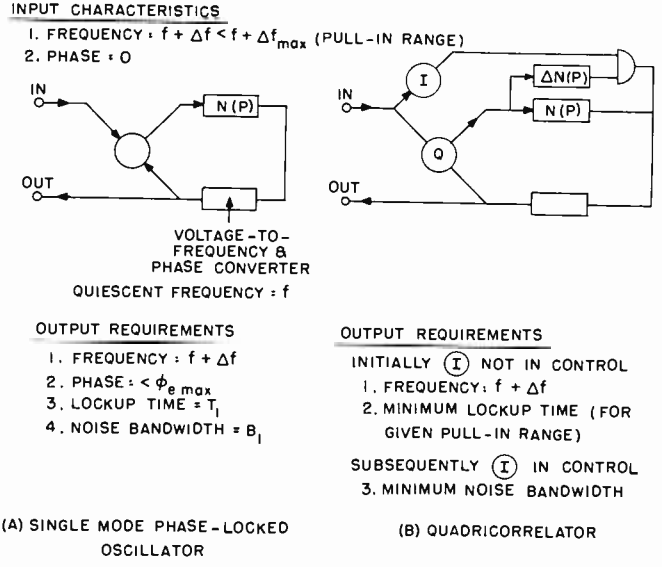


Fig. 6—Single-mode phase-locked oscillator circuit (A) compared to quadricorrelator circuit (B).

Richman[6] has shown that a two-mode a-p-c loop, known as a quadricorrelator, can be designed in which the first mode is used to give a large pull-in range and short lockup time, and once lockup is achieved, the second mode, in which one or more loop parameters have been changed, is used to narrow the noise bandwidth. A comparison of single mode and quadricorrelator circuits is illustrated in Figure 6.

From the equations for lockup time, pull-in range, and noise bandwidth, it is evident that either the loop gain, f_c, or the ratio of ac-to-dc loop gain, m, (or both) can be varied to achieve the desired result. Figure 7 shows how these techniques can be implemented. Actually, best performance is realized when both f_c and m are varied such that critical damping is maintained for both operating modes. In this paper, however, only the variable-ratio-type quadricorrelator is considered, because of its simplicity as compared to the variable-gain type.

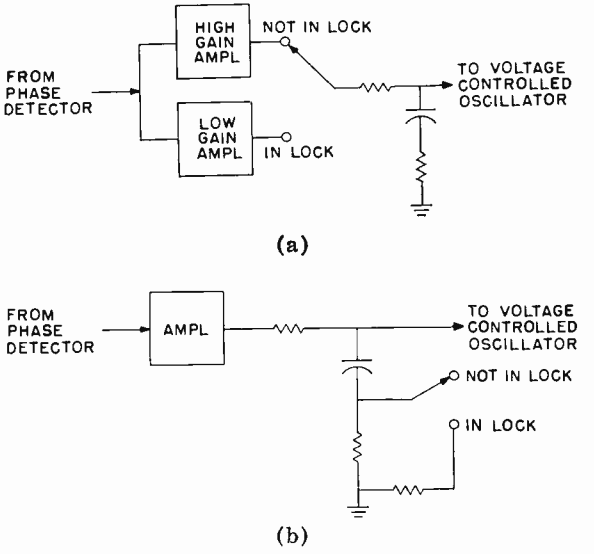


Fig. 7—Quadricorrelator circuits: (a) variable-gain quadricorrelator, and (b) variable-ratio quadricorrelator.

A quadricorrelator employs two phase detectors commonly known as the I (or in-phase) detector and the Q (or quadrature) detector. The output signal from the Q detector controls the local oscillator, and the output signal from the I detector controls the mode of operation. Figure 8 shows a block diagram of a quadricorrelator circuit

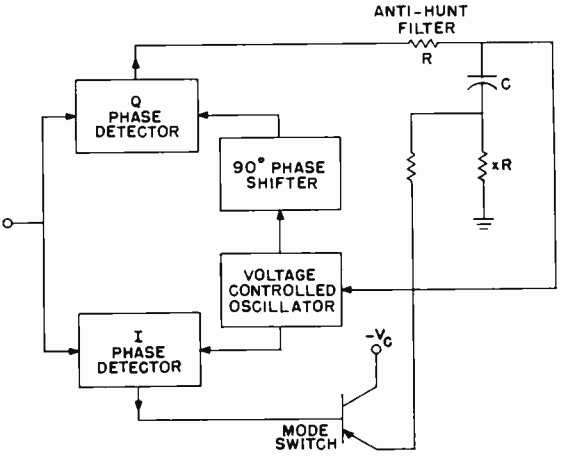


Fig. 8—Variable ratio quadricorrelator circuit.

employing variable-ratio modes of operation. When no signal is applied to the input, the average output signal from both phase detectors is zero; when the system is phase-locked the output voltage from the Q phase detector is virtually zero, and the output voltage from the I phase detector is $-E$ volts. The circuit is designed such that $-E$ volts is more than sufficient to drive the mode switch transistor into saturation, thereby shunting resistor $x\bar{R}$ with another resistor. The result is an effective reduction of m and, therefore, a decrease in the noise bandwidth.

Design of an A-P-C Loop for Time-Division Data-Link Systems

Consider now the design of an a-p-c loop for use in a time-division data-link system.

The particular system under consideration is an airborne time-division data link. Each message consists of a sequence of binary bits in which the words, labels, and parity checks, have a specified format. Usually, messages are not transmitted continuously or in any particular time sequence. For this reason, the data-link receiver must process each message independently.

The lockup time must be a small fraction of the total message duration. Since message length is usually on the order of milliseconds, the short lockup time which must be achieved represents a rather severe design objective for the PSK demodulator.

Another factor which is troublesome, particularly in airborne data links operating in the VHF band, or higher, is frequency error due to oscillator instability and doppler shift. The a-p-c loop must be able to tolerate deviations in frequency of at least 10 kilocycles.

Finally, the noise bandwidth must be a fraction of the message bandwidth.

Example of a Typical Design

Consider a data link subject to the following operating conditions:

1. bit period = 40 microseconds, bit rate = 25 kilobits/second
2. maximum frequency error = 10 kilocycles
3. message length = 15 milliseconds = 375 bits.

First, a suitable noise bandwidth must be established. The desired bandwidth will depend upon the probability of bit error to be satisfied and the compromise between lockup time and pull-in range that can be made. To begin with, let the noise bandwidth equal one-fourth the message bandwidth. Since most of the energy contained in the message is restricted to $\pm 1/\tau$ either side of the carrier frequency (i.e., the

major lobe of the sin x/x spectrum produced by a single message bit), where τ is the message bit duration, the noise bandwidth is

$$f_{nn} = \frac{1}{4} \cdot \frac{2}{\tau} = 12 \text{ kilocycles.}$$

With the noise bandwidth established, the circuit must then be designed to provide the desired lockup time and static phase error. If both lockup time and noise bandwidth cannot be satisfied with the same circuit values (which is the usual case), the quadricorrelator technique must be employed.

As a first approximation, let the maximum pull-in range be 10 times greater than the maximum expected frequency error;

$$\begin{aligned} 10\Delta f &= \Delta f_{\max} \\ &= f_c\sqrt{2m}. \end{aligned} \tag{14}$$

Since a short lockup time and minimum final phase error are of primary interest, the loop gain must be such that, from Equations (3) and (13),

$$f_c \geqq \frac{\Delta f}{\phi_{e\max}} \tag{15}$$

$$f_c \geqq \sqrt{xT}\,\frac{\Delta f}{m\sqrt{T_f}}. \tag{16}$$

If the steady-state phase error is not to exceed 5°, the minimum loop gain specified by Equation (15) is

$$f_c \geqq \frac{10{,}000}{5°\left(\frac{2\pi}{360}\right)} = 110 \text{ kilocycles.}$$

For an a-p-c loop operating at 400 kilocycles, a loop gain of 150 kilocycles can be realized without too much difficulty. With a loop gain of 150 kilocycles, Equation (14) gives

$$m = 0.22,$$

and it follows that

$$x = 0.3.$$

The lockup time must now be specified. It will be assumed that 5 per cent of the message period, or .75 millisecond (18 bits), can be used for lockup time. Since Δf is one-tenth of $f_c\sqrt{2m}$, the lockup time can be approximated as

$$T_f \simeq xT\left(\frac{\Delta f}{mf_c}\right)^2. \tag{13}$$

In turn,

$$T = RC = \frac{T_f}{x}\left(\frac{mf_c}{\Delta f}\right)^2.$$

Substituting the values of m, x, f_c, Δf and T_f,

$$RC = 0.027 \text{ second.}$$

For $R = 4700$ ohms, then, $C = 5.7$ microfarads. All the circuit parameters have thus been specified, and for this particular design the noise bandwidth is

$$f_{nn} \simeq 2mf_c = 66 \text{ kilocycles.}$$

The noise bandwidth so obtained is, obviously, much larger than the desired value of 12 kilocycles; indeed, it exceeds the message bandwidth. To obtain the required noise bandwidth while satisfying the pull-in range or lockup time requirements, therefore, a quadricorrelator will be employed. With the variable-ratio quadricorrelator, m can be reduced after lockup to

$$m = \frac{f_m}{2f_c} = 0.040,$$

with a corresponding value for x of 0.042. Repeating the foregoing procedure to obtain the required RC-time constant after lockup, the two-mode filter shown in Figure 9 results. The magnitudes of the errors introduced by using approximate formulas and circuit nonlinearities are indicated in the following section where experimental results are given for a system using these design values.

Experimental Results

A breadboard model of the system shown in Figure 10 was constructed to measure the performance of a phase-multilock demodulator

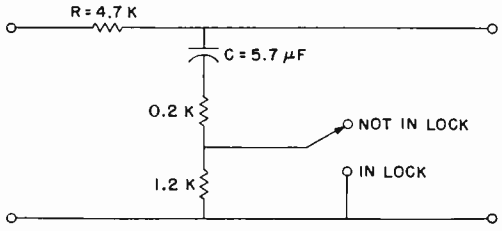


Fig. 9—Two-mode filter.

employing a quadricorrelator to filter the phase-reference signal. For convenience, 200 kilocycles was chosen for the carrier frequency. Figure 11 shows the major components of the demodulator and typical waveforms associated with this circuit. Multiplication by 2 of the input signal is accomplished by the full-wave rectifier circuit shown in Figure 12. The second-harmonic component of the rectified signal is filtered by the phase-locked-oscillator circuit containing the oscillator, the phase detector, and the d-c amplifier shown in Figures 13, 14, and 15, respectively.

As previously mentioned, the quadricorrelator employs two quadrature phase detectors. The signal from the quadrature, or Q, detector controls the local oscillator and the signal from the in-phase, or I, detector controls the mode switch. The mode switch effectively changes the filter characteristics in accordance with the locked and unlocked states of the demodulator. Two quadrature signals are generated by shifting the phase of the signal from the local oscillator by 90° before it is applied to the phase detector. The broadband phase shifter shown in Figure 16 insures that an accurate 90° phase shift will be

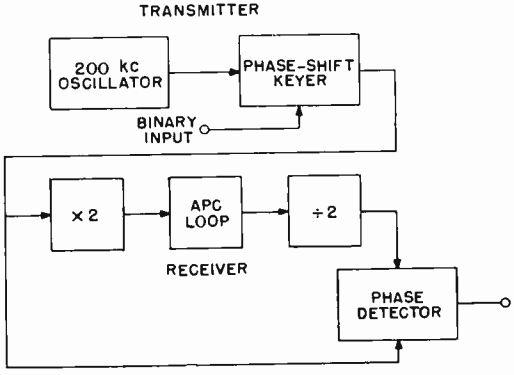


Fig. 10—Experimental system.

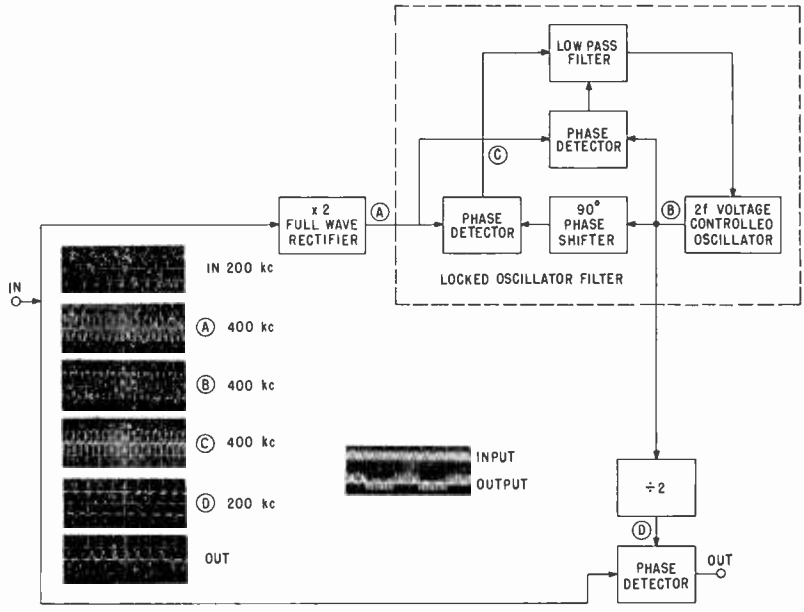


Fig. 11—Waveforms in experimental demodulator.

realized over the entire frequency range of the local oscillator. This broad band 90° phase shift could also be achieved by a phase-locked-oscillator circuit, but the simple passive circuit shown proved to be adequate. (If the phase shift were a function of frequency, the ability of the demodulator to tolerate frequency errors would be restricted.)

The mode switch and anti-hunt filter associated with the quadri-correlator are shown in Figure 17. When phase-lock occurs the output

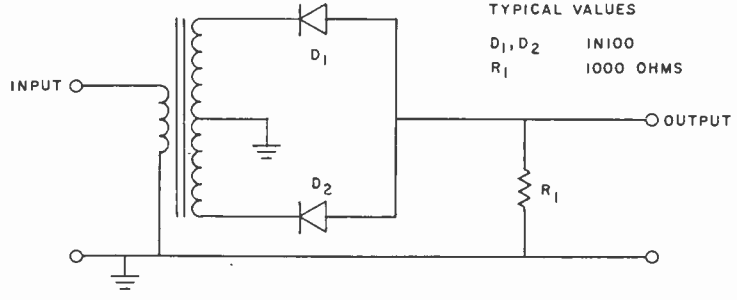


Fig. 12—Full wave rectifier and doubler.

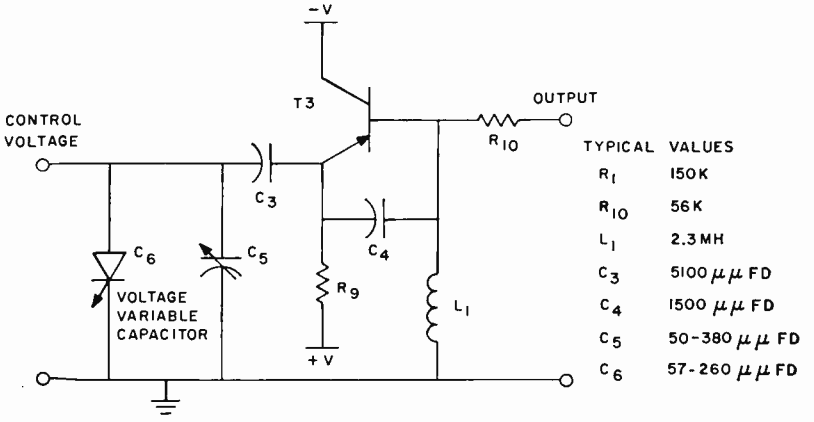


Fig. 13—Voltage-controlled oscillator.

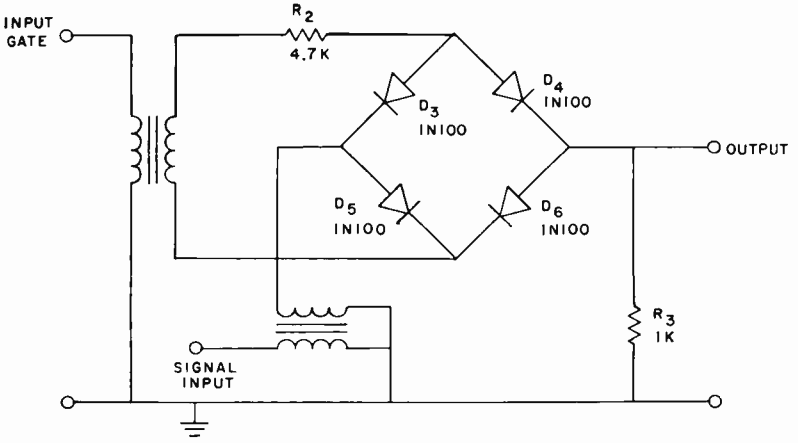


Fig. 14—Phase detector.

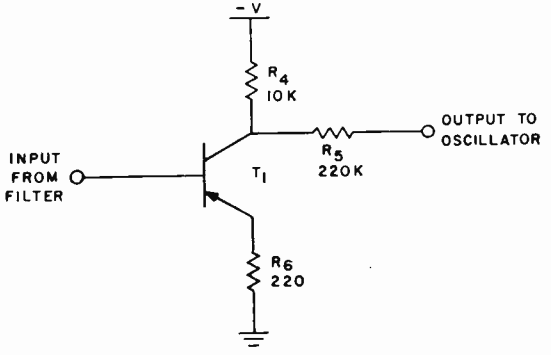


Fig. 15—D-C amplifier.

