

RCA REVIEW

a technical journal

**RADIO AND ELECTRONICS
RESEARCH • ENGINEERING**

VOLUME XIX

SEPTEMBER 1958

NO. 3

RADIO CORPORATION OF AMERICA

DAVID SARNOFF, *Chairman of the Board*

FRANK M. FOLSOM, *Chairman of the Executive Committee*

JOHN L. BURNS, *President*

E. W. ENGSTROM, *Senior Executive Vice-President*

DOUGLAS H. EWING, *Vice-President, Research and Engineering*

JOHN Q. CANNON, *Secretary*

ERNEST B. GORIN, *Vice-President and Treasurer*

RCA LABORATORIES

J. HILLIER, *Vice-President*

RCA REVIEW

C. C. FOSTER, *Manager*

C. H. VOSE, *Business Manager*

PRINTED IN U.S.A.

RCA REVIEW, published quarterly in March, June, September, and December by RCA Laboratories, Radio Corporation of America, Princeton, New Jersey. Entered as second class matter July 3, 1950 at the Post Office at Princeton, New Jersey, under the act of March 3, 1879. Subscription price in the United States and Canada; one year \$2.00, two years \$3.50, three years \$4.50; in other countries: one year \$2.40, two years \$4.30, three years \$5.70. Single copies in the United States, \$.75; in other countries, \$.85.

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 XIX

SEPTEMBER 1958

NUMBER 3

CONTENTS

	PAGE
A Study of 468-Megacycle Tropospheric Scatter Propagation Over a 289-Mile Path	321
J. B. ATWOOD, G. B. MACKIMMIE, D. G. SHIPLEY, AND G. S. WICKIZER	
Design and Development of the 21CYP22 21-inch Glass Color Picture Tube	334
C. P. SMITH, A. M. MORRELL, AND R. C. DEMMY	
Effect of Crystal Growth Variables on Electrical and Structural Prop- erties of Germanium	349
F. D. ROSI	
Detection of Asymmetric Sideband Signals in the Presence of Noise	388
T. MURAKAMI AND R. W. SONNENFELDT	
A Note on the Dispersion of Interdigital Delay Lines	418
F. PASCHKE	
Metallographic Aspects of Alloy Junctions	423
A. S. ROSE	
Design Considerations for Direct-Coupled Transistor Amplifiers....	433
J. E. LINDSAY AND H. J. WOLL	
Harmonic Generation with Nonlinear Reactances	455
K. K. N. CHANG	
A Color "Electrofax" Process	465
J. S. RYDZ AND S. W. JOHNSON	
RCA TECHNICAL PAPERS	487
AUTHORS	489

© 1958 by Radio Corporation of America
All rights reserved

RCA REVIEW is regularly abstracted and indexed by *Industrial Arts Index Science Abstracts* (I.E.E.-Brit.), *Electronic Engineering Master Index*, *Chemical Abstracts*, *Proc. I.R.E.*, and *Electronic & Radio Engineer*.

RCA REVIEW

BOARD OF EDITORS

Chairman

R. S. HOLMES
RCA Laboratories

A. A. BARCO
RCA Laboratories

M. C. BATSEL
Defense Electronic Products

G. L. BEERS
Radio Corporation of America

H. H. BEVERAGE
RCA Laboratories

G. H. BROWN
Industrial Electronic Products

I. F. BYRNES
Industrial Electronic Products

D. D. COLE
RCA Victor Television Division

O. E. DUNLAP, JR.
Radio Corporation of America

E. W. ENGSTROM
Radio Corporation of America

D. H. EWING
Radio Corporation of America

A. N. GOLDSMITH
Consulting Engineer, RCA

A. L. HAMMERSCHMIDT
National Broadcasting Company, Inc.

O. B. HANSON
Radio Corporation of America

E. W. HEROLD
RCA Laboratories

J. HILLIER
RCA Laboratories

D. D. HOLMES
RCA Laboratories

C. B. JOLLIFFE
Defense Electronic Products

E. A. LAPORT
Radio Corporation of America

C. W. LATIMER
RCA Communications, Inc.

H. W. LEVERENZ
RCA Laboratories

G. F. MAEDEL
RCA Institutes, Inc.

H. F. OLSON
RCA Laboratories

R. W. PETER
RCA Laboratories

D. S. RAU
RCA Communications, Inc.

D. F. SCHMIT
Radio Corporation of America

G. R. SHAW
Electron Tube Division

L. A. SHOTLIFF
Radio Corporation of America

I. WOLFF
RCA Laboratories

Secretary

C. C. FOSTER
RCA Laboratories

REPUBLICATION AND TRANSLATION

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

A STUDY OF 468-MEGACYCLE TROPOSPHERIC SCATTER PROPAGATION OVER A 289-MILE PATH*

BY

J. B. ATWOOD,[†] G. B. MACKIMMIE,[#] D. G. SHIPLEY,[†]
AND G. S. WICKIZER[†]

Summary—Daytime field-strength recordings on 468 megacycles were conducted over a 289-mile path from Covey Hill, P. Q., Canada to Riverhead, New York during the period September 1955 to December 1956. The purpose of this investigation was to determine the path loss and its seasonal variations.

The hourly distributions of field strength at the receiving location resembled a Rayleigh distribution and the monthly cumulative distributions of the hourly medians followed a log-normal curve. The lowest hourly median value of received power was 159.2 decibels below one watt, and the highest value was 129.2 decibels below one watt. The value calculated for the plane-wave, free-space, received power was 47 decibels below one watt. The typical fast fading rate over this scatter path was on the order of 30 fades per minute and the slow fading rate was on the order of 5 fades per minute.

A test was made to see if diversity could be obtained by using two separate feeds offset from the axis of the 28-foot parabola. The median value of the correlation coefficient between the outputs of the two feeds was about 0.9.

A simultaneous comparison between two receiving sites showed that the site with the higher elevation, and resulting better foreground clearance, had an 8.8-decibel stronger field.

INTRODUCTION

A S part of the propagation-research program of RCA Laboratories, a tropospheric scatter circuit was set up between two sites spaced 289 miles. The transmitter was located at an RCA Victor Co., Ltd. test site at Covey Hill, P. Q., Canada and the receiver was located at the Radio Reception Laboratory of RCA Laboratories located at Riverhead, Long Island, New York. The frequency used was 468 megacycles. Daylight field-strength recording was started in September 1955 and was terminated fourteen months later in December 1956. In addition to the field-strength recording, two short tests were made. One of these was a test of the diversity action from a parabola with a dual feed, and the other was a comparison of two receiving-site locations.

* Manuscript received March 19, 1958.

[†] RCA Laboratories, Riverhead, New York.

[#] RCA Victor Company, Ltd., Montreal, Quebec, Canada.

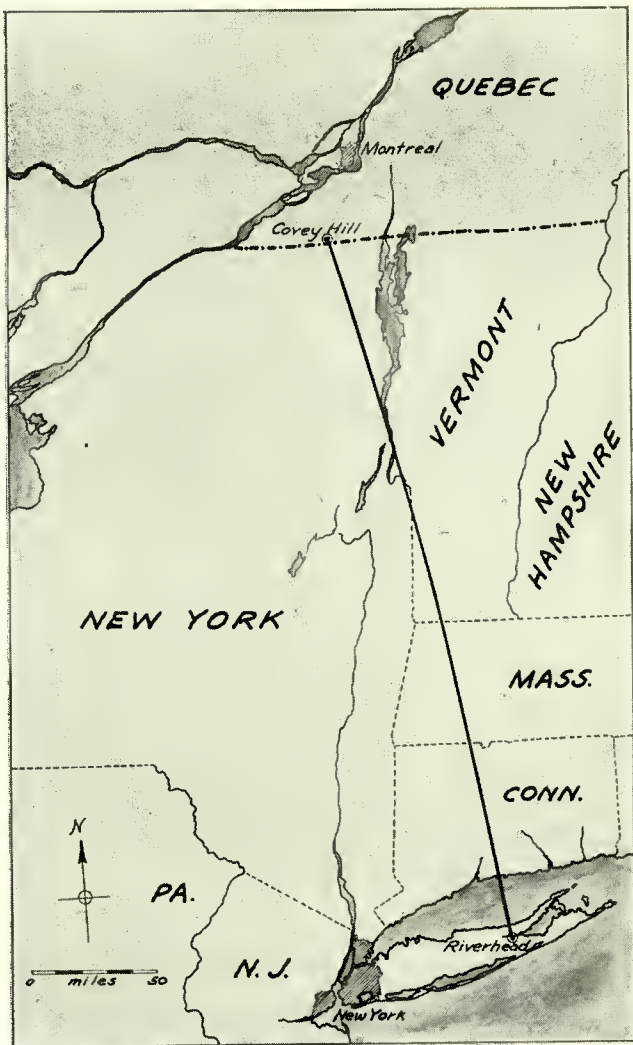


Fig. 1—Location of the propagation path.

PROPAGATION PATH

The location of the propagation path is shown on Figure 1. The path was over land except for the short section over Long Island Sound near the receiving site.

The $\frac{4}{3}$ earth's radius profile of the transmission path is shown on Figure 2. Only the ends of the path profile have been drawn, as

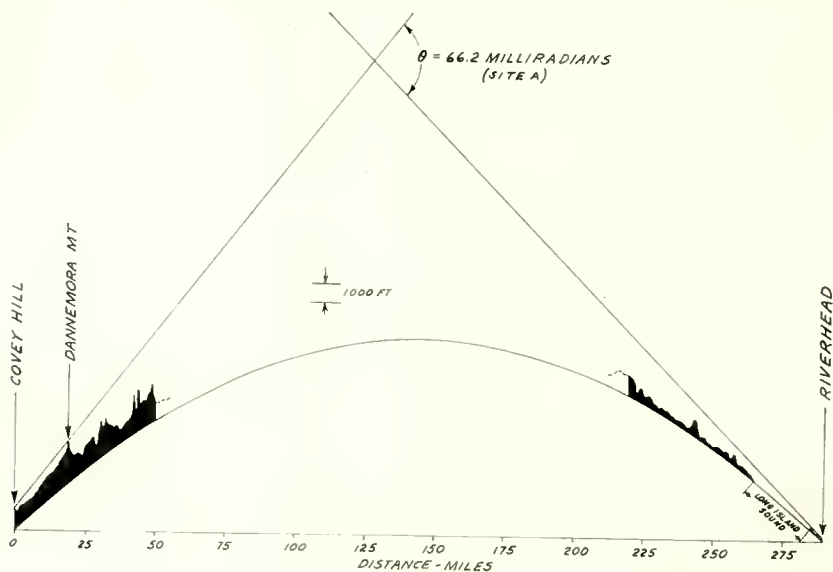


Fig. 2— $4/3$ earth's radius profile of the propagation path.

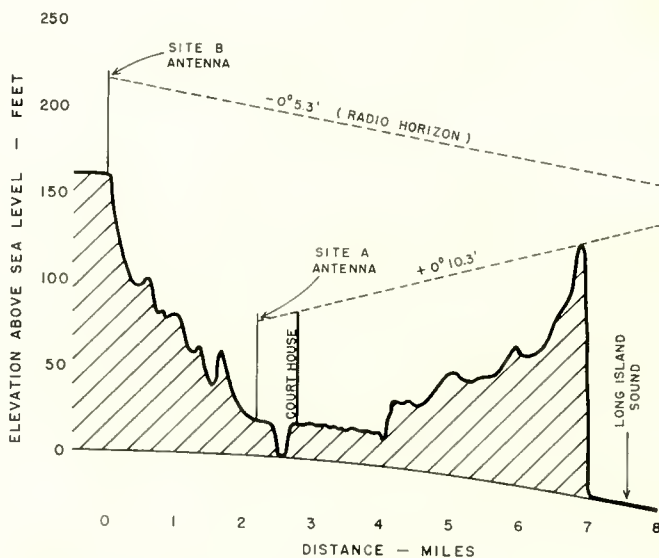


Fig. 3—Enlarged profile of the receiving end of the propagation path based on $4/3$ earth's radius.

an inspection of the path showed no other obstructions. The transmitting antenna was located about 1,150 feet above sea level. There was an obstruction in the path, caused by Dannemora Mountain located 19.5 miles from the transmitter, at an elevation of 2,200 feet above sea level. The receiving end of the path was also partially obstructed. A more detailed profile of the receiving end of the path is shown in Figure 3. The antenna location is marked "Site A" in this figure. There was a certain amount of foreground clutter, but the main obstruction was a cliff on Long Island Sound. The receiving antenna was located at an elevation about 77 feet above sea level. The cliff was about 4.7 miles away with an elevation of about 147 feet above sea level. The result was that the cliff was about 10.3 minutes of angle above the 0° elevation from the receiving antenna on a $4/3$ earth's radius profile. The angle formed by the intersection of the transmitting and receiving rays which grazed the obstructions on the radio horizons was 66.2 milliradians ($3^\circ 48'$). This angle, which is shown in Figure 2, is equivalent to a separation of about 410 miles between the transmitting and receiving sites over a smooth spherical earth.

TRANSMITTER

The transmitter operated on a frequency of 468 megacycles and had a power output of 340 watts using an unmodulated carrier. A photograph of the transmitting antenna installation is shown in Figure 4. The antenna was a 40-foot parabola, with a focal length of 20 feet, whose center was located 50 feet above the ground. The ground elevation at the antenna site was 1,100 feet.

The antenna had a gain of 31 decibels with reference to a half-wave dipole and the beam width was 3.8° between the half-power points. The effective radiated power was equivalent to about 428 kilowatts in a half-wave dipole or 56.3 decibels greater than one watt (dbw). The transmissions were horizontally polarized.

RECEIVER

Most of the field-strength recording was done using a 28-foot parabola at the receiving end. A photograph of the receiving antenna installation is shown in Figure 5. This parabola had a focal length of 12 feet and was fed with a circular-waveguide feed. The E-plane (horizontal) pattern and the gain of the receiving antenna were measured using an oscillator, located about 4.5 miles in front of the antenna, as a source. The beam width was 3.9° between the half-power points and approximately 13.5° between the first nulls. The gain



Fig. 4—Transmitting antenna installation at Covey Hill.

of the receiving antenna was 31.6 decibels with reference to a half-wave dipole. The parabola was mounted with its center 57 feet above the ground on a steel tower adjacent to the Riverhead Laboratory. The ground elevation at this location is 20 feet above sea level.



Fig. 5—Receiving antenna installation at Riverhead.

The receiver was a multiple-conversion superheterodyne using three quartz-crystal oscillators and one variable oscillator that could be tuned manually or operated with a motor-driven automatic frequency control. The final intermediate frequency was 10 kilocycles. The equivalent noise bandwidth of the receiver was 535 cycles, and the noise figure was about 15 decibels.

RECORDING EQUIPMENT

Field-strength recording was accomplished by connecting the amplified automatic-gain-control (a-g-c) voltage from the receiver to a signal-strength analyzer. This analyzer consists of an automatic voltage integrator and counting device with twelve adjustable levels.¹ It contains twelve thyratrons with individually adjustable biases. The output of each thyatron is connected to a relay which is used to start and stop a motor-driven mechanical counter. A photoelectric potentiometer recorder with the chart operating at three inches per hour was also connected to the amplified a-g-c output from the receiver. This was used for monitoring the transmissions and for recording the slow fading characteristics. To examine the fast fading characteristics, as well as aircraft reflections, a high-speed recorder was available with a response essentially flat to 60 cycles per second. The measured response of the a-g-c circuit showed that it was down 3 decibels at 62.5 cycles per second. This corresponds to a time constant of 2.55 milliseconds.

TRANSMISSION SCHEDULE

The transmission schedule varied somewhat from time to time, but in general it enabled recording from 9 A.M. to 4 P.M. on Monday through Thursday and from 9 A.M. to 2 P.M. on Friday. In addition to the regular recording schedule, two continuous runs about 30 hours in length were made during the course of the year. No significant change in scatter propagation was noted during the nighttime hours.

FIELD-STRENGTH RECORDING

The counters on the signal-strength analyzer were read hourly and, from these readings, hourly field-strength distribution curves were drawn. The hourly distributions of field strength in general followed a Rayleigh distribution except that they did not fall off as fast as the Rayleigh curve near the low-level end. Also, the high-level ends of

¹ R. W. George, "Signal Strength Analyzer," *Electronics*, Vol. 24, p. 75, January, 1951.

the distribution curves were affected by the number of aircraft reflections during the recording period.

Each month, the median values from the hourly distribution curves were used to plot a monthly cumulative distribution of the hourly medians. The monthly cumulative distribution of the hourly medians followed a log-normal distribution when there were sufficient data to be representative of the period. If this distribution is assumed for these curves, it is possible to find the standard deviation (σ) for each monthly curve. σ is a measure of the slope of the curve and would be an indication of the spread of the hourly medians for that month. The value of σ for each of the median distribution curves has been

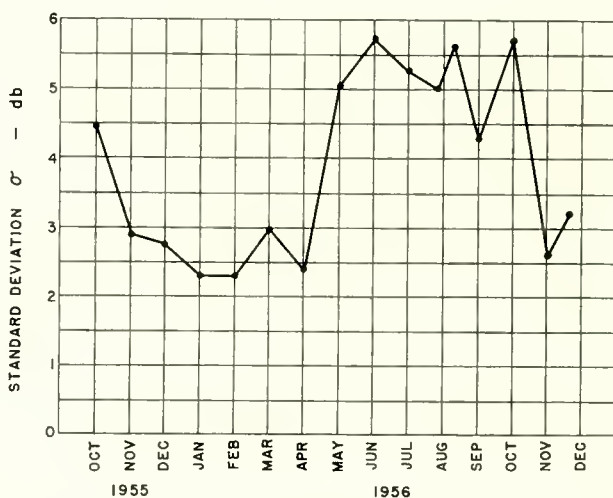


Fig. 6—Standard deviation of the monthly cumulative distribution of hourly medians.

plotted on a yearly basis in Figure 6. In the change-over from winter to summer conditions, the value of σ increased suddenly about two months before there was any large change in the median value of the field strength. In the change-over from summer to winter conditions, this was reversed. The value of σ decreased about two months after the median value of the hourly medians began to decrease.

If, instead of the median distribution, the monthly total field-strength distribution had been plotted, σ would be an indication of the monthly fading range. A plot of σ from the monthly total field-strength distribution curves produced a curve of the same shape as Figure 6. The winter value of σ was about 6.75 decibels and the summer value was about 8.5 decibels.

Also, from the curves showing the monthly cumulative distribution of the hourly medians, the 1, 10, 50, 90, and 99 per cent points were plotted on the yearly graph shown in Figure 7. This yearly graph shows the seasonal variation in the distribution of the hourly medians for each month.

The lowest hourly median value that was recorded was 159.2 decibels below one watt on December 2, 1955. The highest hourly median value recorded was 129.2 decibels below one watt on August 30, 1956. This gave a difference of 30 decibels between the highest and lowest hourly medians.

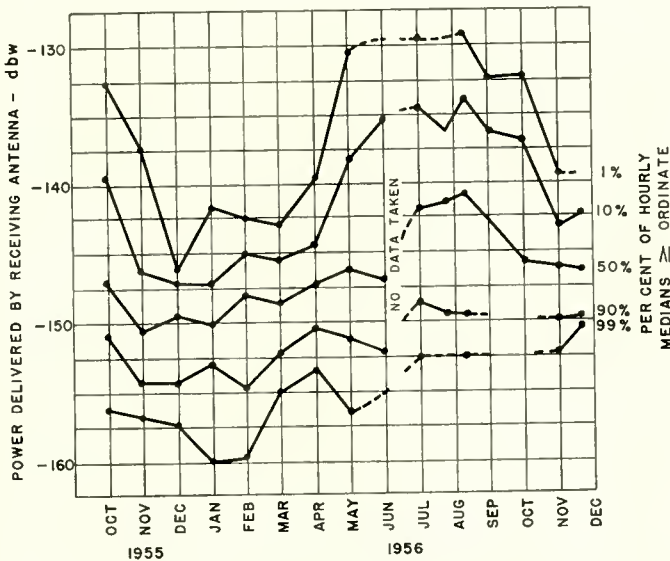


Fig. 7—Summary of the monthly cumulative distribution of hourly medians showing the seasonal variations.

The lowest median value of the curves showing the monthly cumulative distribution of hourly medians was 150.7 decibels below one watt in November 1955. The highest value of 140.8 decibels below one watt occurred in August 1956. This gave a spread of 9.9 decibels between the lowest winter month and the highest summer month. The value calculated for the plane-wave, free-space, received power was 47 decibels below one watt.

Figure 8 is a reproduction of a portion of a typical winter field-strength recording made at three inches per hour. The spikes on this chart sample were caused by aircraft reflections. Aircraft reflections were a common occurrence on this circuit. Listening tests made during

short periods distributed over five days indicated that audible aircraft reflections were present about 38 per cent of the time. The maximum fading rate, as a result of reflections from an aircraft, was about 1,000 fades per minute. The typical fast fading rate over the scatter path was on the order of 30 fades per minute and the slow fading rate was on the order of 5 fades per minute. These figures are based on the number of fades that were 5 decibels or greater.

DIVERSITY TEST

A short test was made to see if it would be possible to obtain diversity reception by using two independent feeds offset from the axis of the 28-foot parabola. Each of these feeds had its own trans-



Fig. 8—Typical winter sample recording of received field strength. The large spikes were caused by aircraft reflections.

mission line and receiver. Since each of the feeds was offset from the focus of the parabola, two beams were produced which diverged from the axis of the parabola. Each of these beams saw a different portion of the scattering volume illuminated in space by the transmitting antenna and it might be expected that the fading produced by these separate scattering volumes would not be coincident in time. If this expectation were realized, the outputs of the two receivers could be combined in diversity for improved performance.

The results of this test showed that a small amount of diversity could be obtained for short periods of time. However, it should be pointed out that the median value of the correlation coefficient between the fields produced by the two different scattering volumes was about 0.9.

SITE COMPARISON

During the month of August 1956, a comparison was made between the site where the 28-foot parabola was located and another site at a higher elevation. The profile near the receiving terminal, Figure 3, shows the location of these two sites. Two identical 76-inch parabolas were used for the comparison, one at each location.

Prior to the site comparison, the gain of one of the 76-inch parabolas was compared with the gain of the 28-foot parabola. The gain difference between the two parabolas was first obtained by comparing the outputs of the two antennas when receiving a local oscillator about 4.5 miles away. This gave a difference of 12.8 decibels. The two antennas were then compared by simultaneously recording over the scatter circuit from Covey Hill for seven days. The average of the differences in the simultaneous hourly medians was 11.2 decibels. It is not known whether this discrepancy was due to aperture-to-medium coupling loss or to local reflections. More extensive measurements would be necessary to separate and evaluate these effects.

Referring to Figure 3, Site A was the one used for the field-strength recording (on the 28-foot parabola) described previously. This site had two obstructions. One was a large courthouse about 0.6 mile from the antenna. The top of this was about 6 feet above the optical 0° elevation. The other obstruction was a cliff on Long Island Sound which was 4.7 miles away. This was about 10.3 minutes of angle above the 0° elevation on a $4/3$ earth's radius profile. On this profile, the line from the antenna to the cliff cleared the roof of the courthouse by a few feet. The other site, marked B in Figure 3, was 140 feet higher than Site A. Site B was 2.3 miles farther from the transmitter than Site A, and all three were on the same great-circle path. The obstruction in the path from Site B was located 32.4 miles away resulting in the radio horizon being 5.3 minutes of angle below the 0° elevation. The cliff was 8.7 minutes of angle below the 0° elevation.

Recordings of the Covey Hill transmissions were made simultaneously at the two locations for seven days. The average difference in the simultaneous hourly medians at the two locations was 8.8 decibels. The location with the higher elevation and the better foreground clearance had the higher field strength.

Figure 9 is a plot of the cumulative distribution of the hourly medians at the two locations. Figure 10 is a plot of the distribution of the differences between the simultaneous hourly median values at each site.

The calculated difference in path loss between the two sites did not

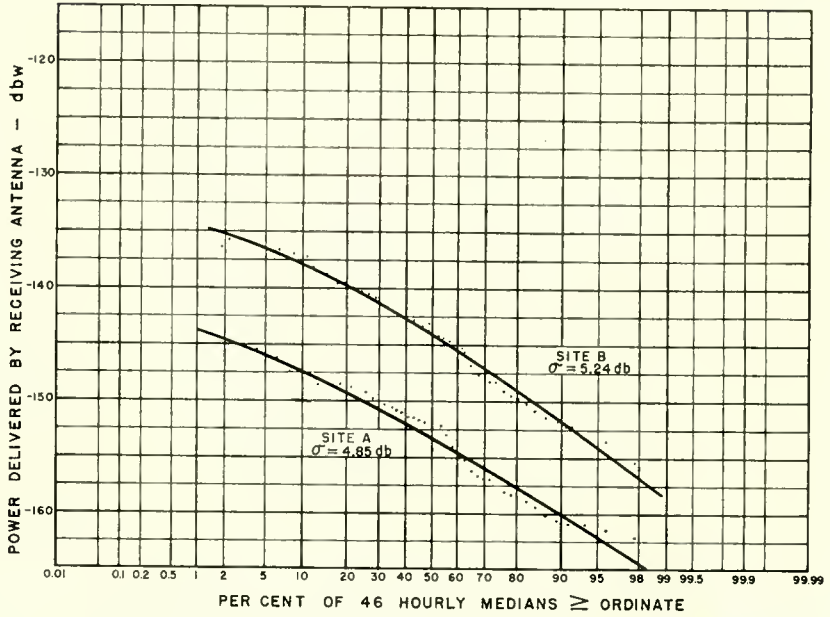


Fig. 9—Cumulative distribution of hourly medians from recordings taken simultaneously at Site A and Site B.

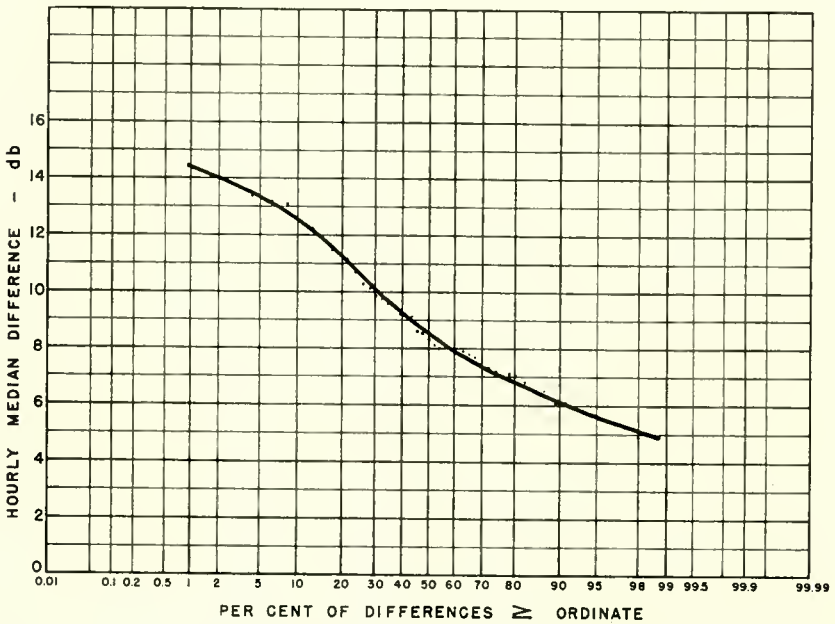


Fig. 10—Distribution of the differences between the hourly-median received field strengths at Site A and Site B.

exceed 2.8 decibels, so it would appear that the remaining 6 decibels was caused by foreground clutter at Site A.

CONCLUSION

In general, the path loss of this circuit was highest during the winter months and lowest during the summer months. The highest hourly median value of the path loss was 112.2 decibels below the value calculated for plane-wave, free-space transmission. The lowest hourly median value of the path loss was 82.2 decibels below the free-space value. This gave a spread of 30 decibels between the highest and lowest hourly median values.

On a monthly basis, the path loss (based on the median value of the distribution of hourly medians for each month) had a high value of 103.7 decibels, and a low value of 93.8 decibels, below the free-space value. Thus there was a seasonal variation of 9.9 decibels on a monthly basis as compared with a 30 decibel variation on an hourly basis.

The hourly distributions of field strength generally followed a Rayleigh distribution and the cumulative hourly median distribution curves followed a log-normal distribution. There was an abrupt change in the monthly fading range between summer and winter. Also, the standard deviation of the cumulative distribution of hourly medians for each month showed an abrupt seasonal change. This value changed from about 5 decibels in the summer to about 2.5 decibels in the winter.

Aircraft reflections were a common occurrence on this circuit and it was observed that, during periods of low field strength, the presence of an aircraft reflection would occasionally increase the field strength by as much as 30 decibels. During periods of higher field strength, the increase caused by aircraft reflections was less.

The test of the dual-beam diversity indicated that a small amount of diversity action was present for short periods of time. With the antenna arrangements used, a median value of the correlation coefficient of about 0.9 was obtained over this circuit.

The site-comparison tests showed the importance of site evaluation for practical applications. The site with the better foreground clearance had a median received field strength 8.8 decibels higher than the poorer site. It is interesting to note that the poorer site with the 28-foot parabola would produce a receiver input only 2.4 decibels higher than the better site utilizing a 76-inch parabola.

The data presented provide useful information on the path losses and the allowances necessary for seasonal and hourly variations in the field strength. These data will be helpful in determining the power

and antenna gain requirements for tropospheric scatter propagation paths on the order of 300 miles in length.

ACKNOWLEDGMENTS

The propagation recording program described in this paper was initiated by E. A. Laport and H. H. Beverage. The project was carried out under the general direction of H. O. Peterson.

In addition, A. J. Herbig assisted in the design and testing of the receiving antenna and W. Leis aided in the taking and reduction of the data. The operation at Covey Hill was in charge of R. Simard and the transmitter was operated by J. Robichaud. The illustrations for the paper were prepared by J. D. Doughty.

DESIGN AND DEVELOPMENT OF THE 21CYP22 21-INCH GLASS COLOR PICTURE TUBE*

By

C. P. SMITH, A. M. MORRELL, AND R. C. DEMMY

RCA Electron Tube Division,
Lancaster, Pa.

Summary—This paper describes the features of the new 21-inch round glass color picture tube, RCA-21CYP22, and evaluates the improvements incorporated in this design as compared to earlier designs. The newly developed glass bulb having a frit closure is described, as well as the bulb-alignment system now used and the special glass properties. The improved design of the shadow mask, which provides increased light output and contrast, and the associated screen processing are discussed. Improvements in the electron gun and their effects on tube performance are also described, as well as the revised components and set-up procedure.

THE RCA-21CYP22 is a 70° shadow-mask color-picture tube in an all-glass envelope. In general, it is similar to the 21AXP22-A except for several important design features which have been developed for this new type and the use of a glass rather than a metal envelope.

The over-all length of the 21CYP22 is 25 13/32 inches, which is the same as that of its metal predecessor. The outline drawing and major dimensions are shown in Figure 1. Comparison of the dimensions of the glass 21CYP22 and the metal 21AXP22-A shows that the major dimensions are the same, and with minor modifications in the mounting system the glass tube may be used as a replacement for the older metal tube; the weight of the glass tube is 33½ pounds. Two terminals are provided on the envelope for an external protective resistor to be used to protect the cathode in case of a momentary internal arc.

PERFORMANCE ADVANTAGES

The performance advantages that result from the new design features incorporated into the glass tube follow:

1. Ease of picture-tube mounting in receiver. Ease of mounting comes about from the use of a glass envelope which reduces insulation problems. Mounting of the 21CYP22 is essentially the same as with a black-and-white glass picture tube.

* Manuscript received May 8, 1958.

2. Ease of set-up. The adjustments unique to color tubes, i.e., purity set-up and convergence, are simplified in the 21CYP22 because the number of corrections needed to obtain optimum performance are reduced.

3. Improved gun performance. Design changes in the internal convergence pole-piece assembly have reduced beam distortion and interaction between convergence pole-pieces and other neck components.

4. Increased light output. The increased light output of the 21CYP22 is due to the use of a graded-hole shadow-mask. The transmission of the mask in the edge regions is the same as that of the shadow-mask used in its predecessor tube; however, by the use of the graded-holes the transmission is gradually increased toward the center so that a 44-per-cent increase in transmission is achieved. The

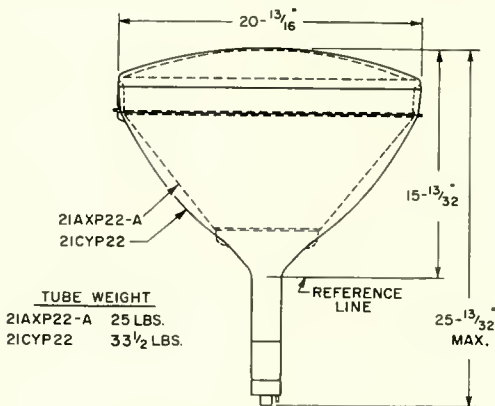


Fig. 1—Comparison of outline drawing and major dimensions for 21CYP22 and 21AXP22-A color picture tubes.

resultant light output is obtained at no sacrifice in register requirements since a lighthouse correction lens has been developed which compensates for the “degrouing” factor which, in previous types, was noncorrected and reduced light output. The “degrouing” factor is explained in more detail later.

5. Improved contrast. This improvement has been achieved by several means. First, the greater electron-beam transmission of the mask allows the use of a filter glass in the faceplate of the 21CYP22 having a slightly lower light transmission. Secondly, the graded-hole shadow-mask also incorporates tapered apertures. Tapering minimizes the scattering of electrons from the sides of the apertures and thereby increases detail contrast and improves color saturation. Third, the use of a masking “electron shield” removes “stray” electrons at the edges and thereby minimizes a source of color dilution.

ENVELOPE FEATURES

The glass envelope used in the 21CYP22 differs from any other envelope used for color-picture tubes to date in that it makes use of a devitrifying frit seal and does not require a reinforcing metal band at the seal edges. As in former color-picture tubes the envelope is made in two parts, the funnel and the panel, and the two parts are sealed together during tube fabrication. The devitrifying frit used in this tube is a material unique to glass envelope manufacture and was developed by Corning Glass Works. The properties of the frit are shown in Table I. The important characteristic of this frit is that it is a soft solder glass until the time of sealing when the glass devitrifies or changes into a ceramic phase. The sealing process is not reversible, and once the seal is made subsequent thermal processing will not soften the seal. The frit, in a paste form held in suspension by an amyl acetate and nitrocellulose binder, is applied from an extrusion machine in a manner similar to toothpaste being squeezed

Table I—Properties of Devitrifying Frit used for Sealing the Funnel and Panel

	VITREOUS	DEVITRIFIED
SOFTENING POINT (°C)	368	510 ± 10
EXPANSION (0-300°C) (10 ⁻⁷ /°C)		96.5 ± 0.5
LOG OF VOLUME RESISTIVITY (OHM-CM.)		
AT 200°C		9.5
AT 300°C		7.8

from a tube. The ribbon of frit is applied to the seal edge of the funnel only. The sealing is accomplished by first allowing the frit to air dry, then positioning the panel in its appropriate position on the funnel, and sending the two parts through a sealinglehr. The thermal cycle involves raising the temperature of the parts to 440°C, holding this temperature for one hour, and then slowly cooling the parts. Upon completion of this process, the panel and funnel are sealed together and subsequently may be put through normal exhaust bakeout.

GLASS PROPERTIES

The properties of the glasses used in the panel and funnel are shown in Table II. In the selection of these glasses, several factors were taken into consideration. First, the use of a glass having a high strain point is necessary so that thermal effects and vacuum loading would cause minimum deformation. Deformations of the faceplate would cause a change of the faceplate contour and result in misregister caused by variations in spacing between mask and faceplate. The

hardness of the faceplate glass used in the 21CYP22 has reduced this distortion and consequent misregister to a negligible value. This reduction of misregister in turn permits better color purity and white uniformity together with greater ease of tube set-up.

The use of two separate glasses, one for the panel and another for the funnel, is necessary so that the faceplate would have a non-browning characteristic and the funnel a high degree of X-ray absorption.

Table II—Properties of Glasses used in 21CYP22

	PANEL	FUNNEL	NECK
SOFTENING POINT (°C)	679-683	668-674	677-681
ANNEALING POINT (°C)	483-488	482-485	482-484
EXPANSION ($10^{-7}/^{\circ}\text{C}$)	97.8-99.4	98.1-98.8	94.7-95.8
LOG OF VOLUME RESISTIVITY (OHM-CM.) @ 250°C	9.18-9.26	9.24-9.28	10.13-10.16

FACEPLATE PANEL FEATURES

In the design of the 21CYP22 envelope a primary aim was an envelope which would permit the design of a tube requiring the least amount of correction for optimum performance. This aim was accomplished by relating the lighthouse referencing to the envelope sealing alignment so that minimum purity correction would be required. Figure 2 illustrates the system used for referencing the panel on the screening lighthouse. It will be noted that the panel rests on the seal edge and is oriented by the V-pad and one of the plain molded-glass pads on the skirt of the panel. These two reference points are positioned during panel-funnel sealing to two ground pads near the seal edge of the funnel. These funnel pads have previously been ground concentric with the inside of the neck. Thus, when the gun is sealed into the neck its position will correspond to the positions of the light sources in the screening lighthouse. This coincidence of the position of the electron beams and the light sources means minimum purity correction will be necessary with this tube. The rotational orientation of the gun is established on the gun sealing machine with reference to the V-pad on the panel.

Another feature of this envelope is the accurate control of the inner contour of the faceplate with respect to the mask mounting studs. The glass envelope has made it possible to obtain a contour, referenced from the studs, accurate to within ± 0.023 inch at any point on the inner surface of the faceplate. This accurate contour together with a similarly accurate contour of the mask results in optimum phosphor-dot size and spacing which again shows up as good color purity and white uniformity.

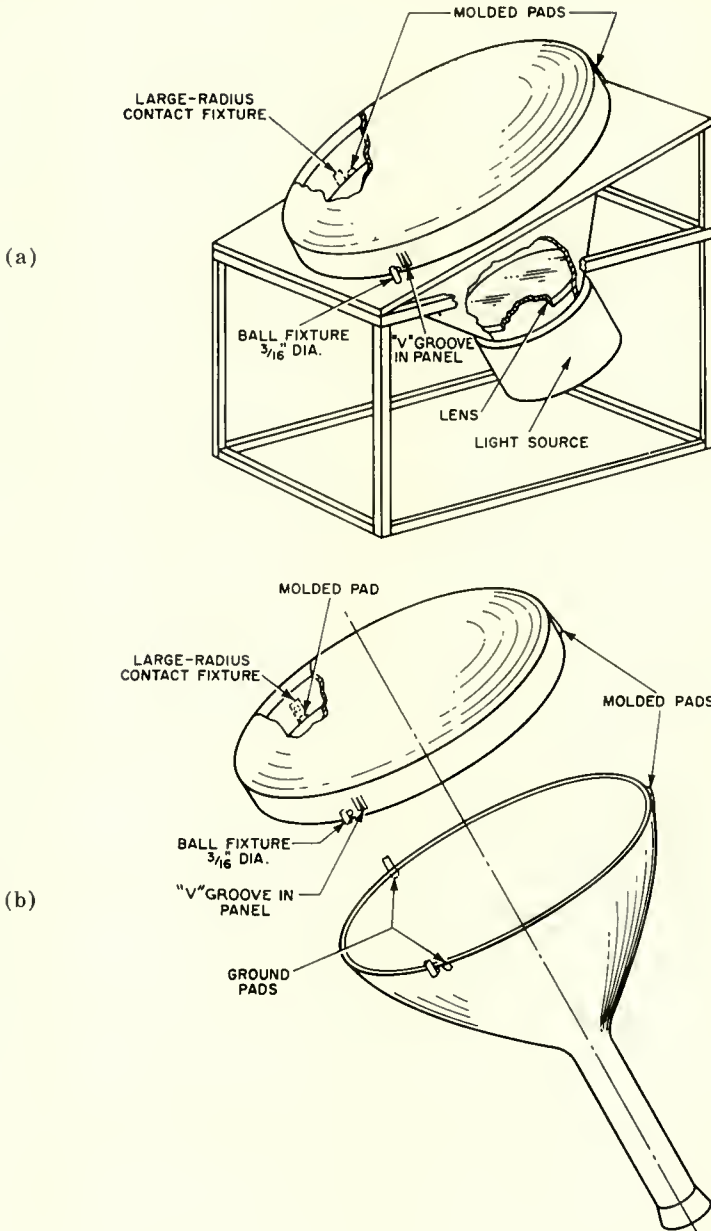


Fig. 2—System used for referencing (a) panel on screening lighthouse and (b) panel on funnel.

SHADOW-MASK CONSIDERATIONS

The shadow-mask is one of the critical parts of the color tube. The methods of etching and forming this part to close tolerances are the basis for the high quality of the 21CYP22. Considerable effort has been expended in research on methods to improve the quality of the mask and reduce its cost.

A significant improvement has been obtained by changing the mask material from 0.0075-inch copper-nickel to 0.006-inch cold-rolled steel. In addition to savings in material cost, the steel mask provides several important advantages. The lower expansion of steel compared to copper reduces the changes during processing and gives a greater degree of stability during tube operation. Although the steel mask is thinner, it is still considerably stronger than the copper-nickel mask. This strength reduces the number of damaged masks due to mishandling during tube fabrication. In addition, the magnetic properties of the steel mask make the demagnetization process, used in tube set-up, more effective.

To eliminate the problem of rusting during tube fabrication and to increase thermal emissivity, the steel mask is steam blackened. This operation is a simplification of the iron plating and chemical blackening formerly used on the copper mask. Processing of the glass tube requires an additional high-temperature cycle because of the funnel-panel sealing. Problems due to flaking of a chemically blackened mask are eliminated by the use of steam blackening.

GRADED-HOLE FEATURE

In the 21CYP22 significant improvements in light output and contrast have been achieved. The gain in light output was made possible by the use of a graded-hole shadow-mask having apertures that increase gradually in diameter from the outer edge of the mask inward to the center. As a result of this gradation, the mask permits increased light output from the screen. Enlargement of the diameter of the apertures was made possible by the adoption of the "degroupping" correction lens in the lighthouse and an associated change in tube geometry. This lens optimizes phosphor-dot placement with respect to electron beam trio.

For a better understanding of why this lens is required, reference is made to Figure 3. For simplicity, the geometry is shown with only two beams in the plane of the axis. As the screen is scanned toward the edge, the convergence angle of the beams must be decreased, thus moving their deflection centers from A to A' and from B to B'. The

triad of fluorescent spots produced by the three beams passing through an aperture is, in effect, a demagnified image of the three beams as cut by the deflection plane. Thus, when dynamic convergence fields are applied, the object in the deflection plane AB is enlarged to A'B' and its image, the fluorescent triad, is correspondingly increased in size. The phosphor screen is exposed from light sources at A and B, and the undeflected beams pass through these color centers. When deflection and dynamic convergence fields are applied, the fluorescent spots are "degroupped" with respect to the phosphor-dot triads.

In the 21CYP22 this effect is greatly reduced by the "degroupping" lens. This lens produces an optical analogy of the action of the electron

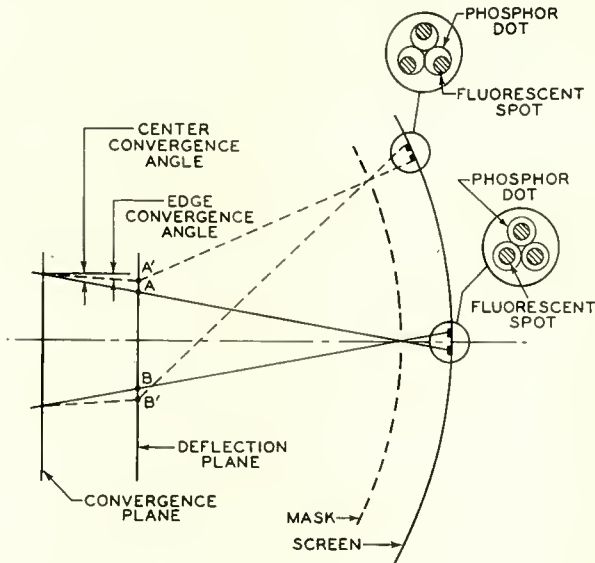


Fig. 3—Geometry of electron beam trio "degroupping."

beams. Thus, the light source does not appear to remain in one position but rather tends to simulate the motion of the beam (A to A' and B to B'). This motion together with the associated change in mask-to-faceplate spacing produces tangent phosphor-dots and proper landing of the electron beam on the dot over the entire screen when dynamic convergence is applied. Removal of this degroupping error allows the diameter of the mask apertures to be increased in size, hence giving greater light output without any decrease in screen purity or uniformity tolerances.

The radial lens previously used in 21AXP22-A to compensate for the axial movement of the deflection plane as a function of the scanning

angle is incorporated into the correction lens for the 21CYP22 lighthouse together with the degrouping feature. Thus, this single lens having two functions provides essentially a complete analogy between the movement of the light source in the screening lighthouse and the apparent motion of the electron beams during tube operation.

Considerable research effort was spent on determining the physical shape of the required degrouping lens, shown in Figure 4, and in obtaining a practical method for manufacturing them. The accuracy with which the lenses have been reproduced permits the use of an interchangeable system of lighthouses, allowing the three colors to be printed on different lighthouses. This interchangeable feature is important in mass production of the tube.

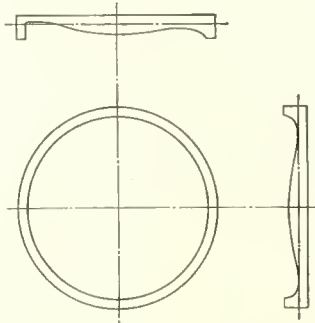


Fig. 4—Contour of "degrouping" lens.

CONTRAST IMPROVEMENT

Another important performance feature of the 21CYP22 is the improvement in contrast. It might be noted that light output and contrast are related features in any picture tube and that a gain in one is as important as a gain in the other. The 21CYP22 incorporates four important design changes that improve contrast. The features are: (1) use of tapered apertures in the mask, (2) elimination of faceplate "haze," (3) introduction of a masking "electron shield," and (4) higher absorption in the faceplate glass.

In former color picture tubes the aperture in the shadow-mask was essentially straight-sided or cylindrically shaped. As the deflection angle increases, thereby causing the beam to approach the mask at some angle other than normal, the electron beam strikes one side of the aperture due to the finite thickness of the mask material. As a result of this striking of the side wall of the aperture, "stray" electrons are scattered and hit phosphor-dots other than those intended.

Because the efficiency of the blue or green phosphor is considerably higher than the red, the "strays" are more noticeable in the red field. The effect of these "strays" is to reduce the saturation of an individual color field and also reduce contrast when the three fields are used to produce white.

In the 21CYP22 this effect has been greatly reduced by the use of conically shaped or tapered apertures. Tapering permits the electron beams to be transmitted through the mask without striking the side wall of the apertures. Figure 5 shows this effect. The degree of taper in the shadow-mask is not the same over the entire mask. It is maximum at the edge and is gradually reduced toward the center. This gradation in taper simplifies mask manufacture while at the same time

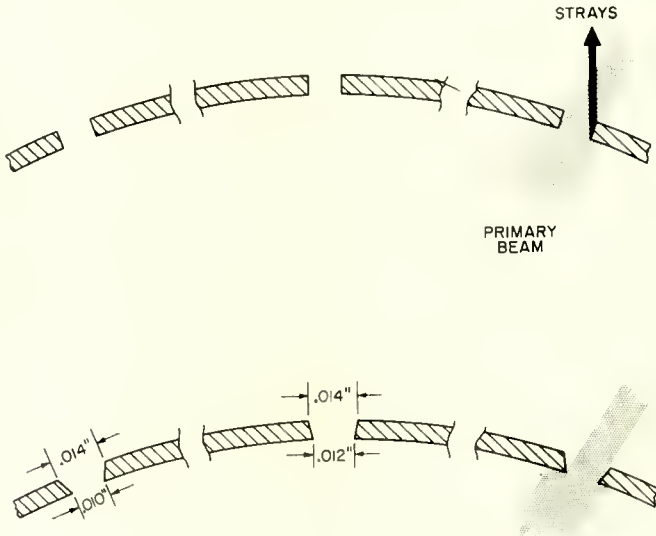


Fig. 5—Advantages of tapered apertures in reducing stray electrons.

provides maximum taper where it is needed, that is, near the edge of the mask where the electron beam approaches the mask at the greatest incident angle. Another advantage of the tapered apertures is the fact that more of the beam is allowed to pass through the aperture as compared to a straight-sided aperture wherein upwards of 20 per cent of the beam is blocked by the side wall. Thus, tapering achieves increased light output as well as improved contrast and color saturation.

STRAY ELECTRONS

Another cause of color dilution in former shadow-mask color picture tubes was stray electrons "sneaking" around the mask-frame assembly,

bouncing off the panel side wall, and striking the phosphor screen. In the glass tube the manufacturing technique requires that the panel side wall be tapered approximately 4 degrees. This taper increases the distance between the frame and side wall. The greater spacing together with the overscan of the raster results in the concentration of a considerable number of stray electrons around the periphery of the screen. In the 21CYP22 an electron masking shield is welded to the under side of the frame closing the gap between the frame and the panel side wall. A small clearance is maintained between the glass wall and the outside edge of the shield to accommodate the differential

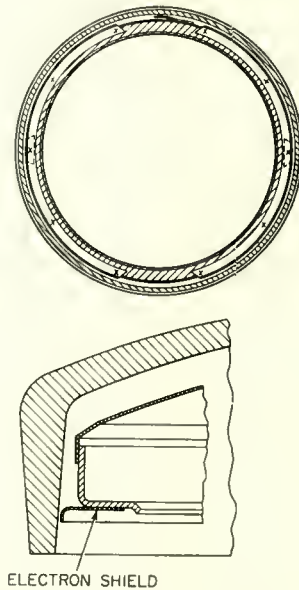


Fig. 6—Electron masking shield.

thermal expansion. The shield is made of thin steel in four sections for ease in manufacture and insertion. An opening in the shield at the top and bottom of the screen permits the required internal electrical connection between funnel and panel. The shield is illustrated in Figure 6.

FACEPLATE HAZE

In the 21AXP22-A a phenomenon which we call faceplate haze is evident. This phenomenon shows up as a tint of specific color when the screen is viewed at wide angles. It was determined that the cause of haze is related to the degrees of optical contact between the

phosphor particles and the inner surface of the faceplate. With a high degree of optical contact, light from the phosphors would be internally reflected within the faceplate glass and appear as a haze. This problem also exists in black-and-white picture tubes. In a color tube, where three different color-emitting phosphors are used, the difference is contact of the three phosphors with the faceplate was such that the haze assumes the color of the phosphor having the best contact. Figure 7 illustrates the mechanism of the internal reflection. In this figure dot No. 1, a phosphor-dot of a particular color, is in poor optical contact with the glass panel. The light rays emitted beyond the critical angle are totally reflected from the glass surface and never enter the glass. Phosphor dot No. 2 is in good optical contact with the glass. As a result, the light rays emitted by this dot do not "see" the interface (vacuum to glass) and all the rays emitted beyond the critical angle are internally reflected within the glass. This internally reflected light is scattered within the glass and leaves the

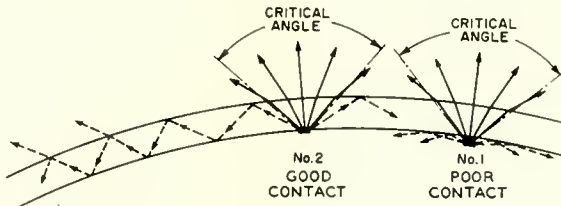


Fig. 7—Mechanism of internal reflections causing "faceplate haze."

glass near the edge of faceplate because of the increasing curvature. If the two dots emit different colors, the color of the dot having the better optical contact will be partially internally reflected and will give a specific color of haze.

In the initial stages of the development of the glass tube it was apparent that the thicker faceplate glass and variations in the curvature near the edge, commonly referred to as "suck up" in the glass panel, caused a haze pattern considerably more intense than that noted in the metal tube with its thinner parallel-surfaced faceplate.

One approach to reduce this effect was the etching of the inner surface of the faceplate to minimize over-all optical contact. This approach had a number of disadvantages both technically and economically. Another approach, which was the solution finally used in the 21CYP22, was to equalize the optical contact between the phosphors by adding a dispersing agent to the red phosphor slurry to increase its contact so that it would be equal to that of the green and blue phosphors.

An additional aid to contrast has been the increased light absorption of the tube faceplate. Although this change tends to decrease light output slightly, the gain in contrast makes up for the slight loss in light output and gives a substantial improvement under high ambient light.

GUN IMPROVEMENTS

Improvements have also been made in the design of the 21CYP22 gun. Although the gun is similar to that used in the earlier metal tube, revisions have been made to improve performance. The radial-converging pole-piece assembly shown in Figure 8 has been revised to provide improved dynamic convergence by means of easing the dynamic convergence adjustments. The shape of the pole-pieces has

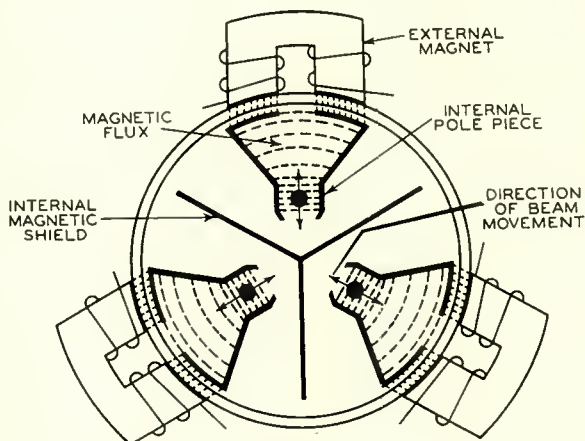


Fig. 8—Schematic of radial-converging pole pieces.

been redesigned and a magnetic "Y" shield has been introduced between them. The purpose of the pole-pieces is to couple the externally provided magnetic flux to the associated beam. In guns of earlier construction interaction occurs so that the field applied to the red gun also influences the green and blue beams. The direction and magnitude of this interaction is such that a field about 30 per cent stronger is required to obtain complete dynamic convergence than would be necessary if no interaction existed.

Another advantage of the "Y" shield is reduction of coupling between the deflecting yoke and the internal pole-pieces. Figure 9 shows how unshielded pole pieces distort the leakage field from the yoke so as to produce a so-called red-green crossover pattern. Formerly this effect had been minimized by shields on the back of the deflecting

yoke. However, with the introduction of the "Y" shield this effect is reduced to a negligible value and yoke shielding is no longer required for this purpose.

To obtain maximum highlight brightness the limiting aperture formerly used in the grid No. 3 is removed. Reduction in the needed magnetic correction applied in the gun region has minimized beam distortion so that the limiting aperture could be removed with no loss in resolution. Since all the electrons leaving the cathode can now reach the mask or screen, highlight brightness is appreciably increased. This condition is particularly important when high grid-

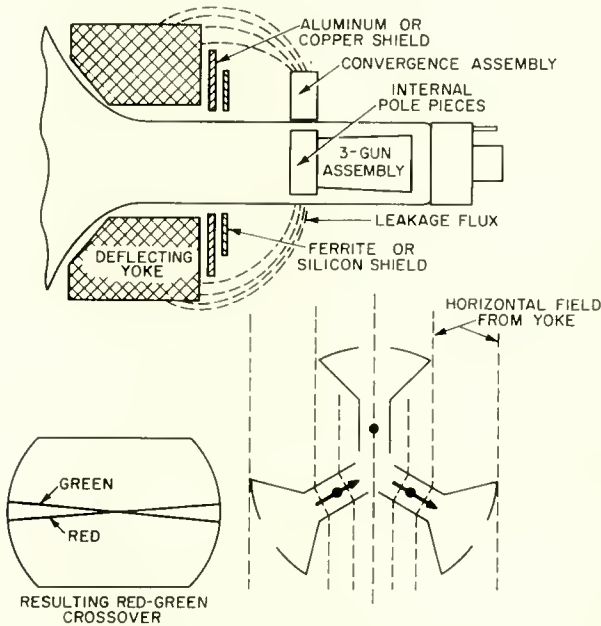


Fig. 9—Red-green crossover pattern caused by unshielded pole pieces.

No. 2 voltages are used because of the increased beam diameter in the grid-No. 3 region.

TUBE MOUNTING

The mounting of the 21CYP22 is simplified, compared to the 21AXP22-A, because of the reduced insulation requirements of the glass envelope. As was noted earlier, the over-all dimensions are similar to those of the metal counterpart and hence the 21CYP22 may be used as a replacement for the metal tube. The mounting of this tube is similar to that of any glass black-and-white picture tube. To

aid in the design of a mounting system specific g ratings have been established and are given in the technical data for this type. Thus, the mounting system may be designed and tested with accelerometers.

It is obvious that the magnetic properties of this tube type are different from the metal tube. To minimize the effects of the earth's field, it is recommended that a magnetic shield made of cold-rolled steel be used around the forward portion of the tube. It is possible to use cold-rolled steel rather than a high-permeability material because a degaussing operation is involved in the set-up of this tube.

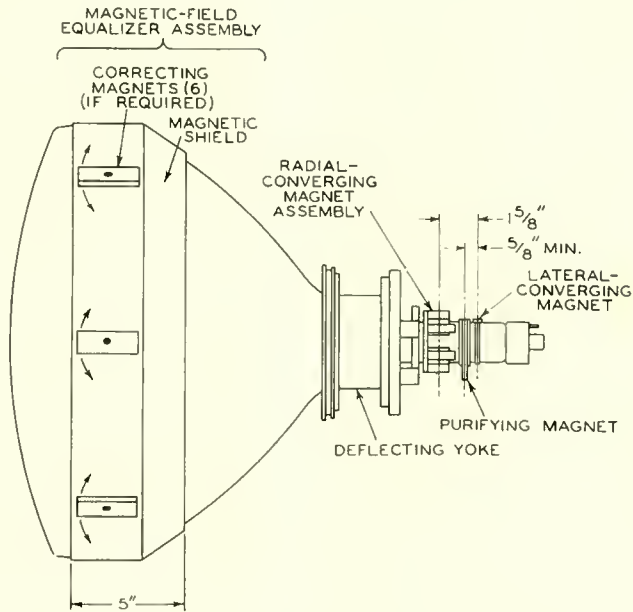


Fig. 10—Sketch showing relative placement of typical components on 21CYP22.

This degaussing operation, which is done when the tube is located in the customer's home, effectively counteracts the earth's magnetic field by setting up an opposing field in the shield as well as in the internal frame and mask of the tube. With this relatively simple shield the effect of the earth's field on the 21CYP22 is not essentially different than on the 21AXP22-A.

The components used on the 21CYP22 (Figure 10) are similar to those of the metal tube with the exception that the magnetic field equalizer assembly is considerably simplified. Because of the mechanical

stability of the glass envelope, the random mechanical errors have been reduced to a point where the need for correction has been greatly minimized. In the 21CYP22, simple snap-on correcting magnets are used. In a majority of cases no magnets are required; however, under adverse field conditions it may be necessary to use several. The components associated with the gun, that is the purifying magnet, the lateral-converging magnet, and the radial-converging magnet assembly, are the same as formerly used. The deflecting yoke is similar except for its mounting and the fact that the "Y" shield used in the pole-piece assembly has reduced the need for shielding on the back of the yoke.

CONCLUSIONS

The 21CYP22 picture tube gives a brighter picture of better contrast than its metal predecessor, with excellent color purity and screen uniformity. The improvements made include: an improved gun and pole-piece assembly, use of a graded-hole shadow-mask, contrast improvement by means of an electron shield, tapered-hole shadow-mask, and increased filtering in the faceplate glass. These design improvements permit superior performance with major simplifications in the tube mounting, receiver circuitry, and set-up technique.

EFFECT OF CRYSTAL GROWTH VARIABLES ON ELECTRICAL AND STRUCTURAL PROPERTIES OF GERMANIUM*†

BY

F. D. ROSI

RCA Laboratories,
Princeton, N. J.

Summary—Examination of dislocation etch-pit distribution revealed plastic deformation by octahedral slip in germanium crystals grown by the Czochralski technique. An intimate association of slip with the total number of dislocations was also found, and this suggested that deformation by slip is the main source of dislocations. Growth conditions which prevent slip by providing a planar solid-liquid interface have yielded 1–40 ohm-cm crystals with dislocation density $<10^2 \text{ cm}^{-2}$. Significant changes in dislocation density and/or carrier lifetime were obtained on varying the temperature gradient in the crystal, growth velocity, crystal orientation, and crystal size. Additions of n- and p-type impurities within their limits of solid solubility had no significant effect on dislocation density. The dependence of carrier lifetime on dislocation density in the range of 10 to 10^5 cm^{-2} was investigated in both high- and low-resistivity crystals. A maximum in carrier lifetime was obtained corresponding to a dislocation density of $3 \times 10^3 \text{ cm}^{-2}$, and this has been attributed to the formation of copper-dislocation recombination centers. The effective cross section for hole capture was determined in 2 ohm-cm crystals for dislocations produced during crystal growth and by plastic deformation. Variations in the number of dislocations produced during crystal growth had significant effects on (1) the precipitation of copper from a supersaturated solid solution, (2) the surface recombination velocity, and (3) the wettability of molten indium on {111} surfaces.

INTRODUCTION

IT IS well recognized that some properties of semiconductors can be greatly influenced by very small concentrations of chemical impurities and structural imperfections. For this reason, rigid requirements are placed on the degree of perfection in semiconductor crystals for use in research and devices. To meet these requirements, it is necessary to understand the effect of variations in the conditions of crystal growth on the density of defects, as well as on structure-sensitive electrical properties.

In this paper, a description will be given of the results obtained

* Presented at AIME symposium on semiconductors held in New York City, February 1956.

† Manuscript received February 19, 1958.

through the specimen, and recording the drop in potential continuously along the specimen. The conductivity type was determined from the sign of the thermoelectric power between the germanium and a metal probe heated to about 50°C.

The pulse photoconductivity method⁶ was used to determine the lifetime of minority carriers. In this technique, the lifetime is obtained directly by measuring the time decay of excess conduction after a pulse excitation of hole-electron pairs by a short flash of light. The specimens were in the form of slabs, $3 \times 3 \times 14$ millimeters, which were cut adjacent to those regions of the crystal used for dislocation studies. Preparation of the specimens for measurement consisted of surface grinding and then mechanically polishing with No. 305 American Optical compound. Electrical contacts were then made by soldering the ends with a 63% Pb-37% Sn alloy. A final etch with a solution of 6 parts HF, 1 part H₂O₂, and 3 parts H₂O removed the cold-worked layer from the mechanical treatment, and any surface film from the soldering process. Moreover, this etch gave a low surface recombination velocity of 13 cm sec.⁻¹, which was determined in order to obtain the true bulk lifetime in the case of measured lifetimes greater than 500 microseconds.

RESULTS AND DISCUSSION

Metallographic Observations

Metallographic examinations were made on a large number of crystals which had been solidified under a wide range of growth conditions. The first notable observation was the occurrence of octahedral slip as evidenced by the alignment of dislocation etch-pits along lines corresponding to the intersection of slip planes with the {111} surface of observation. This feature is strikingly illustrated by the etch-pit patterns in Figure 1, which were observed in crystals grown under conditions of a large thermal gradient in the solid near the solid-liquid interface ($> 50^\circ\text{C cm}^{-1}$). Silicon crystals grown under similar conditions exhibited identical slip patterns, as may be seen from the micrograph of Figure 2. The etchant used to develop dislocation pits on a {111} surface in silicon consisted of 3 parts CP-4 solution with 5 parts of a solution of 200 grams copper nitrate in 2 cm³ nitric acid, 4 cm³ hydrofluoric acid, and 4 cm³ water.⁷ In both germanium and silicon, examination of the other three {111} planes

⁶ D. T. Stevenson and R. J. Keyes, "Measurement of Lifetimes and Diffusion Constants in Germanium," *Bull. Amer. Phys. Soc.*, Vol. 29, p. 18, March, 1954.

⁷ S. M. Christian and R. V. Jensen, "Dislocation Etch Pits on Silicon," *Bull. Amer. Phys. Soc.*, Vol. 1, p. 140, March, 1956.

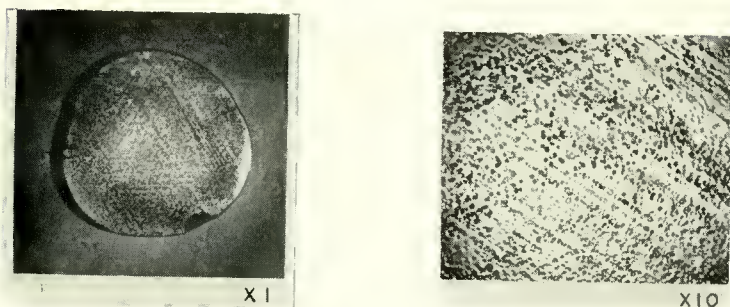


Fig. 1—Appearance of multiple octahedral slip as outlined by chemical etch-pits on a $\{111\}$ surface of a germanium crystal at different magnifications.

revealed the same slip pattern as in Figures 1 and 2 with the noteworthy exception that the pits were never found to align themselves along the trace of the $\{111\}$ plane normal to the axis of growth. Additional evidence of slip was found on examining the solid-liquid interface, which was exposed by rapidly pulling the crystal away from the melt during growth. Figure 3 shows the typical appearance of intersecting slip bands on such an exposed surface.

It may be concluded from the foregoing observations that deformation by slip is an important source of dislocations in Czochralski-grown crystals. This conclusion was, at first, unexpected, since in this growth technique there is no mechanical constraint imposed on the crystal, as in the case of crystal growth in a horizontal boat.

To determine the extent to which slip contributes to the measured dislocation density, it was important to grow crystals under conditions which would inhibit the motion of dislocations and yield a relatively low density of slip lines. It was found that this could be accomplished by growing crystals of different diameter with a steep temperature gradient in the solid. Since the thermal movement of edge dislocations

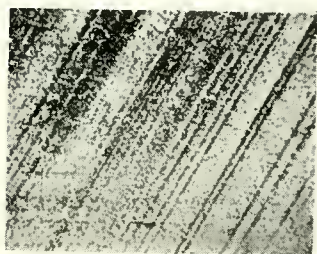


Fig. 2—Appearance of multiple octahedral slip as outlined by chemical etch-pits on a $\{111\}$ surface of a silicon crystal. $\times 15$.