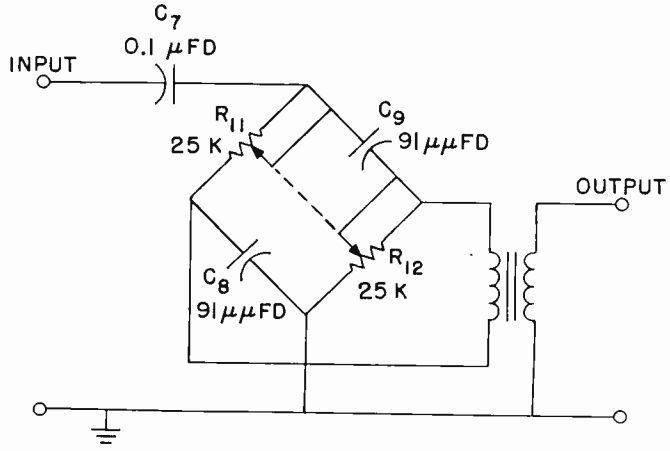


Fig. 16—Broadband 90° phase shifter.

signal from the I phase detector is a negative voltage sufficient to drive the transistor T_2 into saturation, thereby reducing the value of (xR_7), which effectively reduces the noise bandwidth.

To insure that no frequency-dependent phase shift is introduced when the oscillator signal is divided by 2, the sinusoidal output signal from the local oscillator, which is virtually noise free, is sent through a symmetrical clipping circuit where it is converted into a square wave. The transitions of the square wave correspond to the zero axis crossings of the input sine wave. This square wave is fed to the ÷2 multivibrator shown in Figure 18 which is operated such that it "flips" every time the input signal undergoes a positive transition. Thus, the output signal from the local oscillator is divided by 2 without introducing phase shift. The signal from the ÷2 multivibrator is the phase reference used to demodulate the received signal.

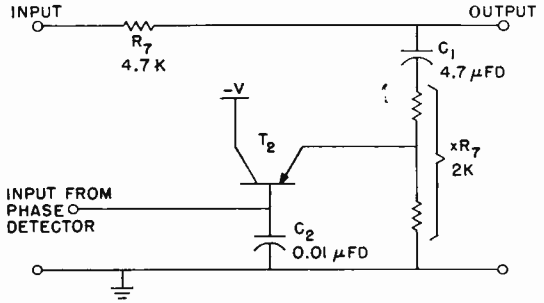


Fig. 17—Anti-hunt filter and mode switch.

Performance Measurements

Gain of A-P-C Loop

The gain of the a-p-c loop is equal to the product of the gains of the phase detector, the d-c amplifier, and the voltage-controlled oscillator. Figure 19 is a plot of the measured characteristics of the phase detector. The gain of the phase detector is equal to the slope of this characteristic curve:

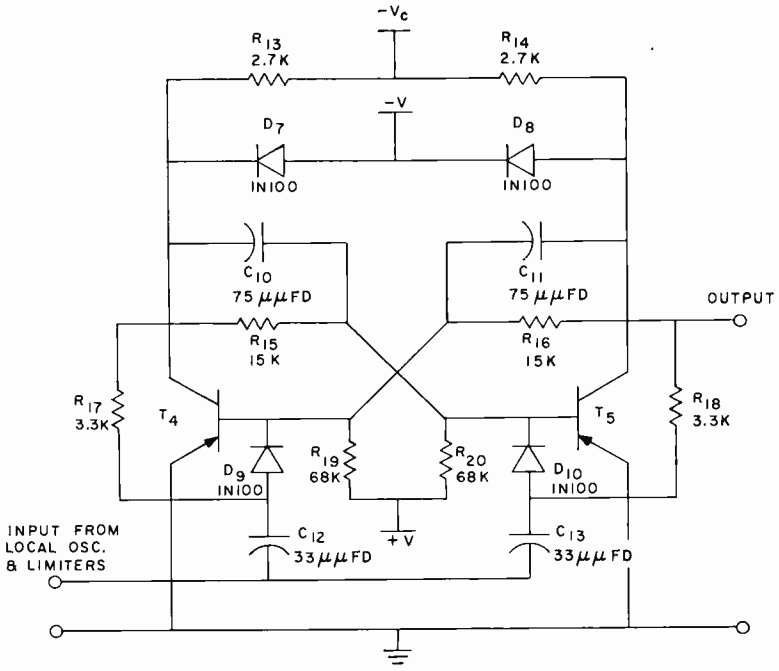


Fig. 18—Divide by 2 multivibrator.

$$\mu = \frac{dE_d}{d\phi}\bigg|_{\phi=90°} = 0.3 \text{ volt/radian.}$$

Similarly, the gain of the voltage-controlled oscillator, β, is equal to the inverse slope of its characteristic curve, Figure 20, which is

$$\beta = \frac{df}{dE_0}\bigg|_{f=400 \text{ kc}} = 10.8 \text{ kilocycle/volt.}$$

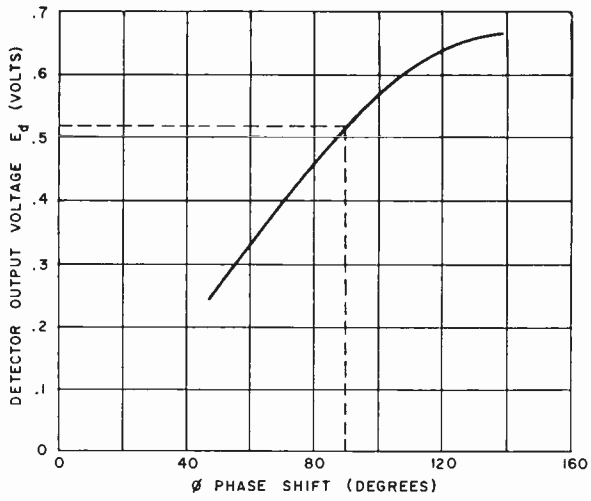


Fig. 19—Phase-detector characteristics.

The gain of the d-c amplifier shown in Figure 15 is 46. Thus, the gain of the a-p-c loop is

$$f_c = 46\,\mu\beta = 149 \text{ kilocycles.}$$

Hold-in and Pull-in Range

Theoretically, the maximum hold-in range is independent of the filter and is equal to the gain of the a-p-c loop. It was not possible

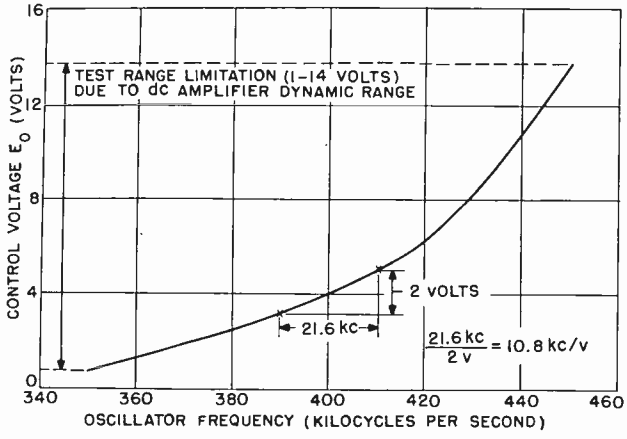


Fig. 20—Voltage-controlled oscillator characteristics.

to realize this range experimentally, however, because the voltage from the d-c amplifier was limited by its power supply to a 13-volt range which, in turn, limited the voltage-controlled oscillator to a 344-to-456 kilocycle frequency range. The measured hold-in range was

$$\frac{f_H}{2} = 56 \text{ kilocycles},$$

where f_H is the total frequency range over which phase lock is maintained.

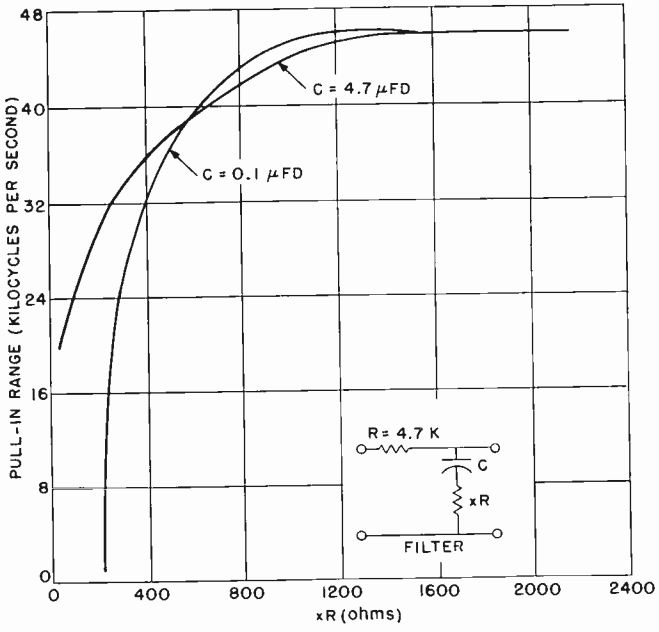


Fig. 21—Pull-in range of experimental system.

The pull-in range is a function of the a-p-c loop filter. Theoretically, it is given by

$$\Delta f_{max} \simeq f_c\sqrt{2m}. \tag{11}$$

Figure 21 gives the measured pull-in range for various combinations of filter constants. Note that the hold-in range is always greater than the pull-in range. Therefore, once an a-p-c loop is locked to an input signal it will be able to follow slow frequency variations that are wider than the pull-in range.

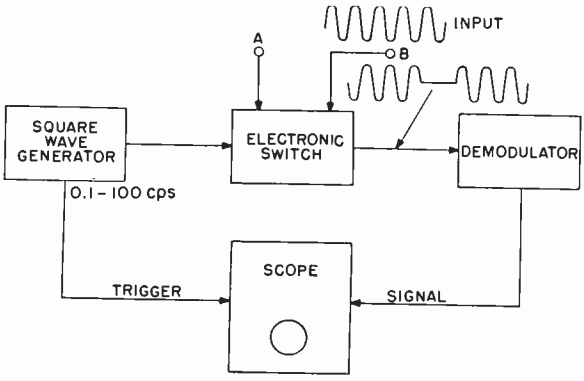


Fig. 22—Pull-in-time test circuit.

Lockup Time

Theoretically, lockup time is given by

$$T_f \simeq xT\left(\frac{\Delta f}{mf_c}\right)^2 \tag{13}$$

The lockup time of the experimental system was measured using the test circuit shown in Figure 22. The following procedure was used to make this measurement. First, the unlocked frequency of the phase-locked oscillator was measured and the frequency of the input signal was adjusted to give a known frequency error. The input signal was switched on and off by an electronic switch driven by a square wave having a repetition rate low enough to insure that the voltage-controlled oscillator reached its steady-state unlocked frequency before each burst of the input carrier signal. Figure 23 is a typical oscillo-

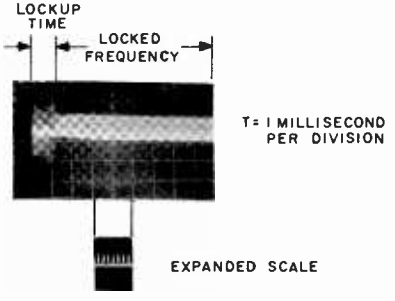


Fig. 23—A-P-C loop pull-in performance.

scope picture of the lockup behavior of the a-p-c loop. This picture shows the signal that appeared at the output of the I phase detector. When there is *no* input signal the output of the phase detector is zero. When an input signal is applied, the output of the phase detector fluctuates in accordance with the relative phase of the input signal and the voltage-controlled oscillator signal. Figure 23 shows that the experimental circuit required approximately 1 millisecond to establish the proper phase reference signal.

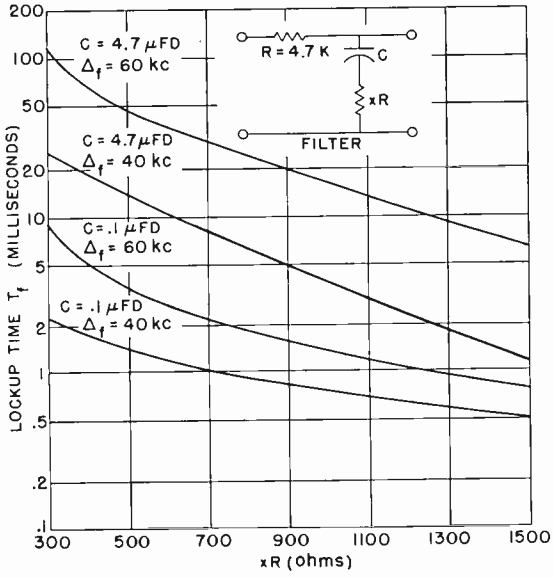


Fig. 24—Lockup time versus filter time constants.

Figure 24 shows the experimental curves which relate lockup time and filter time constants. It can be seen that increasing either the filter time constant or the initial frequency error causes an increase in lockup time. A comparison of the experimental and theoretical curves relating lockup time and initial frequency error is given in Figure 25. Considering the approximations involved in deriving useful design equations and the nonlinearities in the actual circuits, the difference between the experimental and theoretical curves are reasonable.

Figure 26 shows improvement in system performance due to use of the quadricorrelator. The experimental curves show that the a-p-c loop with the quadricorrelator provides a much shorter lockup time

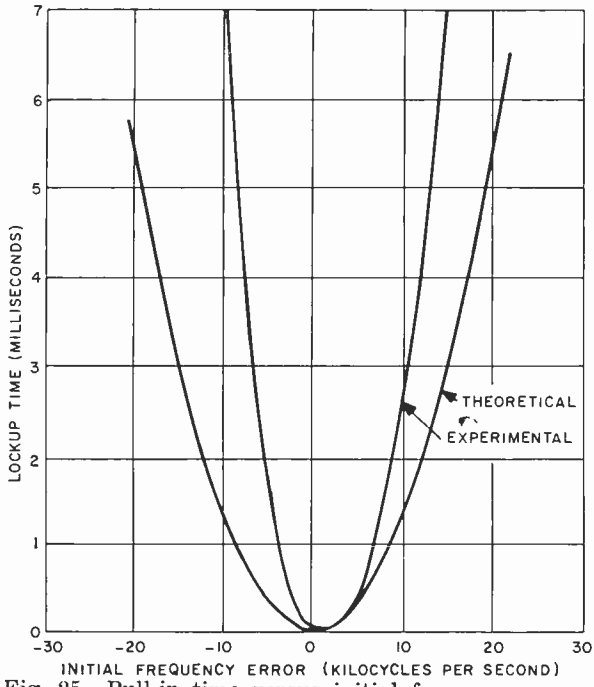


Fig. 25—Pull-in time versus initial frequency error.

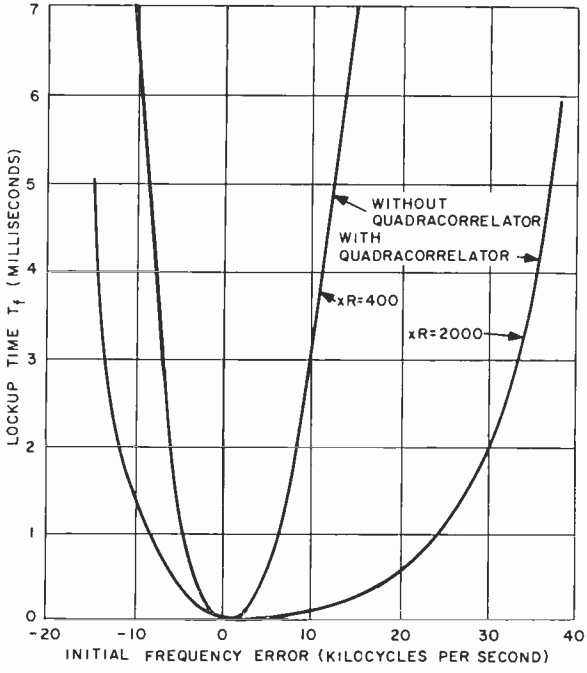


Fig. 26—Mode pull-in times versus initial frequency error.

than the a-p-c loop without the quadricorrelator. In operation, the mode switch effectively reduced the filter resistor xR from the unlocked value of 2000 ohms to the locked value of 400 ohms, thereby providing short lockup time and narrow noise bandwidth.

Probability of Error

Probability of error was measured using the test circuit shown in Figure 27. The generally accepted definition of signal-to-noise ratio,[7] E/N, which is signal energy per bit to noise power per cycle of bandwidth, was used for this measurement.

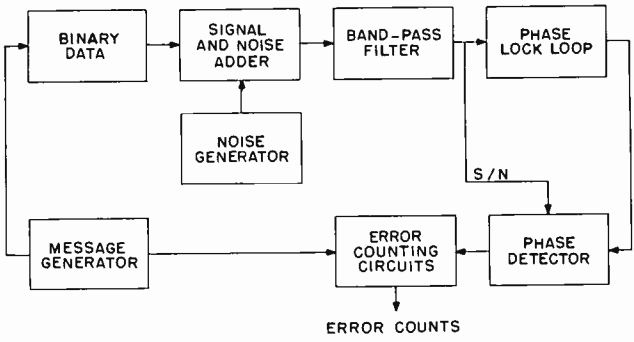


Fig. 27—Circuits for measuring probability of error.

To measure noise power per cycle of bandwidth accurately, a filter having a known bandwidth (which was wider than the spectrum of the input signal) was used to filter the input signal, which was contaminated with noise. N is equal to the noise power at the output of the filter divided by the filter bandwidth. Signal energy per bit, E, is equal to the signal power at the output of the filter times the bit period. In terms of the r-m-s signal, S, and r-m-s noise, N_{rms}, the signal-to-noise ratio is

$$\frac{E}{N} = \left(\frac{S}{N_{rms}}\right)^2 BT \qquad \begin{cases} B = 104 \text{ kilocycles} \\ T = 42.1 \text{ microseconds.} \end{cases}$$

In the experimental demodulator, the constants of the filter were $C = 4.7$ microfarads, $R = 4700$ ohms, and $xR = 400$ ohms, correspond-

[7] J. Lawton, "Comparison of Binary Data Transmission Systems," *Conf. Proc. 2nd National Convention on Military Electronics*, p. 54, 1958.

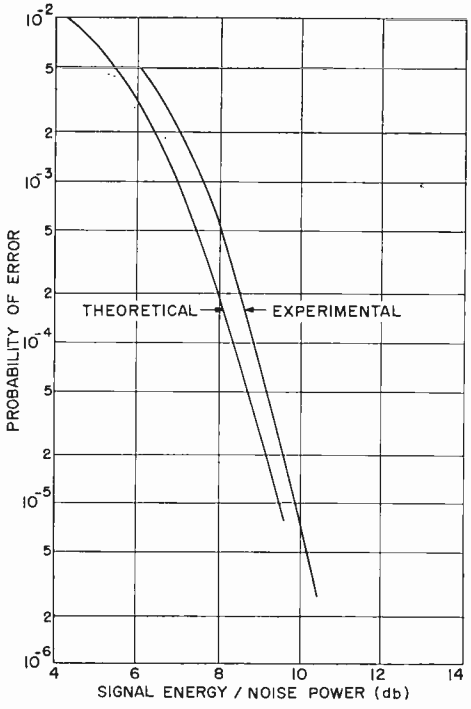


Fig. 28—Probability of error versus ratio of signal energy to noise power.

ing to the curve shown in Figure 25. To evaluate the over-all performance of this demodulator, its performance was compared to the performance of a theoretically ideal PSK demodulator. Figure 28 shows that the performance of the experimental demodulator is within 1 decibel of the theoretical limit.

ANALYSIS OF DOUBLE-STREAM INTERACTIONS IN THE PRESENCE OF A FINITE AXIAL MAGNETIC FIELD*

BY

BAYRAM VURAL

RCA Laboratories,
Princeton, N. J.

Summary—*In a magnetically focused electron beam there are, in addition to the well-known space-charge waves, other modes of propagation which are associated with the cyclotron resonance of the electrons. The present study indicates that coupling, not only between space-charge waves but also between a space-charge wave and a cyclotron wave, and between cyclotron waves, may lead to exponential growth of the signal along the stream. In contrast to the plasma frequency of an unneutralized beam, which is limited by space-charge depression, the cyclotron frequency can easily be adjusted by the external magnetic field to cover the entire microwave range.*

INTRODUCTION

RECENTLY a number of workers[1-6] have investigated the modes of propogation along a plasma column situated in an axial magnetic field. In addition to the well-known waves associated with the plasma resonance, they have found waves which are associated with the cyclotron resonance of the electrons. Some of the authors[1,4]

* Manuscript received 15 September 1961.

[1] R. W. Gould, A. W. Trivelpiece, "A New Mode of Wave Propagation on Electron Beams," *Electronic Waveguides*, Polytechnic Press, p. 215, April 1958, N. Y.

[2] W. O. Schumann, "Über Wellenausbreitung im Plasma Zwischen Zwei unendlich gut leitenden Ebenen in Richtung eines aufgeprägten äusseren Magnetfeldes," *Zeitschrift für Angew. Phys.*, Band VIII, Heft 10, p. 482, Oct. 1956.

[3] B. Agdur, "Notes on the Propagation of Guided Microwaves Through an Electron Gas in the Presence of a Static Magnetic Field," *Electronic Waveguides*, p. 177, Polytechnic Press, April 1958, N. Y.

[4] L. D. Smullin, P. Chorney, "Propagation in Ion-Loaded Waveguides," *Electronic Waveguides*, p. 229, Polytechnic Press, April 1958, N. Y.

[5] H. Wilhelmsson, "On the Properties of the Electron Beam in the Presence of an Axial Magnetic Field of Arbitrary Strength," *Trans. of Chalmers University of Technology*, Gothenburg, Sweden, No. 205, 1958.

[6] Y. B. Fainberg, M. F. Gorbatenko, "Electromagnetic Waves in a Plasma Situated in a Magnetic Field," *Soviet Phys. Tech. Phys.*, Vol. 4, No. 5, p. 487, November 1959.

have pointed out that growing waves can be obtained by interaction of these cyclotron waves with a second beam.

Unfortunately, the use of plasmas for amplification of microwaves is still impractical, because present-day plasmas are never fully ionized and scattering of electrons by neutral molecules causes a large noise modulation on the signal[7] and reduces available gain. However, it has been shown[8-10] that the cyclotron waves also exist in unneutralized electron beams in much the same way as in neutral plasmas.

In the following discussion, a neutral plasma will be assumed, because the plasma analysis is somewhat simpler and the propagation constants do not differ greatly from those of an unneutralized beam. First, the dispersion equation of a system consisting of two intermingled cylindrical beams immersed in an axial magnetic field is derived. Second, the important special case of a single electron beam is discussed. This discussion is concerned with the ω–β diagram, reduction factors, and the excitation problems. Third, the available interaction possibilities are discussed qualitatively using the concepts of coupling of modes and the kinetic power theorem. Fourth, some quantitative results, obtained by solving the dispersion equation by means of a digital computer, are given. The detailed derivations of the dispersion equation, excitation of the beam, and gain of a TWT using the slow cyclotron wave are given in the Appendixes.

Derivation of the Dispersion Equation

In deriving the dispersion equation, the following assumptions are made:

(1) The phase velocity of the waves is much smaller than the velocity of light ($v_{ph} \ll c$).

(2) Only first-order perturbations are of interest (small signals).

(3) The thermal velocity spread is negligible.

(4) The system is a finite symmetrical cylinder containing two intermingled electron beams immersed in a finite axial magnetic field.

---

[7] G. D. Boyd, L. M. Field, R. W. Gould, "Interactions Between an Electron Stream and an Arc Discharge Plasma," *Electronic Waveguides*, p. 367, Polytechnic Press, April 1958, N. Y.

[8] R. H. C. Newton, "On Space Charge Waves," *Journ. Electronics and Control*, Vol. 5, p. 510, December 1958.

[9] F. Paschke, "The Propagation of Perturbations Along Magnetically Focused Electron Beams," *RCA Review*, Vol. XX, No. 2, p. 254, June 1959.

[10] D. H. Trevena, "On Space Charge Waves," *Journ. Electronics and Control*, Vol. 6, p. 50, January 1959.

From Maxwell's equations, the linearized equations of motion, and the equation of continuity, and with the perturbation assumed to be of the form $\exp j(\omega t - n\theta - \beta z)$, the following wave equation for the axial component of the electric field is obtained (the detailed derivation is given in Appendix A):

$$\frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r}\frac{\partial E_z}{\partial r} + \left(-\frac{n^2}{r^2} + \gamma^2\right) E_z = 0, \tag{1}$$

where E_z = the axial component of the electric field,

r = radial coordinate of the cylindrical coordinate system (r, θ, z),

n = an integer,

γ = the radial propagation constant, which is given by

$$\gamma^2 = -\beta^2 \frac{1 - \dfrac{\omega_{p1}^2}{(\omega - \beta v_1)^2} - \dfrac{\omega_{p2}^2}{(\omega - \beta v_2)^2}}{1 - \dfrac{\omega_{p1}^2}{(\omega - \beta v_1)^2 - \omega_c^2} - \dfrac{\omega_{p2}^2}{(\omega - \beta v_2)^2 - \omega_c^2}} \tag{2}$$

where β = axial propagation constant,

ω_{p1} = plasma frequency of beam #1,

v_1 = drift velocity of beam #1,

ω_{p2} = plasma frequency of beam #2,

v_2 = drift velocity of beam #2, and

ω_c = cyclotron frequency of the electrons.

The general solution of Equation (1) can be expressed as follows:

$$E_z = [AJ_n(\gamma r) + BN_n(\gamma r)]e^{-jn\theta}, \tag{3}$$

where J_n is a Bessel function of the first kind of nth order, and N_n is Neumann's Bessel function of the second kind of nth order. The constants A and B in Equation (3) must be determined from the boundary and initial (or excitation) conditions. Three possible configurations for the formulation of the boundary conditions are indicated in Figure 1. The configurations indicated in Figures 1(a) and

1(b) are the main concern of this study, and the investigation is restricted to circularly symmetric modes $\partial/\partial\theta = 0$.

Boundary Conditions

(1) Beams partially filling the drift tube.

Denoting the perturbations inside of the beams with the subscript I and those outside of the beams with the subscript II,

$$E_{zI}(b) = E_{zII}(b), \tag{4}$$

$$E_{rII}(b) = E_{rI} + \frac{\sigma}{\epsilon_0}, \tag{5}$$

or $[H_{\theta II}(b) = H_{\theta I} + i_0\bar{r}(b),]$

$$E_{zII}(a) = 0, \tag{6}$$

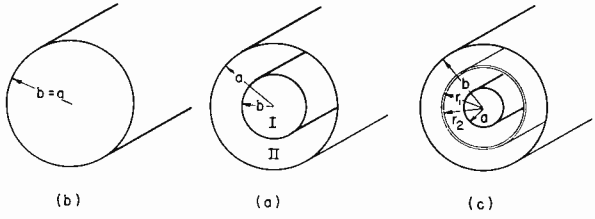


Fig. 1—Possible configuration for the formulation of the boundary value problem: (a) The case of beam partially filling the drift tube ($b/a < 1$); (b) The case of beam filling the drift tube ($b/a = 1$); (c) The case of an annular beam drifting between the inner and outer conductor of a coaxial line.

where $\sigma = \rho r(b)$ is a-c surface-charge density, and $ir(b)$ is an a-c surface current per unit length.

The conditions expressed in Equations (4) to (6), and Equation (2), and the solution of the wave equation for the region II lead to the following equation relating γ and β:

$$\left[1 - \frac{{\omega_{p1}}^2}{(\omega - \beta v_1)^2 - {\omega_c}^2} - \frac{{\omega_{p2}}^2}{(\omega - \beta v_2)^2 - {\omega_c}^2}\right]\left[(\gamma b)\frac{J_1(\gamma b)}{J_0(\gamma b)}\right] = -(\beta b)\frac{I_1(\beta b) + \dfrac{I_0(\beta a)}{K_0(\beta a)}K_1(\beta b)}{I_0(\beta b) - \dfrac{I_0(\beta a)}{K_0(\beta a)}K_0(\beta b)} \tag{7}$$

where I_0, I_1, K_0, and K_1 are modified Bessel functions.

(2) Beams filling the drift tube ($a = b$).

From Equation (7), introducing the condition $a = b$, one obtains

$$J_0(\gamma b) = 0, \tag{8}$$

which yields

$$\gamma b = 2.405. \tag{9}$$

This last condition, which has been obtained as a special case of Equation (7), can also be derived directly from the boundary condition

$$E_z(b) = 0 \tag{10}$$

since

$$E_z = AJ_0(\gamma r). \tag{11}$$

($B = 0$ for convergence of the field at $r \to 0$). At $r = b$, $J_0(\gamma b) = 0$ and $E_z = 0$, which yields Condition (9).

Equation (2) must be solved in connection with Equation (7) or (9) to obtain the dispersion of the system. Further detailed considerations will be restricted to the case of beams filling the drift tube, i.e., to the solution of Equation (2) in conjunction with Equation (9).

The double-stream dispersion equation investigated by Nergaard,[11] Pierce,[12] and Haeff[13] is a special case of the Equations (2) and (9) for $\omega_c = \infty$ and $b = \infty$.

Dispersion of A Single Beam

The dispersion of a single beam filling the drift tube is obtained from Equations (2) and (9) by setting $\omega_{p2} = 0$, and reducing Equation (2) to

$$\gamma^2 = -\beta^2 \frac{1 - \dfrac{{\omega_{p1}}^2}{(\omega - \beta v_1)^2}}{1 - \dfrac{{\omega_{p1}}^2}{(\omega - \beta v_1)^2 - {\omega_c}^2}}. \tag{12}$$

---

[11] L. S. Nergaard, "Analysis of a Simple Model of a Two-Beam Growing-Wave Tube," *RCA Review*, Vol. IX, No. 4, p. 585, December 1948.

[12] J. R. Pierce, "Double-Stream Amplifiers," *Proc. I.R.E.*, Vol. 37, p. 980, Sept. 1949.

[13] A. V. Haeff, "The Electron-Wave Tube—A Novel Method of Generation and Amplification of Microwave Energy," *Proc. I.R.E.*, Vol. 37, p. 4, January 1949.

For $\omega_c = \infty$, Equation (12) reduces to the dispersion equation for the Hahn–Ramo[14,15] space-charge waves.

It is useful to normalize Equation (12) in the following form:

$$\frac{(\gamma b)^2}{(\beta_e b)^2} = -\frac{(\beta)^2}{(\beta_e)^2} \frac{1 - \dfrac{\beta_p^2}{(\beta_e - \beta)^2}}{1 - \dfrac{\beta_p^2}{(\beta_e - \beta)^2 - \beta_c^2}}, \tag{13}$$

where b represents the beam radius. and

$$\beta_e = \frac{\omega}{v_1}, \qquad \beta_p = \frac{\omega_{p1}}{v_1}, \qquad \beta_c = \frac{\omega_c}{v_1}.$$

It is now easy to consider two interesting limiting cases:

$b \to \infty$ (thick-beam case), and

$b \to 0$ (thin-beam case).

Since b is finite and is given by Equation (9), for $b \to \infty$, $\gamma b/\beta_e b \to 0$. Equation (13) yields the following solutions:

$$\beta_{1,2} = 0, \tag{14a}$$

$$\beta_{3,4} = \beta_e \mp \beta_p \text{ (space-charge waves)} \tag{14b}$$

$$\beta_{5,6} = \beta_e \mp \beta_c \text{ (cyclotron waves)}. \tag{14c}$$

For $b \to 0$, $\gamma b/\beta_e b \to \infty$. In this case Equation (13) yields the following solution:

$$\beta_{1,2} = \mp \infty, \tag{15a}$$

$$\beta_{3,4} = \beta_e \text{ (synchronous waves)} \tag{15b}$$

$$\beta_{5,6} = \beta_e \mp \sqrt{\beta_c^2 + \beta_p^2} \text{ (cyclotron waves)}. \tag{15c}$$

---

[14] W. C. Hahn, "Small Signal Theory of Velocity-Modulated Electron Beams," *Gen. Electric Rev.*, Vol. 42, p. 258, June 1939.

[15] S. Ramo, "Space Charge and Field Waves in an Electron Beam," *Phys. Rev.*, Vol. 56, p. 276, August 1, 1939.

In Figure 2 some ω–β diagrams for the general case (b = finite) are indicated.

It is easier first to construct the ω–β diagram for the stationary beam case ($v_1 = 0$). The ω–β diagram for the case $v_1 \neq 0$ can be obtained from the above case by application of the Galilean transformation;

$$\omega' = \omega + \beta v_1, \tag{16a}$$

$$\beta' = \beta. \tag{16b}$$

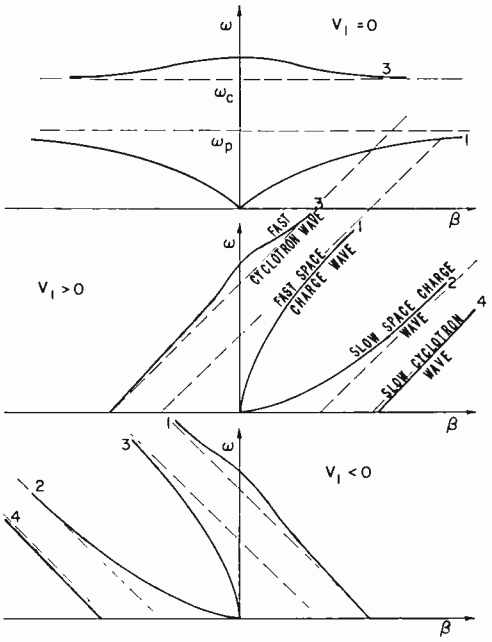


Fig. 2—ω–β diagrams of an electron beam (corresponding to case (b) of Figure 1) immersed into a finite axial magnetic field.

Reduction Factors for a Cylindrical Beam in a Drift Tube Immersed in an Axial Magnetic Field

Defining

$$\frac{(\beta_e - \beta)^2}{\beta_p^2} = p^2$$

and introducing this into Equation (12), one obtains

$$p^2 = \frac{1}{2}\left(\frac{\omega_c^2}{\omega_p^2}+1\right) \mp \sqrt{\frac{1}{4}\left(\frac{\omega_c^2}{\omega_p^2}+1\right)^2 - \frac{\dfrac{\omega_c^2}{\omega_p^2}}{\dfrac{(\gamma b)^2}{(\beta b)^2}+1}}. \qquad (17)$$

In Figure 3, $|p|$ is plotted as function of βb. ω_c/ω_p and (γb) are parameters and they are assumed to be given. The behavior of this reduction factor, for a neutralized beam in a finite magnetic field, as

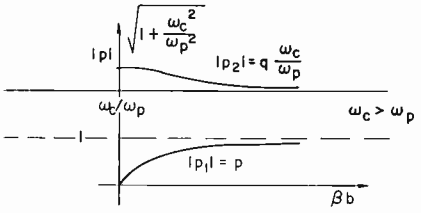


Fig. 3—Reduction factors for a finite cylindrical electron beam in a magnetic field as function of normalized beam radius βb.

a function of the normalized beam radius (βb) does not differ essentially from the behavior of the reduction factor of an unneutralized beam in a finite magnetic field.[9]

Excitation of the Beam

If the dispersion equations of a single beam and of a double-beam system situated in a finite axial magnetic field are rationalized, it is found that the resulting equations are of sixth and tenth degrees in β, respectively. This means that, in general, six and ten variables are necessary to describe the excitation of the beam; e.g., v_r, v_θ, v_z, ρ, E_r and E_z, and v_{r1}, $v_{\theta 1}$, v_{z1}, ρ_1, v_{r2}, $v_{\theta 2}$, v_{z2}, ρ_2, E_r, and E_z.