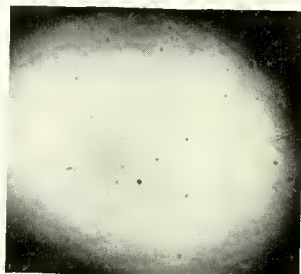


Fig. 3—Appearance of multiple $\{111\}$ slip lines on as-withdrawn solid-liquid interface of germanium crystal. $\times 50$.

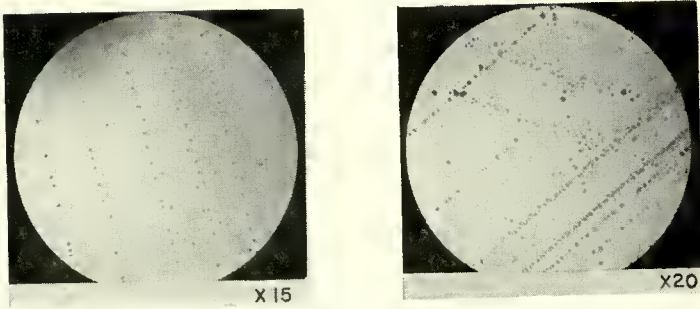


Fig. 4—Association of dislocation etch-pits with octahedral slip of different density on $\{111\}$ surface of germanium crystal.

occurs by a "climb" mechanism involving the diffusion and condensation of excess vacancies on the dislocations, the presence of a steep temperature gradient would allow less time at the high temperatures for the diffusion of vacancies to occur. Thus, the dislocations generated by slip might be expected to remain at the site of their formation or, at least, along their own slip planes.

Examinations of etch-pit distribution were made on a number of crystals which exhibited a low density of slip lines and were grown with a temperature gradient in the solid of $\sim 100^\circ\text{C cm}^{-1}$. Areas of low slip density were found by examining various regions along the crystal and by growing crystals of different diameter. With but few exceptions, it was found that not only did the dislocations appear preferentially along the slip lines, but, more important, the areas between slip lines were devoid of dislocation etch-pits, as may be seen from the micrographs of Figure 4. This intimate association of slip with the total number of edge-type dislocations strongly suggests that slip is the main source of dislocations in germanium crystals

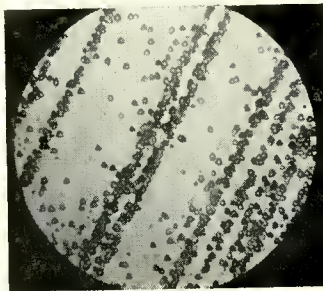


Fig. 5—Association of dislocation etch-pits with octahedral slip on $\{111\}$ surface of silicon crystal. Note that alignment of triangular pits is such that a $\langle 110 \rangle$ edge is parallel to slip lines. $\times 40$

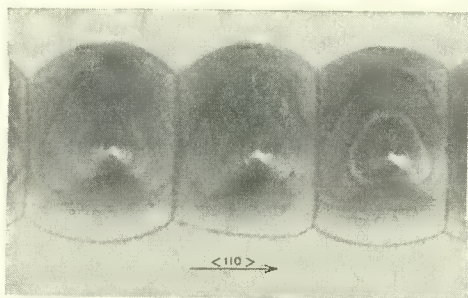


Fig. 6—Alignment of dislocation etch-pits along a slip line in germanium crystal at high magnification. $\times 250$

grown from the melt by the Czochralski technique. Figure 5 shows that this characteristic association is also encountered in silicon crystals under similar conditions of growth. The triangular character of the etch-pits in Figure 5 clearly illustrates their alignment along a $\langle 110 \rangle$ direction consistent with the crystallographic geometry of dislocations along slip lines. This is even more strikingly illustrated in Figure 6 by the magnified etch pits along a slip line in germanium. It is noteworthy that the central dot of the etch-pits is not symmetrical with respect to the triangular terraces, and this could be a manifestation of the extra half-plane characteristic of edge dislocations.

It follows from the above arguments that in crystals grown with a small temperature gradient in the solid, the dislocation movement could be sufficient to yield a random distribution of dislocations in contrast to the crystallographic alignment in Figure 4. In general, this was found to be the case, as evidenced by the micrographs in Figure 7 taken from a germanium and silicon crystal grown with a temperature gradient of $\sim 40^\circ\text{C cm}^{-1}$.

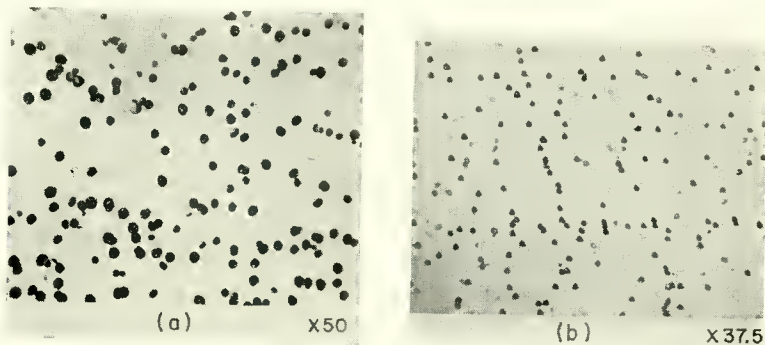


Fig. 7—Random distribution of dislocation etch-pits on $\{111\}$ surfaces of crystals grown with a temperature gradient in the crystal of $\sim 40^\circ\text{C cm}^{-1}$: (a)—germanium, (b)—silicon.

More striking evidence of the increased mobility of edge dislocations in the presence of a low temperature gradient is given by the examples in Figure 8 of both germanium and silicon crystals. In both micrographs the dislocations have aggregated during growth into arrays normal to the slip lines similar to those observed by Vogel⁸

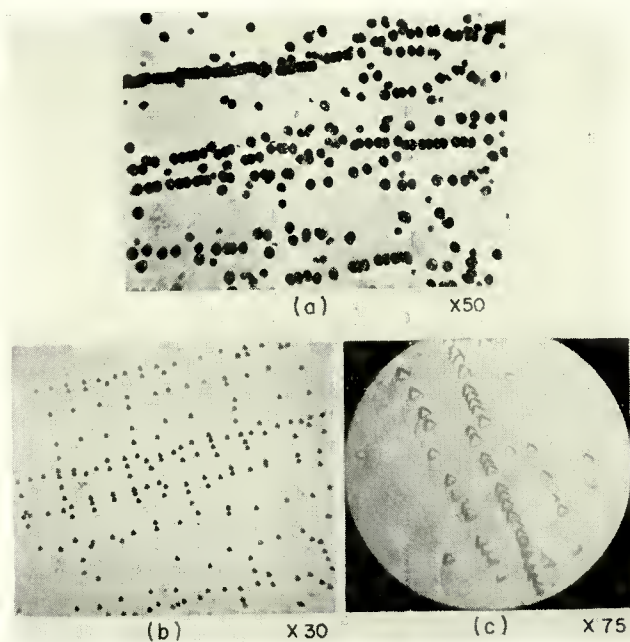


Fig. 8—Formation of low angle boundaries or polygonized structure on {111} surface of crystals grown with a temperature gradient in the crystal of $10\text{--}20^\circ\text{C cm}^{-1}$: (a)—germanium, (b) and (c)—silicon. Note that arrangement of triangular etch-pits along boundaries in silicon crystals is normal to former slip lines (compare with Figure 5).

in bent germanium crystals. These arrays represent low angle boundaries or a polygonized structure,^{9,10} and correspond to the minimum energy configuration that the dislocations can assume. Their formation necessarily involves dislocation climb which was found by metallographic examination to occur in a {110} plane parallel to the [111]

⁸ F. L. Vogel, "Dislocations in Polygonized Germanium," *Acta Metallurgica*, Vol. 3, p. 95, January, 1955.

⁹ R. W. Cahn, "Slip and Polygonization in Aluminium," *Jour. of Metals*, Vol. 19, p. 129, January, 1951.

¹⁰ E. Tegtsoonian and B. Chalmers, "The Macromosaic Structure of Tin Single Crystals," *Canadian Jour. of Phys.*, Vol. 29, p. 370, September, 1951; "Further Observations on the Macromosaic Structure of Tin Single Crystals," *Canadian Jour. of Phys.*, Vol. 30, p. 388, September, 1952.

direction of growth. Since the climb process depends upon self-diffusion to supply the required vacancies, the likelihood of formation of low angle boundaries by this mechanism increases as the length of time the temperature is held close to the melting point increases. Thus, the appearance of dislocations in random distribution across the crystal and at low angle boundaries can be regarded as manifestations of the increased mobility of dislocations in the presence of a low temperature gradient in the solid. The dislocations, however, could still have been generated initially by a slip process. It should be mentioned here that the alignment of the triangular etch pits in the silicon crystal of Figure 8b is in accordance with the geometry of

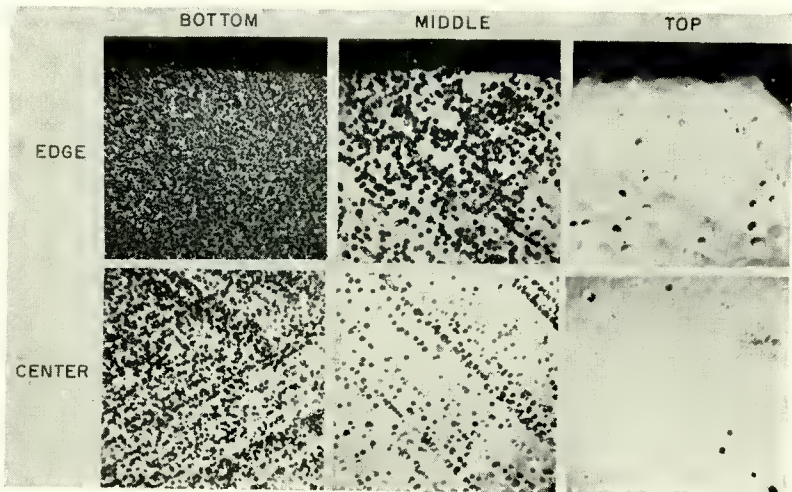


Fig. 9—Radial and longitudinal variations in density of dislocation etch-pits in a 1-inch diameter germanium crystal. $\times 14$

dislocations at low angle boundaries. This should be compared with the alignment of similar etch-pits along the slip lines in Figure 5.

Another interesting feature in connection with slip during growth was the variation in the density of slip lines and, hence, dislocations across the diameter and along the length of the crystal. An extreme example of this variation is shown in Figure 9, which clearly reveals a marked increase in the density of etch-pits on going from the top to the bottom and from the center to the edge of the crystal. This heterogeneity was encountered particularly in large diameter (~ 1 inch) crystals on which a gas jet of helium or hydrogen was directed just above the solid-liquid interface during growth. In such crystals, as in Figure 9, it was not uncommon to find variations in pit density

of 10^2 to 10^5 cm^{-2} from top to bottom of the crystal. A knowledge of the existence of such variations, and their avoidance through selected growth conditions are important to the fabrication and performance of transistors and transistor-like devices.

Figure 10 reveals the radial etch pit pattern in a silicon crystal grown under conditions similar to the germanium crystal in Figure 9. The gradual decrease and disappearance of pits outlining the $\{111\}$ slip traces on approaching the center of the crystal strongly suggest that the slip occurs preferentially near the crystal surface. This slip pattern further indicates that the crystal must contain composite-type dislocations (screw and edge components), which have been discussed by Ellis.¹¹

It should be remembered that in the usual Czochralski technique

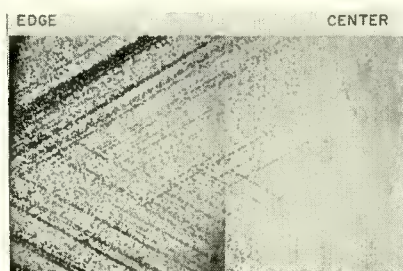


Fig. 10—Radial distribution of dislocation etch-pits and slip lines on $\{111\}$ surface of silicon crystal. $\times 12$

of crystal growth the level of the melt and, hence, the environment of the crystal are constantly changing with growth. This is reflected in a changing temperature gradient in both the crystal and melt which, in turn, affects the nature of the solid-liquid interface. Such variations in temperature gradients can give rise to thermal stresses in the crystal which are sufficiently large to induce slip and, therewith, dislocations. These thermal effects are expected to be exaggerated in large-diameter crystals, where it is most difficult to maintain linear thermal currents both radially and longitudinally under conditions of changing gradients. It would be interesting, therefore, to see whether a direct relationship exists between the shape of the solid-liquid interface and the etch-pit patterns of Figures 9 and 10.

The topography of the solid-liquid interface was studied in most cases by rapidly withdrawing the crystal from the melt. This technique clearly reveals the gross curvature of the interface as well as

¹¹ S. G. Ellis, "Dislocations in Germanium," *Jour. Appl. Phys.*, Vol. 26, p. 1140, September, 1955.

small irregularities. Its singular disadvantage is that a portion of the interface is obscured by the presence of a liquid drop (see Figure 11) retained by surface tension on withdrawal. The drop usually takes on a conical shape, since germanium expands on freezing. In a few instances the curvature of the interface was determined by the formation of p-n electrical junctions within the crystal. This was accomplished by simply adding the desired amount of p- or n-type



Fig. 11—Planar solid-liquid interface as revealed by rapid withdrawal of small diameter crystal from the melt. Conical protuberance is result of liquid drop retained by crystal surface on withdrawal. $\times 3$

impurity to the melt at various times during growth. By sectioning the crystal longitudinally and then copper plating, the interface was outlined at each junction since the p and n regions have different current densities. The usefulness of this technique was first suggested by the work of Camp¹² on resistivity striations in germanium crystals.

Examinations of the solid-liquid interface were made on a large number of crystals exhibiting the type of dislocation distribution

¹² P. R. Camp, "Resistivity Striations in Germanium Crystals," *Jour. Appl. Phys.*, Vol. 25, p. 459, April, 1954.

shown in Figures 9 and 10. The results of this study are summarized in the schematic drawings of Figure 12, which show that a correlation exists between these two physical features. Those parts of the crystal (shaded areas) exhibiting a high pit and slip line density (see, for example, Figures 9 and 10) were invariably associated with the highly curved portion of the interface. This was the case regardless of whether the interface was convex or concave toward the solid. In the

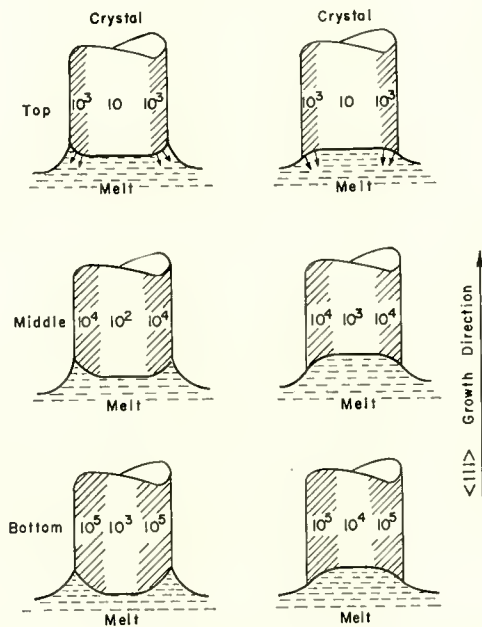


Fig. 12—Schematic representation of radial and longitudinal variations in dislocation density in large diam. germanium and silicon crystals produced from the melt by the Czochralski technique. The numerical figures in the shaded and clear regions indicate the dislocation density per cm^2 , and the arrows in the upper drawings indicate direction of dislocation propagation with advance of the solid-liquid interface.

central portion of the crystal where the interface was relatively flat, the density of pits was very low. Since a curved interface is indicative of nonparallel thermal currents in the crystal, thermal stresses can be expected in the shaded regions of Figure 12 as a result of unequal thermal expansion of adjacent parts of the crystal on cooling. It is believed that the relaxation of those stresses through plastic flow by slip is responsible for the marked gradients in dislocation density shown in Figures 9 and 10. It follows that the higher density of dislocations at the bottom of the crystal is attributed to a more curved

solid-liquid interface, as well as to the propagation of dislocations formed in the earlier stages of growth. The greater curvature results from the expected increased inhomogeneity of heat flow in the crystal near the end of growth, in addition to the contribution of a marked radial gradient of dislocations already present in the crystal.

In the schematic drawings of Figure 12, the notable difference in the dislocation pattern for the convex and concave interfaces is the gradual disappearance with growth of the radial dislocation gradient for the convex case. Since dislocations formed at a solid-liquid interface tend to develop perpendicular to that interface, this difference can be attributed to the difference in the direction of dislocation propagation as indicated by the arrows in the upper diagrams in Figure 12.

It follows from the foregoing that the density of dislocations, as revealed by CP-4 etching, could be greatly reduced by growing crystals under conditions which provide an essentially planar solid-liquid interface where the radial temperature gradient is essentially zero. Such conditions were obtained in $\frac{1}{4}$ -inch-diameter crystals with a temperature gradient along the crystal of $\sim 150^\circ\text{C cm}^{-1}$ (see Figure 11), and in $\frac{3}{8}$ to $\frac{1}{2}$ inch diameter crystals with a temperature gradient of $\sim 10^\circ\text{C cm}^{-1}$. The steep gradient was obtained by directing a jet of helium or hydrogen gas on the crystal just above the solid-liquid interface and the low gradient was obtained by drawing the crystal through a nichrome-wound, quartz tube furnace maintained at $\sim 800^\circ\text{C}$ and positioned just above the crucible. In both cases, the dislocation density along a 4-inch length of crystal was generally $< 10^2 \text{ cm}^{-2}$, and the same low density was found on examining all four $\{111\}$ planes. In the small-diameter crystals with a high thermal gradient, densities lower than 10 cm^{-2} were frequently obtained; and, in fact, by using a seed crystal from the 10 cm^{-2} material, one could obtain crystals with no dislocation etch pits. These low dislocation densities should be compared with the range of values, 9×10^5 to 10^7 cm^{-2} , recently reported by Kurtz and Kulin¹³ also in Czochralski-grown crystals. Since the work of these investigators show agreement between dislocation densities determined by direct counting of CP-4 etch-pits and by X-ray broadening, it would appear that the difference between their values and those found in the present study are due to genuine differences in crystal quality. It is further noteworthy that Kurtz and Kulin reported no evidence of extensive slip, nor marked gradients in dislocation density in any of their crystals.

¹³ S. S. Kulin and A. D. Kurtz, "Effect of Dislocations on Minority Carrier Lifetime in Germanium," *Acta Metallurgica*, Vol. 2, p. 354, March, 1954.

The above results clearly point to the importance of maintaining thermal conditions near the solid-liquid interface constant during growth in order to obtain uniform and controlled dislocation density along the entire length of the crystal. In the Czochralski technique this can be accomplished by a feed-in-pull-out arrangement, whereby the volume and composition of crystal being withdrawn from the melt is fed back into the melt at the same rate from a feed bar.¹⁴ Under such conditions, the level of the melt and the crystal environment do not vary significantly during growth, and, at the same time, a means is provided for continuous crystal growth. With feed-in-pull-out furnaces it has been possible to produce long crystals with uniform dislocation density and resistivity.

In some of the crystals, where the dislocation density was $< 10^2$ cm^{-2} , shallow etch-pits of circular geometry were observed. The

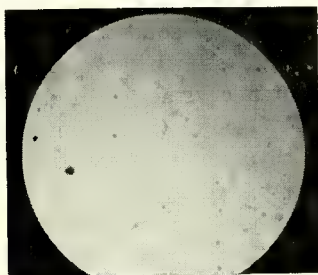


Fig. 13 — Appearance of shallow, circular etch-pits on {111} surface of germanium crystal with a dislocation density $< 10^2$ cm^{-2} . The larger etch-pit on the left is a typical CP-4 edge dislocation pit. $\times 50$

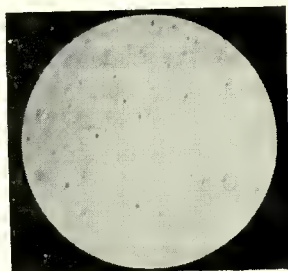


Fig. 14—Protuberances appearing as dark spots on as-withdrawn solid-liquid interface of germanium crystal. Faint lines in background are {111} slip. $\times 50$

micrograph in Figure 13 reveals the character of these pits, which should be compared to the typical CP-4 dislocation etch-pit appearing on the left side of the micrograph. In general, the circular pits do not contain the center dot which is characteristic of the CP-4 pits. In the few instances where a dot was observed, it disappeared on further etching. It is also significant that the density of these pits ranges from 10^2 to 10^3 cm^{-2} , and that progressive etching results in the disappearance of some pits and the appearance of new ones. These observations suggest that, if these are edge-type dislocation pits, then the dislocations are of short length and, hence, of a composite variety, i.e., screw-edge components.

¹⁴H. Nelson, "Continuous-Process Apparatus for Growing Single-Crystal Germanium," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, p. 66.

One final observation of interest in connection with the metallographic work was the appearance of protuberances on the solid-liquid interface when exposed by rapid withdrawal from the melt. The dark areas in the micrograph of Figure 14 outline such protuberances, which were found by shadow casting techniques to have a crater-like geometry. It is conceivable that these small regions growing in advance of the general solid-liquid interface are a manifestation of growth from screw dislocations which have been observed in germanium by Ellis¹¹ and in silicon by Christian and Jensen.⁷

Influence of Growth Conditions

The influence of various growth conditions on the density of dislocation etch-pits and the lifetime of minority carriers was also investigated. The property of carrier lifetime was selected not only because of its importance to device technology,¹⁵ but also because it is sensitive to defect concentrations $< 10^{13} \text{ cm}^{-3}$. Moreover, it was hoped that a correlation could be drawn between this property and dislocation density. In examining a specific effect, extreme care was taken to maintain all other conditions of growth constant, and, whenever possible, those conditions were used which would yield a low dislocation density. Except where otherwise noted, the crystals were grown with a $\langle 111 \rangle$ axis in a purified helium atmosphere, and with a diameter and length of approximately $\frac{3}{4}$ and 4 inches, respectively. The values of dislocation density and carrier lifetime were determined from cross-sectional slabs cut $1\frac{1}{2}$ inches from the seed end.

Effect of Temperature Gradient

The temperature gradient along the crystal axis near the solid-liquid interface was varied from 10°C to $220^\circ\text{C cm}^{-1}$ by changing the position of the induction heating coil relative to the level of the melt, by placing an auxiliary furnace around the crystal just above the crucible, by directing helium gas jets on the crystal in the vicinity of the solid-liquid interface, by using a water-cooled seed holder, and by using combinations of these techniques. The temperature gradient was determined by means of a thermocouple positioned 1 cm above the solid-liquid interface in the center of a test crystal, $\frac{3}{4}$ inch in diameter and $1\frac{1}{2}$ inches long. This was accomplished by supersonically boring a hole diagonally into the test specimen from its side.

¹⁵ P. Rappaport, J. Loferski and E. G. Linder, "The Electron-Voltaic Effect in Germanium and Silicon P-N Junctions," *RCA Review*, Vol. XVII, p. 100, March, 1956.

The data in Table I summarize the effect of temperature gradient on dislocation density and carrier lifetime in crystals which were grown at a speed of 1 millimeter per minute and rotated at 25 revolutions per minute. It may be seen that the dislocation density generally increases as the temperature gradient is increased. This might be expected, since it became more difficult to obtain uniform thermal currents in these $\frac{3}{4}$ -inch diameter crystals at the higher temperature gradients, as indicated by examining the shape of the solid-liquid interface.

The results in Table I also show an increase in carrier lifetime with increasing temperature gradient in spite of the accompanying rise in dislocation density from 1×10^2 to 4×10^3 cm^{-2} .

Table I—Effect of Temperature Gradient in Crystal, $(dT/dz)_s$, on Edge Dislocation Density and Carrier Lifetime in n-type Germanium.

Crystal No.	$\frac{dT}{dz_s}$, °C cm^{-1}	Edge Dislocation Density, cm^{-2}	Lifetime, μ sec.	Resistivity ohm-cm
1	10	1×10^2	800	38
2	40	2×10^2	1000	39
3	110	4×10^2	1100	41
4	160	5×10^2	1500	40
7	200	2×10^3	1800	40
8	220	4×10^3	1900	43

Effect of Growth Rate

The influence of growth rate on dislocation density and carrier lifetime was investigated in crystals grown with a rotational speed of 25 revolutions per minute and a temperature gradient along the crystal axis near the solid-liquid interface of $\sim 200^\circ \text{cm}^{-1}$. A steep temperature gradient was desirable in order to study a wide range of growth speed. This follows from the knowledge that the rate-determining factor in the growth of a crystal from the melt is the dissipation of the latent heat of fusion, L , and the heat, H , arriving at the solid-liquid interface from the melt, and that the removal of this heat is largely controlled by the temperature gradient in the crystal, $(dT/dz)_s$, near the solid-liquid interface. Hence, a cylindrical crystal of uniform cross section A will grow at a rate dz/dt , given by

$$DLA \left(\frac{dz}{dt} \right) = A K_s \left(\frac{dT}{dz} \right)_s - H,$$

where D = density of the crystal and K_s = thermal conductivity of the crystal near its melting point. It also follows that the maximum growth rate is obtained when the temperature gradient in the melt is zero and all the latent heat of fusion is conducted into the solid. Thus,

$$\left(\frac{dz}{dt} \right)_{\max} = \frac{K_s}{DL} \left(\frac{dT}{dz} \right)_s ;$$

Table II — Effect of Growth Velocity on Edge Dislocation Density and Carrier Lifetime in n-type Germanium.

Crystal No.	Growth Velocity, mm min ⁻¹	Edge Dislocation Density, cm ⁻²	Lifetime μ sec.	Resistivity ohm-cm
18	0.1	8×10^3	1300	41
19	0.5	6×10^2	1500	42
7	1.0	3×10^3	1800	40
34	2.0	9×10^2	2700	42
35	4.0	5×10^3	2000	41
36	6.0	2×10^4	1200	38

and for a $(dT/dz)_s$ value of $\sim 200^\circ\text{C cm}^{-1}$ the maximum rate of growth is 7.9 mm min^{-1} , if one takes $K_s = 0.034 \text{ cal sec}^{-1} \text{ cm}^{-1} \text{ }^\circ\text{C}^{-1}$ *, $D = 5.29 \text{ gm cm}^{-3}$, and $L = 98.3 \text{ cal gm}^{-1}$.

Growth speeds in the range 0.1 to 8 mm min⁻¹ were studied with a $(dT/dz)_s$ of $\sim 200^\circ\text{C cm}^{-1}$. The results in Table II show a marked increase in carrier lifetime with no significant change in the dislocation density as the growth velocity is increased from 0.1 to 2 mm min⁻¹. At higher growth rates, however, the lifetime decreases sharply, and this is accompanied by a systematic increase in the dislocation density. It is noteworthy that at these higher rates of growth etch-pit examinations along the crystal revealed evidence of a lineage structure, which became more pronounced as growth proceeded. This was particularly evident in the crystals grown at speeds of 6 mm min⁻¹. The

* The thermal conductivity of germanium at its melting point was determined by using Goldsmid's¹⁶ value of $0.1 \text{ cal sec}^{-1} \text{ cm}^{-1} \text{ }^\circ\text{C}^{-1}$ at 100°C , and then extrapolating to the melting point by assuming a T^{-1} dependence.

¹⁶ H. J. Goldsmid, "On The Thermal and Electrical Conductivity of Semiconductors," *Physical Society Proc., Section B*, Vol. 67, p. 360, 1954.

micrographs in Figure 15 show typical examples of this lineage structure, which is characterized by a layered hexagonal network consisting of a high density of etch-pits. In the case of the growth at 8 mm min^{-1} , some difficulty was encountered in maintaining crystal contact with the melt, and this condition would indicate that

$$K_s \left(\frac{dT}{dz} \right)_s > DLA \left(\frac{dz}{dt} \right).$$

Moreover, metallographic examination of the crystal at this growth speed revealed definite evidence of grain boundary formation, as shown in Figure 16. Thus, it would appear that for a $(dT/dz)_s$ of $200^\circ\text{C cm}^{-1}$, the limiting growth velocity is between 6 and 8 mm min^{-1} , which is in agreement with the above calculated value of 7.9 mm min^{-1} .

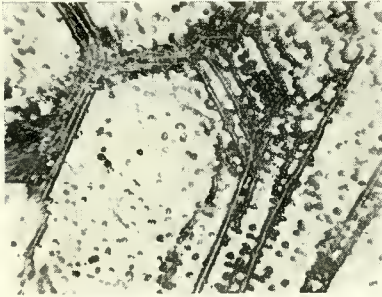


Fig. 15—Hexagonal lineage structure on $\{111\}$ surface of germanium crystal grown at 6 mm min^{-1} . $\times 50$

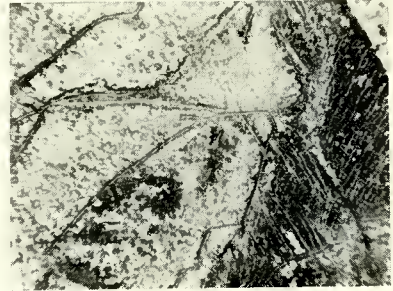


Fig. 16—Grain boundary formation in germanium crystal grown at 8 mm min^{-1} . $\times 50$

X-ray analysis of the geometric etch-pit pattern in Figure 15 showed that the rows of etch-pits correspond to $\langle 110 \rangle$ directions. This would suggest that the growth at these speeds is of the equilibrium form discussed recently by Ellis.¹¹ Hence, the solid-liquid interface is bounded by $\{111\}$ planes whose intersections correspond to the $\langle 110 \rangle$ hexagonal etch-pit alignment. Thus, the lineage structure in Figure 15 represents the troughs between macromosaic equilibrium-growth structures, as shown schematically in Figure 17. The mechanism leading to the preferential concentration of dislocations in the troughs to form small-angle boundaries has been discussed by Ellis,¹¹ and is attributed to the tendency for the dislocations to shorten their length and, hence, their energy by moving in the direction indicated in Figure 17 with advance of the interface. The layered hexagonal pattern could then be a result of the growth of one macromosaic,

equilibrium structure at the expense of adjacent, smaller structures with advance of the interface, as schematically shown in Figure 17. A somewhat similar competitive growth process was described in detail by Weinberg and Chalmers¹⁷ in analyzing cellular structures in metals.

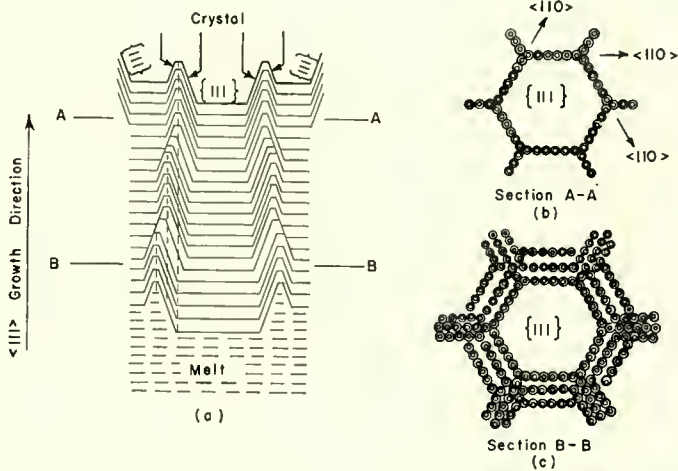


Fig. 17—Schematic illustration of the hexagonal lineage structure of Fig. 15: (a)—Dislocations migrate, as indicated by arrows, in the troughs between $\{111\}$ equilibrium-growth structures to form low angle boundaries. Position of troughs changes with crystal growth, and leads to formation of more low angle boundaries which propagate along dashed lines. This results in the layered hexagonal structure shown in (c). The concentric circles in (b) and (c) represent CP-4 dislocation etch-pits.

Effect of Rotational Speed

The influence of the speed of crystal rotation on carrier lifetime and dislocation density was investigated in crystals grown at a rate of 1 mm min^{-1} and with a temperature gradient in the crystal near the interface of approximately 40°C cm^{-1} . In addition to the usual high purity material (35-40 ohm-cm), the study was extended to include n-type arsenic-doped material of approximately 4 ohm-cm. The data in Table III show that in the range of speeds from 0 to 100 revolutions per minute, no variation in either dislocation density or carrier lifetime was observed. The distribution of dislocation etch-pits also showed no systematic variation.

¹⁷ F. Weinberg and B. Chalmers, "Further Observations on Dendritic Growth in Metals," *Canadian Jour. of Phys.*, Vol. 30, p. 488, October, 1952.

Table III — Effect of Rotational Speed on Edge Dislocation Density and Carrier Lifetime in n-type Germanium.

Crystal No.	Rotational Speed, r.p.m.	Edge Dislocation Density, cm^{-2}	Lifetime, μ sec.	Resistivity ohm-cm
20	0	3×10^2	1400	36
13	5	3×10^2	1300	39
16	50	2×10^2	1300	38
22	0	4×10^2	300	4
14	5	5×10^2	280	3
17	50	3×10^2	270	3.5

Effect of Atmosphere

In areas of material research both in this country and abroad, crystals of germanium are grown from the melt under different atmospheres. To determine whether this variable has any effect on carrier lifetime and dislocation density, crystals of both high and low resistivity were grown in purified argon, helium, and hydrogen and in a vacuum of 1×10^{-5} mm Hg under similar conditions of growth velocity, rotational speed, and temperature gradient in the crystal. From the data in Table IV it is apparent that no significant variation in dislocation density was obtained for both resistivities. Similarly, the carrier lifetime also appears to be independent of the type of atmosphere used with the possible exception of the high resistivity, vacuum-grown crystal which exhibits a somewhat lower lifetime.

Table IV — Effect of Atmosphere During Crystal Growth on Edge Dislocation Density and Carrier Lifetime in Germanium.

Crystal No.	Atmosphere	Edge Dislocation Density, cm^{-2}	Lifetime, μ sec.	Resistivity ohm-cm	Type
6	Argon	3×10^3	1200	45	n
15	Helium	1×10^3	1400	43	n
12	Hydrogen	7×10^2	1400	41	n
41	Vacuum	2×10^3	900	40	n
21	Argon	8×10^2	350	5	n(As)
26	Helium	2×10^3	400	4	n(As)
24	Vacuum	1×10^3	350	6	n(As)

Effect of Crystal Orientation

It was shown in the section on metallographic observations that {111} slip is largely responsible for edge-type dislocations as outlined by CP-4 etch-pits, and that the slip is most likely caused by stresses resulting from unequal thermal expansion of contiguous areas of the crystal. Furthermore, examinations of etch-pit patterns on the four {111} planes of the crystal revealed no evidence of slip on the {111} plane normal to the growth direction. This suggested that the principal stress responsible for slip is in the direction of growth or perpendicular to it, since in either condition the resolved shear stress factor on this plane is a minimum.

In view of these considerations, the number of available slip planes and, hence, the density of dislocations would be expected to vary with crystal orientation. Based on this argument and on the assumption that the principal stress is in the direction of growth, the density

Table V—Effect of Crystal Orientation on Edge Dislocation Density and Carrier Lifetime in n-type Germanium

Crystal No.	Crystal Axis	Edge Dislocation Density, cm^{-2}	Lifetime, μ sec.	Resistivity ohm-cm
33	$\langle 100 \rangle$	3×10^3	1900	38
7	$\langle 111 \rangle$	2×10^3	1800	40
32	$\langle 110 \rangle$	5×10^2	1300	36

of dislocations should increase as the growth axis changes from a $\langle 110 \rangle \rightarrow \langle 111 \rangle \rightarrow \langle 100 \rangle$ direction. It may be seen from the data in Table V that the $\langle 110 \rangle$ orientation does give a lower dislocation density than either the $\langle 111 \rangle$ or $\langle 100 \rangle$ under identical conditions of growth. In these crystals the density of dislocations was determined for the four {111} planes, and then averaged to give the values in Table V.

It is noteworthy that the carrier lifetime is somewhat higher in the $\langle 100 \rangle$ and $\langle 111 \rangle$ crystals in spite of their higher dislocation density. It will be remembered that the same relationship was observed in studying the effect of thermal gradient (see Table I).

It should be mentioned here that in the $\langle 100 \rangle$ crystal the dislocation density, as measured by counting etch-pits on the {100} surface normal to the growth axis, was similar to that obtained by counting CP-4 etch pits on the {111} surfaces. The method used to develop the pits on a {100} surface consists of a 3-minute etch in a

solution of 1 part HF, 1 part H_2O_2 , and 4 parts H_2O followed by a 15 second etch in a solution of 1 part H_2O , 5 parts HF, and 5 parts HNO_3 . Figure 18 shows the typical appearance of these etch-pits, which have a pyramidal shape in contrast to the conical shape of the CP-4 pits on a $\{111\}$ surface.

Effect of Crystal Size

The effect of crystal diameter was investigated in crystals grown with a speed of 1 millimeter per minute, a rotational speed of 25 revolutions per minute, a temperature gradient in the crystal near the interface of $\sim 150^\circ \text{ cm}^{-1}$, and a jet of helium directed on the crystal near the interface. Under these conditions, it was found that the shape of the solid-liquid interface changed appreciably with crystal

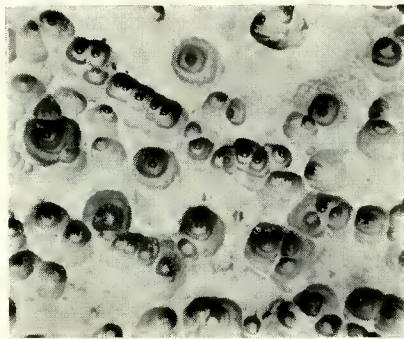


Fig. 18—Appearance of dislocation etch-pits on $\{100\}$ surface of germanium crystal. $\times 50$

diameter. For a 1-inch-diameter crystal, the interface was convex toward the solid. However, the degree of convexity decreased as the diameter was reduced, and at 0.25 inch the interface was essentially planar. In view of the metallographic observations discussed earlier in connection with the topography of the interface, a marked reduction in dislocation density might be expected with a decrease in crystal diameter. It is apparent from the data in Table VI that such a dependence was observed. More important, the data suggest that it is possible to virtually eliminate those dislocations revealed by CP-4 etch-pits on $\{111\}$ surfaces. In this connection, it should be noted that the low dislocation density in crystal No. 31 was observed on all four $\{111\}$ planes. Moreover, it persisted for approximately 6 inches of growth. This was probably due to the negligible change in growth conditions as a result of the small volume of material withdrawn from the melt.

Table VI—Effect of Crystal Diameter on Edge Dislocation Density and Carrier Lifetime in n-type Germanium

Crystal No.	Diameter, cm.	Edge Dislocation Density, cm^{-2}	Lifetime, μ sec.	Resistivity, ohm-cm
10	2.9	7×10^4	1000	41
9	2.1	2×10^3	1900	42
37	1.2	7×10^2	1800	39
31	0.6	<10	900	37

As regards the influence of crystal diameter on carrier lifetime, the data in Table VI reveal no systematic pattern. It is of considerable interest, however, that crystal No. 31 has the lowest lifetime in spite of its very low dislocation density. Furthermore, on comparing the values of carrier lifetime and dislocation density for the different crystals, it would appear that a maximum in lifetime occurs in the vicinity of $1 \times 10^3 \text{ cm}^{-2}$ under the specified conditions of growth.

Effect of Impurities

The role of impurities was investigated primarily in connection with dislocation density, since their effect on carrier lifetime had already been studied in detail by Burton et al.¹⁸ Crystals doped with arsenic were grown with resistivities as low as 0.8 ohm-cm, which corresponds to a concentration of approximately 5×10^{-8} atom frac-

Table VII—Effect of Arsenic Additions on Edge Dislocation Density.

Crystal No.	Resistivity ohm-cm	Edge Dislocation Density, cm^{-2}
41	0.8	400
42	7.2	200
43	22.7	300
40*	39.6	200
23	34.6	<10
30	1.5	<10

* No arsenic addition.

¹⁸ J. A. Burton, C. W. Hull, F. J. Morin and J. C. Severiens, "Effect of Nickel and Copper Impurities on the Recombination of Holes and Electrons in Germanium," *Jour. Phys. Chem.*, Vol. 57, p. 853, November, 1953.

tions. Growth conditions were selected which would yield both moderate and very low dislocation densities in the undoped crystal. The results of this study are shown in Table VII.

From Table VII it may be seen that the dilute alloying with arsenic has no significant effect on dislocation density for both the moderate and low density conditions of growth. Since similar results were also obtained for additions of antimony, gallium, and indium, it should be possible to produce low dislocation density material ($<10^2 \text{ cm}^{-2}$) with electrical resistivities suitable for most transistor applications.

Figures 19 and 20 summarize the results of the lifetime measurements for both n-type (arsenic and antimony additions) and p-type (gallium and indium additions) material. The measurements were made in the same region of the crystals, which were grown with dislocation densities ranging from 2 to $5 \times 10^3 \text{ cm}^{-2}$. The number of carriers per cm^3 , n , was calculated from measurements of the Hall constant, R_H , using the equation¹⁹

$$n = \frac{\mu_H}{\mu_d} \frac{1}{R_H e}.$$

The ratio of Hall to drift mobility, μ_H/μ_d was taken as 1.05 for electrons and 1.8 for holes in accordance with the data of Prince.²⁰ The dependence of lifetime on carrier concentration and resistivity, shown in Figures 18 and 19, is in general agreement with the Shockley-Read²¹ theory of lifetime in semiconductors. The greater decrease in minority carrier lifetime for the n-type material is attributed to the larger capture cross section for holes. The results also support the earlier work of Burton et al.,¹⁸ although there are significant differences in the absolute values of lifetime for comparable carrier concentrations. This could be due to the different techniques used in the measurement of lifetime.

The effect of nickel additions was also investigated, since nickel is known²² to have a very limited solid solubility in germanium corresponding to approximately 0.8 ohm-cm at 900°C and 15 ohm-cm at 700°C. In the present study it was not possible to obtain room tem-

¹⁹ W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand Company, Princeton, N. J., 1950.

²⁰ M. Prince, "Drift Mobilities in Semiconductors, II Silicon," *Phys. Rev.*, Vol. 93, p. 1204, March, 1954.

²¹ W. Shockley and W. T. Read, Jr., "Statistics of the Recombinations of Holes and Electrons," *Phys. Rev.*, Vol. 87, p. 835, September, 1952.

²² F. van der Maesen and J. A. Brenkman, "The Solid Solubility and the Diffusion of Nickel in Germanium," *Philips Research Reports*, Vol. 9, p. 225, June, 1954.

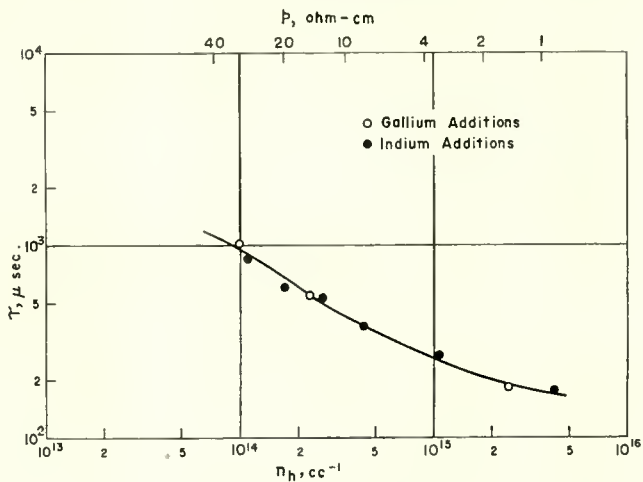


Fig. 19—Dependence of minority carrier lifetime, τ , on carrier concentration, n_h , in p-type germanium crystals.

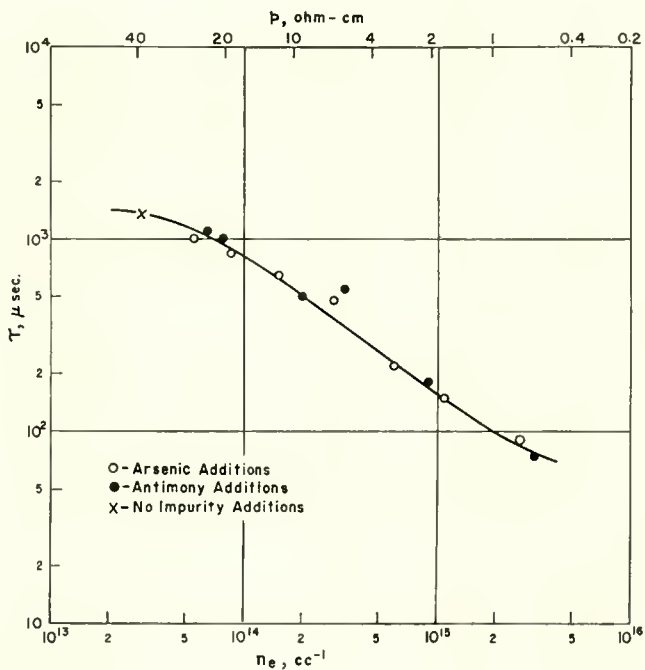


Fig. 20—Dependence of minority carrier lifetime, τ , on carrier concentration, n_e , in n-type germanium crystals.

perature resistivities less than 10 ohm-cm in single crystals by nickel doping. The results listed in Table VIII show that an increase in the nickel content (reflected in decreasing resistivity) produces no significant change in the dislocation density, provided the maximum solid solubility limit of nickel is not exceeded. However, upon the addition of nickel in amounts which slightly exceed this limit (crystal No. 49) the dislocation density increases by several orders of magnitude. Moreover, metallographic examination of CP-4 etch-pits on the {111} surface normal to the growth direction reveals a hexagonal pattern very similar to that shown in Figure 15, but with a much smaller average pit diameter.

This implies that even for such small concentrations diffusion of im-

Table VIII—Effect of Nickel Additions on Edge Dislocation Density and Minority Carrier Lifetime.

Crystal No.	Resistivity ohm-cm	Lifetime μ sec.	Edge Dislocation Density, cm^{-2}
49	10.4	<1	$\sim 10^6$
51	13.9	1	700
52	24.8	3	500
40*	39.6	1500	600

* No nickel addition.

purities in the liquid zone adjacent to the growing crystal by convection is not sufficiently rapid to avoid the constitutional supercooling discussed by Chalmers.²³ For nickel additions greater than that in crystal No. 49, excessive lineage occurred and, subsequently, polycrystallinity.

The lifetime data in Table VIII also show that nickel, like copper, is especially effective as a recombination center in germanium. Nickel concentrations of only 10^{-9} atom fractions lowers the lifetime from 1500 to 3 microseconds, which is in general support of the work of Burton et al.¹⁸ and Maesen and Brenkman.²² This effect is one reason for the severe requirements on the purity of germanium crystals for use in transistor devices.

Effect of Dislocations on Some Electrical and Physical Properties

In this section results will be presented to show the influence of dislocations produced during crystal growth and revealed by CP-4 etch pits on selected electrical and physical properties of germanium.

²³ J. W. Rutter and B. Chalmers, "A Prismatic Substructure Formed During Solidification of Metals," *Canadian Jour. of Phys.*, Vol. 31, p. 15, January, 1953.

Minority Carrier Lifetime

It was found in the present study that by varying the crystal diameter, crystal orientation, and the concentration of dislocations in the seed crystal, a wide range of dislocation density could be obtained under similar conditions of growth velocity and temperature gradient in both the crystal and melt. It seemed desirable, therefore, to study the effect of dislocations on the lifetime of minority carriers. Several investigators^{13,24,25} have already reported on such effects, but a more generalized treatment was attempted here by extending the study to

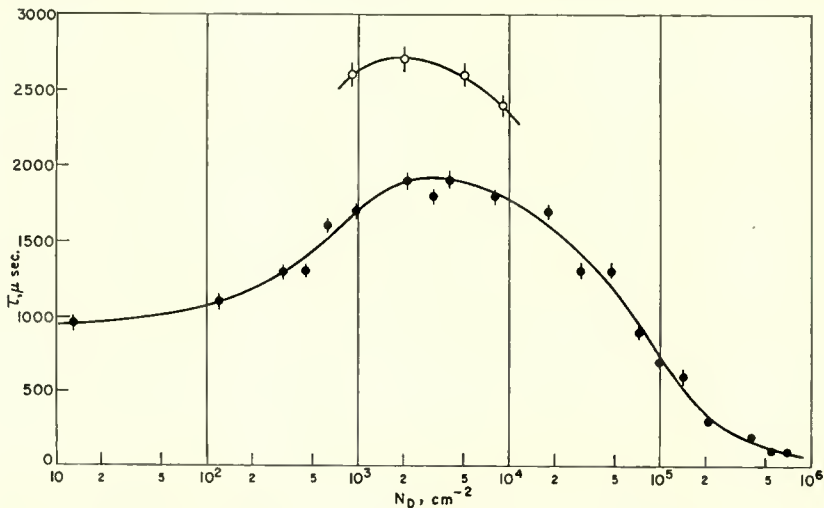


Fig. 21—Effect of edge dislocation density, N_D , on carrier lifetime, τ , in n-type, 40 ohm-cm germanium crystals produced at different growth rates: ● = 1 mm min⁻¹, ○ = 2 mm min⁻¹.

include effects over a much wider range of dislocation density and in crystals of both high (40 ohm-cm) and low (2 ohm-cm) resistivity. Only those dislocations revealed by CP-4 etch-pits on {111} surfaces were considered, and, except where otherwise noted, the dislocations were distributed randomly and/or along slip lines.

The dependence of minority carrier lifetime, τ , on edge dislocation density, N_D , in the range 10 to 10⁶ cm⁻² is shown by the solid circles in the semi-log plot of Figure 21 for the 40 ohm-cm crystals grown at a velocity of 1 millimeter per minute. The vertical line through the points indicates an estimated error in the high lifetime measurements

²⁴ F. L. Vogel, W. T. Read, Jr., and L. C. Lovell, "Recombination of Holes and Electrons at Lineage Boundaries in Germanium," *Phys. Rev.*, Vol. 94, p. 1791, June, 1954.

²⁵ J. Okada, "Effects of the Dislocations on Minority Carrier Lifetime in Germanium," *Physical Society of Japan*, Vol. 10, p. 1110, December, 1955.

of ± 50 microseconds. It may be seen that the lifetime at first increases with increasing number of dislocations to a maximum value of 1900 microseconds at a dislocation density of approximately 3×10^3 cm^{-2} . It then decreases at an increasing rate up to a density of 10^5 cm^{-2} ; in the range 5×10^4 to 10^6 cm^{-2} , τ is approximately proportional to $1/N_D$.

Kulin and Kurtz¹³ have also studied the variation in lifetime with dislocation density in 2-8 ohm-cm crystals grown from the melt. Their dislocation densities were determined by measuring the breadth of X-ray diffraction lines, and these determinations were substantiated by chemical etch-pit counting. Typical values were reported in the range 10^6 to 10^7 cm^{-2} , which is considerably higher than that in Figure 20. In this connection, it is noteworthy that X-ray measurements similar to those of Kulin and Kurtz for selected dislocation densities in the range shown in Figure 21 revealed no significant variation in the breadth of diffraction lines (mean breadth = 25 minutes) and, hence, in total dislocation density. This strongly suggests that in these germanium crystals there is no direct correspondence between dislocations as determined by CP-4 etch-pit counting on $\{111\}$ surfaces, and those determined by the double crystal X-ray rocking curve technique. It further suggests a background imperfection density in these crystals of 10^6 cm^{-2} or greater, which has no relation to the marked lifetime variations shown in Figure 21. It is possible that these imperfections are dislocations of the composite variety discussed recently by Ellis,¹¹ and found to be present in amounts $>10^6$ cm^{-2} in germanium.

It is generally believed that the carrier lifetime in high resistivity germanium is controlled by residual amounts of copper and by vacant lattice sites at concentration less than 10^{13} cc^{-1} , in addition to dislocations. Thus, any analysis of the lifetime variation shown in Figure 21 should take into account the presence of these defects and their possible interaction. It has been shown by Cottrell²⁶ that impurity atoms are preferentially attracted to edge-type dislocations in order to relieve the strain energy around the misfit impurity atom. Moreover, this interaction would be expected to decrease the stress field of the distorted lattice associated with the dislocation. Recently, Kurtz and Kulin²⁷ have calculated values of this binding energy for various

²⁶ A. H. Cottrell, "Effect of Solute Atoms on the Behaviour of Dislocations," Conference on the Strength of Solids, *Phys. Society of London*, p. 30, 1948.

²⁷ A. D. Kurtz and S. A. Kulin, "The Interaction of Impurity Atoms with Dislocations in Germanium," *Acta Metallurgica*, Vol. 2, p. 352, March, 1954.

impurity atoms in germanium, and for copper the value is 0.12 ev. It is proposed, therefore, that the observed increase in lifetime on going from 10 to 10^3 cm^{-2} dislocations is due to the interaction of residual copper atoms with the dislocations to form copper-dislocation atmospheres which are less effective recombination centers than the free copper atoms. This is possible since the bonding at the core of the dislocation may be changed, and this would affect the electronic levels of both the dislocation and the copper atom. Certainly the capture cross section of the dislocation will be reduced, since its stress field is lowered by the condensation of copper atoms. Based on this argument, the maximum in lifetime should occur at a dislocation density whose volume density of sites corresponds to the residual content of copper atoms in germanium. The volume density of sites associated with a dislocation density N_D , is given by N_D/b , where b is the spacing between neighboring atoms on a $\{111\}$ plane in a $\langle 110 \rangle$ direction. For germanium, $b = 4\text{\AA}$. Thus the dislocation density, 3×10^3 cm^{-2} , associated with the lifetime maximum in Figure 21, gives an atom concentration of 7.5×10^{10} cm^{-3} , which is a reasonable value for the residual copper content in germanium. This, of course, assumes a strongly condensed copper atmosphere, which can be expected at room temperature.

Figure 21 also shows the variation in lifetime over a small dislocation density range for crystals grown at a velocity of 2 millimeters per minute. The higher lifetime of these crystals for comparable dislocation densities with the crystals grown at 1 millimeter per minute has been attributed to a lower residual copper content. This follows from the argument that at the faster growth rate there is less time at a given temperature for any copper present in the heating chamber to diffuse into the growing crystal, or for copper to diffuse from the bottom to the top of the crystal as a result of a concentration gradient. In these crystals, therefore, the maximum in lifetime should occur at a somewhat lower dislocation density than those grown at 1 millimeter per minute and examination of the data does indicate a small shift to a lower dislocation density of 2×10^3 cm^{-2} , which corresponds to an atom concentration of 2.25×10^{10} cm^{-3} . According to the theory of recombination at traps by Shockley and Read,²¹ this small change in residual copper content is enough to account for the observed increase in lifetime at the faster growth rate.

Any analysis of the observed variation in carrier lifetime in the dislocation density range 10 to 10^4 cm^{-2} must also take into account the possibility of a vacancy concentration in excess of equilibrium.

It was shown recently by both Logan²⁸ and Mayburg²⁹ that the quenching-in of vacancies from temperatures near the melting point of germanium is not so time dependent in crystals of low dislocation density. This follows from the argument that the vacancy concentration in germanium on cooling down from high temperatures decreases by a preferential migration and condensation on dislocations. Thus, in the very low dislocation density crystals of the present study, one may have an excess vacancy concentration which will decrease with increasing density of dislocations. It is noteworthy that this process of vacancy condensation results in an increase in the length and, hence, energy of the dislocation, through the formation of jogs or steps³⁰ in the dislocation line. Consequently, the net effect is the removal of one lattice imperfection at the expense of increasing the energy of another. This defect interaction might then be expected to cause either no change in lifetime, or a small increase or decrease depending on the difference in the capture cross section of the two defects. Since the dislocation probably has the larger capture cross section because of the postulated space charge surrounding it,³¹ it is unlikely that the condensation process could account for the observed increase in lifetime from 10 to 10^3 cm^{-2} dislocations. However, it could explain the surprisingly small decrease in lifetime between 2×10^3 to 9×10^3 cm^{-2} dislocations, since the concentration of excess vacancies would be expected to decrease at an increasing rate as more dislocations are introduced into the crystal. This is particularly true in view of the low self-diffusion coefficient of germanium.³²

The dependence of lifetime on dislocation density was also studied in n-type, arsenic-doped germanium with a resistivity of 1.8 to 2.1 ohm-cm. From the data in Figure 22 it may be seen that the lifetime increases with increasing dislocation density in the range, 10 to 5×10^3 cm^{-2} in a manner similar to the 40 ohm-cm material in Figure 21. In the dislocation density range 10^4 to 10^6 cm^{-2} , the lifetime is inversely proportional to the number of dislocations, as shown by the plot of $1/\tau$ versus N_D in Figure 23.

²⁸ R. A. Logan, "Thermally Induced Acceptors in Germanium," *Phys. Rev.*, Vol. 101, p. 1455, March, 1956.

²⁹ S. Mayburg, "Acceptors Quenched into Germanium," *Phys. Rev.*, Vol. 103, p. 1130, August, 1956.

³⁰ A. H. Cottrell, *Dislocations and Plastic Flow in Crystals*, Clarendon Press, Oxford, England, 1953.

³¹ W. T. Read, Jr., "Theory of Dislocations in Germanium," *The Philosophical Magazine*, Vol. 45, p. 775, August, 1954.

³² H. Letaw, Jr., W. M. Portnox, and L. M. Slifkin, "Self-Diffusion in Germanium," *Phys. Rev.*, Vol. 102, p. 636, May, 1956.

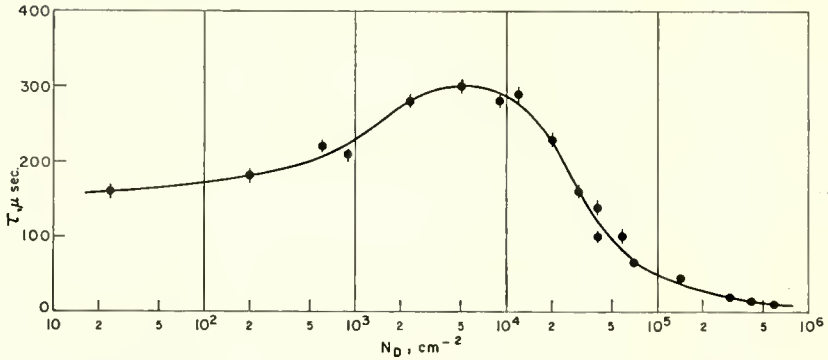


Fig. 22—Effect of edge dislocation density, N_D , on carrier lifetime, τ , in n-type (arsenic-doped), 2 ohm-cm germanium crystals.

For low resistivity samples, the lifetime, τ_D , due to randomly distributed edge dislocations is related to the effective cross section, A_p , of the occupied traps for capture of a hole by the approximation

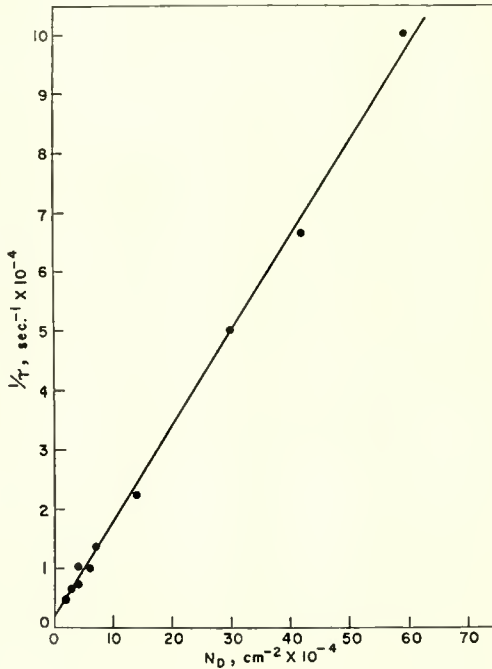


Fig. 23—Dependence of minority carrier lifetime, τ , on edge dislocation density, N_D , in 2 ohm-cm, n-type (arsenic-doped) germanium crystals.

$$\tau_D \approx \frac{1}{N_D N_S f v A_p}$$

where v is the thermal velocity of the conduction electrons or holes (1.9×10^7 cm sec⁻¹), N_D is the density of dislocations, N_S is the linear density of sites along the dislocation line (2.5×10^7 cm⁻¹), and f is the fraction of full sites, or occupied acceptors, on a dislocation line. Thus, the values of $N_S f$ gives the density of active recombination centers for n-type material.

It is apparent from the above expression that if reasonable values are assigned to the fraction of full sites, it should be possible to calculate the cross section of the dislocation for the capture of holes from the slope of the curve in Figure 23. If it is assumed that the recombination centers are the dangling bonds along the dislocation, then according to Read³³ the value of f is approximately 0.03 for the case where $N_D = 10^{15}$ cc⁻¹ (1.7 ohm-cm) and the dislocation acceptor level is 0.225 ev below the conduction band. This gives a value for the capture cross section for holes, A_p , of 1.1×10^{-14} cm², which is smaller than the 4×10^{-14} cm² value of Okada,²⁵ who also measured dislocation density by counting CP-4 etch-pits. Thus, the cross section of recombination for dislocations is larger than that of impurity recombination centers such as copper and nickel,¹⁹ and this is not unreasonable if each dislocation is surrounded by a space charge as postulated by Read.³¹

It is noteworthy that the above capture cross section is presumably for dislocations with a condensed arsenic atmosphere. It would be of interest to compare this cross section with that for dislocations with no impurity atmosphere. To make such a comparison, n-type (arsenic-doped) 2 ohm-cm crystals containing an initial dislocation density of 4 to 5×10^3 cm⁻² were plastically deformed by bending at approximately 575°C, where the diffusion of arsenic in germanium is very low.³⁴ Moreover, the deformation was done in molten sodium cyanide to avoid conversion to p-type by copper contamination and the formation of copper atmospheres around the dislocations. The dependence of lifetime on dislocations produced by plastic flow under these conditions is shown in Figure 24 by the $1/\tau$ versus N_D curve, which has a slope of 0.51 sec⁻¹ cm². This gives a capture cross section

³³ W. T. Read, Jr., "Statistics of the Occupation of Dislocation Acceptor Centres," *The Philosophical Magazine*, Vol. 45, p. 1119, November, 1954.

³⁴ W. C. Dunlap, Jr., "Diffusion Impurities in Germanium," *Phys. Rev.*, Vol. 94, p. 1531, June, 1954.

of $3.6 \times 10^{-14} \text{ cm}^2$, which is significantly larger than that for dislocations produced during crystal growth in similarly doped material. These data serve to emphasize the importance of considering impurity interactions in determining the electrical effects of dislocations.

Surface Recombination Velocity

Since the surface recombination of minority carriers often plays an important role in determining the characteristics of semiconductor devices, it seemed of interest to investigate this property in germanium crystals with different dislocation density.

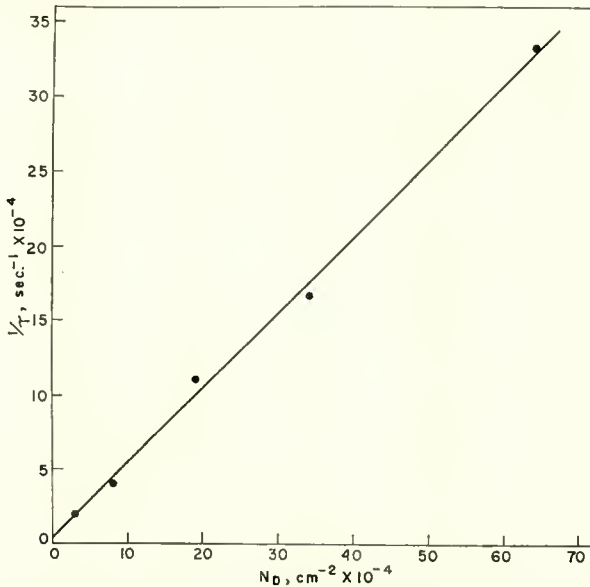


Fig. 24—Dependence of minority carrier lifetime, τ , on edge dislocation density, N_D , in 2 ohm-cm, n-type (arsenic-doped) germanium crystals plastically deformed at 575°C .

The surface recombination can be expressed quantitatively through the surface recombination velocity, s , which is usually obtained by measuring the actual lifetime, τ_m , of minority carriers in a sample of known volume lifetime, τ_v , and dimensions. For the case of a rectangular bar of cross-sectional dimensions, B and C , Shockley¹⁹ gives, for small values of s , the formula

$$\frac{1}{\tau_m} = \frac{1}{\tau_v} + 2s \left(\frac{1}{B} + \frac{1}{C} \right).$$

Thus, a plot of $1/\tau_m$ versus $(B^{-1} + C^{-1})$ should be linear with a slope of $2s$.

Measurements of τ_m for different values of $(B^{-1} + C^{-1})$ were made on high resistivity (~ 40 ohm-cm) n-type crystals with dislocation densities ranging from <10 cm^{-2} to 4×10^4 cm^{-2} . The initial sample dimensions were approximately $3 \times 3 \times 13$ millimeters, and the chemical surface treatment consisted of etching with a solution of 6 parts HF, 1 part H_2O_2 and 3 parts H_2O . This etch was selected, since it usually gave a surface recombination velocity <20 cm sec^{-1} . The results of these lifetime measurements are summarized in Figure

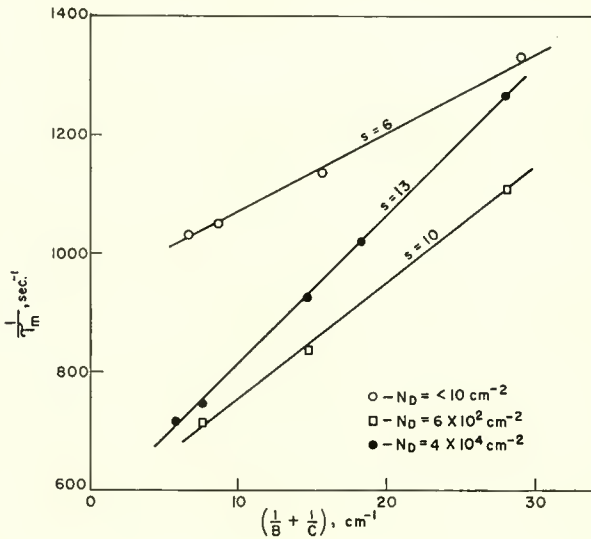


Fig. 25—Variation of measured carrier lifetime, τ_m , with change in cross-sectional dimensions, B and C , in 40 ohm-cm germanium crystals of different dislocation density. The surface recombination velocity, s , is obtained from slope of the curves.

25, from which it may be seen that the $1/\tau_m$ versus $(B^{-1} + C^{-1})$ curves are linear. Each point on the curves represents the average value of measurements on two crystals. Figure 25 also shows that the value of s , determined from the slope of the curves, decreases systematically with decreasing dislocation density, and it is of interest that s can be reduced to the unusually low value of 6 cm sec^{-1} by virtually eliminating dislocation as revealed by CP-4 etch-pits. These results suggest that any analysis of surface recombination processes must take into account differences in the number of dislocations intersecting the surface.