Since the slow-wave assumption is made, the cutoff waveguide modes can be neglected, and the excitation of the beam can be described by four variables, such as v_r, v_θ, v_z, ρ. These three a-c velocities and the a-c space-charge density are given by the following relations:

$$v_r = \sum_{i=1}^{4} A_i \frac{\gamma}{\beta_i} \frac{\eta(\omega - \beta_i v_0)}{(\omega - \beta_i v_0)^2 - \omega_c^2} J_1(\gamma r) \exp\{j(\omega t - \beta_i z)\}, \qquad (18a)$$

$$v_\theta = \sum_{i=1}^{4} -j A_i \frac{\gamma}{\beta_i} \frac{\eta\omega_c}{(\omega - \beta_i v_0)^2 - \omega_c^2} J_1(\gamma r) \exp\{j(\omega t - \beta_i z)\}, \qquad (18b)$$

$$v_z = \sum_{i=1}^{4} j A_i \frac{\eta}{(\omega - \beta_i v_0)} J_0(\gamma r) \exp\{j(\omega t - \beta_i z)\}, \tag{18c}$$

$$\rho = \sum_{i=1}^{4} \left[-j A_i \frac{\epsilon_0 \omega_p^2 \beta_i}{(\omega - \beta_i v_0)^2} J_0(\gamma r) - j A_i \frac{\gamma^2}{\beta_i} \frac{\epsilon_0 \omega_p^2}{(\omega - \beta_e v_0)^2 - \omega_c^2} J_0(\gamma r) \right] \exp\{j(\omega t - \beta_i z)\}, \tag{18d}$$

where

A_1 = amplitude of the fast space-charge wave,

A_2 = amplitude of the slow space-charge wave,

A_3 = amplitude of the fast cyclotron wave,

A_4 = amplitude of the slow cyclotron wave,

$\beta_1 = \beta_e - p\beta_p$ = propagation constant of the fast space-charge wave,

$\beta_2 = \beta_e + p\beta_p$ = propagation constant of the slow space-charge wave,

$\beta_3 = \beta_e - q\beta_c$ = propagation constant of the fast cyclotron wave,

$\beta_4 = \beta_e + q\beta_c$ = propagation constant of the slow cyclotron wave.*

Two excitation cases have been investigated:

(1) The case of radial-velocity modulation:

at $z = 0$: $v_r = v_{r0}$,

$v_\theta = 0$,

$v_z = 0$,

$\rho = 0$. (19)

(2) The case of axial velocity modulation:

at $z = 0$: $v_r = 0$,

$v_\theta = 0$,

$v_z = v_{z0}$,

$\rho = 0$. (20)

* In general, reduction factors for fast and slow waves are different. Here it is assumed that $p_f = p_s = p$ for space-charge waves, and $q_f = q_s = q$ for cyclotron waves.

Introducing, in turn, the conditions (19) and (20) into Equations (18) and solving these equations for the wave-amplitudes A_i, the following results are obtained.

In both radial and axial velocity-modulation cases, all four waves of the beam are excited. In both cases, the space-charge waves are excited with equal amplitudes, i.e., $A_1 = -A_2$. The same is also true for the cyclotron waves, i.e., $A_3 = -A_4$.

It is further found that in the case of radial velocity modulation the cyclotron waves are dominant, while in the case of axial velocity modulation the space-charge waves are dominant (See Appendix B).

Qualitative Discussion of the Interaction Problem

The interaction problem can most conveniently be discussed by the theory of coupling of modes,[16] in connection with the kinetic power theorem.[17-20] The kinetic power theorem expresses the power flow along an electron beam associated with signals. It has been shown[17-20] that the kinetic power flows associated with the fast waves of the beam (i.e., the fast space-charge wave and the fast cyclotron wave) are positive. On the other hand, the kinetic power flows associated with the slow waves of the beam (i.e., the slow space-charge wave and the slow cyclotron wave) are negative. This means that it is possible to remove energy from the electron beam by increasing the amplitude(s) of the slow wave(s). Similarly, increasing the amplitude of the fast waves requires the addition of kinetic energy to the electron beam.

The basic idea of the theory of coupling of modes is that a system can be divided into a number of subsystems, such as a beam and a structure or a beam and a second beam, whose propagation characteristics can first be studied for no coupling. When a slight coupling is introduced, interaction will occur if the field distributions of the sub-

---

[16] R. W. Gould, "A Coupled Mode Description of the Backward-Wave Oscillator and the Kompfner Dip Condition," *Trans. I.R.E. PGED*, Vol. ED-2, No. 4, p. 37, October 1955.

[17] W. H. Louisell, J. R. Pierce, "Power Flow in Electron Beam Devices," *Proc. I.R.E.*, Vol. 43, p. 425, April 1955.

[18] H. A. Haus, D. L. Bobroff, "Small Signal Power Theorem For Electron Beams," *Journ. Appl. Phys.*, Vol. 28, No. 6, p. 694, June 1957.

[19] P. A. Sturrock, "In What Sense Do Slow Waves Carry Negative Energy?," *Jour. Appl. Phys.*, Vol. 31, No. 11, p. 2052, November 1960.

[20] A. E. Siegman, "Waves on a Filamentary Electron Beam in a Transverse-Field Slow-Wave Circuit," *Journ. Appl. Phys.*, Vol. 31, No. 1, p. 17, January 1960.

systems are similar. Thus the phase velocities of the subsystems have to be approximately the same. A growing wave may result if the power flows of the interacting modes have opposite signs.[16,17,20]

Figure 2 shows the dispersion (ω–β diagram) of an electron beam situated in an axial magnetic field. It has been indicated earlier that the kinetic power flow is positive for the fast waves and negative for the slow waves. Thus the slow waves are capable of delivering energy when coupled to the fast waves of another beam or to a proper circuit.

Figure 4 shows ω–β diagrams of two intermingled electron beams of different drift velocities situated in a finite axial magnetic field. These dispersion curves have been obtained through the use of the

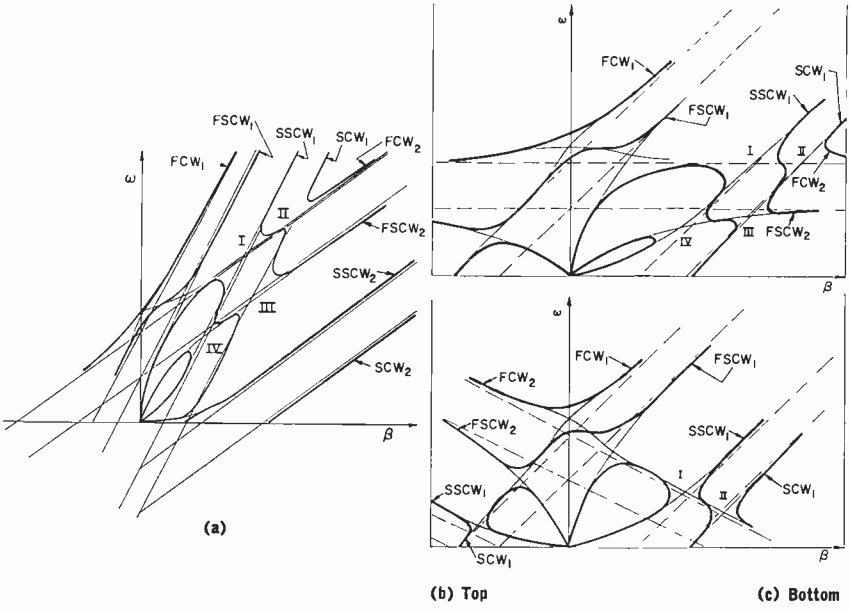


Fig. 4—Dispersion diagrams of two intermingled electron beams of different drift velocities and situated in a finite axial magnetic field: (a) $v_1 > 0$ and $v_2 > 0$; (b) $v_1 > 0$ and $v_2 = 0$; (c) $v_1 > 0$ and $v_2 < 0$. Interaction areas: (I) Interaction of the fast cyclotron wave of beam 2 with the slow space charge wave of beam 1; (II) Interaction of the fast cyclotron wave of the beam 2 with the slow cyclotron wave of the beam 1; (III) Interaction of the fast space charge wave of the beam 2 with the slow cyclotron wave of the beam 1; (IV) Interaction of the fast space charge wave of the beam 2 with the slow space charge wave of the beam 1. FCW_1 = fast cyclotron wave of the faster electron beam; $FSCW_1$ = fast space-charge wave of the faster electron beam; SCW_1 = slow cyclotron wave of the faster electron beam; $SSCW_1$ = slow space-charge wave of the faster electron beam; FCW_2 = fast cyclotron wave of the slower electron beam; $FSCW_2$ = fast space-charge wave of the slower electron beam; SCW_2 = slow cyclotron wave of the slower electron beam; $SSCW_2$ = slow space-charge wave of the slower electron beam.

Table I

Interacting Waves	v_{phase}	v_{power}	v_{group}	Resulting Device
Two beams: Fast cyclotron wave 2 and slow cyclotron wave 1	Same	Opposite	Same	Cyclotron wave double-stream amplifier —Type A
Two beams: Fast cyclotron wave 2 and slow space-charge wave 1	Same	Opposite	Same	Cyclotron and space-charge wave double-stream amplifier—Type AB
Two beams: Fast space-charge wave 2 and slow cyclotron wave 1	Same	Opposite	Same	Cyclotron and space-charge wave double-stream amplifier—Type BA
Two beams: Fast space-charge wave 2 and slow space-charge wave 1	Same	Opposite	Same	Space-charge wave double-stream amplifier—Type B
Beam and plasma: Slow cyclotron wave of the beam and the cyclotron wave of the plasma	Same	Same	Opposite	Cyclotron wave backward-wave amplifier or oscillator—Type A
Beam and plasma: Slow space-charge wave of the beam and the cyclotron wave of the plasma	Same	Same	Opposite	Cyclotron and space-charge wave backward-wave amplifier or oscillator—Type AB
Beam and plasma: Slow cyclotron wave of the beam and the plasma wave	Same	Opposite	Same	Plasma—cyclotron wave forward-wave amplifier
Beam and plasma: Slow space-charge wave of the beam and the plasma wave	Same	Opposite	Same	Plasma — space-charge wave forward-wave amplifier
Two beams: Slow cyclotron wave of beam 1 and fast cyclotron wave of beam 2	Same	Same	Opposite	Cyclotron wave backward-wave oscillator.
Two beams: Slow space-charge wave of beam 1 and fast cyclotron wave of beam 2	Same	Same	Opposite	Cyclotron and space-charge wave backward-wave oscillator.

theory of coupling of modes and the kinetic power theorem, and they indicate the general behavior to be expected from the double-stream system under consideration.

Table I summarizes the interaction possibilities existing in a double-stream system situated in a finite axial magnetic field and gives the devices based on these interaction possibilities.

Quantitative Results

In the first section the dispersion equation of a double-stream system was derived and in the third section the interaction problem was discussed in a qualitative manner. In this section quantitative solutions to the dispersion equation are given. It has been shown[21] that the solution of the dispersion equation by itself, without inquiring into the initial conditions, will yield areas of amplifying and evanescent waves.

In order to obtain a quantitative solution, it is useful to normalize the dispersion Equation (3) in the following form:

$$\left(\frac{\gamma b}{\beta_c b}\right) = -\left(\frac{\beta}{\beta_c}\right)^2 \frac{1 - \dfrac{\left(\dfrac{\omega_{p1}}{\omega_c}\right)^2}{\left(\dfrac{\omega}{\omega_c} - \dfrac{\beta}{\beta_c}\right)^2} - \dfrac{\left(\dfrac{\omega_{p2}}{\omega_c}\right)^2}{\left(\dfrac{\omega}{\omega_c} - \dfrac{v_2\beta}{v_1\beta_c}\right)^2}}{1 - \dfrac{\left(\dfrac{\omega_{p1}}{\omega_c}\right)^2}{\left(\dfrac{\omega}{\omega_c} - \dfrac{\beta}{\beta_c}\right)^2 - 1} - \dfrac{\left(\dfrac{\omega_{p2}}{\omega_c}\right)^2}{\left(\dfrac{\omega}{\omega_c} - \dfrac{v_2\beta}{v_1\beta_c}\right)^2 - 1}}, \quad (21)$$

where

ω/ω_c = normalized frequency (normalized with respect to cyclotron frequency of electrons),

β/β_c = normalized propagation constant ($\beta_c = \omega_c/v_1$),

ω_{p1}/ω_c = normalized plasma frequency of beam 1,

ω_{p2}/ω_c = normalized plasma frequency of beam 2,

---

[21] P. A. Sturrock, "Kinematics of Growing Waves," *Phys. Rev.*, Vol. 112, No. 5, p. 1488, December 1, 1958.

$\alpha = v_2/v_1 =$ normalized velocity parameter.

$\gamma b/(\beta_c b)$ is a parameter which is related to beam diameter:

$$\beta_e b = \frac{\gamma b}{\left(\frac{\gamma b}{\beta_c b}\right)} \frac{\omega}{\omega_c}.$$

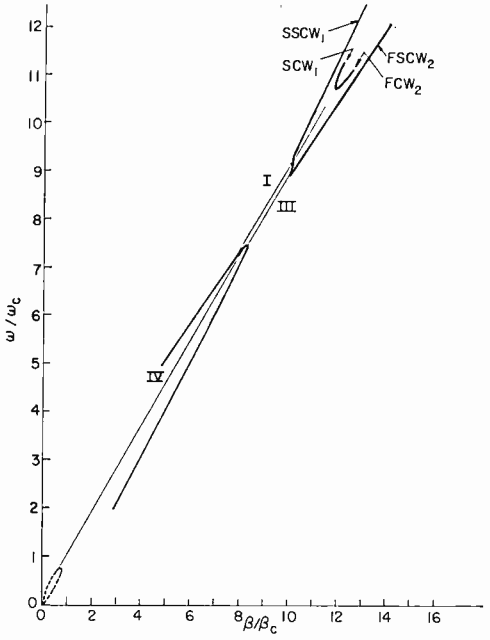


Fig. 5—Calculated ω–β diagram of two intermingled electron beams, showing different interaction areas: $\alpha = v_2/v_1 = 0.8$; $(\omega_{p1}/\omega_c)^2 = 0.5$; $(\omega_{p2}/\omega_c)^2 = 0.7$; $(\gamma b/\beta_c b)^2 = 10.8$. (See Figure 4 for definitions of curve designations.)

Using (ω_{p1}/ω_c), (ω_{p2}/ω_c), (v_2/v_1), and $(\gamma b/\beta_c b)$ as parameters, the dispersion equation has been solved with the help of a digital computer.

Figures 5 through 16 show some of the results obtained for various sets of parameters. These calculations (Figures 5, 9, 13) confirm the qualitative picture developed on the basis of the theory of coupling of modes and the kinetic power theorem.

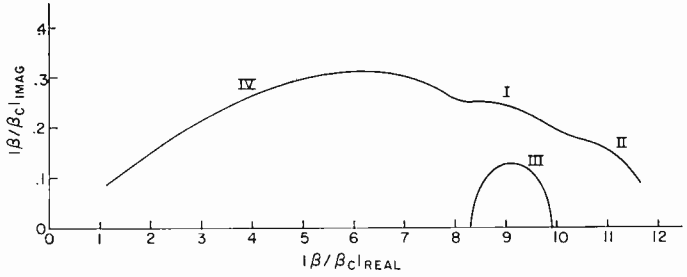


Fig. 6—Normalized gain per unit length as function of the normalized propagation constant (which is proportional to frequency) for the parameters: $v_2/v_1 = 0.8$; $(\omega_{p1}/\omega_c)^2 = 0.5$; $(\omega_{p2}/\omega_c)^2 = 0.7$; $(\gamma b/\beta_c b)^2 = 10.8$.

Figures 5, 9 and 13 represent calculated dispersion diagrams (or parts thereof) for various ratios of drift velocities ($\alpha = 0.8$; 0.5, and 0). For $\alpha = 0.8$ (Figure 5) all four possible interactions can take place. On the other hand, for $\alpha = 0.5$ or 0 (Figures 9 and 13) the interaction between the space-charge waves (interaction area IV) does not occur. This indicates the critical effect of finite beam size and of the ratio of drift velocities for this type of interaction.

Depending on α, the various interactions occur over different frequency ranges. In Figures 6, 10, and 16, the imaginary parts of the propagation constants (gain parameter) for various interaction areas and various sets of parameters are plotted as functions of the real part of the corresponding propagation constants. These latter quantities are proportional to the frequency. The proportionality factors can be obtained from Figures 5, 9, and 13. Therefore Figures 6, 10, and

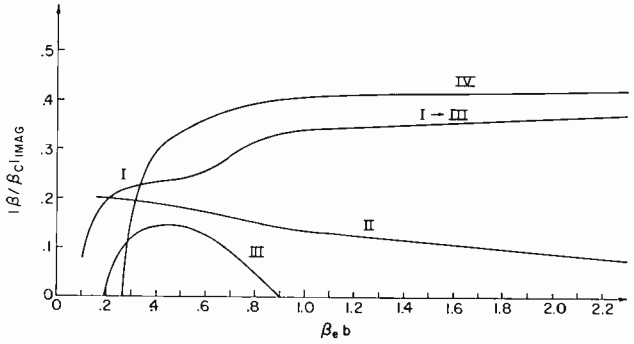


Fig. 7—Normalized gain per unit length as function of $\beta_e b$ for different interaction areas. (I) $\omega/\omega_c = 7.9$; (II) $\omega/\omega_c = 9.9$; (III) $\omega/\omega_c = 8.1$; (IV) $\omega/\omega_c = 5.8$; common parameters for all the curves: $(\omega_{p1}/\omega_c)^2 = 0.5$; $(\omega_{p2}/\omega_c)^2 = 0.7$; $v_2/v_1 = 0.8$.