Precipitation of Copper

In supersaturated solid solutions, dislocations can be expected to act as nuclei for the precipitation of the excess solute atoms. Thus, variations in dislocation density in as-grown crystals should have a significant effect on the kinetics of the precipitation process. To check this hypothesis, the precipitation of copper from a supersaturated solid solution in germanium crystals of variable dislocation density was investigated by observing changes in resistivity on annealing at 500°C. The results of this study are summarized in Table IX.

The samples used in these experiments were in the form of bars $3 \times 3 \times 13$ millimeters. The bars were copper plated in a solution of

Table IX—Effect of Edge Dislocation Density on Precipitation of Copper in Germanium

Crystal No.	Edge Dislocation Density, cm^{-2}	Resistivity, ohm-cm		
		Original	Cu Diffusion 1 hr. at 825°C.	4 hrs. at 500°C.
23	<10	35	0.39	6
8B	10^6	37	0.23	35
11B	10^6	37	0.21	34
11T	10^2 (C), 10^4 (E)	41	0.25	9(C), 19(E)

copper sulphate to a plate thickness of ~ 5000 Å. The copper plate was diffused into the germanium by gradually heating the specimens up to 825°C in evacuated quartz tubes, and annealing at this temperature for 1 hour. The specimens were then cooled rapidly to room temperature by quenching the quartz tubes in water in order to obtain a supersaturated solid solution of the copper in germanium. This was possible, since the solid solubility of copper in germanium is known to decrease sharply with decreasing temperature.^{35,36} The data in Table IX show that after the diffusion and quenching treatments, the resistivity of the germanium samples ranges from 0.2 to 0.4 ohm-cm. This corresponds to an atom fraction of copper of $\sim 3 \times 10^7$, which

³⁵ C. D. Thurmond and J. D. Struthers, "Equilibrium Thermochemistry of Solid and Liquid Alloys of Germanium and of Silicon. II The Retrograde Solid Solubilities of Sb in Ge, Cu in Ge, and Cu in Si," *Jour. Phys. Chem.*, Vol. 57, p. 831, November, 1953.

³⁶ C. S. Fuller, J. D. Struthers, J. A. Ditzenberger, and K. B. Wolfstirn, "Diffusivity and Solubility of Copper in Germanium," *Phys. Rev.*, Vol. 93, p. 1182, March, 1954.

is just within the solid solubility limit at 825°C ($\sim 5 \times 10^7$ atom fraction of copper). After quenching to room temperature, the specimens were given a final anneal for 4 hours at 500°C in order to precipitate the copper, whose solubility in germanium is reduced to only 10^{-9} atom fraction at this temperature. This final anneal, therefore, should result in an increase in the resistivity of the sample to near the original value of 37 ohm-cm, provided the precipitation process is complete. It is apparent from the data in Table IX that this recovery of resistivity was essentially complete only in those crystals (Nos. 8B and 11B) with high dislocation density. On the other hand, crystal No. 23, with a low dislocation density of $<10 \text{ cm}^{-2}$, had

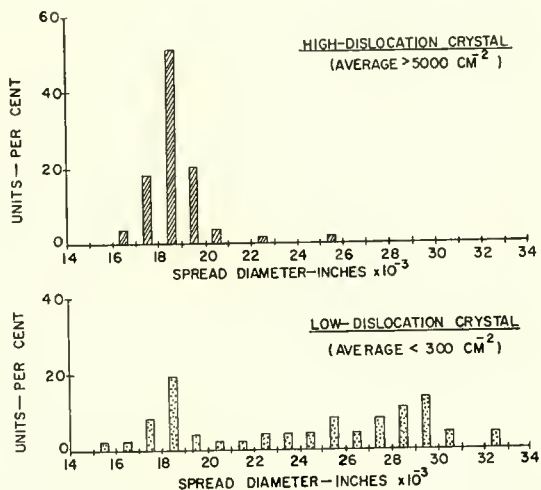


Fig. 26—Distribution plots of spread diameter on {111} surface of germanium crystals of different dislocation density (After Kettering).³⁷

a final resistivity of only 6 ohm-cm. Equally significant, crystal No. 11T, which had a sharp dislocation concentration gradient from the center to the edge of the crystal, also exhibited a gradient in resistivity after the 500°C anneal such that the regions of higher dislocation density had higher resistivity.

The above results clearly indicate that the dislocations introduced during crystal growth and revealed by CP-4 etch-pits significantly increase the rate of copper precipitation from a supersaturated solution in germanium. This is to be contrasted with similar effects recently reported by Logan²⁸ in germanium crystals where variations in dislocation density were produced by plastic deformation.

Wettability of Molten Indium

In connection with the role of dislocations produced during crystal growth on the physical properties of germanium, recent experiments by Kettering³⁷ relating to the wettability of molten indium pellets on germanium single crystal surfaces are of direct interest. In the fabrication of alloy junction transistors, it was found that under identical conditions of heat treatment for forming electrical (p-n) junctions variations in the bulk dislocation density had a marked effect on the degree of indium spreading on {111} surfaces. This is clearly indicated in Figure 26 by the distribution plots of the spread diameter on two sets of crystals with different dislocation density. It may be inferred from this statistical data that the intersection of dislocations with the {111} surface limits the indium spreading during the alloying process.

In an effort to substantiate the effect observed by Kettering more definitively, Pankove,³⁸ through exacting experimental techniques, placed indium pellets on CP-4-etched {111} germanium surfaces directly over (1) dislocation etch-pits, (2) dislocation-free regions, and (3) dislocations produced by mechanically deforming regions otherwise free of dislocation pits. On alloying, it was found that all the indium pellets over dislocations exhibited a limited spread of 20 to 21 mils, while those over dislocation-free regions spread excessively (24 to 30 mils).

It is apparent from the above studies that the dislocations produced during crystal growth have a significant effect on the extent to which molten materials will spread over germanium surfaces. Since the extent of this spreading influences the functional geometry of the alloy junction transistor (i.e., the junction area, spacing, and topography), then control over the dislocation density in semiconductor crystals is important to the fabrication of transistor devices with uniform electrical characteristics.

SUMMARY

Variations in the conditions of crystal growth by the Czochralski technique were found to have significant effects on some structural and electrical properties of germanium. The principle observations and conclusions are summarized below along with brief comments for completeness.

³⁷ H. V. Kettering, private communication.

³⁸ J. I. Pankove, "Effect of Edge Dislocations on the Alloying of Indium to Germanium," *Jour. Appl. Phys.*, Vol. 28, p. 1054, September, 1957.

1. Plastic deformation by octahedral slip was observed in most crystals grown with a temperature gradient in the solid near the solid-liquid interface $>40^{\circ}\text{C cm}^{-1}$. This was indicated by the crystallographic alignment of chemical etch-pits outlining edge-type dislocations.

2. Examinations of etch-pit distribution showed an intimate association of slip with the total number of dislocations revealed by CP-4 etching. This suggested that deformation by slip is the main source of dislocations in germanium crystals grown from the melt.

3. The appearance of dislocation etch-pits in random distribution across the crystal, and at low-angle polygonized boundaries, can be regarded as manifestations of the increased mobility of dislocations in the presence of a low temperature gradient in the crystal.

4. Examinations of the topography of the solid-liquid interface showed that a high density of slip and, hence, dislocations is associated with a curved interface indicative of nonparallel thermal currents in the crystal adjacent to the interface. The occurrence of slip, therefore, is attributed to the relaxation of thermal stresses resulting from unequal thermal expansion of adjacent parts of the crystal.

5. Growth under conditions which provide an essentially planar solid-liquid interface has resulted in crystals with dislocation densities $<10^2\text{ cm}^{-2}$, and the use of seeds with such low dislocation densities has yielded crystals with virtually no dislocation etch-pits over several inches of growth.

6. Marked radial and longitudinal gradients in dislocation density can be expected in crystals grown by the conventional Czochralski technique where the environment of the crystal near the solid-liquid interface is constantly changing with growth. A modification of this technique is discussed which permits the production of long crystals with uniform dislocation density.

7. The radial pattern of slip lines in large diameter crystals indicated that (1) slip originates near the surface of the crystal and (2) composite-type dislocations (screw-edge components) are produced.

8. An increase in the rate of crystal growth from 0.1 to 2 millimeters per minute resulted in a marked increase in minority carrier lifetime, and this has been attributed to a decrease in the residual copper content.

9. Variations in the rotational speed of the crystal and in the ambient during growth had no significant effect on either dislocation density or carrier lifetime.

10. Crystals grown in a $\langle 110 \rangle$ direction have generally lower dislocation densities as compared to those grown in the $\langle 111 \rangle$ and

$\langle 100 \rangle$ directions. This orientation effect has been attributed to variations in the number of available slip systems for relief of thermal stresses with crystal orientation.

11. Significant reductions in dislocation density were effected by decreasing the diameter of the crystal. With a large temperature gradient in the crystal, dislocation densities $< 10 \text{ cm}^{-2}$ were obtained in $\frac{1}{4}$ -inch diameter crystals.

12. Additions of n- and p-type impurities within their limits of solid solubility had no significant effect on dislocation density. It is possible, therefore, to produce crystals of low dislocation density ($< 10^2 \text{ cm}^{-2}$) with electrical properties suitable for most transistor applications.

13. The dependence of carrier lifetime on edge dislocation density in the range 10 to 10^6 cm^{-2} was investigated in both high- and low-resistivity crystals. The carrier lifetime increased with increasing dislocation density to a maximum value corresponding to a dislocation density of approximately $3 \times 10^3 \text{ cm}^{-2}$. This has been attributed to the formation of copper-dislocation recombination centers.

14. Metallographic and complementary X-ray data suggest that there is no direct correspondence between dislocations as determined by CP-4 etch-pit counting on $\{111\}$ surfaces and those determined by the double crystal, X-ray rocking curve technique.

15. The effective cross section for capture of a hole was determined for dislocations produced by crystal growth and by plastic deformation in n-type, 2 ohm-cm crystals. The value for those produced by crystal growth is $1.1 \times 10^{-14} \text{ cm}^2$, while that for dislocations produced by plastic deformation is $3.6 \times 10^{-14} \text{ cm}^2$.

16. Variations in the number of dislocations produced by crystal growth and outlined by CP-4 etch-pits on a $\{111\}$ surface had significant effects on (1) the precipitation of copper from a supersaturated solid solution in germanium, (2) the surface recombination velocity in high-resistivity crystals, and (3) the wettability of molten indium on $\{111\}$ surfaces of germanium wafers.

ACKNOWLEDGMENTS

The author is especially indebted to J. Gannon for the carrier lifetime measurements and sample preparations for metallographic examination. He is also grateful to R. Dalven and S. Wagner for assistance in the study of impurity effects, to A. Rose for several photomicrographs, and to S. G. Ellis for helpful discussions during the course of the work.

DETECTION OF ASYMMETRIC SIDEBAND SIGNALS IN THE PRESENCE OF NOISE*

BY

TOMOMI MURAKAMI and RICHARD W. SONNENFELDT

RCA Victor Television Division,
Camden, N. J.

Summary—The first part of this paper analyzes three methods of detection of signals contaminated by fluctuation noise:

- (1) Linear Envelope Detectors
- (2) Product Detectors
- (3) Exalted Carrier Detectors.

A new ratio, the "video-to-noise-error ratio" is proposed for a more adequate quantitative evaluation of detector performance. Its use is justified by a large number of theoretical curves and experimental results. The theoretical and practical results show that an improvement in the output video-to-noise-error ratio is obtained by substitution of a product detector for an envelope detector. Improvements of 11 decibels have been measured. The second part of this paper analyzes the effects of impulse noise on the detection of asymmetric sideband signals. It is shown that product detectors eliminate the rectification of the noise envelope found in envelope detectors. This is of particular interest in asymmetric sideband systems where the noise output will be at relatively high frequencies. These can often be separated from lower frequency signals, making possible efficient impulse noise suppression.

INTRODUCTION

THE presence of noise in communication systems sets an ultimate upper limit to their effective usable range. For this reason extensive efforts have been made in the past to study and understand the nature of electrical noise, and to minimize its effects on communication systems.¹⁻⁵

The reduction of the effects of noise at the receiver is of great importance in a television system. By improving the performance of television receivers on weak signals in the presence of noise, many

* Manuscript received February 25, 1958.

¹ Vernon D. Landon, "Impulse Noise in FM Reception," *Electronics*, Vol. 14, pp. 26-30, 73-76, February, 1941.

² J. R. Ragazzini, "The Effect of Fluctuation Voltages on the Linear Detector," *Proc. I.R.E.*, Vol. 30, p. 277, June, 1942.

³ S. O. Rice, "Mathematic Analysis on Random Noise," *Bell Sys. Tech. Jour.*, Vol. 23, p. 282, July, 1944; Vol. 24, p. 46, January, 1945.

⁴ E. G. Fubini and D. C. Johnson, "Signal-to-Noise Ratio in AM Receivers," *Proc. I.R.E.*, Vol. 36, p. 1461, December, 1948.

⁵ David Middleton, "Rectification of a Sinusoidally Modulated Carrier in the Presence of Noise," *Proc. I.R.E.*, Vol. 36, p. 1467, December, 1948.

distant viewers, previously excluded, can be reached without new demands upon transmitter power.

In television and other pulse modulation systems the usual definition of signal-to-noise ratio does not adequately describe the effects of noise due to the type of amplitude modulation employed. Therefore, a new and more suitable definition of signal-to-noise ratio called the "video-to-noise-error" ratio is proposed, and experimental evidence is included to justify its use. The video-to-noise-error ratio permits a much more realistic quantitative evaluation of different detectors.

New techniques have become available for improving the noise performance of video detectors.⁶⁻¹⁴ It is the purpose of this paper to give a detailed analysis of the detection of asymmetric-sideband signals in the presence of noise. Three different signal detectors suitable for video detection are analyzed. These detectors are (1) the envelope (or linear diode) detector, (2) the synchronous detector, and (3) the exalted-carrier detector. In the first part of this paper the effect of fluctuation noise on the detected output is considered. The effect of impulsive noise on the detection process is investigated in the second part of the paper.

SIGNAL-TO-NOISE RATIO OF TELEVISION VIDEO SIGNAL

In communication systems it is customary to express the noisiness of signals in ratios relating the r-m-s value or power equivalents of the detected signal to the r-m-s value or power equivalent of the noise

⁶ Y. W. Lee, "Detection of Periodic Signals in Random Noise by the Method of Correlation," Technical Report No. 157, Res. Lab. of Electronics, M.I.T.

⁷ Y. W. Lee, T. P. Cheatham, and J. B. Wiesner, "Application of Correlation Function in the Detection of Small Signals in Noise," *Proc. I.R.E.*, Vol. 38, p. 1165, October, 1950.

⁸ R. M. Fano, "Signal-to-Noise Ratio in Correlation Detectors," Technical Report No. 186, Res. Lab. of Electronics, M.I.T., February, 1951.

⁹ R. A. Smith, "The Relative Advantages of Coherent and Incoherent Detectors: A Study of Their Output Noise Spectra Under Various Conditions," *Proc. I.E.E.*, Part IV, Vol. 98, p. 43, 1951.

¹⁰ J. P. Costas, "Synchronous Detection of AM Signals," *Tele-Tech*, Vol. 11, p. 55, July, 1952.

¹¹ D. G. Tucker, "The Synchrodyne," *Electronic Engineer*, Vol. 19, p. 75, March, 1947; p. 241, August, 1947; p. 276, September, 1947.

¹² M. G. Crosby, "Exalted-Carrier Amplitude- and Phase-Modulation Reception," *Proc. I.R.E.*, Vol. 33, p. 581, September, 1945.

¹³ J. E. Bridges, "Detection of Television Signals in Thermal Noise," *Proc. I.R.E.*, Vol. 42, p. 1396, September, 1954.

¹⁴ D. B. Harris, "Selective Demodulation," *Proc. I.R.E.*, Vol. 35, p. 565, June, 1947.

contaminating it. In television signals, r-m-s or power represent neither the nature nor the amount of information being transmitted due to the d-c modulation process employed. The d-c component is required to transmit correct values of average brightness. For instance, a blackline drawing on a white background conveys just as much information as the same drawing with white lines on a black background. Obviously, r-m-s and power values associated with signals formed from these two drawings are entirely different. What is of primary concern in television is how much the noise in the signal will vary the brightness from its correct value. The quantity that determines the brightness in a television signal is the rectified carrier; this will be used in the formulation of expressions for the noisiness of television signals.

There is another reason for choosing the rectified carrier here. In television, modulation of the carrier with a single sine wave of a given video frequency is so rare an occurrence in normal pictures as to have no practical importance. Information is transmitted simply by changing the carrier amplitude in step fashion as a result of the abrupt changes in brightness encountered in scanning the scene. Each one of these transitions normally involves the entire available frequency spectrum; therefore the noise should be considered over the entire transmission spectrum.

In the video signal output from the second detector in a television receiver, the d-c component which determines the brightness of the picture is a function of both the rectified carrier and the rectified noise. Since the signal-to-noise ratio of the detected output from the second detector is influenced by the presence of rectified noise, a definition for the signal-to-noise ratio of the detected television signal which includes the rectified noise can be written as

$$\frac{\text{video}}{\text{noise}} = \frac{Y}{N} = \frac{\text{peak rectified carrier}}{\text{r-m-s detected a-c noise} + \text{d-c of rectified noise}} \quad (1)$$

Y/N will be termed the video-to-noise-error ratio. Y/N can be measured directly by use of a synchronous detector which produces no d-c due to the noise components. In that case,

$$Y/N = \frac{\text{d-c}}{\text{a-c}} = \frac{\text{peak rectified carrier}}{\text{r-m-s detected a-c noise}} \quad (2)$$

When envelope detection is used, Y/N has the disadvantage that it cannot be measured directly over the entire range of signal-to-noise

ratios possible. It is possible to calculate this ratio at the output of the detector as a function of the carrier-to-noise ratio at the input to the detector. The difficulty with envelope detection is that the values of rectified carrier and rectified noise are measurable as independent quantities only at the extremes of the range; that is to say, either when the carrier amplitude is much larger than the noise, or when the carrier is very small compared to the noise. A ratio showing the threshold effect is given by

$$N_{\text{rms}} = \frac{\text{peak rectified carrier} + \text{d-c of rectified noise}}{\text{r-m-s detected a-c noise}} \quad (3)$$

The ratio given by Equation (3) can always be measured irrespective of the method of detection.

The d-c term of Equation (2) is the peak rectified carrier, and the d-c term of Equation (3) is the sum of the peak rectified carrier and the rectified noise. Therefore, the d-c of rectified noise may be found as the difference between the respective outputs of synchronous and envelope detectors. From these measurements, the ratio Y/N of Equation (1) may then always be calculated.

It is worth noting that several alternative definitions of video-to-noise-error ratio have been proposed¹⁵⁻¹⁷ and are useful in certain applications. Unfortunately none of these definitions is entirely satisfactory for all types of detectors. The d-c term in the denominator of Equation (1) has been included to account for the threshold effect of envelope detectors. This d-c term is zero for synchronous detectors and also is very small for large input carrier-to-noise ratios for envelope detectors. It becomes significant only for carrier-to-noise ratios in the threshold region.

Consider a television signal of a given carrier amplitude at the peak of the synchronizing pulses. In the picture portion of the signal, the video-to-noise-error ratio can change during the line scan since the depth of modulation of the picture carrier can vary while the amount of thermal noise remains constant. Figure 1 shows a single line of the video signal with a staircase modulation of the picture carrier from white to black. The amount of thermal noise associated with the signal is constant during the scanning interval, but since

¹⁵ W. H. Jordan, "Action of Linear Detector on Signals in the Presence of Noise," MIT Rad. Lab., Report 61-23, July, 1943.

¹⁶ H. Urkowitz, "Filters for Detection of Small Radar Signals in Clutter," *Jour. Appl. Phys.*, Vol. 24, p. 1024, August, 1953.

¹⁷ J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," MIT Rad. Lab. Series, No. 24, McGraw-Hill, 1950, p. 163.

the carrier amplitude varies in the manner shown, the video-to-noise-error ratio changes during the scan. The white portions of the line have poorer video-to-noise-error ratios than the darker parts of the line.

If the second detector in the television receiver is not linear, the video-to-noise-error ratio is further deteriorated for the white areas of the picture. The increase of noise is due to additional noise produced by carrier-to-noise and noise-to-noise products in the non-linear detecting process. Various portions of the television picture on the face of the kinescope have different video-to-noise-error ratios depending on the content of the picture being transmitted.

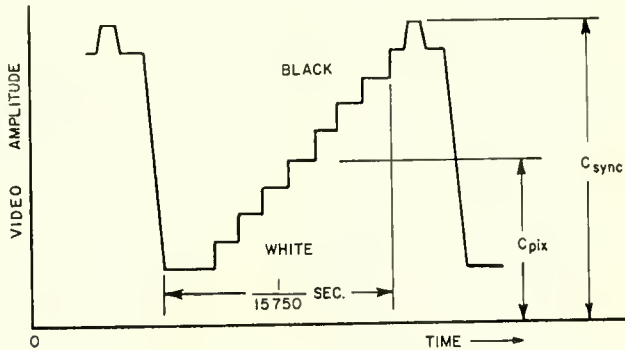


Fig. 1—Staircase modulated picture carrier.

VIDEO-TO-NOISE-ERROR RATIO OF ENVELOPE DETECTOR

The method used in this paper to calculate the video-to-noise-error ratio of an envelope detector preceded by an intermediate frequency amplifier follows that used by Ragazzini² and Bridges.¹³ In this method, the intermediate-frequency response characteristic is divided into a finite number of channels each b radians per second in width. The noise in each channel is represented by a sinusoidal generator having a randomly phased signal output, the amplitude of which is proportional to the intermediate-frequency response at the channel frequency. All the individual noise signals representing the entire intermediate-frequency passband are referenced to a common frequency and phase, and combined vectorially with the desired signal. The magnitude of this resultant quantity gives the detected output from an envelope detector.

The video-to-noise-error ratio of an envelope detector as defined by Equation (1), as a function of the input carrier-to-noise ratio, is given by

$$\frac{Y}{N} = \frac{\beta s^2}{[(sQ)^2 + R^2]^{\frac{1}{2}} + 1}, \quad (4)$$

as shown in Appendix I, Equation (32). The quantity s is the input carrier-to-noise ratio prior to detection, and the constants Q and R are quantities related to the r-m-s noise. Q represents a constant proportional to the r-m-s value of the detected noise when a very large carrier is present and R is a constant proportional to the r-m-s value of detected noise when no carrier is present. For an idealized intermediate-frequency response characteristic, the values of

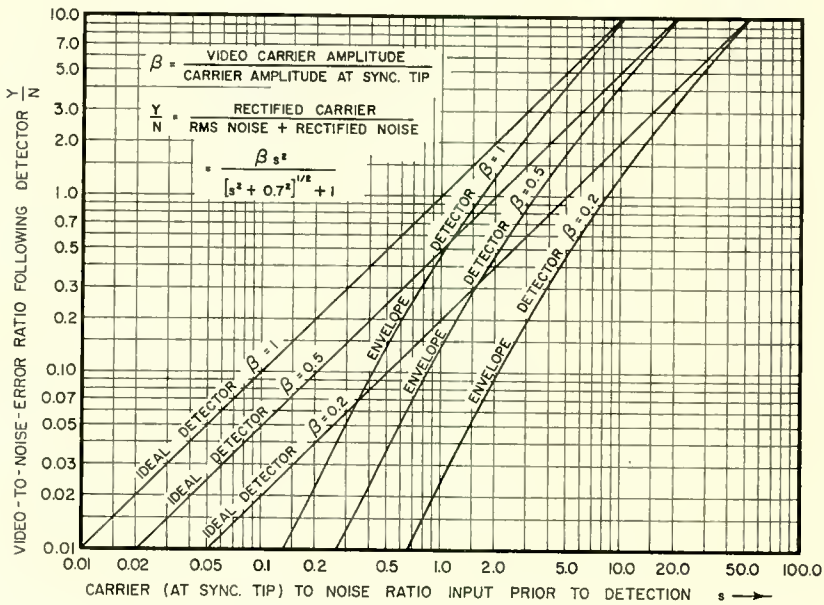


Fig. 2—Calculated video-to-noise-error in the video channel.

Q and R are approximately 1.0 and 0.7 respectively as shown in Appendix I. β is the ratio of the amplitude of the carrier in the picture portion of the signal to the amplitude of the carrier at the peak of the synchronizing pulses. In Figure 1, β is given by the ratio of C_{pix} to C_{sync} .

Using Equation (4), a family of curves for the video-to-noise-error ratio of an envelope detector has been plotted in Figure 2 for different values of β . As β is increased, which corresponds to a reduction in the depth of modulation, the output video-to-noise-error ratio is increased for a given input carrier-to-noise ratio. If positive modulation were used as in the British system instead of negative modulation

this effect would be reversed, that is, for small percentages of modulation the video-to-noise-error ratio would be worse than for high modulation percentages. Since the video-to-noise-error ratio of the envelope detector changes quite rapidly for small values of s , as shown in Figure 2, a small improvement in the noise figure ahead of the detector results in a larger over-all improvement at the output of the detector. For instance, at an input carrier-to-noise ratio of $s = 0.5$, with $\beta = 0.5$, a 3 decibel improvement in the r-f noise figure would result in a 6 decibel over-all improvement in the video-to-noise-error ratio at the output of the envelope detector.

The curves labeled "ideal detector" correspond to detectors in which the output video-to-noise-error ratio is linearly related to the input carrier-to-noise ratio and are included for comparison purposes. The largest difference between the curves for the ideal detector and the envelope detector are seen to occur for small values of β . The relative video-to-noise-error ratio of the envelope detector compared to the ideal detector is shown in Figure 3 for different values of β . When the input carrier-to-noise ratio is unity ($s = 1.0$) and the ratio of picture carrier to the carrier amplitude at the peak of the synchronizing pulses is 0.5 ($\beta = 0.5$), the theoretical advantage of the ideal detector is approximately 11 decibels in video-to-noise-error ratio over that for the envelope detector. The curve of Figure 3 for $\beta = 1$ is very closely that given by Urkowitz¹⁶ (p. 1031 Fig. 8).

The threshold effect of an envelope detector can readily be seen by use of Equation (3) for the ratio of total d-c to the r-m-s noise output of the detector. The specific equation for envelope detection as derived in Appendix I is given by Equation (30) with $Z = 1$,

$$\frac{\text{d-c}}{N_{\text{rms}}} = \frac{s^2 + 1}{[s^2 Q^2 + R^2]^{1/2}} \quad (5)$$

where Q and R have the same values as before. A plot of Equation (5) is shown in Figure 4. The curve labeled envelope detector shows that for small values of input carrier-to-noise ratios the rectified d-c output from the detector approaches a constant value, which corresponds to the value of the rectified noise. This d-c component in the output of the envelope detector determines the maximum brightness level of the display kinescope so that carrier variations below this level are not seen.

VIDEO-TO-NOISE-ERROR RATIO OF SYNCHRONOUS DETECTOR

The video-to-noise-error ratio of the output from a synchronous

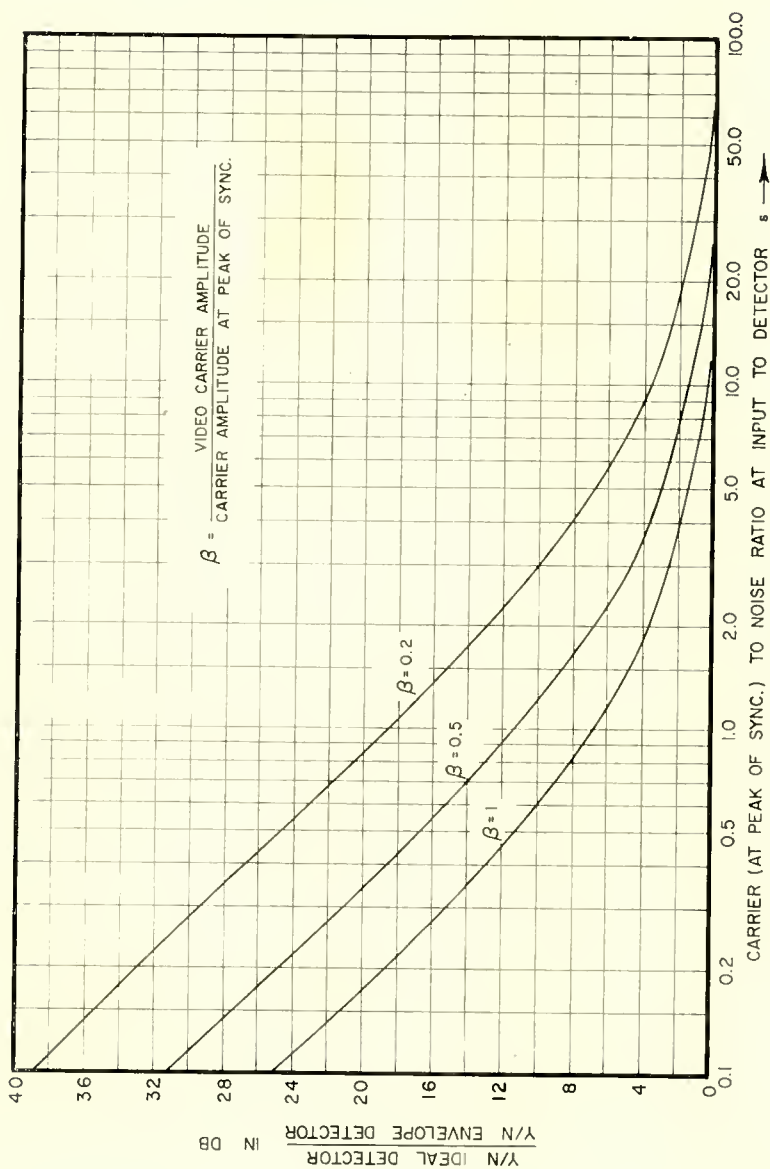


Fig. 3—Relative video-to-noise-error ratio of envelope compared to the ideal detector.

detector is the same as that at the input to the detector. Equivalence of the input and output signal-to-noise ratios can be shown by noting that a synchronous detector is a product-type demodulator. In a double-sideband system, consider a synchronous detector fed from an intermediate-frequency amplifier. The input to the synchronous detector consists of a modulated carrier and noise components which can be expressed by

$$e_1 = C [1 + \lambda F(t)] \cos [\omega_0 t + \theta(\omega_0)] + \sum_{k=0}^q a_k \cos x_k, \quad (6)$$

where

- C = carrier amplitude,
- $\lambda F(t)$ = modulation on the carrier,
- ω_0 = angular carrier frequency,
- $\theta(\omega_0)$ = phase shift of carrier frequency through i-f amplifier,
- a_k = amplitude of the noise of frequency ω_k after passing through i-f amplifier,
- $x_k = \omega_k t + \psi_k$,
- ψ_k = random phase of noise of frequency ω_k .

In the synchronous detector the input signal e_1 is multiplied by the function $e_2 = 2 \cos [\omega_0 t + \theta(\omega_0)]$ so that the output becomes

$$\begin{aligned} 2 e_1 \cos [\omega_0 t + \theta(\omega_0)] &= \\ 2C [1 + \lambda F(t)] \cos [\omega_0 t + \theta(\omega_0)] \cos [\omega_0 t + \theta(\omega_0)] &+ \\ + 2 \cos [\omega_0 t + \theta(\omega_0)] \sum_{k=0}^q a_k \cos x_k & \\ = C [1 + \lambda F(t)] \{1 + \cos [2 \omega_0 t + 2 \theta(\omega_0)]\} & \\ + \sum_{k=0}^q a_k \{ \cos [x_k - \omega_0 t - \theta(\omega_0)] + \cos [x_k + \omega_0 t + \theta(\omega_0)] \}. & \end{aligned}$$

After removal of terms with frequencies higher than the carrier frequency by passing the output through a low-pass filter, the resultant output is given by

$$e_0 = C [1 + \lambda F(t)] + \sum_{k=0}^q a_k \cos [x_k - \omega_0 t - \theta(\omega_0)]. \quad (7)$$

Equation (7) is the output from the synchronous detector when the input consists of a double-sideband signal. A similar result would be obtained when the signal is single sideband and a synchronous detector is used. Equation (7) shows that the demodulated output has a signal component the same as the original, plus noise components with the same amplitude as at the detector input but shifted down in frequency by ω_0 . Thus the output video-to-noise-error ratio is the same as the input carrier-to-noise ratio for all values of the input carrier-to-noise ratio. This equivalence is shown in Figure 2 by the straight lines at

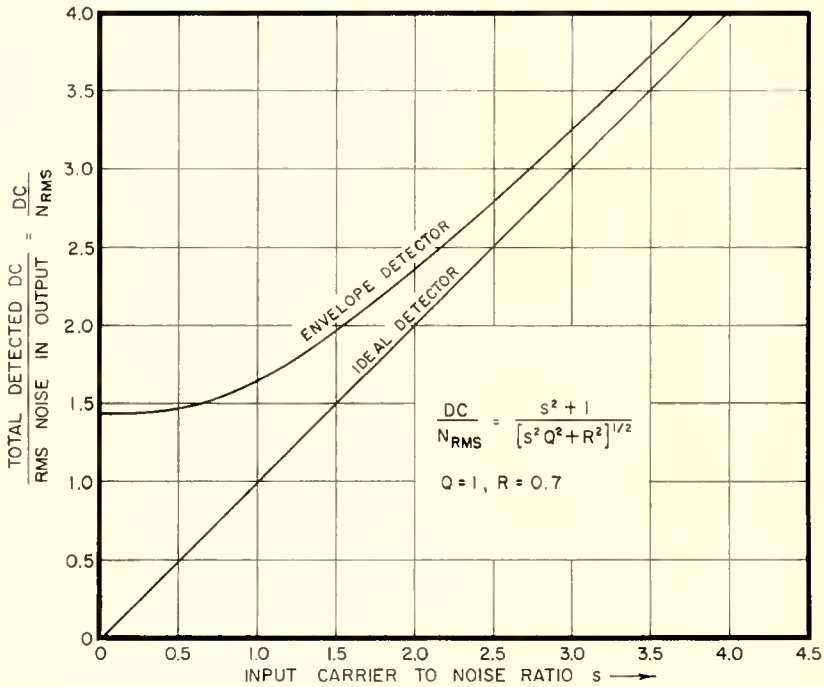


Fig. 4—Ratio of total d-c to r-m-s noise at output of envelope detector following standard television i-f amplifier.

45 degrees to the horizontal, labeled ideal detector for different values of β . In Figure 4 the corresponding curve passes through the origin at 45 degrees. In Equation (7) it is noted that the noise contributes no d-c component, so that the d-c level of the signal is preserved. Thus the synchronous detector has no threshold signal, because noise does not suppress signal detection.

The curves shown in Figure 2 labeled "ideal detector" apply to the synchronous detector. The relative video-to-noise-error advantage of the synchronous detector over the envelope detector is therefore given by the curves of Figure 3.

VIDEO-TO-NOISE-ERROR RATIO OF EXALTED CARRIER DETECTION SYSTEM

Improvement in the video-to-noise-error performance of a detector for low signal levels can also be obtained by use of an exalted carrier system. In the exalted carrier system the carrier-to-noise ratio is increased prior to detection by means of a high Q circuit tuned to the carrier frequency. Such a system is shown diagrammatically in Figure 5. The detected output has a sharply peaked low-frequency response due to peaking of the bandpass response characteristic in the vicinity of the carrier frequency. The peaking in the output response is eliminated by use of a de-emphasis circuit following the detector.

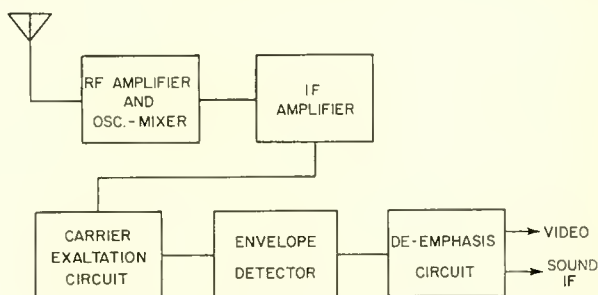


Fig. 5—Exalted carrier detection system.

When the carrier is exalted Z times over the normal carrier level, the output video-to-noise-error ratio as defined by Equation (1) is given by

$$\frac{Y}{N} = \frac{Zs^2}{\left\{ (ZQs)^2 + R^2 \right\}^{1/2} + \frac{1}{Z}}, \quad (8)$$

as is shown in Appendix I, Equation (31). The expression for the video-to-noise-error ratio given by Equation (8) differs from that given by Bridges in that the term $1/Z$ has been added to the denominator. From the definition of video-to-noise-error ratio given by Equation (1), the denominator of Equation (8) represents the r-m-s noise plus the rectified noise; the term $1/Z$ in this case is due to the rectified noise.

The effect of exalting the carrier is shown in Figure 6 for various values of the exaltation ratio Z . The relative improvement in the video-to-noise-error ratio in an exalted carrier detector is approximately 6 decibels over that for an envelope detector when $Z = 4$ and

the input carrier-to-noise ratio is unity. In the limit as Z approaches infinity, the video-to-noise-error ratio of the exalted carrier detector approaches that for the synchronous detector.

The ratio of total rectified d-c to r-m-s noise for the exalted carrier system of detection is given by

$$\frac{\text{d-c}}{N} = \frac{Zs^2 + \frac{1}{Z}}{\{(ZQs)^2 + R^2\}^{1/2}}, \quad (9)$$

as derived in Appendix I, Equation (30). Figure 7 shows the family of curves defined by Equation (9) for various values of the exaltation ratio Z . The curves of Figure 7 show that the threshold signal of the exalted carrier system decreases as the exaltation is increased. The threshold is reduced to zero when Z is made infinite.

EXPERIMENTAL RESULTS FOR THERMAL NOISE PERFORMANCE OF DETECTORS

A noise source and a carrier generator were fed to the i-f input terminals of a television receiver. The d-c output of an envelope detector connected to the i-f amplifier was measured as a function of carrier input amplitude with the noise voltage fixed. This curve is plotted in Figure 8 with the d-c output at zero input carrier normalized to the theoretical value of 1.4. The theoretical curve has been plotted on Figure 8 and is labeled "total d-c output from envelope detector." The discrepancies between curve (A) and the dotted curve (B) are caused by contact potential and the nonlinear small-signal characteristic of the actual diode used in the experiment.

For further illustration, in Figure 8 the total d-c from the envelope detector has been resolved into a d-c component due to the envelope-detected carrier and a d-c component representing rectified noise. These components are obtained by consideration of the d-c term of Equation (25) in Appendix I.

$$\text{d-c} = \frac{C}{s} \left(Z^2 s^2 + 1 \right)^{1/2} = \left(C^2 Z^2 + \frac{C^2}{s^2} \right)^{1/2}. \quad (10)$$

In Equation (10), $s = C/P_n^{1/2}$, where C is the carrier amplitude and

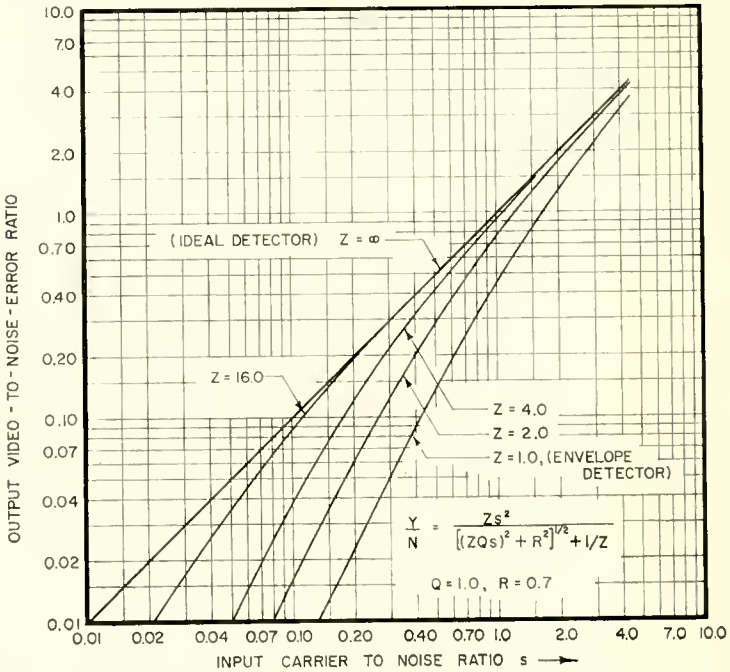


Fig. 6—Video-to-noise-error ratio of exalted-carrier detector following standard television i-f amplifier.

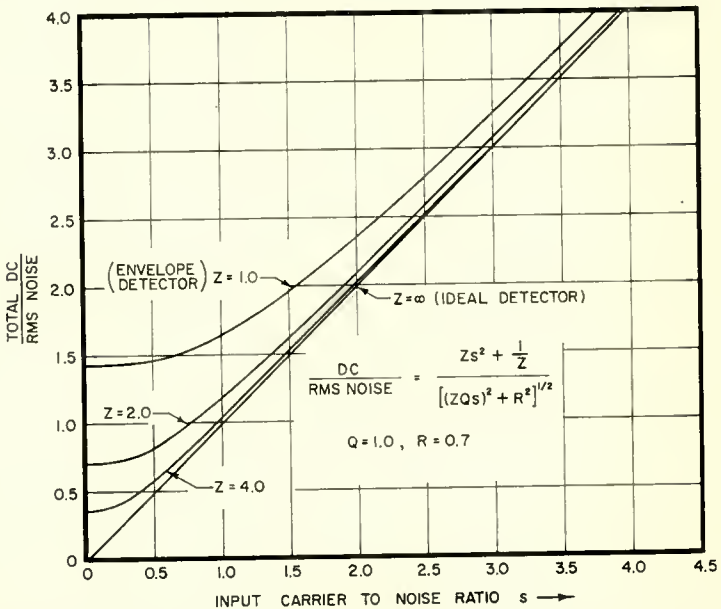


Fig. 7—Ratio of total d-c to r-m-s noise at output of exalted carrier detector after de-emphasis following standard television i-f amplifier.

P_n is the noise power at the output of the i-f amplifier. Since $C^2/s^2 = P_n$ and C^2Z^2 is proportional to the carrier power, assuming that the d-c divides as the ratio of the powers, the d-c due to the rectified noise is given by

$$\begin{aligned} d\text{-}c_{\text{noise}} &= \frac{C^2/s^2}{C^2Z^2 + C^2/s^2} \left(C^2Z^2 + \frac{C^2}{s^2} \right)^{1/2} \\ &= \frac{C^2/s^2}{(C^2Z^2 + C^2/s^2)^{1/2}}. \end{aligned} \quad (11)$$

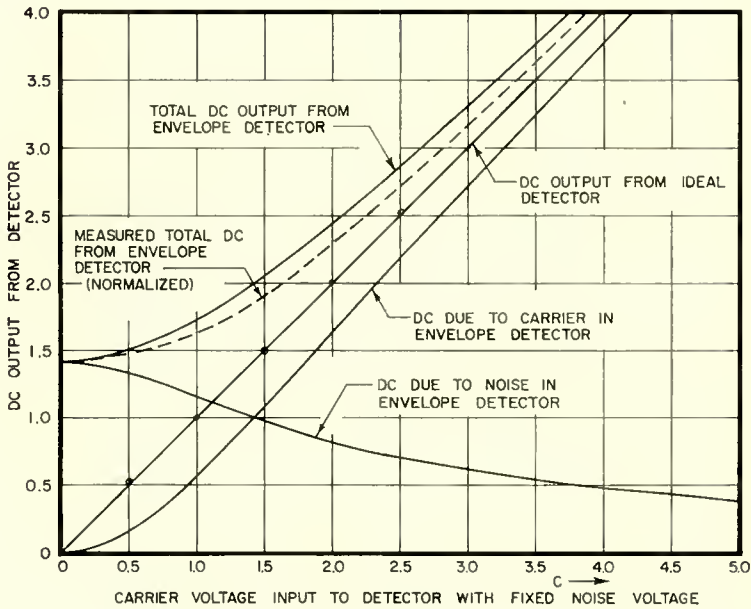


Fig. 8—D-C output from envelope detector with a variable carrier and a fixed noise voltage input.

The d-c resulting from the rectified carrier is then

$$\begin{aligned} d\text{-}c_{\text{carrier}} &= \frac{C^2Z^2}{C^2Z^2 + C^2/s^2} \left(C^2Z^2 + \frac{C^2}{s^2} \right)^{1/2} \\ &= \frac{C^2Z^2}{(C^2Z^2 + C^2/s^2)^{1/2}}. \end{aligned} \quad (12)$$

The d-c output from a synchronous detector connected to the same i-f strip was then measured and was found to be a perfect linear func-

tion. It is plotted by the circles coincident with the line marked "output from ideal detector" in Figure 8.

Next the r-m-s noise output from the synchronous detector was measured. The carrier generator was then adjusted to give a d-c output equal to the r-m-s noise output from the synchronous detector, establishing a carrier-to-noise ratio of unity since

$$\frac{\text{d-c output}}{\text{r-m-s noise output}} = \frac{\text{input carrier amplitude}}{\text{r-m-s noise input}} \quad (13)$$

for a synchronous detector.

The carrier generator was then modulated with a standard video test pattern giving a unity video-to-noise-error ratio at sync tips. A panel of observers was asked to view a display of the synchronous detector output on a video monitor. An envelope detector was then substituted for the synchronous detector. All observers agreed that it was necessary to raise the carrier level to restore the display to its previous readability. Table I summarizes the results, giving averages for five observers.

Table I

Type of Pattern	Required Increase in Input Carrier (db) When Envelope Detector is Substituted			
	$\beta = 0.2$	$\beta = 0.2$ Theoretical	$\beta = 0.5$	$\beta = 0.5$ Theoretical
Indian Head Test Pattern	11	9.5	8	6.4
Close-Up of Model	8	9.5	7	6.4

These results indicate some divergence between the theory and observed results. However, this divergence is not unexpected since no allowance was made for the nonlinear characteristics of the kinescope. Also, there is obviously a difference between what constitutes readability in a test pattern and a more general scene such as the close-up of a model used in the test.

It is believed that a unity carrier-to-noise ratio represents a sufficiently degraded video display to be considered limiting. That is to say, even though a detector could handle signals of lower s , the resultant display would have little or no entertainment value. The data of Table I thus show that, in the limit, substitution of a synchronous detector for an envelope detector can result in improvements in the d-c/noise output ratio of approximately 10 decibels maximum.

Relatively greater improvements are, of course, available where even lower d-c/noise output ratios can be made to produce useful information.

IMPULSE RESPONSE OF ENVELOPE DETECTORS IN ASYMMETRICAL SIDEBAND SYSTEMS

The response of various detectors to impulsive type noise is considered next. The intermediate-frequency amplifier preceding the detector is represented by a flat-staggered triple-tuned bandpass filter in this analysis. The approximate form of the expression for the flat-staggered triple-tuned filter which assumes arithmetic symmetry is used in this analysis since the bandwidth-to-center-frequency ratio in the intermediate-frequency amplifier of most television receivers is approximately 10 per cent.

The envelope of the impulse response as derived in Appendix II is given by

$$h(\alpha) = \frac{\Delta\omega}{2} \left[e^{-2\alpha} + \frac{2}{\sqrt{3}} e^{-\alpha} \sin \left(\sqrt{3}\alpha - \frac{\pi}{3} \right) \right], \quad (14)$$

where

$$\alpha = \frac{\pi\Delta\omega t}{2}$$

$\Delta\omega = 3$ -decibel bandwidth of flat-staggered triple-tuned circuit.

A plot of Equation (14) is shown in Figure 9 for a bandwidth of two units. The curve in Figure 9 is the response of an envelope detector to an impulse with no signal present.

The impulse response of a detector preceded by an intermediate-frequency amplifier in the presence of a carrier is found by calculating the envelope produced by the vector addition of the impulse response of the bandpass system and the carrier. The output of the detector under these conditions is dependent on the relative phase of the impulse and the carrier. The impulse response of an envelope detector preceded by a flat-staggered triple-tuned bandpass circuit in the presence of a carrier at the 50 per cent response point on the amplitude characteristic is given by

$$h(\alpha) = \{ [P(\alpha)]^2 + [Q(\alpha)]^2 \}^{1/2} \quad (15)$$

where

$$P(\alpha) = A - \left[e^{-2\alpha} + \frac{2}{\sqrt{3}} e^{-\alpha} \sin \left(\sqrt{3}\alpha - \frac{\pi}{3} \right) \right] \cos (2.402\alpha + 19^\circ 34' - \theta), \quad (16)$$

$$Q(\alpha) = - \left[e^{-2\alpha} + \frac{2}{\sqrt{3}} e^{-\alpha} \sin \left(\sqrt{3}\alpha - \frac{\pi}{3} \right) \right] \sin (2.402\alpha + 19^\circ 34' - \theta), \quad (17)$$

A = carrier amplitude,

θ = relative phase between carrier and applied impulse as is shown in Appendix II.

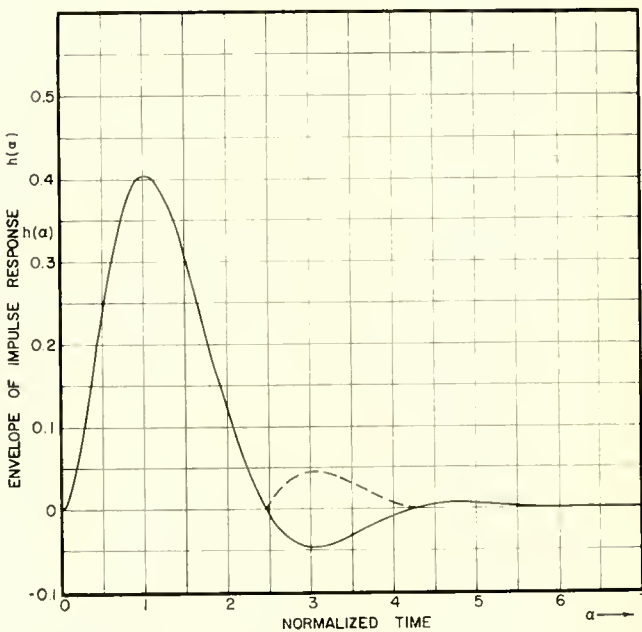


Fig. 9—Envelope of impulse response of a flat-staggered triple-tuned circuit.

Figure 10 shows a plot of Equation (15) for various values of the relative phase between the carrier and the applied impulse for a unit carrier amplitude. For a d-c component to exist in the output, the average of a number of pulses must be taken, since an individual impulse cannot produce a d-c component. As all phases are possible from

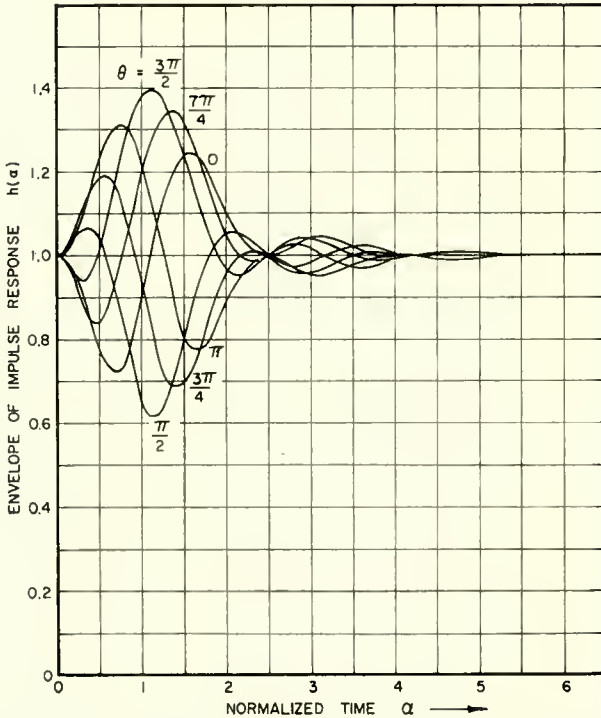


Fig. 10—Envelope of impulse response of a flat-staggered triple-tuned circuit in the presence of a carrier at 50 per cent point on selectivity curve (carrier amplitude = 1.0).

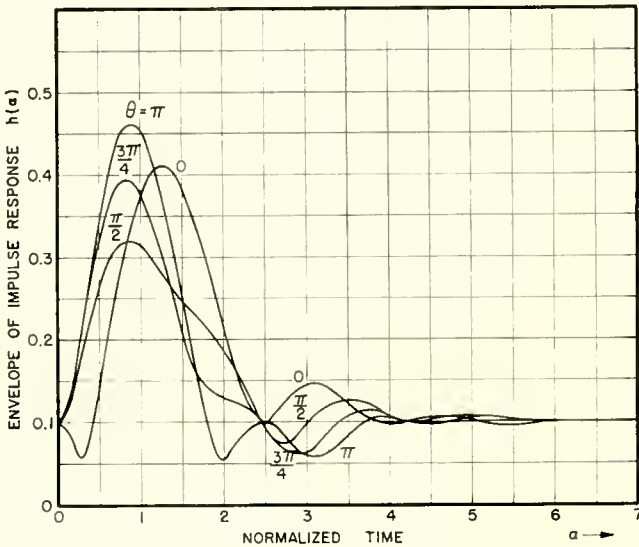


Fig. 11—Envelope of impulse response of a flat-staggered triple in the presence of a carrier at 50 per cent point on selectivity curve.

a source which produces impulses random in time, there is no d-c component in the output when the carrier level is high relative to the impulsive noise amplitude. However, when the carrier amplitude is made small, a d-c component is produced in the envelope detector since each of the individual impulses has a d-c component independent of the relative phase angle θ . The existence of a d-c component is shown by the family of curves in Figure 11 which is drawn for various values of the relative phase angle θ with a carrier amplitude of 0.1.

IMPULSE RESPONSE OF SYNCHRONOUS DETECTOR

When a synchronous detector is used, either the envelope of the in-phase component or the envelope of the quadrature component of the impulse noise plus carrier output from the bandpass filter can be detected. The envelopes of the in-phase and quadrature components are given by Equations (16) and (17), respectively. The in-phase and quadrature components of the envelope of the impulse response of a flat-staggered triple-tuned bandpass circuit in the presence of a carrier at the 50 per cent response point on the selectivity curve is shown in Figure 12. A plot of Equation (16) for various phase angles, θ , between the impulse and carrier is shown in Figure 13 for $A = 0.1$. From Equation (16) it is seen that the shape of the envelope of the impulse response is independent of the carrier amplitude. Thus there can never be a d-c component resulting from noise in the output of a synchronous detector since the average d-c caused by a number of randomly phased impulses is zero. When the carrier amplitude A is unity or larger, the output from the envelope detector is nearly the same as that obtained from the synchronous detector.

The envelope of the quadrature component of the impulse response of a flat-staggered triple-tuned bandpass circuit in the presence of a carrier is similar to that for the in-phase component. However, there is no d-c component caused by the carrier so that the $Q(\alpha)$ curves are similar to the $P(\alpha)$ curves shown in Figure 13 except for a 90° phase lag and an axis shift to zero.

EXPERIMENTAL RESULTS FOR IMPULSE RESPONSE OF DETECTORS

The impulse response of a stagger-tuned triple was experimentally measured. Figure 14 shows the experimental setup.

A carrier generator and an impulse generator were fed to an adder circuit to permit adjustment of the carrier-to-impulse ratio. The combination signal was then fed to a stagger-tuned triple amplifier with center frequency 43.75 megacycles and a 6-decibel bandwidth of 4 megacycles. An envelope detector and a carrier-driven product detector were connected to the output of the triple-tuned amplifier. Further

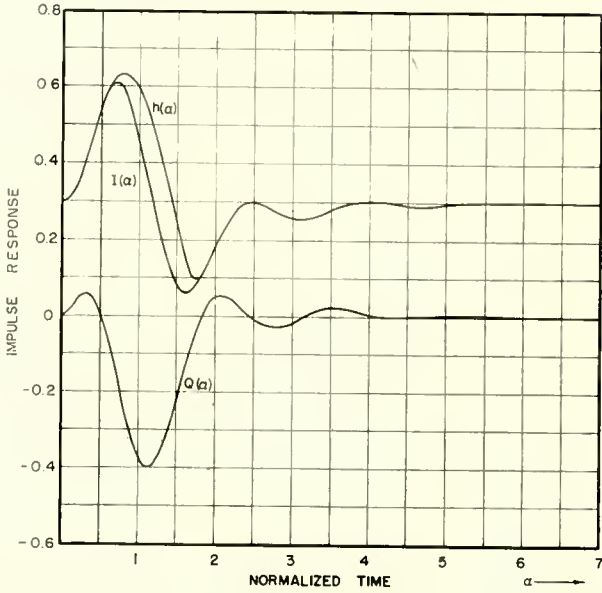


Fig. 12—In-phase and quadrature components of the impulse response of a flat-staggered triple in the presence of a carrier at 50 per cent point on the selectivity curve.

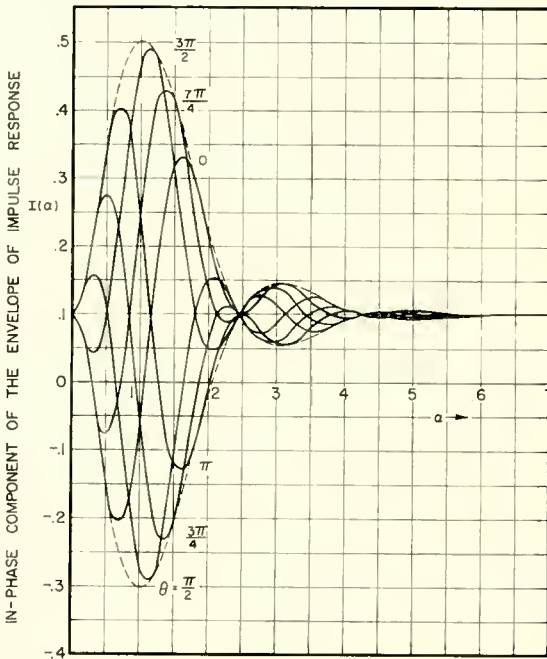


Fig. 13—In-phase component of impulse response of flat-staggered triple in the presence of a carrier at 50 per cent point on selectivity curve.

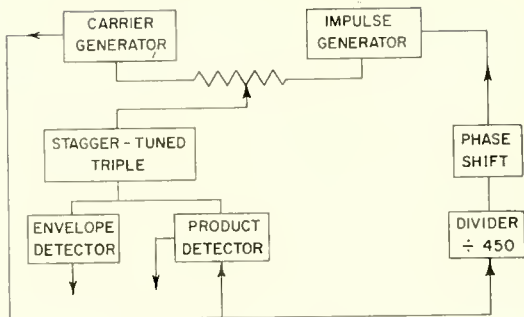


Fig. 14—Equipment used for impulse response test.

a divider chain was used to obtain a 116.7-kilocycle signal from the 45.75-megacycle carrier by division by 450. The 116.7-kilocycle sine wave was clipped and differentiated to obtain periodic impulses. A phase shifter in the divider chain permitted adjustment of the phase of the 116.7-kilocycle pulses relative to the 45.75-megacycle carrier wave.

Curves (a), (b), (c) and (d) in Figure 15 show the product detector output for impulse phases relative to the carrier of 0° , 90° , 180° , and 270° respectively for very high impulse-to-carrier ratios. The smeary appearance of the curves is due to some phase jitter in the divider chain.

Curve (e) of Figure 15 shows the product detector output when the divider was unlocked from the 45.75-megacycle carrier so that the

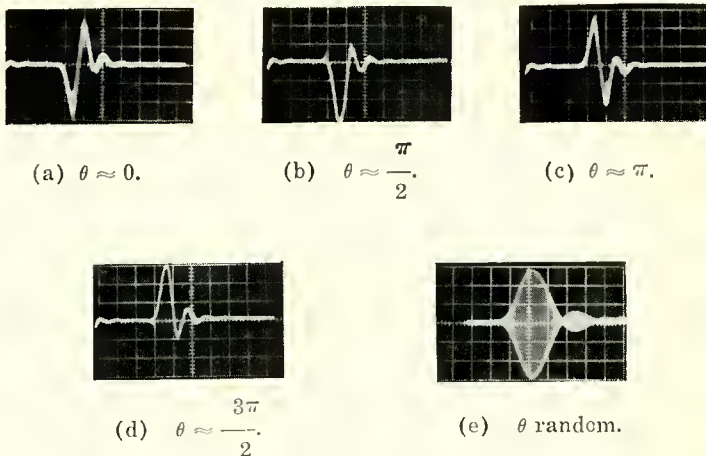


Fig. 15—Impulse response of synchronous detector. θ = relative phase between impulse and carrier at $t = 0$.

impulses assumed all possible phases in turn at the divided beat-frequency rate.

Curves (a) to (e) should be compared to Figure 13 where the theoretical curves are plotted. The results are in good agreement.

Curve (a) of Figure 16 shows the envelope detector response at a phase of 0° , and curve (b) shows the envelope detector response when the divider was unlocked. These results are in good agreement with Figure 11, the theoretical curve for the envelope response at 0.1 carrier. It is to be especially noted that the envelope is practically phase insensitive as shown in Figure 16 and as predicted by the theory.



Fig. 16—Impulse response of envelope detector.

CONCLUSIONS

These studies indicate that with respect to fluctuation noise, the video-to-noise-error ratio at the output of an envelope detector is less than the input carrier-to-noise ratio. The loss in video-to-noise-error ratio increases very rapidly as the input carrier-to-noise ratio is lowered, and ultimately becomes so severe that a virtual threshold exists, below which signal detection is impossible. This threshold is established by the rectification of noise which results in a minimum d-c output voltage from the detector, below which the output cannot fall.

When synchronous detection is used, the video-to-noise-error ratio at the detector output is always found to be the same as the input carrier-to-noise ratio. No detection threshold exists in this case, and the signal will always be detected, although, of course, the output will contain the noise intermingled with the signal. There is no noise rectification and thus no d-c component in the output caused by the noise.

The exalted carrier detector is found to perform like an envelope detector when the exaltation is unity, and like a synchronous detector when the exaltation is infinite. With intermediate exaltation values the performance is found between these limits.

With respect to impulse noise, envelope detectors are found to be responsive to the timing of impulse noise (i.e., the relative phasing

of the impulse with respect to the received carrier) when the noise amplitude is small with respect to the carrier amplitude. However, when the noise is large with respect to the received carrier the envelope detector is not responsive to the timing of the impulse and signal detection ceases for the duration of the impulse. In this case low-frequency and d-c components are developed in the output.

Synchronous detectors, in the presence of impulse noise, are always responsive to the timing of the impulse with respect to the carrier. No low frequency or d-c components are developed in the output of a synchronous detector when the impulses are not coherent with the carrier. The output spectrum for an impulsive type input has its predominant frequency components clustered around the frequency equal to the difference between the geometric mean frequency of the bandpass system and the carrier frequency. This frequency difference can be made large in asymmetric sideband systems. For certain applications such as in television, the synchronous detector provides a means for obtaining noise-free synchronizing signals since the detected impulsive noise which occurs at a relatively high beat frequency can be eliminated by a low-pass filter.

APPENDIX I — SIGNAL-TO-NOISE-ERROR RATIO OF DETECTORS

This analysis proceeds in the same manner as that given by Bridges¹³ except that the signal considered is the rectified carrier and not the modulation on the carrier. The thermal noise voltage which has passed through the intermediate frequency passband is represented as a sum of voltages due to a finite number of noise sources within the passband. The passband is divided into a finite number of channels, each b radians per second wide. The noise in each channel is represented by a sinusoidal generator with a constant amplitude, randomly phased output. The amplitude of the output voltage in a given channel is proportional to the amplitude characteristic of the passband at the channel frequency. Using this concept, the signal plus thermal noise prior to detection can be written as

$$V_c + V_n = C Z \cos x_c + \sum_{k=0}^q a_k \cos x_k \quad (18)$$

where

V_c = carrier voltage,

V_n = thermal noise voltage,

C = carrier amplitude without exaltation,

Z = carrier exaltation,

$$x_c = \omega_c t,$$

ω_c = carrier frequency,

$$a_k = a(\omega_k) [P_n b / F_{eq}]^{1/2},$$

$a(\omega_k)$ = amplitude response of intermediate-frequency characteristic at ω_k ,

P_n = thermal noise power at output of intermediate-frequency amplifier,

$b = \omega_{k+1} - \omega_k$ = channel width in radians per second,

F_{eq} = equivalent noise bandwidth of intermediate-frequency passband,

$$x_k = \omega_k t + \psi_k,$$

ψ_k = random phase.

Equation (18) can be rewritten as

$$V_c + V_n = [A^2 + B^2]^{1/2} \cos \left[x_c + \tan^{-1} \frac{B}{A} \right], \quad (19)$$

where

$$A = CZ + \sum_{k=0}^q a_k \cos(x_k - x_c), \quad (20)$$

$$B = \sum_{k=0}^q a_k \sin(x_k - x_c). \quad (21)$$

When the voltage output of the intermediate-frequency amplifier represented by Equation (19) is detected, the output from the detector is given by

$$\begin{aligned} e_0 &= [A^2 + B^2]^{1/2} \\ &= \left\{ C^2 Z^2 + 2CZ \sum_{k=0}^q a_k \cos(x_k - x_c) + \sum_{k=0}^q a_k^2 \right. \\ &\quad \left. + 2 \sum_{r=0}^{q-1} \sum_{k=r+1}^q a_r a_{k+r} \cos(x_r - x_{k+r}) \right\}^{1/2}. \end{aligned} \quad (22)$$

Let the input carrier-to-noise ratio be given by

$$s = \frac{C}{P_n^{1/2}} = \left[\frac{C^2}{\sum_{k=0}^q a_k^2} \right]^{1/2}. \quad (23)$$

Using the definition of a_k and s , Equation (18) can then be written as

$$e_0 = \left[C^2 Z^2 + \frac{C^2}{s^2} \right]^{1/2} \left\{ 1 + \frac{2Zs}{Z^2 s^2 + 1} \left[\frac{b}{F_{cq}} \right]^{1/2} \sum_{k=0}^q a(\omega_k) \cos(x_k - x_c) \right. \\ \left. + \frac{2}{Z^2 s^2 + 1} \left[\frac{b}{F_{cq}} \right] \sum_{r=0}^{q-1} \sum_{k=1}^{q-r} a(\omega_r) a(\omega_{k+r}) \cos(x_r - x_{k+r}) \right\}^{1/2} \quad (24)$$

If the channel width, b , is made small compared to the equivalent noise bandwidth, F_{cq} , of the intermediate-frequency amplifier, Equation (24) can be represented by the first two terms of the binomial expansion. Use of the first two terms of the binomial expansion gives

$$e_0 = \frac{C}{s} (Z^2 s^2 + 1)^{1/2} \left\{ 1 + \frac{Zs}{Z^2 s^2 + 1} \left[\frac{b}{F_{cq}} \right]^{1/2} \sum_{k=0}^q a(\omega_k) \cos(x_k - x_c) \right. \\ \left. + \frac{1}{Z^2 s^2 + 1} \left[\frac{b}{F_{cq}} \right] \sum_{r=0}^{q-1} \sum_{k=1}^{q-r} a(\omega_r) a(\omega_{k+r}) \cos(x_r - x_{k+r}) \right\} \quad (25)$$

The first term in the brace of Equation (25) represents the d-c component of the output. The second and third terms represent the carrier-and-noise products and the noise-and-noise products, respectively.