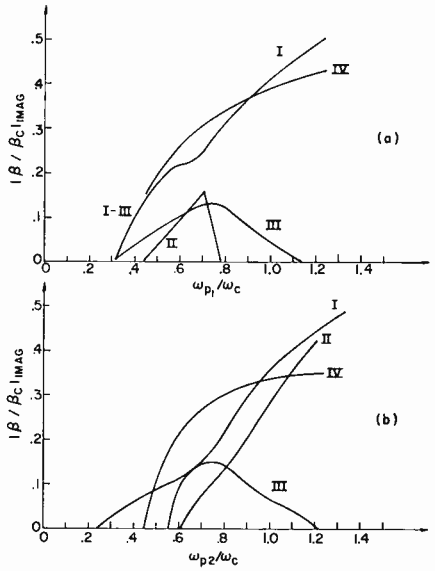


Fig. 8—Normalized gain per unit length as function of normalized beam densities (ω_{p1}/ω_c and ω_{p2}/ω_c): (I) $\omega/\omega_c = 7.9$; (II) $\omega/\omega_c = 9.9$; (III) $\omega/\omega_c = 8.1$; (IV) $\omega/\omega_c = 5.8$. In both (a) and (b), $v_2/v_1 = 0.8$ and $(\gamma b/\beta_c b)^2 = 10.8$; in (a) $(\omega_{p2}/\omega_c)^2 = 0.7$ and in (b) $(\omega_{p1}/\omega_c)^2 = 0.7$

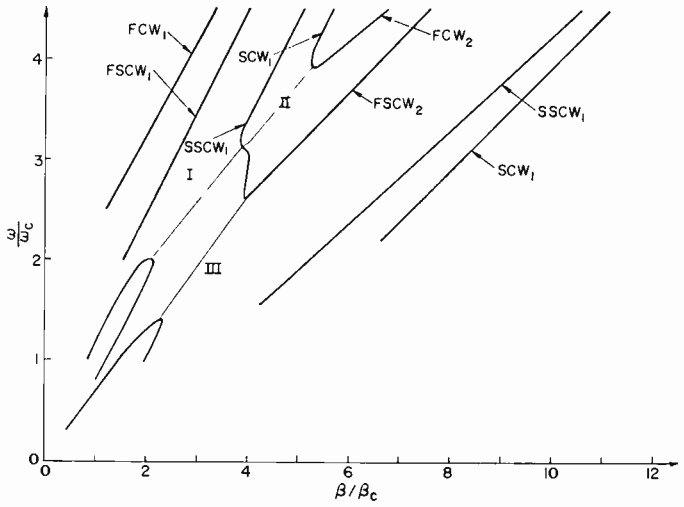


Fig. 9—Dispersion diagram of two intermingled electron beams: $v_2/v_1 = 0.5$; $(\omega_{p1}/\omega_c)^2 = 0.5$; $(\omega_{p2}/\omega_c)^2 = 0.7$; $(\gamma b/\beta_c b)^2 = 10.8$. (See Figure 4 for definitions of curve designations.)

16 are essentially gain-versus-frequency curves for various interaction types.

As indicated earlier, for $\alpha = 0$ the interaction areas I and II become backward-wave type. In these areas the solutions of the dispersion equation are no longer complex. On the other hand, the interaction of the fast wave of the electron beam with the backward-wave mode of the nondrifting plasma lead to an evanescent type of coupling (area V in Figure 13) and the dispersion equation yields complex roots in this area. The imaginary part of this root is also given as a function of

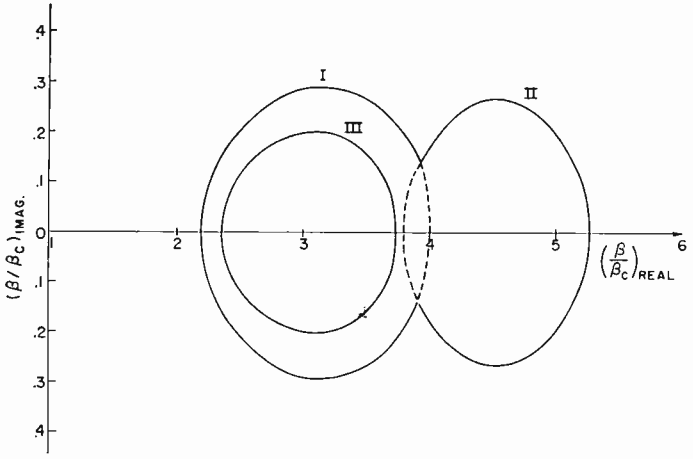


Fig. 10—Normalized gain per unit length as function of the normalized propagation constant, for different interaction areas: $v_2/v_1 = 0.5$; $(\omega_{p1}/\omega_e)^2 = 0.5$; $(\omega_{p2}/\omega_e)^2 = 0.7$; $(\gamma b/\beta_e b)^2 = 10.8$.

various parameters in Figures 14, 15, and 16. These curves, which are designated V or V′ should not be confused with the gain parameters due to interaction types I to IV.

In Figures 7, 12, and 15, the normalized gain parameter is represented as a function of the normalized beam radius $\beta_e b$ for various sets of parameters. Figures 8, 11, and 14 show the same quantity as function of normalized densities ω_{p1}/ω_c or ω_{p2}/ω_c.

Since the frequencies and propagation constants corresponding to interaction areas I and III are very close and to a great extent overlapping, in calculating gain parameter as a function of another parameter in one area, one can easily go into the other area. This happened several times in the course of the calculations, the results are indicated in Figures 7, 8, and 12.

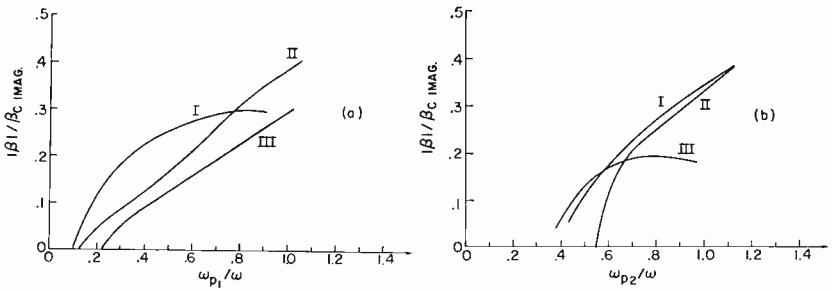


Fig. 11—Normalized gain per unit length as functions of normalized beam densities for different interaction areas: (I) $\omega/\omega_c = 2.7$; (II) $\omega/\omega_c = 3.4$; (III) $\omega/\omega_c = 2.1$. In both (a) and (b), $v_2/v_1 = 0.5$ and $(\gamma b/\beta_c b)^2 = 10.8$; in (a) $(\omega_{p2}/\omega_c)^2 = 0.7$ and in (b) $(\omega_{p1}/\omega_c)^2 = 0.5$.

Discussion and Conclusions

The analysis of double-stream interactions in the presence of a finite axial magnetic field indicates the possibility of a number of double-stream amplifiers and oscillators. These are based on the interaction of the fast waves (fast cyclotron and fast space-charge waves) of the slower electron beam with the slow waves (slow space-charge and slow cyclotron waves) of the faster electron beam. Depending on the relative velocities of the interacting electron beams, the possible devices are either of the forward-wave type or of the backward-wave type.

The currently very popular problem of an electron beam interacting with a stationary plasma is included in the analysis as a special case.

Some of the double-stream devices which the analysis predicts seem

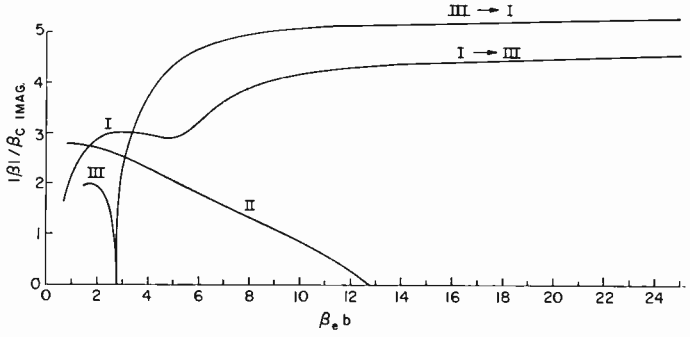


Fig. 12—Normalized gain per length as function of $\beta_e b$ for different interaction areas: $v_2/v_1 = 0.5$; $(\omega_{p1}/\omega_c)^2 = 0.5$; $(\omega_{p2}/\omega_c)^2 = 0.7$; (I) $\omega/\omega_c = 2.7$; (II) $\omega/\omega_c = 3.4$; (III) $\omega/\omega_c = 2.1$.

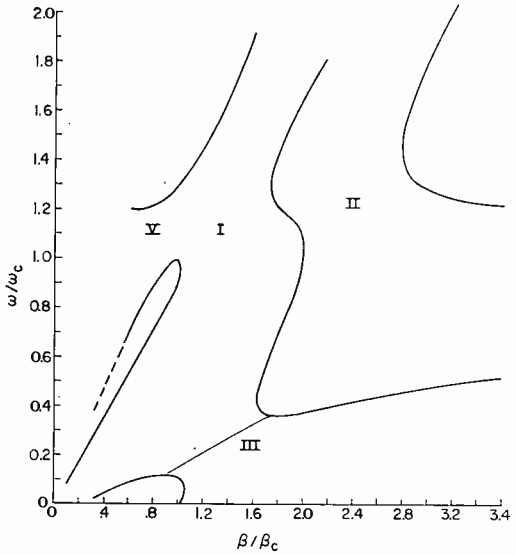


Fig. 13—Part of the dispersion diagram of two intermingled electron beams: $v_2/v_1 = 0$; $(\omega_{p1}/\omega_c)^2 = 0.5$; $(\omega_{p2}/\omega_c)^2 = 0.7$; $(\gamma b/\beta_c b)^2 = 10.8$.

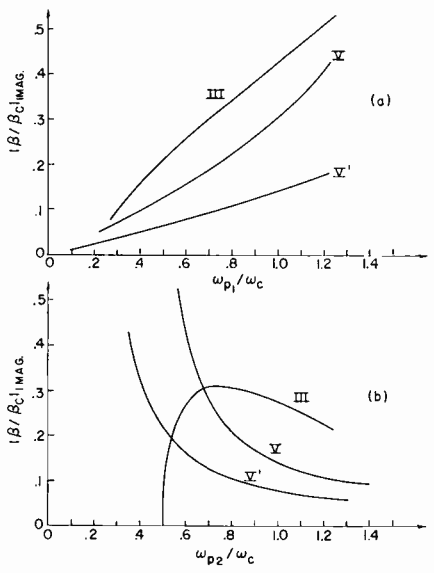


Fig. 14—Normalized gain per unit length for interaction area III and the normalized imaginary part of the propagation constant for interaction area V (interaction of the fast-charge wave of electron beam with the backward-wave mode of the stationary plasma) as function of normalized beam and plasma densities: (III) $\omega/\omega_c = 0.26$; (V) $\omega/\omega_c = 1.11$; (V′) $\omega/\omega_c = 1.05$. In both (a) and (b), $v_2/v_1 = 0$ and $(\gamma b/\beta_c b)^2 = 10.8$; in (a) $(\omega_{p2}/\omega_c)^2 = 0.7$ and in (b) $(\omega_{p1}/\omega_c)^2 = 0.5$.

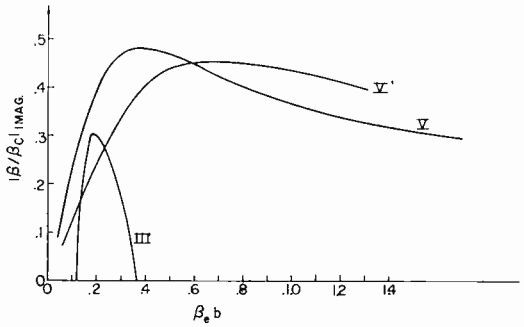


Fig. 15—Normalized gain per unit length for interaction area III, and normalized imaginary part of the propagation constant for interaction area V (see Figure 14 as function of $\beta_e b$: $v_2/v_1 = 0$; $(\omega_{p1}/\omega_e)^2 = 0.5$; $(\omega_{p2}/\omega_e)^2 = 0.7$; (III) $\omega/\omega_e = 0.26$; (V) $\omega/\omega_e = 1.11$; (V') $\omega/\omega_e = 1.05$.

to be promising for high-power amplification and generation at microwave and millimeter-wave frequencies. The calculations presented in Figure 6 indicate the possibility of extremely wide-band double-stream amplification. Several modes take part in this amplification. The fast cyclotron wave and the fast space-charge wave of the slower beam play the role of the "circuits." The slow space-charge wave and the slow cyclotron wave of the faster beam supply the energy for gain. The interaction of these four modes results in four interaction areas. If the beam parameters are chosen properly, the transition from one interaction area into the next can be made smooth. In addition, the

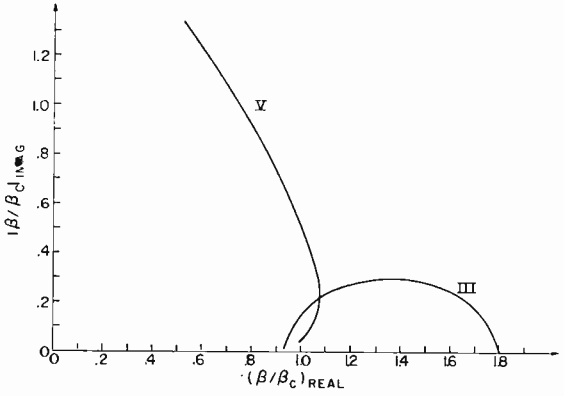


Fig. 16—Normalized gain per unit length for interaction area III, and normalized imaginary part of the propagation constant for interaction area V (see Figure 14) as function of the normalized propagation constant (or frequency: $v_2/v_1 = 0$; $(\omega_{p1}/\omega_e)^2 = 0.5$; $(\omega_{p2}/\omega_e)^2 = 0.7$; $(\gamma b/\beta_e b)^2 = 10.8$.

realization of this wide-band amplification scheme requires proper wide-band multimode input and output couplers.

It is interesting to compare the dependence of gain parameter in various interaction areas:

(a) On the normalized beam radius $\beta_e b$, and

(b) On the normalized densities ω_{p1}/ω_c and ω_{p2}/ω_c. Figures 7, 8, 11, 12, and 15 show the dependence of the normalized gain parameter on either $\beta_e b$ or ω_{p1}/ω_c and ω_{p2}/ω_c. The data contained in these figures indicate the usefulness and the practical limitations of devices which might be based on various interaction possibilities.

ACKNOWLEDGMENT

The author is greatly indebted to F. Paschke for many valuable discussions and suggestions. The assistance of N. L. Gordon and S. W. Kahng in carrying out computations and the assistance of L. S. Nergaard, S. Bloom and M. Lampert in writing the manuscript is also gratefully acknowledged.

APPENDIX A — DISCUSSION AND DERIVATION OF THE DISPERSION EQUATION

The assumptions made have been stated in the text. In this appendix a heuristic derivation of the dispersion equation is given.

Several authors[1,3,4] have derived the dispersion equation for a stationary nonthermal plasma. For instance, starting with Equation (16) of Reference (1) and writing it in the following form

$$\gamma^2 = -\beta^2 \frac{1 - \dfrac{\omega_p^2}{\omega^2}}{1 - \dfrac{\omega_p^2}{\omega^2 - \omega_c^2}}, \tag{22}$$

and applying to it the Galilean transformation ($\omega = \omega' + \beta' v_0$ and $\beta = \beta'$), one obtains the dispersion equation of a drifting beam in an axial magnetic field;

$$\gamma^2 = -\beta^2 \frac{1 - \dfrac{\omega_p^2}{(\omega - \beta v_0)^2}}{1 - \dfrac{\omega_p^2}{(\omega - \beta v_0)^2 - \omega_c^2}}. \tag{23}$$

Then, for a system consisting of n intermingled electron streams, the following dispersion equation results:

$$\gamma^2 = -\beta^2 \frac{1 - \sum_{i=1}^{n} \frac{\omega_{pi}^2}{(\omega - \beta v_i)^2}}{1 - \sum_{i=1}^{n} \frac{\omega_{pi}^2}{(\omega - \beta v_i)^2 - \omega_{ci}^2}}. \tag{24}$$

The case of $n = 2$ yields the dispersion equation of a double-stream system in the presence of a finite axial magnetic field;

$$\gamma^2 = -\beta^2 \frac{1 - \frac{\omega_{p1}^2}{(\omega - \beta v_1)^2} - \frac{\omega_{p2}^2}{(\omega - \beta v_2)^2}}{1 - \frac{\omega_{p1}^2}{(\omega - \beta v_1)^2 - \omega_c^2} - \frac{\omega_{p2}^2}{(\omega - \beta v_2)^2 - \omega_c^2}}. \tag{25}$$

If the left-hand side of Equation (25) is written in the form $(\gamma b/b)^2$ (where b = beam radius and γb = constant from Equation (9), the limiting cases result:

(1) $b \to \infty$, $\omega_c \to \infty$ (thick beam in an infinite magnetic field). This yields the dispersion equation investigated by Nergaard, Pierce, and Haeff for purely space-charge wave interaction;

(2) $b \to 0$, ω_c = finite (thin beam in a finite magnetic field). This results in the following dispersion equation for cyclotron wave interaction:

$$\frac{\omega_{p1}^2}{(\omega - \beta v_1)^2 - \omega_c^2} + \frac{\omega_{p2}^2}{(\omega - \beta v_2)^2 - \omega_c^2} = 1. \tag{26}$$

Thus in the case of thick beams, the space-charge waves are dominant, whereas in case of thin beams, the cyclotron waves are dominant.

If the slow-wave assumption is omitted, the dispersion equation takes the following form (for the case $v_0 = 0$ and a single beam):

$$\gamma^2 = -\frac{\epsilon_{zz}}{\epsilon_{rr}} \beta^2 + \omega^2 \mu_0 \epsilon_{zz}, \tag{27}$$

where

$$\epsilon_{rr} = \epsilon_0 \left(1 - \frac{\omega_p^2}{\omega^2 - \omega_c^2} \right) \tag{28}$$

$$\epsilon_{zz} = \epsilon_0 \left(1 - \frac{\omega_p^2}{\omega^2} \right). \tag{29}$$

Thus

$$\gamma^2 = -\beta^2 \frac{1 - \dfrac{\omega_p^2}{\omega^2}}{1 - \dfrac{\omega_p^2}{\omega^2 - \omega_c^2}} + \frac{\omega^2}{c^2} \left(1 - \frac{\omega_p^2}{\omega^2} \right). \tag{30}$$

The more general dispersion equation, considering both TE and TM waves for the above case is given in References (2), (3), and (5).