The r-m-s value of the noise components in the output is given by

$$\text{r-m-s noise} = \frac{C}{s} (Z^2 s^2 + 1)^{1/2} \left\{ \left[\frac{ZsQ}{Z^2 s^2 + 1} \right]^2 + \left[\frac{R}{Z^2 s^2 + 1} \right]^2 \right\}^{1/2} \quad (26)$$

where

$$Q = \left[\frac{b}{F_{cq}} \sum_{k=0}^q a^2(\omega_k) \right]^{1/2} \cong 1 \quad (27)$$

$$\text{since } \lim_{\substack{b \rightarrow 0 \\ q \rightarrow \infty}} b \sum_{k=0}^q a^2(\omega_k) = F_{cq}$$

$$R = \left\{ \left[\frac{b}{F_{cq}} \right]^2 \sum_{r=0}^{q-1} \sum_{k=1}^{q-r} a^2(\omega_r) a^2(\omega_{k+r}) \right\}^{1/2} \quad (28)$$

The exalted d-c component is given by

$$\text{d-c} = \frac{C}{s} (Z^2 s^2 + 1)^{1/2}$$

The ratio of d-c to r-m-s noise at the output of the detector is then given by

$$\frac{\text{exalted d-c}}{\text{r-m-s noise}} = \frac{Z^2 s^2 + 1}{[(ZsQ)^2 + R^2]^{1/2}}. \quad (29)$$

After de-emphasis of the exalted d-c component by the factor Z , the ratio becomes

$$\frac{\text{d-c}}{\text{r-m-s noise}} = \frac{Zs^2 + \frac{1}{Z}}{[(ZsQ)^2 + R^2]^{1/2}}. \quad (30)$$

If the output video-to-noise-error ratio Y/N is defined as the ratio of rectified carrier to r-m-s noise plus rectified noise, then the video-to-noise-error ratio for the exalted carrier system can be written as

$$\frac{Y}{N} = \frac{Zs^2}{[(ZsQ)^2 + R^2]^{1/2} + \frac{1}{Z}}. \quad (31)$$

When $Z = 1$, Equation (31) reduces to

$$\frac{Y}{N} = \frac{s^2}{[(sQ)^2 + R^2]^{1/2} + 1}. \quad (32)$$

which is the expression for the video-to-noise-error ratio of the envelope detector.

When $Z = \infty$, Equation (31) reduces to

$$\frac{Y}{N} = s, \quad (33)$$

which is the video-to-noise-error ratio of the synchronous detector, and shows that the output video-to-noise-error ratio is equal to the input carrier-to-noise ratio.

EVALUATION OF Q AND R

Consider the idealized television intermediate-frequency response

of the receiver shown by curve A in Figure 17. The equivalent noise bandwidth for this characteristic is given by¹⁸

$$F_{eq} = \frac{1}{G} \int G_f df, \tag{34}$$

where the square of the amplitude characteristic G_f is given by curve B in Figure 17. The value of G , the available power gain, is taken as unity for this case. The value of F_{eq} for this ideal characteristic is 3.26 megacycles as calculated by use of Equation (34).

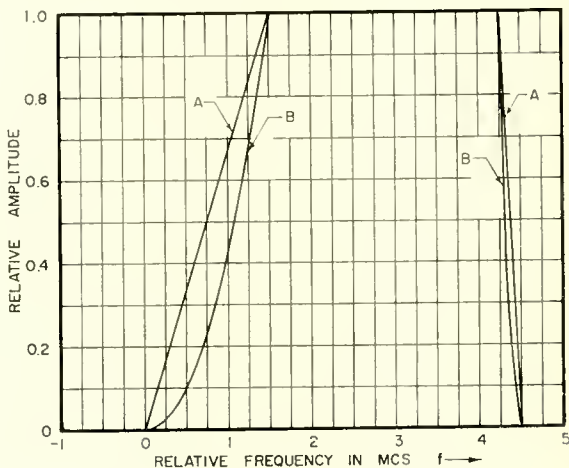


Fig. 17—Idealized receiver i-f response.

Dividing the passband into 18 different channels, each 0.25 megacycle wide, and using Equation (27) for Q , gives

$$Q = \left[\frac{b}{F_{eq}} \sum_{k=0}^{18} a^2(\omega_k) \right]^{1/2} = \left[\frac{.25}{3.26} (13.53) \right]^{1/2} = 1.017. \tag{35}$$

Using Equation (28) for R gives

$$R = \left\{ \left[\frac{b}{F'_{eq}} \right]^2 \sum_{r=0}^{17} \sum_{k=1}^{18-r} a^2(\omega_r) a^2(\omega_{k+r}) \right\}^{1/2}$$

¹⁸ S. Goldman, *Frequency Analysis, Modulation and Noise*, McGraw-Hill Book Co., New York, N. Y., 1948, p. 254.

$$= \left\{ \left[\frac{.25}{3.26} \right]^2 (85.2) \right\}^{1/2}$$

$$= 0.707. \quad (36)$$

APPENDIX II — IMPULSE RESPONSE OF FLAT-STAGGERED TRIPLE IN THE PRESENCE OF A CARRIER

The exact form of the normalized flat-staggered triple-tuned band-pass impedance function can be written as¹⁹

$$R = \frac{\delta^3}{\left\{ \left[\frac{j\omega}{\omega_0} + \frac{\omega_0}{j\omega} \right]^2 + \delta \left[\frac{j\omega}{\omega_0} + \frac{\omega_0}{j\omega} \right] + \delta^2 \right\} \left[\frac{j\omega}{\omega_0} + \frac{\omega_0}{j\omega} + \delta \right]}, \quad (37)$$

where

$$\delta = \frac{\text{3-decibel bandwidth}}{f_0},$$

$$f_0 = \frac{\omega_0}{2\pi} = \text{center frequency of system.}$$

For the narrow-band case $\left[\frac{j\omega}{\omega_0} + \frac{\omega_0}{j\omega} \right]$ can be approximated by

$j \frac{2}{\omega_0} (\omega - \omega_0)$, so that Equation (37) can be written as

$$R = \frac{\delta^3}{\left\{ \left[\frac{2}{\omega_0} (p - j\omega_0) \right]^2 + \delta \left[\frac{2}{\omega_0} (p - j\omega_0) \right] + \delta^2 \right\} \left\{ \frac{2}{\omega_0} (p - j\omega_0) + \delta \right\}}, \quad (38)$$

where $p = j\omega$.

The unit impulse response of this bandpass system is given by $L^{-1} \left[\frac{Z(p)}{R} \right]$ and since $L^{-1} [F(p + \alpha)] = e^{-\alpha t} L^{-1} [F(p)]$, the impulse response can be written as

¹⁹ G. E. Valley and H. Wallman, *Vacuum Tube Amplifiers*, McGraw-Hill Book Co., Inc., New York, N. Y., 1948, p. 183.

$$\begin{aligned}
 H(t) &= \text{Re} \left[e^{j\omega_0 t} L^{-1} \left\{ \frac{\frac{\omega_0^3}{8} \delta^3}{\left[p^2 + \frac{\delta\omega_0}{2} p + \frac{\omega_0^2 \delta^2}{4} \right] \left[p + \frac{\delta\omega_0}{2} \right]} \right\} \right] \\
 &= \frac{\delta\omega_0}{2} \left[e^{-\frac{\delta\omega_0 t}{2}} + \frac{2}{\sqrt{3}} e^{-\frac{\delta\omega_0 t}{4}} \sin \left(\frac{\sqrt{3}}{4} \omega_0 \delta t - \frac{\pi}{3} \right) \right] \cos \omega_0 t. \quad (39)
 \end{aligned}$$

The carrier is assumed to be at the frequency where the amplitude response is equal to one half. For the flat-staggered triple-tuned circuit this frequency can be found from the equation for the absolute value of the amplitude given by

$$\left| \frac{Z}{R} \right| = \frac{\delta^6}{\left[\delta^6 + \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)^6 \right]^{1/2}} = \frac{\delta^6}{\left\{ \delta^6 + \left[\frac{2}{\omega_0} (\omega - \omega_0) \right]^6 \right\}^{1/2}}. \quad (40)$$

Equating $\left| \frac{Z}{R} \right|$ to one half,

$$\frac{\delta^6}{\left\{ \delta^6 + \left[\frac{2}{\omega_0} (\omega_c - \omega_0) \right]^6 \right\}^{1/2}} = \frac{1}{2}. \quad (41)$$

Solving Equation (41) for ω_c gives

$$\omega_c = \omega_0 \pm \frac{(3)^{1/6}}{2} \delta \omega_0. \quad (42)$$

The phase at this frequency is found by use of Equation (38) and is given by

$$\Psi_c = \tan^{-1} \frac{x_c^3 - 2\delta^2 x_c}{\delta^3 - 2\delta x_c^2}, \quad (43)$$

where

$$x_c = \frac{2}{\omega_0} (\omega_c - \omega_0) = \pm (3)^{1/6} \delta.$$

When the unit impulse is applied to a triple-tuned bandpass system in the presence of a steady carrier, the input to the detector can be expressed as

$$f(t) = A \cos \omega_0 t + B \cos (\omega_c t + \Psi_c + \theta), \quad (44)$$

where

$$A = \frac{\delta \omega_0}{2} \left[e^{-\frac{\delta \omega_0 t}{2}} + \frac{2}{\sqrt{3}} e^{-\frac{\delta \omega_0 t}{4}} \sin \left(\frac{\sqrt{3}}{4} \omega_0 \delta t - \frac{\pi}{3} \right) \right],$$

B = steady carrier amplitude,

θ = relative phase between impulse and carrier.

Equation (44) can be written as

$$\begin{aligned} f(t) &= \text{Re} \{ A e^{j\omega_0 t} + B e^{j(\omega_c t + \Psi_c + \theta)} \} \\ &= \text{Re} \{ e^{j(\omega_c t + \Psi_c + \theta)} [A \cos [(\omega_0 - \omega_c)t - \Psi_c - \theta] + B \\ &\quad + j A \sin [(\omega_0 - \omega_c)t - \Psi_c - \theta]] \}. \end{aligned} \quad (45)$$

Equation (45) can be rewritten as

$$f(t) = [P^2(t) + Q^2(t)]^{1/2} \cos \left[\omega_c t + \Psi_c + \theta + \tan^{-1} \frac{Q(t)}{P(t)} \right] \quad (46)$$

where

$$P(t) = A \cos [(\omega_0 - \omega_c)t - \Psi_c - \theta] + B, \quad (47)$$

$$Q(t) = A \sin [(\omega_0 - \omega_c)t - \Psi_c - \theta]. \quad (48)$$

If $\alpha = \frac{\delta \omega_0 t}{4}$, then Equations (47) and (48) become

$$P(\alpha) = \frac{\delta \omega_0}{2} \left[e^{-2\alpha} \right. \quad (49)$$

$$\left. + \frac{2}{\sqrt{3}} e^{-\alpha} \sin \left(\sqrt{3} \alpha - \frac{\pi}{3} \right) \right] \cos [2(3)^{1/6} \alpha - \Psi_c - \theta] + B,$$

$$Q(\alpha) = \frac{\delta \omega_0}{2} \left[e^{-2\alpha} \right. \quad (50)$$

$$\left. + \frac{2}{\sqrt{3}} e^{-\alpha} \sin \left(\sqrt{3} \alpha - \frac{\pi}{3} \right) \right] \sin [2(3)^{1/6} \alpha - \Psi_c - \theta].$$

A NOTE ON THE DISPERSION OF INTERDIGITAL DELAY LINES*

BY

F. PASCHKE

RCA Laboratories,
Princeton, N. J.

Summary—It is shown that the effect of the backwall on the dispersion of an interdigital delay line can be taken into account by a lumped susceptance which periodically loads the "ideal" line. Experimental results are in good agreement with the theory.

INTERDIGITAL delay lines have found widespread use in traveling-wave amplifiers and oscillators. The most important characteristic of a delay line in a traveling-wave device is its dispersion, for the dispersion determines the electron velocity required to obtain interaction. Theoretical prediction of the dispersion is therefore highly desirable. Although the theory of Leblond and Mourier¹ yields satisfactory results, a further improvement in the theory is possible.

The structure to be considered is depicted in Figure 1. It consists of an interdigital delay line mounted in a channel with the line spaced a distance b from the bottom of the channel. For lines with relatively large b , it has been shown² that the field configuration is very much like that of a TEM-wave traveling with the velocity of light along a "folded-strip line" with strip spacing t , and strip width d . It has further been shown that this model can be successfully applied to the computation of rather complicated structures.³ In this paper, this model will be used to compute the effect of the backwall distance on the dispersion of the interdigital line shown in Figure 1.

It is assumed that the field between each bar and the backwall is that of a "strip-above-ground" transmission line which is short-

* Manuscript received March 24, 1958.

¹ A. Leblond and G. Mourier, "Etude des lignes à barreaux a structure périodique pour tubes électroniques UHF," *Annal. de Radioélectr.*, Vol. 9, p. 180, April, 1956 and p. 311, October, 1956.

² R. Warnecke, P. Guenard and O. Döhler, "Phénomènes fondamentaux dans les tubes à onde progressive," *L'Onde élect.*, Vol. 34, p. 323, April, 1954.

³ F. Paschke, "Untersuchung einer Interdigital-Verzögerungsleitung," *Arch. elektr. Übertr.*, Vol. 10, p. 195, May, 1956.

circuited at one end. Neglecting fringing fields, the characteristic impedance of this line is given by

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{b}{L-t}. \quad (1)$$

Seen from the center of the gap between the tip of a bar and the sidewall (Figure 1), the fields between the digit and the backwall can be taken into account by the susceptance Y of the short circuited strip-above-ground transmission line of length $s + t/2$ and characteristic impedance Z . This susceptance is determined by the well-known equation

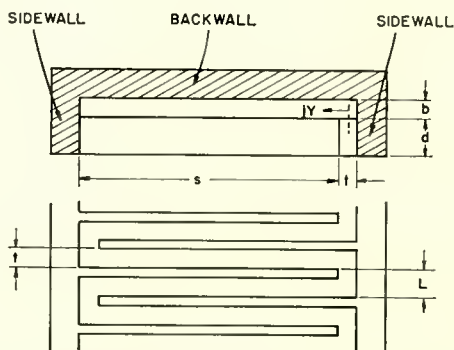


Fig. 1—Interdigital delay line. The fields between backwall and digits can be taken into account by the susceptance Y of the short-circuited strip-above-ground transmission line of length $s + t/2$.

$$Y = -\frac{1}{Z_0} \cot \frac{2\pi}{\lambda} \left(s + \frac{t}{2} \right). \quad (2)$$

Thus the delay line of Figure 1 can be represented by an interdigital line without backwall but with every gap between bar and sidewall shunted by a susceptance Y (Figure 2). Such a structure was treated in an earlier paper.³ The delay for the fundamental wave is

$$\frac{c}{v_p} = \frac{\lambda}{2L} \left(1 - \frac{\phi}{\pi} \right), \quad (3)$$

where

$$\psi = \phi - \pi \quad (4)$$



Fig. 2—Equivalent loaded line. The susceptance Y which represents the fields between digits and backwall appears in each gap between the tip of each digit and the sidewall.

is the phaseshift per cell of length L . ϕ is the phaseshift of the loaded detour line shown in Figure 3. The second term in Equation (4) originates from the folding of the line which reverses the sign of the electric field. The following expression for ϕ is given in Reference (3):

$$\tan \frac{\phi}{2} = \tan \frac{\pi L}{\lambda} \left(1 + \frac{s}{L} \right) \sqrt{\frac{1 + \frac{YZ}{2} \cot \frac{\pi L}{\lambda} \left(1 + \frac{s}{L} \right)}{1 - \frac{YZ}{2} \tan \frac{\pi L}{\lambda} \left(1 + \frac{s}{L} \right)}} \quad (5)$$

Y is the susceptance defined by Equation (2) and Z is the characteristic impedance of the (folded) strip-line

$$Z = \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{t}{d} \quad (6)$$

Equation (5) yields an infinite number of passbands in which ϕ has real values. For practical purposes only the first band at the low-frequency side is of interest. The π -mode wavelength, λ_π , is the longer cutoff wavelength where the phaseshift per cell equals $-\pi$ or, according to Equation (4), $\phi = 0$. From Equation (5) one finds, for λ_π ,

$$1 + \frac{YZ}{2} \cot \frac{\pi L}{\lambda_\pi} \left(1 + \frac{s}{L} \right) = 0 \quad (7)$$

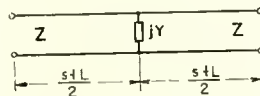


Fig. 3—Detour line. The delay of the lines shown in Figures 1 and 2 is achieved by folding up of such "detour lines."

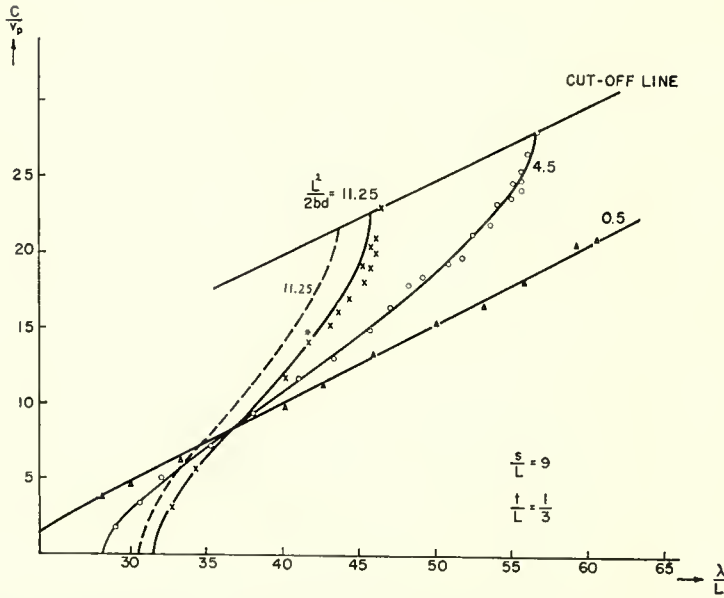


Fig. 4—Measurement of the backwall effect on the dispersion of the interdigital line. The solid curves are computed and the marked points are measured. The dashed curve is computed for the condition $s/L \gg 1$ and corresponds to the theory of Leblond and Mourier.

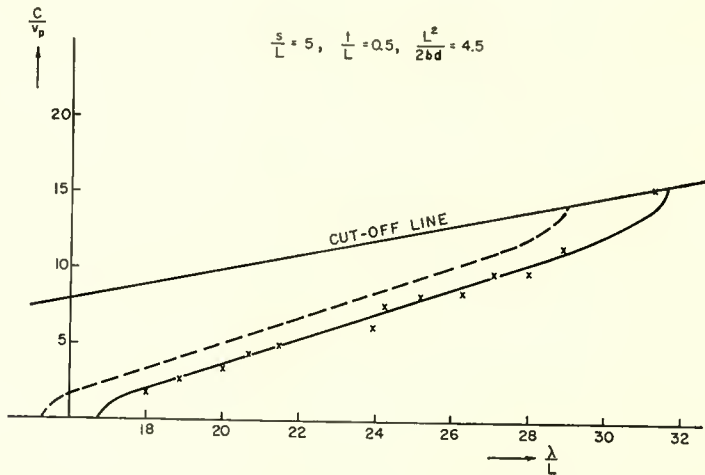


Fig. 5—Measurement of the dispersion of an interdigital line with backwall. The solid curve is computed and the marked points are measured. The dashed curve is computed for the condition $s/L \gg 1$ and corresponds to the theory of Leblond and Mourier.

Here the delay has the value

$$\frac{c}{v_p} = \frac{\lambda_\pi}{2L}. \quad (8)$$

The band is limited on the high-frequency side by $\psi = 0$ or $\phi = \pi$. Equation (5) yields, for this lower cutoff wavelength, λ_0 ,

$$1 - \frac{YZ}{2} \tan \frac{\pi L}{\lambda_0} \left(1 + \frac{s}{L} \right) = 0. \quad (9)$$

Here the delay is zero as in a waveguide at cutoff. For the condition

$$\frac{s}{L} \gg 1, \quad (10)$$

the present equations become those of Leblond and Mourier,¹ who used a method of Fletcher⁴ for a general analysis of cross-bar delay lines. It can easily be verified that in their method the fields between the tips of the digits and the sidewalls are neglected. This is justifiable only as long as Expression (10) holds.

A number of interdigital structures were built and cold-tested. Figure 4 shows experimental results and the theoretical values of the delay computed from Equations (3) and (5) for an L-band line with various backwall-distances, b . The agreement is excellent. Another experiment was performed at a lower s/L value and higher frequency (S-band) and also showed a very good agreement with the theory (Figure 5). For comparison, curves using Leblond's and Mourier's¹ analysis are also given in Figures 4 and 5.

ACKNOWLEDGMENT

The assistance of S. Skillman in the numerical evaluation of Equation (5) is gratefully acknowledged.

⁴R. C. Fletcher, "A Broad-Band Interdigital Circuit for Use in Traveling-Wave Amplifiers," *Proc. I.R.E.*, Vol. 40, p. 951, August, 1952.

METALLOGRAPHIC ASPECTS OF ALLOY JUNCTIONS*

BY

ARNOLD S. ROSE

RCA Semiconductor and Materials Division,
Somerville, N. J.

Summary—Defects in the junction structure of large area alloy junction transistors have been found to be associated with the presence of oxide films on the indium and germanium surfaces. Nonplanar alloying, irregular germanium regrowth, and unwetted areas result. When very low dislocation density germanium crystal is used, these defects are largely minimized. With dislocation densities of the order of 5,000 per square centimeter, alloying following prior thermal deoxidation in a reducing atmosphere has a similar effect in eliminating these types of defects.

PHYSICAL CHARACTERISTICS OF LARGE-AREA JUNCTIONS

EXAMINATION of typical indium-germanium, p-n-p junctions with diameters larger than 0.050 inch reveals several characteristic defects which may be classified into one or more of three general types:

1. Incomplete junctions resulting from unwetted areas
2. Irregularly contoured junction fronts
3. Poorly deposited regrown germanium

Each of these defects must be avoided in order to obtain ideal junctions, such as those of Figure 1, which are essentially planar both at the germanium p-n junction and at the indium-germanium interface, and are also free of unwetted areas.

An unwetted area, as the term implies, occurs where the molten indium has been prevented locally from contacting and alloying with the germanium as shown in cross section in Figure 2. The second type of defect, similar in effect to that of an unwetted area, occurs when the indium is alloyed on the underside of the germanium pellet. As shown also in Figure 2, the recrystallizing germanium is prevented from depositing upon the base germanium in isolated locations and as a result, the p-n junction front is irregular in appearance. The possibility of shorting between the indium dot and the n-type base germanium is evident in both instances, and in addition, it is felt that spurious electrical surface phenomena may be associated with these two types of junction defects.



Fig. 1—Cross section of transistor (75 \times). Both junctions are planar and parallel. The flatness of each junction is that of a crystallographic plane.

The flatness of large-area junctions in transistors is generally conspicuous by its absence. The most prevalent junction fronts have a small planar area at their deepest penetration and irregularly curved boundaries extending to the surface. Such irregularly shaped junctions result from the variability in the wetting of the molten indium^{dot} to the germanium pellet during alloying in a reducing atmosphere furnace. At 155°C, the indium melts and balls up into spherical shape which is maintained by a surface film; the film is not reduced until some higher temperature, depending upon the extent and character of the film, is reached. At this higher temperature, the molten indium wets down with the disappearance of the restraining film to form a typical shallow spherical cap. In the interval between the indium melting point and the film breakdown, solution of germanium proceeds in accordance with the requirements of the constitution diagram and penetration into the pellet at the narrow-diameter contact area between the balled dot and the pellet is taking place. The depth of this penetration is proportional to the temperature at which the film-free dot assumes an equilibrium shape and contact area with the pellet. If this temperature is high (e.g., 480°C), there will be little germanium left to be dissolved on heating to the alloying temperature of 550°C. As a result of this, the junction front has a deep, small-diameter, planar penetration surrounded by an outward-curving receding boundary which is not concentric with the dot if the dot extension has been eccentric. If the film is reduced and the dot wets down at a low temperature (e.g., 350°C), a maximum area of contact is maintained over a longer temperature span and larger increments of germanium will

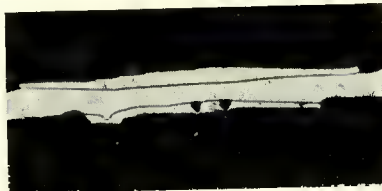


Fig. 2—Cross section of transistor (25 \times). The irregular contour of both junction fronts is evident. The emitter junction (lower) has an unmet area and two voids in the recrystallized germanium.

be dissolved. The result of this is a relatively larger area planar junction as shown in Figure 3.

The difficulty in achieving reproducible spacing between emitter and collector, and the danger of shorting between these junctions when the maximum penetrations are directly opposed gives added emphasis to the desirability of uniformly planar junctions. One further aspect of the variation in mode of penetration of the indium, is that of the appearance of the regrown germanium. When the penetration has been essentially planar, as with a light surface film in Figure 3, the regrowth extending inward from the peripheral {111} planes generally has a smooth appearance. When the penetration has a curving peripheral boundary, consisting of many {111} planes, the regrowth appears as a multi-layered irregularly tiered deposit; this causes adverse effects

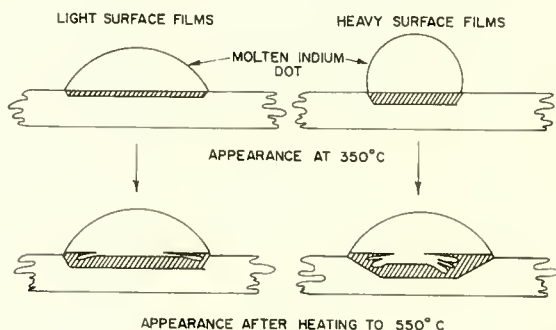


Fig. 3—Effect of surface films on junction front. The junction front of the dot whose surface film is light enough to permit the dot to wet down at 350° C will be planar over a larger area than that of a dot which is maintained as a sphere by a heavy surface film at this temperature. The regrown germanium in the latter possesses an irregular surface after heating to 550° C.

on those transistor characteristics affected by leakage current, such as I_{co} and V_{co} .

The spreading of indium dots can be minimized and a flat alloying penetration obtained by the use of a graphite restraining washer to contain the indium as it melts. If a cylindrical indium dot is used in conjunction with a weighted pressure washer, the molten dot will be maintained in the same cylindrical form as when solid. With this type of jigging, an indium reservoir of constant volume and uniform height is available for alloying to the germanium throughout the alloying temperature range. It is essential that the weight of the pressure washer be so selected that the liquid indium will neither be squeezed out under, nor allowed to lose contact with the restraining washer. This method yields alloy-junction diameters identical with the diameter of the restraining washer and dot.

RELATION BETWEEN DISLOCATION DENSITY AND UNWETTED AREAS

When dislocation-free pellets are alloyed as described above, it is found that the junction front is extremely flat, as shown in Figure 4, and that there are no unwetted areas. The flatness achieved here over a large junction area (0.100 inch diameter) is that of a crystallographic plane, and is what would be expected to occur when successive {111} planes are "peeled-off" by the inwardly advancing molten alloy pool. The flatness is unaffected by the increased penetration associated with edge dislocation sites. Because there are no unwetted areas, completely planar junctions result.

However, since dislocation-free germanium is not normally used for transistor production, the junction fronts obtained on higher dislocation density crystal are of interest. The characteristic appearance of the unwetted areas found on such crystal is shown in Figure 5. After the indium has been removed, the surface of the unwetted area generally appears as a small circular area surrounded by portions of



Fig. 4—Cross section of an alloyed junction (15 \times). The p-n junction is crystallographically flat as defined by a {111} plane.

hexagonal or triangular regrowth which are elevated with respect to the circle. When sectioned, this regrowth is similar in nature to the growth found at the edge of the junction. In the instance of the unwetted spot, the {211} plates grow from the {111} slopes of the unwetted germanium rather than inward from the limiting {111} planes of the junction periphery. As the germanium recrystallizes at the unwetted spot, the growth may be so rapid in some instances as to cover completely and conceal the base germanium.

- MICRORADIOGRAPHIC EXAMINATION OF UNWETTED AREAS

It was felt that some understanding as to the origin of the unwetted areas was a prerequisite to an effort to eliminate them. Two likely causes for incomplete wetting are: (1) debris which would mechanically separate the germanium from the indium, or (2) trapped gas which would behave in a similar manner. Because deliberate inclusions of debris such as graphite, dirt, or filings were found to have no

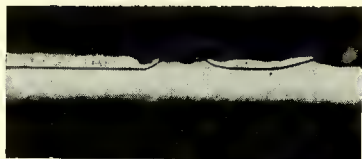


Fig. 5—Cross section of an alloyed junction (25 \times). The p-n junction is interrupted by an unwetted area.

specific correlation with the occurrence of unwetted spots, attention was focused on the possible causes and effects of trapped gas. An indication that gas bubbles might prevent wetting was found in the generally circular nature of the typical unwetted area. It was thereupon decided to utilize radiography in this investigation.

The X-ray technique involves placing the alloyed pellet and dot above an extremely fine-grain emulsion photographic plate, and exposing the plate to X-rays for a sufficient length of time to disclose any inhomogeneity in the density of the indium. Any porosity present in the dot will be visible as a dark area on a light field. Because the units are so small, examination of the film is made under a microscope using transmitted light so that disclosure of minute porosity is possible. Dots as small as 0.010 inch in diameter have been X-rayed and voids as small as 0.001 inch in diameter were detected. Further, close correlation was noted as in Figure 6 between the size and location of the porosity disclosed by X-rays and the corresponding unwetted areas at the germanium-indium interface.



Radiograph (12 \times)



View of In-Ge interface (12 \times)

Fig. 6—Comparison between radiograph and appearance of unwetted areas. The dark spots in the radiograph correspond to porosity in the indium dot following alloying. After removal of the indium, as viewed at right, the location of the unwetted areas corresponds to that of the porosity.

With the establishment experimentally of this correlation between unwetted areas and voids in the indium dot, and the availability of microradiography as a tool for observation of the latter, a series of tests was planned to further this investigation. The results of these tests are summarized in Table I.

Table I—Alloying Tests
0.100-Inch Diameter Indium Dots

Test Number	Germanium Dislocation Density*	Alloying Temperature °C	Degree of Porosity and Unwetted Areas
1	Medium	200	None
2	Medium	300	Many
3	Medium	400	Many
4	Medium	500	Many
5	Medium	550	Few
6	High	400	Many
7	High	500	Many
8	High	550	Many
9	Zero	400	Many
10	Zero	500	None
11	Zero	550	None
12	Zero	{400 Topside {550 Upside Down	Many
13	High	550	Many
14	High	{400 Topside {550 Upside Down	Many

* Medium = 5,000 to 8,000 per square centimeter
High = 30,000 to 40,000 per square centimeter
Zero = 0 to 10 per square centimeter

The following conclusions may be drawn from the data:

1. Porosity at the germanium-indium interface causes unwetted areas.
2. This porosity is caused by gas generated and trapped at the germanium-indium interface between 200° and 300°C. (Refer to Tests 1 and 2)
3. The number of gas bubbles generated is independent of the germanium dislocation density. (Refer to Tests 3, 6 and 9)
4. The gas bubbles are liberated from the interfacial surface and escape from the molten dot with increasing temperature. (Refer to Tests 1 to 5, 6 to 8, and 9 to 11)
5. The amount of gas which is liberated from the germanium-indium interface varies with the dislocation density. (Refer to Tests 5, 8 and 11)

Tests 12 and 14 were designed to verify these observed aspects of the formation and subsequent liberation of gas bubbles as a function of dislocation density and temperature. In these tests, the alloying process was a two-step procedure; first, the pellets were heated to 400°C with the dots in the normal topside position, then they were

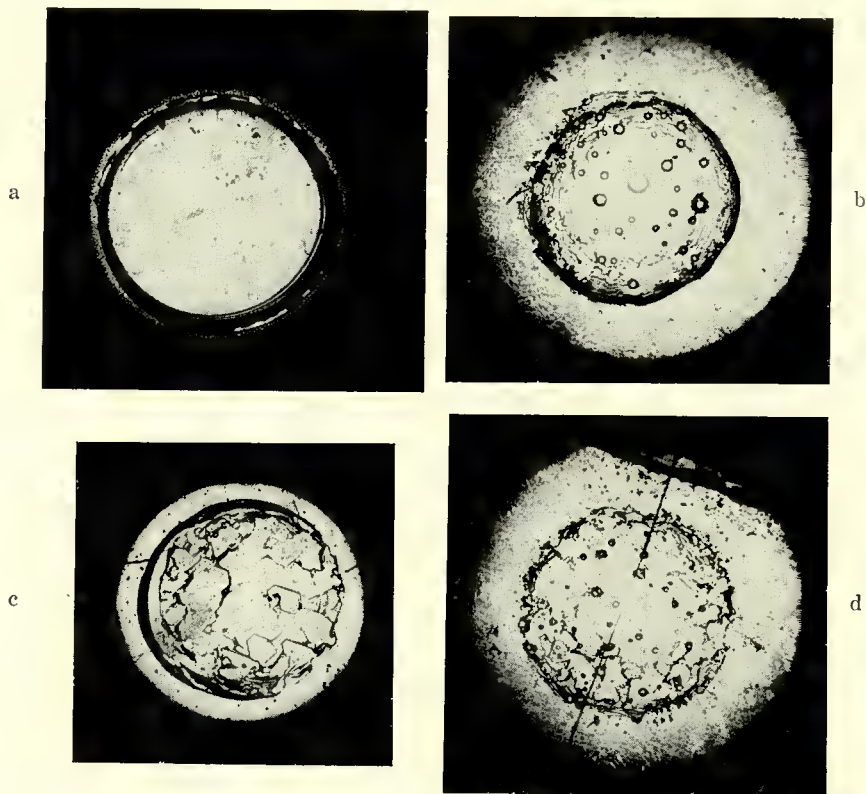


Fig. 7—Views of In-Ge interfaces (8 \times). (a) Completely alloyed interface of zero dislocation density germanium after normal topside alloying; (b) Following underside alloying, many unregrown areas caused by gas entrapment are visible; (c) Unwetted areas are noted on alloyed interface of high dislocation density germanium following normal topside alloying; (d) Following underside alloying both unregrown and unwetted areas are visible.

turned over and alloyed to 550°C with the dots on the underside of the pellets.