Appendix B—Excitation of the Beam

Neglecting the cutoff wave-guide modes and considering only circularly symmetric mode TM_{01}, there are four variables to describe the excitation of the beam. These are v_r, v_0, v_z, ρ as given in Equations (18). Assuming at $z = 0$ and $t = 0$

$$v_r = v_{r0},$$

$$v_\theta = v_{\theta 0},$$

$$v_z = v_{z0},$$

$$\rho = \rho_0,$$

for the space-charge wave,

$$\beta_1 = \beta_e - p\beta_p, \tag{31}$$

$$\beta_2 = \beta_e + p\beta_p; \tag{32}$$

for the cyclotron wave,

$$\beta_3 = \beta_e - q\beta_c, \tag{33}$$

$$\beta_4 = \beta_e + q\beta_c. \tag{34}$$

The analysis will be restricted to two important excitation cases:

(1) The case of radial-velocity modulation:

at $z = 0$

$$v_r = v_{r0}, \quad v_\theta = 0, \quad v_z = 0, \quad \rho = 0. \tag{35}$$

(2) The case of axial-velocity modulation:

at $z = 0$:

$$v_r = 0; \quad v_\theta = 0; \quad v_z = v_{z0}; \quad \rho = 0. \tag{36}$$

Introducing these conditions into Equations (18), the following result is obtained:

In the case of radial velocity modulation,

$$-A_2 = A_1 = \frac{v_{r0} p \beta_p}{2\eta \dfrac{\gamma J_1(\gamma r)}{v_0}} \cdot \frac{\beta_e}{\left(\dfrac{1}{1 - \dfrac{\beta_c^2}{p^2 \beta_p^2}} - \dfrac{q^2}{q^2 - 1} \right)} \tag{37}$$

$$-A_3 = A_4 = \frac{v_{r0} q \beta_c}{2\eta \dfrac{\gamma J_1(\gamma r)}{v_0}} \cdot \frac{\beta_e}{\left(\dfrac{1}{1 - \dfrac{\beta_c^2}{p^2 \beta_p^2}} - \dfrac{q^2}{q^2 - 1} \right)} \tag{38}$$

or

$$A_1 = -A_2 = F_1 p \beta_p \tag{39}$$

$$A_3 = -A_4 = F_1 q \beta_c. \tag{40}$$

Since $q > p$ ($q > 1$, $p < 1$) and assuming $\omega_c > \omega_p$ ($\beta_c > \beta_p$), the conclusion is then clear that, in the case of radial velocity modulation, the cyclotron waves are excited more strongly than the space-charge waves.

Similarly, in the case of axial velocity modulation, $A_1 = -A_2$, $A_3 = -A_4$, and $A_1 > A_3$. In this case the space-charge waves are excited more strongly than the cyclotron waves.

Appendix C—Interaction of the Slow Cyclotron Wave with a Slow-Wave Structure

An idealized simple model[22] consisting of an inductive wall and an electron beam filling half infinite space in front of the inductive wall is investigated. The whole system is immersed in a uniform magnetic field in the direction of z (see Figure 17). Assuming the validity of the applicable assumptions made in the first section and in addition assuming $E_y = 0$ and $\partial/\partial y = 0$, one obtains a wave equation for E_z;

$$\frac{\partial^2 E_z}{\partial x^2} - \beta^2 \frac{1 - \dfrac{\omega_p^2}{(\omega - \beta v_0)^2}}{1 - \dfrac{\omega_p^2}{(\omega - \beta v_0)^2 - \omega_c^2}} E_z = 0. \tag{41}$$

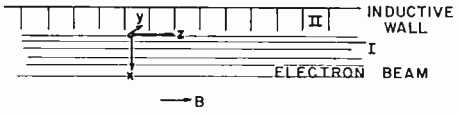


Fig. 17—A simple model of a traveling-wave tube consisting of an inductive wall interacting with the slow cyclotron wave of an electron beam which fills the half infinite space.

The solution of this equation can be expressed as

$$E_z = E_0 e^{-\gamma x} \tag{42}$$

where $\gamma > 0$, because of the required convergence of the field for $x \to \infty$.

Equations (41) and (42) yield

$$\gamma^2 = \beta^2 \frac{1 - \dfrac{\omega_p^2}{(\omega - \beta v_0)^2}}{1 - \dfrac{\omega_p^2}{(\omega - \beta v_0)^2 - \omega_c^2}}. \tag{43}$$

[22] F. Paschke, "Die Wechselseitigkeit der Kopplung in Wanderfeldröhren," *Archiv a elektr. Übertragung*, Vol. 11, p. 137, April 1957.

Transverse boundary conditions:

At $x = 0$,

$$E_{zII} = E_{zI}, \tag{44}$$

$$H_{yII} = H_{yI} + \tilde{K}, \tag{45}$$

$$\frac{E_{zII}}{H_{yII}} = j\,X, \tag{46}$$

where $\tilde{K}$ is the surface current and is given by

$$\tilde{K} = -i_0\,\tilde{x} = -\frac{\epsilon_0\omega}{\beta_e}\,\frac{\beta_p{}^2}{(\beta_e-\beta)^2-\beta_c{}^2}\,E_x. \tag{47}$$

$\tilde{x}$ is the transverse a-c displacement and is obtained from the equation of motion. H_{yI} is obtained from Curl $H = i + \epsilon_0 \partial E/\partial t$ and is given by

$$H_{yI} = \frac{\epsilon_0\omega}{\beta}\left[1 - \frac{\beta_p{}^2}{(\beta_e-\beta)^2-\beta_c{}^2}\right]E_x + \frac{\epsilon_0\omega}{\beta_e}\,\frac{\beta_p{}^2}{(\beta_e-\beta)^2-\beta_c{}^2}\,E_x. \tag{48}$$

Equations (47) and (48) yield

$$H_{yII} = \frac{\epsilon_0\omega}{\beta}\left[1 - \frac{\beta_p{}^2}{(\beta_e-\beta)^2-\beta_c{}^2}\right]E_x. \tag{49}$$

From the slow-wave assumption,

$$E_x = \frac{j}{\beta}\,\frac{\partial E_z}{\partial x} = -j\,\frac{\gamma}{\beta}\,E_z. \tag{50}$$

At $x = 0$,

$$E_x(0) = -j\,\frac{\gamma}{\beta}\,E_{zI} = -j\,\frac{\gamma}{\beta}\,E_{zII}. \tag{51}$$

From Equations (49) and (51)

$$\frac{H_{yII}}{E_{zII}} = jX = \frac{1}{-j\dfrac{\epsilon_0\omega\gamma}{\beta}\left[1 - \dfrac{\beta_p{}^2}{(\beta_e-\beta_1)^2-\beta_c{}^2}\right]}, \tag{52}$$

where X is the impedance of the inductive wall at $x = 0$. Defining

$$\epsilon_0\omega X \equiv \beta_1, \tag{53}$$

where β_1 is the propagation constant of the slow-wave structure, the following dispersion equation is obtained by combining Equations (43) and (52):

$$\beta_1{}^2\left[1 - \frac{\beta_p{}^2}{(\beta_e-\beta)^2}\right]\left[1 - \frac{\beta_p{}^2}{(\beta_e-\beta)^2-\beta_c{}^2}\right] = \beta^2. \tag{54}$$

Assuming synchronism between the circuit wave and the slow cyclotron wave of the beam (i.e., $\beta_1 = \beta_e + \beta_c$) and $\beta_p \ll \beta_c$, and restricting $\beta \cong \beta_e + \beta_c + \delta$, the following equation and solutions are obtained:

$$(\xi + \beta_c)\,[\xi^2\beta_c + (\beta_c + \beta_e)\beta_p{}^2] = 0, \tag{55}$$

$$\xi_1 = -\beta_c,$$

$$\xi_{2,3} = \mp j\beta_p\sqrt{1 + \frac{\beta_e}{\beta_c}}, \tag{56}$$

where $\xi = 2\delta$, or

$$\beta_1 = \beta_e + \frac{\beta_c}{2},$$

$$\beta_{2,3} = \beta_e + \beta_c \mp j\frac{\beta_p}{2}\sqrt{1 + \frac{\omega}{\omega_c}}. \tag{57}$$

Gain per unit length:

$$|\beta_{im}| = \frac{\omega_p}{2v_0}\sqrt{1 + \frac{\omega}{\omega_c}}. \tag{58}$$

A DYNAMIC-CAPACITOR ELECTROMETER SUITABLE FOR MEASURING ELECTROPHOTOGRAPHIC RECORDING MEDIA*

BY

E. C. GIAIMO

RCA Laboratories,
Princeton, N. J.

Summary—A dynamic-capacitor electrometer has been designed for accurate and convenient measurement of the charge acceptance, dark decay of charge, and the decay of charge during illumination on Electrofax and related electrophotographic materials. The apparatus provides means for obtaining data under controlled conditions of temperature and relative humidity.*

With the apparatus, existing and proposed machine operating cycles can be simulated. The electronic behavior of Electrofax paper and metal coated plates can be studied over short and long time intervals. Such data has direct applicability to high-speed print-out systems and to office copier and relief-plate applications operating on relatively long processing cycles. In addition to its research applications, the apparatus should be useful in establishing test procedures and material specifications.

INTRODUCTION

AT the present time, commercially available electrometers are not well suited for measuring the complete charge acceptance and decay of electrophotographic media on a continuous basis over a wide range of time intervals. The dynamic-capacitor electrometer described herein is specifically designed to provide a means for measuring the complete characteristic curve of such media. It includes means for controlled corona charging and for measurement of the charge acceptance and subsequent decay, both in the dark and during illumination. The charge-decay measurements are made in an electrostatically shielded enclosure under controlled ambient conditions of temperature and relative humidity in order to provide reproducible ambient conditions.

The electrometer and its associated circuitry provide a means for obtaining data which simulates the environmental conditions and processing time ranges encountered in a wide variety of applications. The device provides a means for investigating postulated physical models describing the mechanisms of operation. With it, test procedures can be established to form the basis for materials' standardization.

* Manuscript received 15 September 1961.

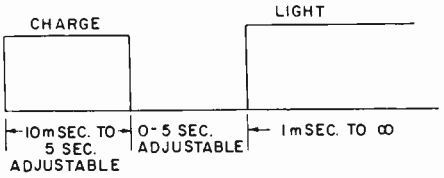


Fig. 1—Timing intervals.

General Description of Apparatus

The apparatus consists of a dynamic-capacitor electrometer incorporating a corona charging source. The peripheral equipment includes timing circuits for impressing a high voltage on the corona unit to obtain precisely timed charging intervals. Means are also included for synchronizing a recorder so that charge build-up as well as charge decay can be observed. The timing circuitry also produces synchronizing signals which can be used to initiate a fixed interval of illumination at a prescribed and adjustable time interval after charging ceases. The timing intervals available are shown in Figure 1. In order to record short-time-interval events, a cathode-ray oscilloscope and camera are used.

The electrometer is housed in a hermetically sealed electrostatically shielded enclosure so that studies can be made under controlled conditions of ambient temperature and relative humidity. Under standardized test conditions, this provides means for obtaining reproducible electrostatic measurement not possible with commercial electrometers and other measuring techniques. A block diagram of the electrometer and its associated equipment is shown in Figure 2.

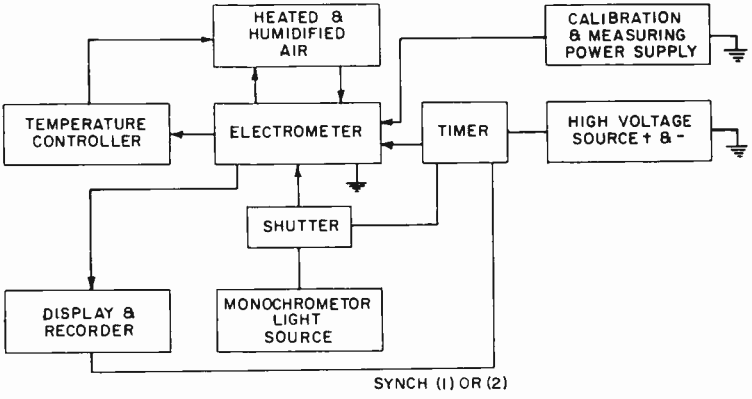


Fig. 2—Electrometer block diagram.

Electrometer Details

The electrometer is based on the dynamic-capacitor principle, i.e., a time-varying potential can be developed at the terminals of a capacitor by periodically inserting and withdrawing a charged dielectric. Detailed operation of similar devices are discussed in the literature.[1,2] A partial derivation of the potential appearing at the terminals of the electrometer as a function of time is given in the Appendix.

The electrometer consists of a parallel-plate capacitor; one plate is a ½-inch-square fixed transparent electrode, and the other is a rotatable metal disk which carries the charged dielectric. The fixed

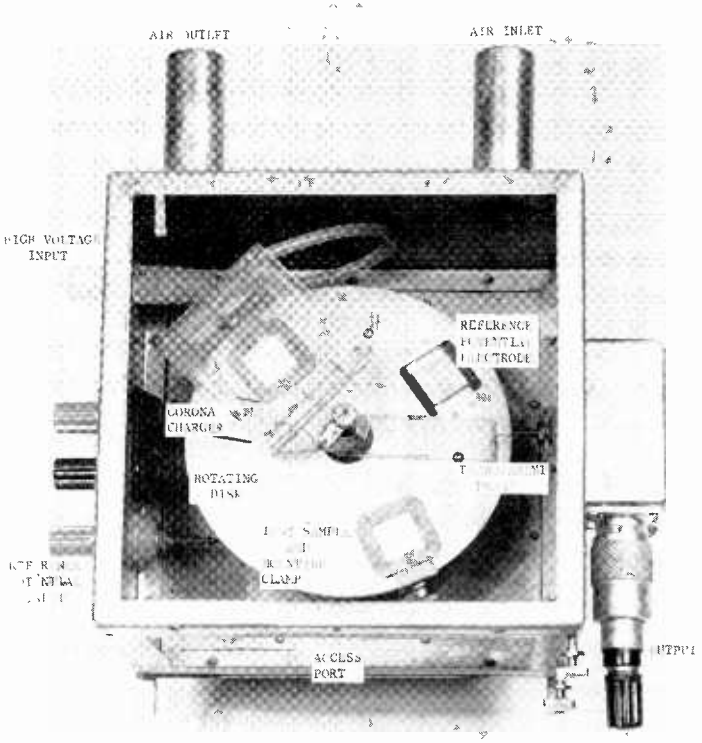


Fig. 3—Top view of electrometer.

[1] H. Palevsky, R. K. Swank, and R. Grenchik, "Design of Dynamic Condenser Electrometers," *Rev. Sci. Instr.*, Vol. 18, No. 5, p. 298, May 1947.

[2] J. R. MacDonald and D. E. Edmonson, "Exact Solution of a Time-Varying Capacitance Problem," *Proc. I.R.E.*, Vol. 49, No. 2, p. 453, February 1961.

electrode is separated from the disk by an air space of approximately .030 inch. The fixed electrode is transparent, so that during light-decay measurements the sample can be continuously and uniformly illuminated. The exact light-decay curve can thus be obtained without introducing light pulses which might give results which are difficult to interpret, since most of the materials investigated show nonlinear behavior. The corona charging unit is transparent for the same reason.

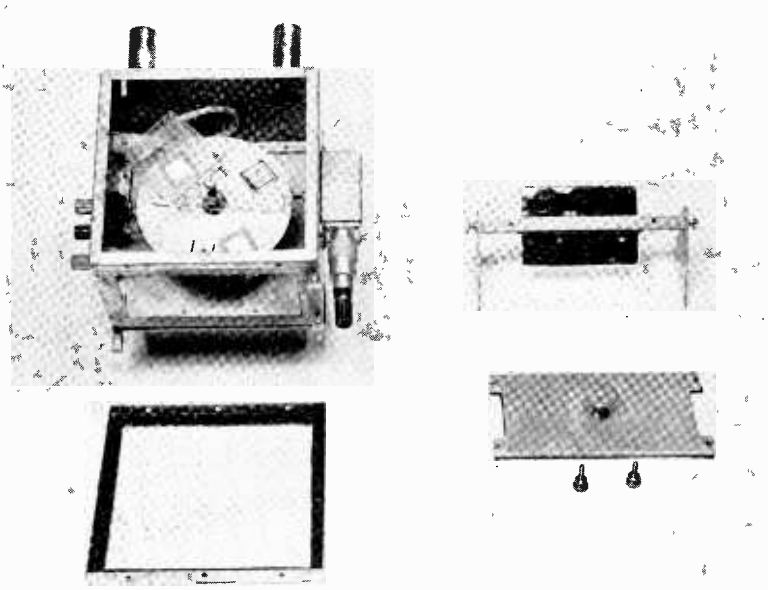


Fig. 4—Component parts of electrometer.

A top view of the electrometer with its transparent conductively coated cover and its metal access-port cover removed is shown in Figure 3. The cover is part of the electrostatic shielding. Figure 4 shows the cover and access port as well as a front-surface mirror for directing a horizontal beam of light into the unit. Figure 5 shows the completely assembled unit which measures approximately 6 inches in each dimension. It is fully electrostatically shielded and can be sealed from the existing ambient atmosphere.

In operation the sample is mounted on the 4-inch-diameter brass rotatable disk by a metal clamp having a 1/2-inch-square aperture. Provision has been made so that two samples can be measured simultaneously. Under some test conditions, it is desirable to compare an unknown directly with a standard under identical conditions. The

number of samples that might be carried by the disk is limited only by the space available around the disk. The drive motor for the disk is located in the motor compartment beneath a base plate on which is mounted a fixed shaft which carries the rotating disk on ball bearings. The fixed shaft also supports the transparent charging unit and the fixed condenser electrode. The disk is driven by a friction wheel at

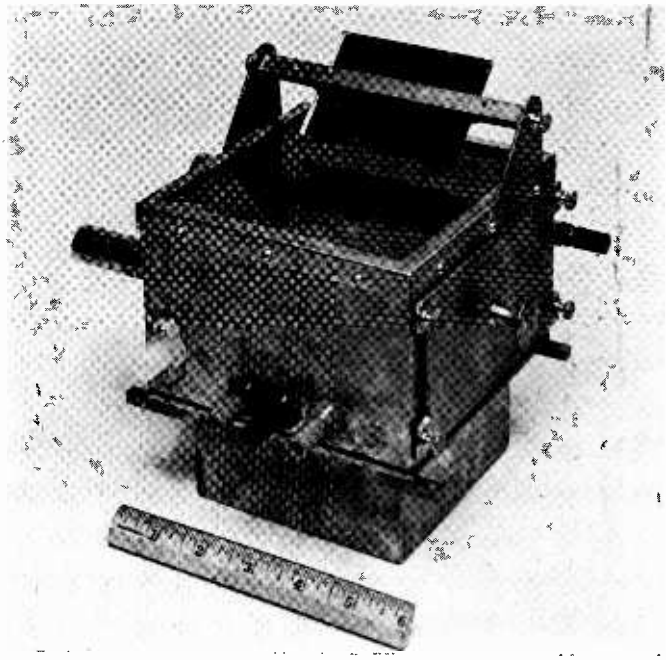


Fig. 5—Electrometer assembly.

approximately 900 rpm. In addition to the samples, a reference electrode is carried by the disk. The reference electrode is insulated from the disk and is connected through a slip ring to a variable-potential source. The electrode area and the distance it protrudes above the disk is the same as that of the sample. The reference electrode provides a means for measuring the amplitude of the voltage pulse generated by the sample as it passes under the probe. A reference voltage pulse, which is identical to that from the sample, is obtained from the reference electrode and appears between successive sample pulses. A typical pulse from either the sample or the reference electrode is shown in Figure 6.

The sample is charged as it rotates beneath the corona charging

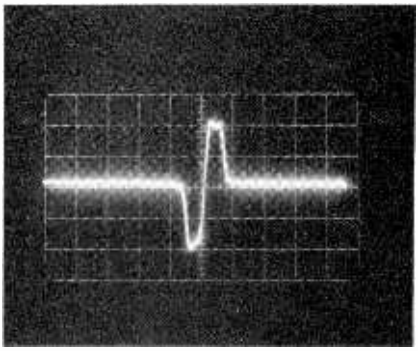


Fig. 6—Typical voltage-output pulse appearing at grid of oscilloscope input amplifier.

unit. The charging interval can be adjusted from 0.01 to 5 seconds. The sample receives a charging pulse each time it passes beneath the charger. A typical charging curve is shown in Figure 7 for an Electrofax coating on paper at 60% relative humidity and 25°C, and with the charging voltage set at 6 kilovolts.

The charger consists of three .0022-inch tungsten wires spaced 1/2 inch apart. The spacing between the corona charger and the disc is adjustable between 1/4 and 1 inch by sliding the charger along the shaft to which it is attached. The electrometer and the corona charging source are housed in a neoprene gasketed electrostatically shielded enclosure as shown in Figure 3.

The sample under test can be heated with hot air when temperature studies are required. A thermocouple mounted on the fixed shaft of

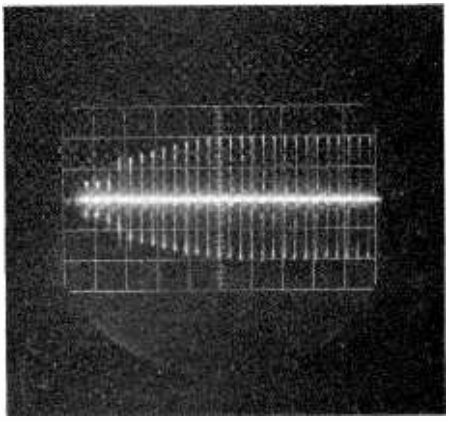


Fig. 7—Typical charging curve of Electrofax layer.

the sample disk is used to measure the sample temperature. The measured temperature difference between the sample and the shaft is 1° over the range 20° to 65°C. The thermocouple output is used to provide the input control signal to a temperature controller which controls the power input to the heating element in the heating chamber.

The entire electrometer assembly, exclusive of the hot air chamber, is compact enough that it can be rigidly fastened to a monochromator light source for light-decay measurements.

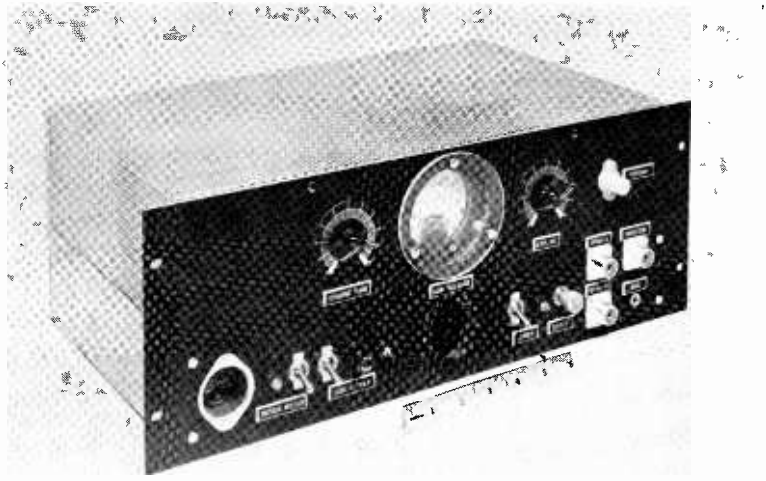


Fig. 8—Control circuit chassis (front panel).

Time and Control Details

The high-voltage interval timer and synchronizing circuits are shown in Figures 8, 9, and 10.

The timing circuits consist of adjustable one-shot multivibrators. The sequence of charging and charge measurement is manually started by pressing a push-button switch on the front panel of the unit.

The timing circuits, corona power supply, and motor for rotating the electrometer sample disk are all controlled from the front panel of the time and control chassis.

The corona voltage is adjustable from 0 to 10 kilovolts and is set on the panel meter. The polarity can be changed by switching the plug and jack located behind the front panel. The interconnections are shown in Figure 9. This photograph also shows the modified relay for switching the high voltage.

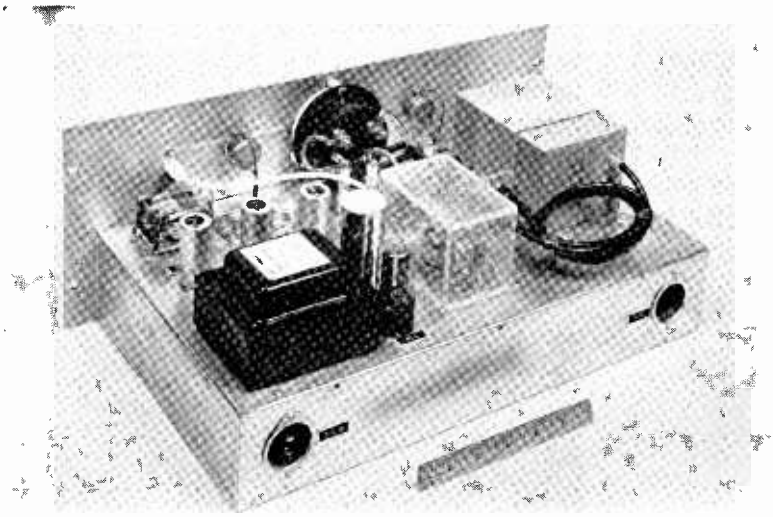


Fig. 9—Control circuit chassis (rear view).

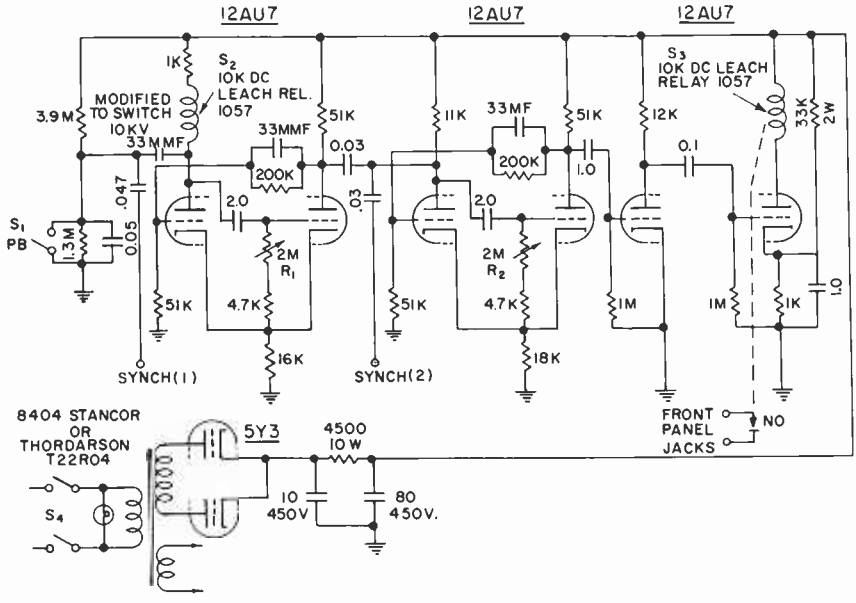


Fig. 10—Timing circuit.

Typical Data

A typical complete characteristic curve of an Electrofax layer consisting of a photoconductive zinc oxide powder in a silicone insulating resin is shown in Figure 11.

The characteristic curve is the envelope of the data shown taken above or below the horizontal axis. The photograph shows the way in which the layer acquires charge during each of its transits beneath the corona discharge operating at 6 kilovolts. The cathode-ray-oscilloscope sweep was initiated when the discharge was started. The sweep

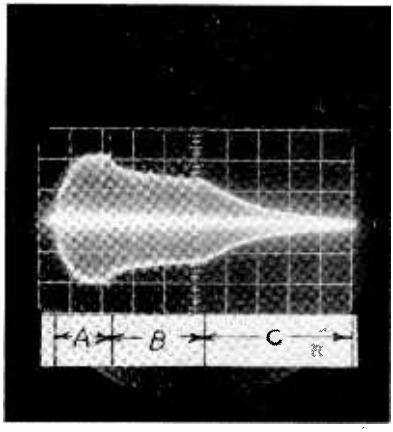
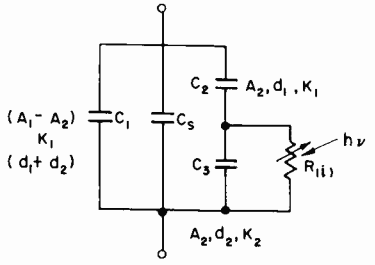


Fig. 11—Typical electrofax characteristic curve: A is charge-buildup time, B dark-decay time, and C light-decay time.

used for obtaining the data of Figure 11 was 1½ seconds per division, and the deflection sensitivity was 300 volts per vertical division. The reference potential electrode was at zero potential at the time the data was recorded. The charging interval is designated A and is equivalent to 3 seconds. It will be observed that the sample reaches a saturation potential of 600 volts in approximately 1½ seconds. The interval B is the dark-decay time. During the interval C, the decay of charge occurs with light set at some arbitrary level.

By comparing the dark-decay curve alone on a given sample with one including the light decay, the differential of the surface-charge density can be determined for the image and nonimage areas for all times after the charging interval. This data is essential in establishing optimum processing cycles for printing devices utilizing such materials. The data can also be used for indirectly studying various electrophotographic developing systems. Charge-density levels needed for a given development means can be measured.

It will be observed from the recordings in Figures 6, 7, and 11 that both positive and negative pulses appear in the trace. This is caused by the input circuit of the oscilloscope which differentiates the approximately triangular waveshape of the voltage appearing at the electrometer terminals as the sample passes under the probe. The exact waveshape of the voltage appearing at the terminals is determined by the shape of the probe, the shape of the aperture in the clamp holding the sample to the disk, and the charge distribution on the sample. Since the charge distribution must be uniform with the type of corona charging used, the pulse is governed by the geometry of the probe and sample.



A_1 = AREA OF THE TRANSPARENT PROBE
A_2 = AREA OF SAMPLE UNDER PROBE AT ANY INSTANT IN TIME
d_1 = PROBE TO SAMPLE SPACING
d_2 = SAMPLE THICKNESS
K_1 = DIELECTRIC CONSTANT OF AIR
K_2 = DIELECTRIC CONSTANT OF SAMPLE
C_S = SHUNT CAPACITY OF PROBE AND LEADS OUTSIDE SAMPLE AREA
$R_{(i)}$ = PHOTOCONDUCTOR LEAKAGE IN THE DARK AND WITH LIGHT

Fig. 12—Equivalent circuit.

APPENDIX

The potential appearing at the electrometer terminals can be calculated using the equivalent circuit shown in Figure 12. C_2 is the capacitance of the sample in series with the probe-to-sample capacitance as the sample passes beneath the probe. C_1 is the capacitance of the probe-to-sample-disk, and C_s is the shunt capacitance of the probe and leads outside the sample area.

It will be assumed that the sample is charged to a uniform charge density q_s.

Using the notation shown in Figure 12, the voltage as a function of time is given by

$$V(t) = \frac{q_s A_2(t)}{C(t)}.$$

The time variation of capacitance, $C(t)$, is

$$C(t) = \frac{K_1 A_1}{d_1 + d_2} + A_2(t)\left[\frac{K_1 K_2}{K_1 d_2 + K_2 d_1} - \frac{K_1}{d_1 + d_2}\right] + C_s.$$

Since $d_1 \gg d_2$ and $K_2 > K_1$,

$$C(t) \approx \frac{K_1 A_1}{d_1 + d_2} + C_s.$$

Therefore, the approximate potential as a function of time is given by

$$V(t) \approx \frac{q_s A_2(t)}{\frac{K_1 A_1}{d_1 + d_2} + C_s}.$$

The value of $A_2(t)$ depends on the geometry of the probe and sample.

Three different probe-sample shapes, square, circular, and circular sectors, have been used to evaluate $A_2(t)$. In each, identical probe and test sample shapes have been used. Perfect congruency between the sample and the probe is assumed when one is directly over the other. Rather complex expressions result for the square and circular geometries. The simplest results are obtained using circular sectors.

RCA TECHNICAL PAPERS†

Third Quarter, 1961

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Acoustic Considerations in the Design of Recording Studios," M. Rettinger, *Audio Eng. Soc. Jour.* (July) 1961
"Analog and Threshold Building Blocks for Variable-Radix Adders and Other Logic," J. Sklansky, *Communication and Electronics* (July) 1961
"An Electric-Fence Control Circuit," C. Turner and J. Eimbinder, *Electronics Illustrated* (July) 1961
"The Elimination of Intersymbol Interference by Input Signal Shaping," I. Gerst and J. Diamond, *Proc. I.R.E.* (July).... 1961
"Full Binary Adder with One Tunnel Diode," B. Rabinovici and C. A. Renton, *Proc. I.R.E.* (July) (Correspondence) 1961
"Internal 'Disappearing Filament' for Maintaining Constant Temperature in a Vacuum," I. S. Solet, *Rev. Sci. Instr.* (July) (Notes) 1961
"100-Mc Tunnel-Diode Ring Counter," F. P. Heiman, *Proc. I.R.E.* (July) (Correspondence) 1961
"A Modern Acoustic Missile Launch Locator, the AN/TNS-5," R. M. Carrell and R. Richter, *Audio Eng. Soc. Jour.* (July) .. 1961
"Porous Nickel Oxide," I. S. Solet, *Nature* (July) 1961
"The Propagation of Wide-Band Signals Through the Ionosphere," H. Staras, *Proc. I.R.E.* (July) (Correspondence) 1961
"Shear-Compensated Hysteresigraph for Thin Magnetic Films," M. J. Schindler, *Rev. Sci. Instr.* (July) (Notes) 1961
"A Standard for Positioning the Vacuum Guide in Transverse Track Video-Tape Recorders," A. H. Lind, *Jour. S.M.P.T.E.* (July) 1961
"A Waveguide Quadruplexer System," P. Foldes and Coauthor, *Trans. I.R.E. PGMTT* (July) 1961
"Zeros of Nonlinear Functions," R. W. Klopfenstein, *Jour. Association for Computing Machinery* (July) 1961
"Average vs. RMS Meters for Measuring Noise," J. J. Davidson, *Trans. I.R.E. PGA* (July-August) 1961
"Physically Short High-Power Balun Is Continuously Variable," J. T. Coleman, *Electronics* (July 28) 1961
"Computers for Industrial Control," R. W. Sonnenfeldt, *Trans. I.R.E. PGIE* (August) 1961
"Dependence of Segregation of Impurities on the Crystallinity of Gallium," L. R. Weisberg and P. R. Celmer, *Trans. Metallurgical Society of AIME* (August) 1961
"An Improved Loudness Indicator," J. L. Hathaway, *Trans. I.R.E. PGB* (August) 1961
"Increase in Dielectric Constant During Switching in Lithium Selenite and Triglycine Sulfate," E. Fatuzzo, *Jour. Appl. Phys.* (August) 1961
"The 'Law of Ohm'," R. E. Schell, *Electronic Industries* (August) (Letters to the Editor) 1961
"Monotonic Trap Distributions," R. H. Bube, *Jour. Appl. Phys.* (August) (Letters to the Editor) 1961
"Recent Advances in Low-Noise Traveling-Wave Tubes," C. L. Cuccia and H. J. Wolkstein, *Microwave Journal* (August).. 1961

† Report all corrections to *RCA Review*, RCA Laboratories, Princeton, N. J.

"Stable Low-Noise Tunnel-Diode Frequency Converter," F. Sterzer and A. Presser, *Proc. I.R.E.* (August) (Correspondence) ... 1961
"Waveguides and the Receptor Mechanism in Color Vision," A. C. Schroeder, *Jour. Opt. Soc. Amer.* (August) (Letters to the Editor) 1961
"Determining Maximum Permissible Dissipation for Silicon Power Transistors," J. Eimbinder and C. R. Turner, *Electronic Design* (August 16) 1961
"Aid to Music Composition Employing a Random Probability System," H. F. Olson and H. Belar, *Jour. Acous. Soc. Amer.* (September) 1961
"Coincident-Current Superconductive Memory," L. L. Burns, Jr., G. A. Alphonse, and G. W. Leck, *Trans. I.R.E. PGEC* (September) 1961
"Concept for an Intercontinental Satellite Communication System," E. A. Laport and S. Metzger, *RCA Review* (September) 1961
"Cross-Section Ratios of Sensitizing Centers in Photoconductors," R. H. Bube, *Jour. Appl. Phys.* (September) 1961
"The Effect of Large Pump Voltage on Tunnel Diode Down Converter Performance," H. J. Prager and K. K. N. Chang, *RCA Review* (September) 1961
"Generation of Linear Binary Sequences," H. Charney and C. Mengani, *RCA Review* (September) 1961
"High-Speed Printing on Electrofax," R. G. Olden, *RCA Review* (September) 1961
"An Integrated Semiconductor Shift Register," J. T. Wallmark and S. M. Marcus, *Trans. I.R.E. PGED* (September) 1961
"An Investigation of Sequential Decoding," W. R. Wadden, D. M. Jones, and J. C. Pennypacker, *RCA Review* (September) 1961
"Korotkoff Sounds in Humans," J. D. Wallace and Coauthors, *Jour. Acous. Soc. Amer.* (September) 1961
"New Method of Compensating Network Design for Feedback Systems," H. Amemiya, *Trans. I.R.E. PGAC* (September) (Correspondence) 1961
"The Noise-Power Probability Distribution in a Multihop FM Radio-Relay System," J. Dutka, *RCA Review* (September) 1961
"Nuvistor Preamplifier for Amateur Receivers," M. R. Adams and P. B. Boivin, *RCA Ham Tips* (September) 1961
"Picture Characteristics of Image-Orthicon and Vidicon Camera Tubes," R. G. Neuhauser, *Jour. S.M.P.T.E.* (September) 1961
"Properties of High-Resistivity Gallium Arsenide Compensated with Diffused Copper," J. Blanc, R. H. Bube, and H. E. MacDonald, *Jour. Appl. Phys.* (September) 1961
"Quantization in Coherent and Quadrature Reception of Orthogonal Signals," G. Lieberman, *RCA Review* (September) 1961
"A Radial-Waveguide Antenna and Multiple Amplifier System for Electronic Scanning," C. P. Clasen, J. B. Rankin, and O. M. Woodward, Jr., *RCA Review* (September) 1961
"Recent Advances in Low-Noise Traveling-Wave Tubes," Part II, C. L. Cuccia and H. J. Wolkstein, *Microwave Journal* (September) 1961
"Reverse Characteristics of Low-Lifetime Germanium Diodes," A. Blicher and I. H. Kalish, *Proc. I.R.E.* (September) (Correspondence) 1961
"A Slide Calculator for Determination of Transistor Dissipation," J. Eimbinder, *Semiconductor Products* (September) 1961
"Some Factors Affecting Applicability of Optical-Band Radio (Coherent Light) to Communication," D. G. C. Luck, *RCA Review* (September) 1961
"Some Thoughts on Digital Components and Circuit Techniques," A. W. Lo, *Trans. I.R.E. PGEC* (September) 1961

"Spectral-Line Interference in X-Ray Fluorescence Spectrometry," E. P. Bertin and R. J. Longobucco, *RCA Scientific Instrument News* (September) 1961

"Spectrophotometric Determination of Vanadium with 3,3'-Diaminobenzidine," K. L. Cheng, *Talanta* (September) 1961

"Statistical Analysis of Multipath Jitter," J. J. Brandinger and H. Goldman, *RCA Review* (September) 1961

"Statistical Properties of M-Ary Frequency-Shift-Keyed and Phase-Shift-Keyed Modulated Carriers," H. Zabronsky, *RCA Review* (September) 1961

"Surface Layer in $BaTiO_3$ Single Crystals," E. Fatuzzo and W. J. Merz, *Jour. Appl. Phys.* (September) 1961

"Synchronization of Pulse Trains," J. Dutka and A. A. Meyerhoff, *RCA Review* (September) 1961

"Temperature Dependence of the Peak Current of Germanium Tunnel Diodes," A. Blicher, R. M. Minton, and R. Glicksman, *Proc. I.R.E.* (September) (Correspondence) 1961

"Temperature Effects on GaAs Switching Transistors," E. D. Haidemenakis, J. A. Mydosh, M. Almeleh, R. Bharat, and E. L. Schork, *Proc. I.R.E.* (September) (Correspondence) 1961

"Crystal-Stabilized Tunnel-Diode Oscillators," J. J. Nagle, *Electronics* (September 1) 1961

"High Electric Fields in Cadmium Sulfide: Field-Effect Constriction of Current Flow and Dielectric Breakdown," R. Williams, *Phys. Rev.* (September 1) 1961

"Multicomponent Magnetoplasma Resonance in p-Type Germanium," R. E. Michel and B. Rosenblum, *Phys. Rev. Letters* (September 15) 1961

"Hot-Switching Surge-Current Nomograph for Diodes," W. Austin, *Electronics* (September 22) (Electronics Reference Sheet) .. 1961

"Comparison of Short Arc Tracking Systems for Orbital Determination," S. Shucker and R. Lieber, *Conf. Proc. Fifth National Convention on Military Electronics* 1961

"The Energy of Electron-Hole Pair Formation by X-Rays in PbO," F. Lappe, *Journal of the Physics and Chemistry of Solids*, Vol. 20, Nos. 3/4, p. 173, Pergamon Press, London 1961

"A New Digital Communication System — Modified Diphase," D. Douglas, G. Aaronson, and G. Meslener, *Conf. Proc. Fifth National Convention on Military Electronics* 1961

"A Position Difference System Using Satellite Doppler Signals," L. Farkas, *Conf. Proc. Fifth National Convention on Military Electronics* 1961

"A Precision Doppler Generator for Radar Target Simulation," C. Faustini, *Conf. Proc. Fifth National Convention on Military Electronics* 1961

"The Problem of Exciton Induced Photoemission from CdS," W. E. Spicer, *Journal of the Physics and Chemistry of Solids*, Vol. 20, No. 1/2, Pergamon Press, London 1961

AUTHORS

JUAN J. AMODEI received his B.S. in EE from Case Institute of Technology in 1956 and an M.S. in EE from the University of Pennsylvania in 1961. In June 1956 he joined Philco Corp., where he worked on transistor radio development. He joined RCA in March 1957 and spent the following three years in the Electronic Data Processing Advanced Development Department. His assignments have consisted of the design of a high-accuracy analog-to-digital converter, design of the circuitry for a medical pressure telemetering pill, and a sampling oscilloscope system for viewing pulses with fraction of a nanosecond rise times. Mr. Amodei transferred to RCA Laboratories at Princeton in December 1959, where he has been engaged in research in the field of high speed digital circuits. Mr. Amodei is a member of the Institute of Radio Engineers, Eta Kappa Nu and Phi Kappa Psi.

R. A. BRADEN graduated from the University of Minnesota with B.S. and M.S. degrees in electrical engineering. He was employed at the Technical and Test Department of RCA from 1926 to 1930, and transferred to RCA Manufacturing Company in 1930. In 1942 he joined the research staff of RCA Laboratories. He worked in the early years on radio receiver and circuit problems, and since 1935 has been engaged in various phases of microwave research. During the last few years, he has been investigating the properties of magnetic materials for microwave applications.

JOHN CARRONA received his B.S. degree in Mechanical Engineering from the Newark College of Engineering in 1950, and his M.S. degree in Metallurgy from Stevens Polytechnic Institute in 1958. He joined RCA in 1954 and since then has worked on the metallurgical problems relating to tube structures and materials, such as grid wires, plate and cathode materials, and glass sealing alloys. In 1959 Mr. Carrona was appointed an Engineering Leader of Tube Materials and Techniques. In his current work, he is concerned with product problems, particularly those related to cathodes and to sealing, exhaust, and aging cycles. Prior to joining RCA, he was Plant Manager for the North American Research Laboratory.
Mr. Carrona is a member of the Institute of Radio Engineers Electrochemical Soc., American Electroplaters Soc., and Sigma Xi.

LEO DAVNE received the B.S.M.E. degree from Louisiana State University in 1947 after having been employed for several years in various mechanical design capacities. He did graduate work in advanced machine and structural design subsequent to graduation and was a Design Engineer at Philco Corp. from 1947 to 1949, working on high-speed radar scanners. From 1949 to 1952, he attended Drexel Institute of Technology Evening School, obtaining the B.S.C.E. degree. During this time he was Project Engineer with Yale and Towne Manufacturing Co. responsible for design of material handling equipment. He was Staff Design Producibility Engineer with Kaiser Fleetwings, Inc., working on airframe components from 1952 to 1955. From 1955 to 1958, he served as a consultant to IBM Corporation and Curtiss Wright Corp. for Allstates Engineering Corp. on mechanical and structural design problems. He joined the technical staff of RCA Laboratories in 1958 as a mechanical design and development engineer.

Mr. Davne has been a registered Professional Engineer since 1949, is a member of the Pennsylvania Society of Professional Engineers, Tau Beta Pi, and the American Society of Mechanical Engineers.

E. C. GIAIMO served in the U.S. Navy as an Electronic Technician from 1943 to 1946. In this capacity, he was a laboratory assistant in microwave research at the Naval Research Laboratory, Washington, D. C., from 1944 to 1946. He received a B.S.E. in 1950 and an M.S.E. in 1955 in electrical engineering from Princeton University. He joined RCA Laboratories in 1951, where he made contributions to ruggedization of traveling-wave-tube structures, germanium purification and testing devices. He has been actively associated with electrophotographic research since 1952. From 1957 to 1959 he was assigned to the technical staff of C Stellarator Associates to perform system specification and coordination duties in the design of a thermonuclear fusion reactor. Mr. Giaimo is a member of the Institute of Radio Engineers and Sigma Xi.

IRWIN GORDON received the B.S. degree in Ceramics from Rutgers University in 1948, the M.S. degree in 1951, and the Ph.D. degree in 1952 from the same institution. From 1948 to 1952 he held a research assistantship in the School of Ceramics at Rutgers University working on single crystal synthesis and studies of various silicates and rare earth aluminates. In 1952 Dr. Gordon joined the technical staff of RCA Laboratories, where he has worked on the development of magnetic materials for various applications. This has included materials for permanent magnets, high-speed storage, multifunction composite materials, and small size high-frequency antennas. Most recently he has investigated non-spinel type magnetic materials for use in the VHF and UHF regions.

He is a member of the American Ceramic Society, the N. J. Ceramic Association, Sigma Xi, Keramos, and a subcommittee (C-21-IIA5, nonmetallic magnetic materials) of the A. S. T. M.

W. J. HANNAN served in the U.S. Navy from 1946 to 1948 as an Aviation Radioman. He graduated from the RCA Institutes in 1951 and was hired by RCA. He received his B.S. in E.E. from Drexel Institute in 1954. He obtained the M.S. in E.E. from the Polytechnic Institute of Brooklyn in 1955, after which he continued graduate studies for another year. At RCA, Mr. Hannan has contributed to the design and development of closed circuit television systems, transistorized television circuits, frequency-shift-keyed and phase-shift-keyed receivers, ground equipment for processing data from a satellite. He taught a transistor circuit theory course at RCA and Rutgers University. He has also contributed to study programs for the application of television to missile tracking, for analysis of digital communication systems, and for improving the jamming immunity of military communication systems.

ROBERT L. HARVEY received the B.S. degree in Electrical Engineering from the University of Tennessee in 1928. After graduation he became a test engineer, and later a member of the receiver engineering group, of the General Electric Company. In 1930 he joined the RCA Manufacturing Company and engaged in receiver production design, receiver advanced development, and later research on receiver circuits and component parts. Since 1942 he has been with RCA Laboratories in Princeton, N. J., where he has engaged in ferrite research, including synthesis of materials, test evaluation, and the application of high-frequency magnetic materials to electronic circuitry. Mr. Harvey is a member of Sigma Xi and a senior member of the Institute of Radio Engineers.

WALTER F. KOSONOCKY was awarded the B.S. and M.S. degrees in electrical engineering by Newark College of Engineering, Newark, N. J., in 1955 and 1957, respectively. Presently, he is a candidate for Sc.D. degree in engineering at Columbia University, New York, N. Y.; for this program he has been awarded the David Sarnoff Fellowship for the academic year 1958-1959. He was a lecturer at Newark College of Engineering during 1958-1959. Since 1955 he has been employed at the RCA Laboratories in Princeton, N. J., where he has been engaged in computer components and systems research. Mr. Kosonocky is a member of Tau Beta Pi and Eta Kappa Nu.

HAROLD B. LAW received the B.S. degree in liberal arts and the B.S. degree in education, both in 1934, from Kent State University at Kent, Ohio. He received the M.S. and Ph.D. degrees in physics from the Ohio State University, Columbus, in 1936 and 1941, respectively. Dr. Law taught elementary mathematics at Maple Heights, Ohio, in 1935-6 and at Toledo, Ohio, for the period 1937-9. He joined the Radio Corporation of America at Camden, N. J., in 1941. In 1942 he transferred to the David Sarnoff Research Center, Princeton, N. J., where he is now a Fellow of the Technical Staff. Dr. Law is a Fellow of The Institute of Radio Engineers, and a member of the American Physical Society and Sigma Xi. In 1946 he was given a joint award, with Dr. A. Rose and Dr. P. K. Weimer, by the Television Broadcasters Association. In 1955 he was presented the Vladimir K. Zworykin Television Prize by the Institute of Radio Engineers.

WILLIAM M. MAZER received the B.E.E. degree from the College of the City of New York in 1947, and the M.E.E. and D.E.E. degrees in 1950 and 1958 from the Polytechnic Institute of Brooklyn, Brooklyn, New York, where he was a member of the research staff of the Microwave Research Institute from 1956 through 1957. From 1950 to 1952 he was a project engineer with the Sperry Gyroscope Company, Great Neck, New York, where he designed guided missile automatic control systems. From 1954 to 1956 he was a development engineer with the Federal Telecommunications Laboratory, Nutley, New Jersey, where he was responsible for the development of an analog computer for a missile guidance system. He was a consultant, engaged in simulator and weapons system analysis, from 1957 to May, 1958, when he joined RCA Defense Electronic Products in New York, N. Y., where he worked on error correcting codes, PCM-FM system studies, and communication system topology. He recently left RCA to join the Advanced Systems Center, U. S. Defense Group, of the ITT Corporation. Dr. Mazer is a member of Sigma Xi, the American Rocket Society, the AAAS, and a senior member of the IRE.

ALAN A. MEYERHOFF (See *RCA REVIEW*, Vol. XXII, No. 3, Sept. 1961, p. 598.)

T. W. OLSON served in the United States Navy as an Aviation Guided Missileman from 1952 to 1955. He graduated from the University of Florida with a B.S. in E.E. degree in 1959. Mr. Olson came to the Applied Research Section of RCA in 1959 under the Graduate Study Program and has completed the course requirements for the M.S.E.E. degree. He is presently writing his thesis. He has worked on the Combat Surveillance Data Link Study Program, phase-shift-keyed demodulation systems, and fiber optics film scanners. For the past year he has been associated with the Electronic Text Reader program. He has been responsible for developing special electronic circuits for the RCA Print Reader. Mr. Olson is a member of Phi Kappa Phi, Sigma Tau, and the Institute of Radio Engineers.

EDWARD G. RAMBERG received the A.B. degree from Cornell University in 1928 and the Ph.D. degree in theoretical physics from the University of Munich in 1932. After working on the theory of x-ray spectra as research assistant at Cornell, he joined the Electronic Research Laboratory of the RCA Manufacturing Company in Camden in 1935. He has been associated with the RCA Laboratories in Princeton since their establishment in 1942. He has worked primarily on electron optics as applied to electron microscopy and television, various phases of physical electronics, thermoelectricity, and optics. In 1949 he was visiting professor in physics at the University of Munich and, in 1960 and 1961 Fulbright Lecturer at the Technische Hochschule, Darmstadt.

Dr. Ramberg is a Fellow of the Institute of Radio Engineers and the American Physical Society and a member of Sigma Xi and the Electron Microscope Society of America.

N. RUDNICK received a B.S. in Mathematics from Queens College in 1942, a B.S. in Engineering Physics from McGill University in 1950, and an M.S. in Physics from M.I.T. in 1952. He joined RCA in 1958 and worked on the high-temperature properties of ceramic insulation for electron-tube heaters. Prior to joining RCA, he was employed by Gulton Industries where he headed up the test and research department of the Glenco Division. In 1961 he rejoined Gulton Industries.

ANATOL G. SAMUSENKO received the B.E.E. degree in 1955 from Polytechnic Institute of Brooklyn. He is currently studying at Moore School of Electrical Engineering, University of Pennsylvania, for the M.S. degree. From 1955 to 1957, he served in the U.S. Army, Ordinance Corps, Frankford Arsenal, Philadelphia, Penna., where he was engaged in the development and the feasibility studies of various fire control systems. Upon discharge from the U.S. Army he joined RCA Industrial Advance Development, Camden, N. J. and later RCA Laboratories, Princeton, N. J. At RCA, he has engaged in the research and the development of novel circuits. He is currently working on the application of the tunnel diode as a high-speed logic element in the digital computer.

S. M. THOMSEN received the M.S. degree in 1930 and the Ph.D. degree in 1937, both in Chemistry and from the University of Wisconsin. He was head of the chemistry department at Dana College, Blair, Nebr., from 1930 to 1943, and Visiting Professor of Physics at the University of Nebraska from 1943 to 1944. He joined RCA Laboratories in 1944, and has worked principally with phosphors and photoconductors.

Mr. Thomsen is a member of the American Chemical Society, Sigma Xi, and of the American Radio Relay League.

B. VURAL obtained the Electrical Engineering degree in 1949 and the Dr. of Technical Sciences degree in 1952 from the Swiss Federal Institute of Technology, Zurich, Switzerland. From 1951 to 1953 he was associated with Brown Boveri and Company, Baden, Switzerland, working on microwave relay communication and on application of a special magnetron as communication transmitter. From 1953 to 1959 he was mainly associated with the Electronic Equipment and Tube Department of Canadian General Electric, Toronto, Ontario, Canada, working mainly on microwave problems in connection with radar and communication.

Dr. Vural joined RCA Laboratories as a Member of the Technical Staff in 1959 and is now engaged in microwave tube research, especially in double-stream interaction, high-power and low noise problems. Dr. Vural is a Senior Member of the Institute of Radio Engineers and a member of the American Physical Society.

RCA REVIEW

a technical journal

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

INDEX

VOLUME XXII

TABLE OF CONTENTS

March

AUTHORS, VOLUME XXII