The results, as shown in Figure 7, reveal no essential difference between the incidence at 400°C of unwetted areas on the zero and the high dislocation density germanium. Because the pellets were fired upside down to 550°C, the gas initially present from the 400° alloying

was trapped and was unable to escape. Close inspection of the two samples at high magnification shows that the nature of the unwetted areas differs. The bubbles in the dot of the dislocation-free sample correspond to depressed spherical cavities in the germanium regrowth. Probing of the bottom of these pits reveals that the n-type base is exposed. By contrast, the regrowth on the high-dislocation-density crystal shows both this type of porosity and the typical unwetting noted previously.

The mechanism by which the depressed cavities occur in the recrystallized germanium when alloying the indium dot on the underside of the pellet involves trapping of the gas at the interface as before. On the zero dislocation density crystal, with equilibrium heating conditions, the indium preferentially dissolves germanium by lateral attack in the [110] direction rather than inwardly on the more difficult [111] axis. This action causes an undercutting type of solution above the adhered gas bubble allowing the bubble to rise. As the alloying proceeds further in the [111] direction, dissolving germanium in planar fashion, the trapped bubble continues to rise, alternately contacting and being undercut by the advancing alloy front. When cooling starts, the trapped bubble in contact with the n-type base physically prevents the recrystallizing p-type germanium from growing at the bubble location.

It is apparent that during alloying on the top side of dislocation-free germanium, the undercutting mechanism causes gas liberation between 400° and 550°C and furnishes an explanation for the uniform wetting obtained at the higher alloying temperature.

With high dislocation density crystals, satisfaction of the germanium-indium phase-equilibrium requirements is obtained more readily by inward attack at the many dislocation sites in the [111] direction. Therefore, considerably less undercutting on the {111} planes takes place. As noted, only a few bubbles are released between 400° and 550°C in normal (topside) alloying and many unwet areas persist. With upside-down alloying, only few bubbles will rise with the alloy front to cause unregrown spots. In addition there are many typical unwetted areas observed on the sample.

ORIGIN OF THE GASEOUS POROSITY

As a result of these observations, the following factors were evaluated as possible sources for the bubble formation:

1. Dissolved gas in the germanium and/or indium
2. Mechanically trapped gas from the furnace atmosphere
3. Surface films on the germanium and/or indium

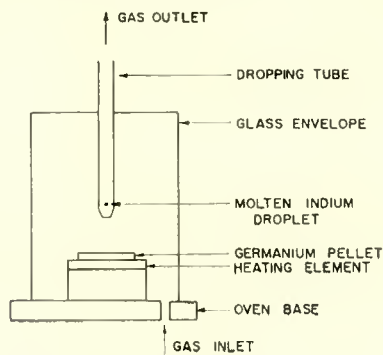


Fig. 8—Dropping tube apparatus for clean dot application.

Following several vacuum annealing and alloying experiments, it became apparent that the first two sources were not important and could safely be dismissed as major contributors to the porosity at the germanium indium interface. The effects of the surface films in this respect were demonstrated by the contrast in porosity obtained by alloying film-free versus oxidized materials.

When an indium dot which has been exposed to air for several hours is alloyed to a germanium surface in a reducing atmosphere, the dot melts at 155°C but, as described previously, is restrained from wetting-down by a sac-like oxide film. On further heating, as the hydrogen becomes effective in reducing this film, the dot assumes a spherical cap shape and traps the gaseous products of this reaction. The porosity visible radiographically and as unwetted areas results.



Fig. 9—View of In-Ge interface ($15\times$). Following alloying procedure in which both In and Ge are preheated there is no evidence of any unwetted areas.

If film-free alloying takes place, no such porosity is evident. This was established experimentally by the use of an apparatus having a modified glass bell, as shown in Figure 8. The pellet was preheated in hydrogen to 600°C, held for five minutes on the graphite hearth and then allowed to cool to 300°C. The indium dot was dropped into the glass tube and held over the hot germanium by the restriction at the end of the dropping tube. The reducing atmosphere rising through this tube deoxidized the surface of the molten drop, which was held suspended (like a flowmeter float) in the gas stream. When the droplet surface was uniformly bright and shiny, indicating complete deoxidation, it was allowed to drop onto the germanium by stopping the gas flow through the tube. Alloying to 550°C followed, and, after cooling, the indium dot was removed. As shown in Figure 9, no unwetted areas were present and the flat contour of the junction front similar to that illustrated in Figure 4 confirmed the observation that the surface films on both germanium and indium have deleterious effects on junction interface geometry.

DESIGN CONSIDERATIONS FOR DIRECT-COUPLED TRANSISTOR AMPLIFIERS*

BY

J. E. LINDSAY[†] AND H. J. WOLL[‡]

Summary—An outstanding source of difficulty in d-c amplifier design is drift. With drift, as with noise, the use of degenerative feedback does not improve the ratio of the signal to the undesired disturbance. By analogy to an approach used with the noise problem, the procedure to be described is to seek a relationship between the driving source and the amplifier which will yield optimum performance with respect to signal-to-drift ratio.

It is shown that

1. The maximum obtainable signal-to-drift ratio of an amplifier is achieved by choosing the optimum source resistance.
2. The signal-to-drift capability of a transistor and the optimum source resistance are nearly the same for the three configurations: common emitter, common collector and common base.
3. Negative feedback can change the optimum source resistance and generally degrades the signal-to-drift ratio.
4. Improved signal-to-drift ratio can be obtained at reduced emitter currents.

INTRODUCTION

PROBABLY the outstanding source of difficulty in d-c amplifier design is drift. With transistors the changes caused by temperature variation are particularly troublesome in this respect.¹ On this account much effort has been expended to realize transistor circuits which achieve a stable operating point over a wide range of temperature. In many cases the use of d-c feedback accomplishes this result very successfully and may appear to offer a remedy for drift in the d-c amplifier. However, while degenerative feedback does indeed reduce drift, the result is deceptive. The desired signal is also reduced.

This aspect of the problem of drift is much like that of noise in a-c amplifiers. In both cases the signal and the undesired disturbance are inseparable and feedback does not distinguish the two. By analogy

* Manuscript received April 1, 1958.

[†] Formerly, RCA Defense Electronic Products, Camden, N. J.; now with Cornell Aeronautical Laboratory Inc., Buffalo, N. Y.

[‡] RCA Defense Electronic Products, Camden, N. J.

¹ H. C. Lin and A. A. Barco, "Temperature Effects in Circuits Using Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, p. 369.

to the approach used with the noise problem²⁻⁴ the procedure described here is to seek a relationship between the driving source and the amplifier which will lead to optimum performance with respect to signal-to-drift ratio.

That such an approach might prove rewarding can be inferred by considering the chief sources of drift in the transistor. These are the changes in saturation current and d-c input conductance due to temperature variation. If the change of saturation current, I_{c0} , were the only source of drift it would be desirable to drive a d-c transistor amplifier from a constant-voltage source. Increasing saturation current would then have no effect at the input and drift would be minimized. If, on the other hand, only changes in the input conductance were associated with drift, it would be desirable to drive the amplifier from a constant-current source.

These ideas are now set forth in detail. Proceeding from fundamental definitions, the concepts of optimum source resistance and obtainable signal-to-drift ratio for a direct-coupled amplifier are developed. These principles are then applied to certain aspects of the problem of drift of transistors operated over an extended temperature range.

AVAILABLE SIGNAL POWER⁵

The signal source of Figure 1 has an internal resistance R_s and supplies an electromotive force of e_s volts. It delivers $e_s^2 R_l / (R_l + R_s)^2$ watts to a load resistance of R_l ohms. This power is maximum and equal to $e_s^2 / 4R_s$ when the load resistance is matched to the source resistance; i.e., when $R_l = R_s$. The quantity $e_s^2 / 4R_s$ is called the available signal power of the source, since it represents the maximum power the source is capable of delivering.

This available signal power is to be regarded as a defined quantity which provides another useful means of specifying the source. As such, the available signal power is characteristic of the source alone independent of the external circuit. This is the case even though available signal power is defined by reference to a matched load.

² H. T. Friis, "Noise Figures of Radio Receivers," *Proc. I.R.E.*, Vol. 32, p. 419, July, 1944.

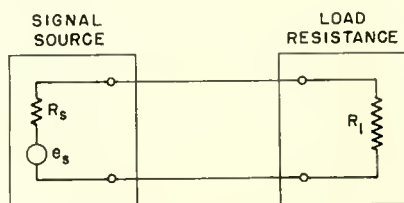
³ H. Goldberg, "Some Notes on Noise Figures," *Proc. I.R.E.*, Vol. 36, p. 1205, October, 1948.

⁴ E. W. Herold, "An Analysis of the Signal-to-Noise Ratio of Ultra High Frequency Receivers," *RCA Review*, Vol. VI, p. 302, January, 1942.

⁵ "Standards on Transducers; Definitions of Terms," *Proc. I.R.E.*, Vol. 39, p. 897, August, 1951.

GAIN^{5,6}

The amplifier of Figure 2 delivers output signal to a load resistance, R_l . A signal generator of source resistance R_s drives the amplifier. The power gain of the amplifiers is the ratio of the signal power delivered to the load to the signal power supplied to the input terminals; e.g., $G_p = P_{out}/P_{in}$. As such, power gain is a property of the amplifier and the load alone, in that it can be specified apart from the source. However, to determine either the power output of the amplifier or the power input to the amplifier, it is necessary to specify



$$\text{Output power} = \frac{e_s^2 R_l}{(R_l + R_s)^2}$$

$$\text{Available power} = \frac{e_s^2}{4 R_s}$$

Fig. 1—Available signal power.

in addition both the source parameters, e_s and R_s , and the input resistance of the amplifier, R_i . The formula for the power input to the amplifier is

$$P_{in} = \frac{e_s^2 R_i}{(R_s + R_i)^2}$$

The expression for the power output, then, is

$$P_{out} = G_p P_{in} = G_p \left[\frac{e_s^2 R_l}{(R_s + R_i)^2} \right]$$

The gain of the circuit of Figure 2, called transducer gain, is defined as the ratio of the power delivered to the load, e_o^2/R_l , to the power available from the signal source, $e_s^2/4R_s$; that is, $G_T = P_{out}/P_{av}$. It is significant that this available power is not, generally, the power

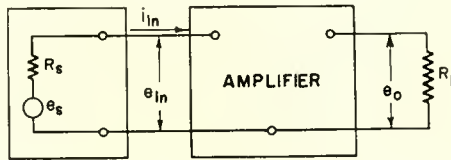
⁶ R. L. Wallace and W. J. Pietenpol, "Some Circuit Properties and Applications of n-p-n Transistors," *Proc. I.R.E.*, Vol. 39, p. 753, July, 1951.

being delivered to the input terminals of the amplifier. Instead, this definition of gain can be regarded as the ratio of the power actually obtained using the amplifier to the power which could be obtained from the source alone. Such a ratio is meaningful to the designer as a measure of what the use of the amplifier accomplishes.

Transducer gain, G_T , is a property of the amplifier and the source and as such must be specified for a given source. The relationships involved are

$$G_p = \frac{P_{out}}{P_{in}} \qquad P_{in} = \frac{e_s^2 R_i}{(R_s + R_i)^2},$$

$$G_T = \frac{P_{out}}{P_{av}} \qquad P_{av} = \frac{e_s^2}{4R_s},$$



$$G_p = \frac{P_{out}}{P_{in}} = \frac{e_o^2 / R_i}{e_{in}^2 / i_{in}}$$

$$G_T = \frac{P_{out}}{P_{av}} = \frac{e_o^2 / R_i}{e_s^2 / 4R_s}$$

Fig. 2—Amplifier gain.

from which

$$G_T = G_p \frac{P_{in}}{P_{av}} = G_p \left[\frac{4R_s R_i}{(R_s + R_i)^2} \right].$$

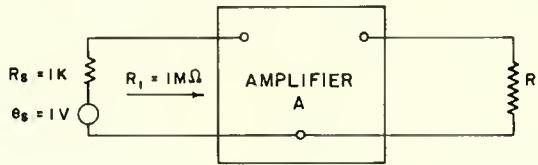
Thus, if transducer gain is specified, only the available source power is needed to determine the power output. This permits a comparison of amplifiers driven by the same source.

The utility of the definition of transducer gain is illustrated by the following example. In Figure 3, amplifier A has an input resistance of one megohm. The power gain (not the transducer gain) of this amplifier is 10^5 or 50 decibels. Amplifier A is driven by a source with internal resistance of 1,000 ohms. If the open-circuit source voltage is one volt, the power delivered to the input terminals of the amplifier is calculated to be

$$P_{in} = \left[\frac{e_s}{R_l + R_s} \right]^2 R_l = \frac{1,000 \times 10^3}{(1,001 \times 10^3)^2} \approx 10^{-6} \text{ watt.}$$

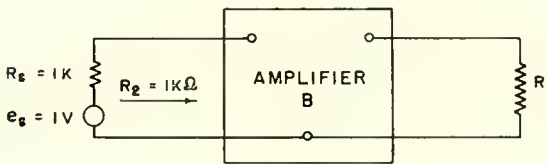
The power output is then 1/10 watt.

Amplifier B, on the other hand, has an input resistance of 1,000 ohms. The power gain, G_p , of this amplifier is 10^4 or 40 decibels. If



$$G_P = \frac{P_{out}}{P_{in}} = 10^5$$

$$G_T = \frac{P_{out}}{P_{av}} = 400$$



$$G_P = \frac{P_{out}}{P_{in}} = 10^4$$

$$G_T = \frac{P_{out}}{P_{av}} = 10^4$$

Fig. 3—Comparison of amplifiers.

amplifier B is driven by the same source as amplifier A, the power delivered to the input terminals is

$$P_{in} = \left[\frac{e_s}{R_2 + R_s} \right]^2 R_2 = \frac{1,000}{(2,000)^2} = 2.5 \times 10^{-4} \text{ watt.}$$

The resulting power output is 2.5 watts. Thus, for the specified source the signal power delivered by amplifier B exceeds that of A even though its power gain is lower.

Specifying available source power as input power takes into account the relationship between the amplifier and the driving source. Defined in this way the transducer gain of amplifier A driven by the given

source is 400, or 26 decibels; the transducer gain of amplifier B driven by the given source is 10^4 , or 40 decibels. Thus the definition of transducer gain provides a basis for comparison of amplifiers driven by the same source.

The transducer gain of the amplifier of Figure 2 depends upon the load resistance and upon the source resistance. If power gain is the sole consideration, it is desirable to match the source resistance to the input of the amplifier. This is the condition required to utilize fully the available power of which the source is capable. For this matched condition the power input, defined as available power from the source, is actually driving the amplifier, and power gain and transducer gain are the same.

Likewise, for constant input the amplifier delivers maximum power to a matched load resistance. The ratio of this power available to the load to power available from the generator is called available gain. If both the source and the load are matched to the amplifier the maximum available gain is obtained. Under these conditions of matched input and matched output, power gain, transducer gain, available gain and maximum available gain are identical.

A given signal source of e_s volts with an internal resistance R_s can be matched to the input of an amplifier by means of an ideal transformer of turns ratio n . The combination is equivalent to a signal source of ne_s volts with an internal resistance n^2R_s . The available power of the source is unchanged by the transformer, and is delivered to the amplifier by setting the equivalent resistance n^2R_s equal to the input resistance of the amplifier. This is also the condition for obtaining at the input terminals of the amplifier the maximum current and the maximum voltage from the given source. Thus, whether drive for the amplifier is considered to be current, voltage, or power, the condition for optimum drive from the given source is the same.

DRIFT

For a direct-coupled amplifier, both gain and the problem of drift are important considerations. Drift appears as an error which is not distinguishable from the desired signal in the output of the direct-coupled amplifier. The drift magnitude alone does not adequately specify the significance of drift in an amplifier. A given drift magnitude in a high-gain amplifier may represent less relative error than a much lower value where the gain is low. The signal-to-drift ratio is used to express the relative error represented by drift.

In order to specify the significance of drift as a relative error it is convenient to refer the drift to the input of the amplifier. For the

amplifier and signal source of Figure 2 this is accomplished by determining the input power (available power from the source) which is sufficient to correct the drift.* This gives the error due to drift in terms of equivalent input power and independently of the amplifier gain.

Now by definition $S = G_T S_{av}$ where S is the signal in the output, S_{av} is input signal power from the source and G_T is the transducer gain of the amplifier and the signal source.

Likewise, $D = G_T D_{av}$ where D is the magnitude of drift in the output and D_{av} is the equivalent input power sufficient to correct for drift. It is thus possible to express the ratio of signal to drift in terms of values at input of the amplifier, independent of the gain since

$$\frac{S}{D} = \frac{S_{av}}{D_{av}}$$

This value of the equivalent input power sufficient to correct for drift is an important performance characteristic of a direct-coupled amplifier with its driving source. This value alone is a criterion of the performance obtained. It can be used as a basis for comparison among amplifiers. It follows, therefore, that a prime objective of direct-coupled amplifier design is to minimize the available power from the source needed to correct for drift.

OPTIMUM SOURCE RESISTANCE AND OBTAINABLE SIGNAL-TO-DRIFT RATIO

The problem of drift in direct-coupled amplifiers has been attacked in a number of ways. The methods range from chopper techniques used with complex high-gain amplifiers⁷ to carefully designed compensation schemes for transistors operated over an extended temperature range.^{8,9} In each case it may be possible to realize improved performance up to the maximum obtainable signal-to-drift ratio for a given amplifier. It is possible to define this capability in terms of measurable values and to prescribe precisely the conditions for achieving the limit. The relationships involved are developed below.

* The device of referring drift to the input by determining the power to correct or eliminate drift is convenient for measurement. The measurement can always be made, even though the drift is so great as to take the amplifier out of the region of linear operation.

⁷ F. H. Blecher, "Transistor Circuits for Analog and Digital Systems," *Bell Sys. Tech. Jour.*, Vol. 35, p. 661, March, 1956.

⁸ E. Keonjian, "Temperature-Compensated DC Transistor Amplifier," *Proc. I.R.E.*, Vol. 42, p. 661, April, 1954.

⁹ J. W. Stanton, "A Transistorized D-C Amplifier," *Trans. I.R.E. PGCT*, p. 65, March, 1956.

Matching, which has been discussed in connection with available power and maximum available gain, is the design procedure for achieving maximum power transfer. The idea of matching for best noise figure is also a well-established concept.⁴ It turns out that the same sort of approach yields a useful result for the signal-to-drift ratio of a direct-coupled amplifier. For a given amplifier there is an optimum source resistance for which the equivalent input power sufficient to correct for drift is a minimum. With this optimum source resistance the obtainable signal-to-drift ratio for the amplifier is achieved.

An expression for the optimum source resistance can be developed from Figure 4. Figure 4a represents a driftless condition of the ampli-

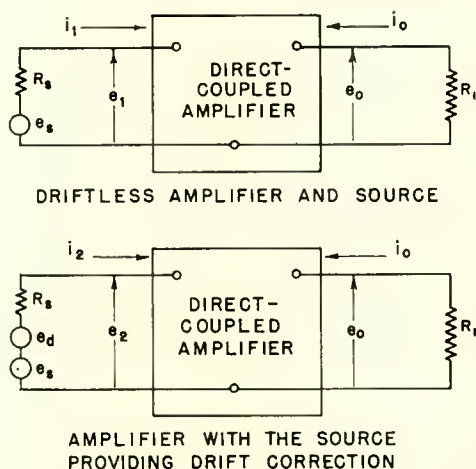


Fig. 4—Drift referred to the input of an amplifier.

fier and source. Figure 4b represents a condition of the amplifier wherein the source contributes input drive to correct the drift.

From Figure 4,

$$e_s = i_1 R_s + e_1,$$

$$e_d + e_s = i_2 R_s + e_2,$$

where

R_s = source resistance,

e_s = open-circuit source signal voltage,

e_d = drift correction voltage,

i_1, e_1 = input current and voltage for the driftless condition of the amplifier,

i_2, e_2 = input current and voltage with drift in the output corrected.

Then

$$e_d = (i_2 - i_1) R_s + (e_2 - e_1).$$

Now it is desired to select a value of source resistance, R_{so} , which yields a minimum for the available power, $D_{av} = (e_d)^2/4R_s$, sufficient to correct for drift. It follows that

$$D_{av} = \frac{(i_2 - i_1)^2 R_s^2 + 2(i_2 - i_1) R_s (e_2 - e_1) + (e_2 - e_1)^2}{4R_s},$$

and that the minimum is obtained for*

$$\frac{\partial D_{av}}{\partial R_s} = 0 = \frac{(i_2 - i_1)^2}{4} - \frac{(e_2 - e_1)^2}{4R_s^2}.$$

This gives, for the optimum source resistance, R_{so} , simply

$$R_{so} = \frac{e_2 - e_1}{i_2 - i_1},$$

where $(e_2 - e_1)$ and $(i_2 - i_1)$, respectively, represent that change of voltage and current at the input terminals of the amplifier which corrects for drift. The corresponding value of the equivalent input power to correct for drift is

$$D_{av} = (e_2 - e_1) (i_2 - i_1).$$

These simple relations specify both the obtainable signal-to-drift ratio of a given amplifier and the condition for achieving the result.

* It may be objected that the e_1, i_1 and e_2, i_2 of Figure 4 are functions of the source resistance, R_s , and that therefore the operation $\partial D_{av}/\partial R_s$ is not valid as shown. That e_1, i_1, e_2 and i_2 vary with source resistance is indeed the case. However, it should be noted that the quantities $(e_2 - e_1)$ and $(i_2 - i_1)$ represent the drift referred to the input, and for a linear amplifier are independent of the signal level.

From another point of view, the question of signal level can be eliminated altogether. This can be done by providing for the amplifier a bias source adjusted to maintain e_1 and i_1 constant for various values of source resistance, but without contributing to the amplifier drift. This latter procedure was adopted for the measurements to be described.

CONSIDERATIONS IN THE DESIGN OF DIRECT-COUPLED TRANSISTOR AMPLIFIERS

General Design Considerations

These considerations have been applied to the problem of the drift of transistor direct-coupled amplifiers due to temperature change. The measurements are described in the next section. The results are summarized below and then treated in detail individually. These are

1. The optimum source resistance and the obtainable signal-to-drift ratio as measured by the power needed to correct for drift are closely the same for the three configurations: common emitter, common base, and common collector.

2. Feedback does not improve the signal-to-drift ratio but generally increases somewhat the available power required to correct the drift. Feedback may increase or decrease the optimum source resistance.

3. Silicon transistors of otherwise comparable performance capabilities, as contrasted to germanium, can be expected to provide 10 decibels or more improvement of obtainable signal-to-drift ratio for the same temperature range at elevated temperatures. The optimum source resistance for silicon transistors is an order of magnitude greater than for germanium transistors operated comparably.

4. The obtainable signal-to-drift ratio can be improved by operating at lower emitter currents. The results realized at $I_e = 0.1$ milliampere as contrasted to those at $I_e = 1.0$ milliampere represent 3 to 5 decibel improvement.

5. Compensation is useful in improving the signal-to-drift ratio. If applied at the output, the optimum source resistance is unchanged. By using compensation at the input it is possible to alter the optimum source so as to improve the match for obtainable signal-to-drift ratio.

The Three Configurations

The optimum source resistance and the obtainable signal-to-drift ratio are closely the same for the three configurations: common emitter, common base, and common collector. Experimental results obtained for a 2N79 and a 2N139 transistor are shown in Figure 5. In each case values for the available power required to correct for drift are plotted against values of the source resistance used. In each case optimum values for the three configurations are nearly the same.

That this result is the valid one is demonstrated by Figure 6. Comparison of the common-base and common-emitter configurations shows that the same output terminal is involved and that the input terminals are simply reversed. If the output currents are held constant and

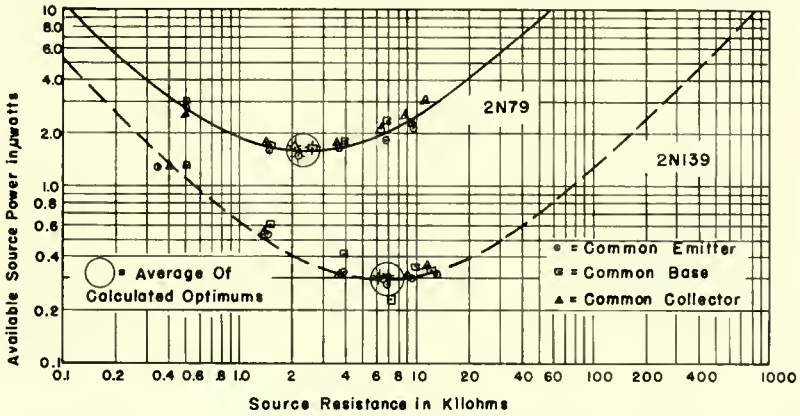


Fig. 5—Available power required to correct for drift due to temperature change plotted as a function of source resistance for the three configurations; common emitter, common base, common collector. The temperature range is 30°C to 50°C .

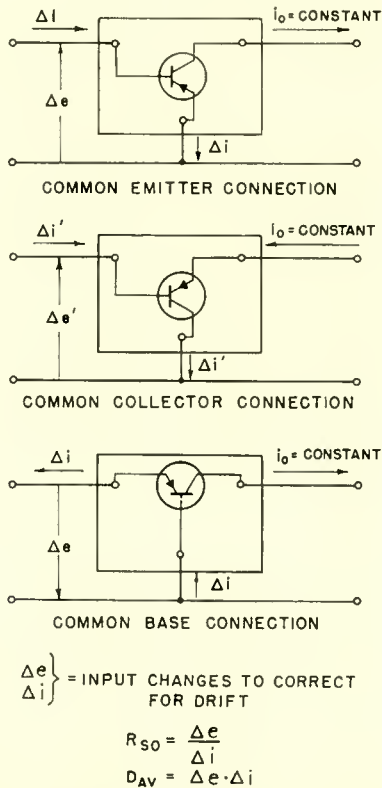


Fig. 6—The optimum source resistance and obtainable signal-to-drift ratio are closely the same for the three connections.

identical for the two cases and if the device remains unchanged for measurements at a given temperature, then the changes at the input terminals must be the same. Slightly smaller current and voltage changes are required to hold emitter current constant, so the obtainable signal-to-drift ratio is slightly greater for the common-collector connection.

Feedback

Negative feedback does not improve the signal-to-drift ratio for uncompensated transistors but rather increases somewhat the available power required to correct for drift. Results obtained using emitter feedback are shown with those for the common-emitter connection in Figure 7 for a 2N79 transistor. Both the optimum source resistance and the power to correct for drift have increased.

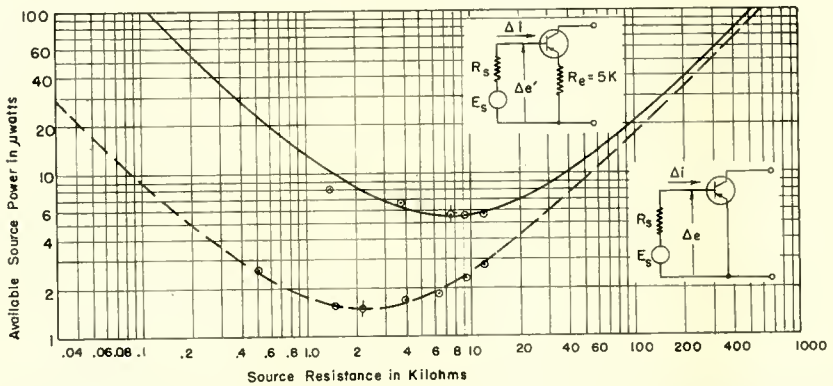


Fig. 7—Available power required to correct for drift due to temperature change plotted as a function of source resistance for the common-emitter connection with and without emitter feedback. The temperature range is 30°C to 50°C . The transistor is a 2N79.

Results obtained using collector feedback are shown with those for the common-emitter connection in Figure 8 for a 2N139 transistor. In this case the power to correct for drift has increased but the optimum source resistance has decreased.

It can be seen from Figure 9 that these are results to be expected. In the case of emitter feedback, the correction current flowing in the feedback resistor results in a voltage drop which appears in the input loop. This increases the input power required and the optimum source resistance. In a like manner, for collector feedback the input correction voltage results in a current flow in the feedback resistor which drains additional current and power from the source. In this case the optimum source resistance decreases.

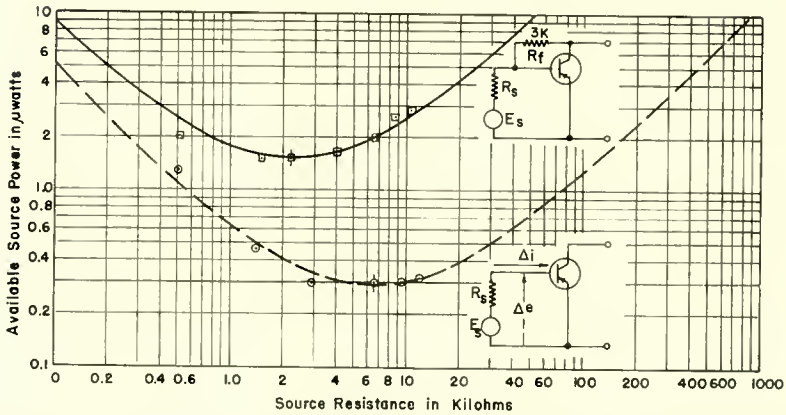


Fig. 8—Available power required to correct for drift due to temperature change plotted as a function of source resistance for the common-emitter connection with and without collector feedback. The temperature range is 30°C to 50°C . The transistor is a 2N139.

This increase of power to correct for drift is a general result of feedback. Figure 10 is a basic circuit of four fundamental methods of feedback applied simultaneously. In Figure 10 the resistor R_1 provides current feedback proportional to the output voltage; R_2 provides current feedback proportional to the output current; R_3 provides voltage feedback proportional to the output current; and R_4 provides voltage feed-

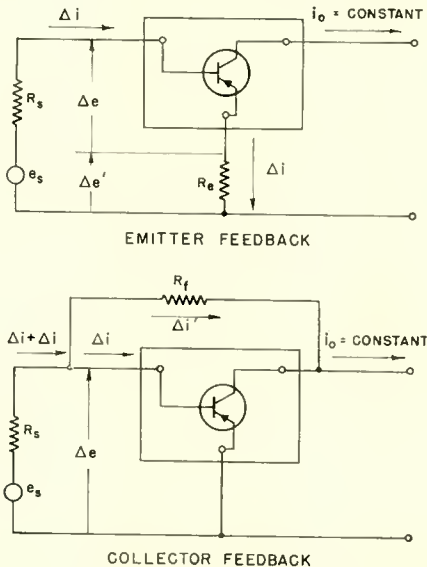


Fig. 9—Feedback increases the power to correct for drift reducing the signal-to-drift ratio.

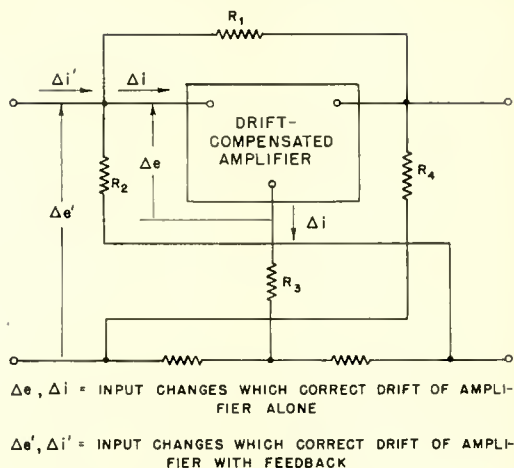


Fig. 10—Basic circuit of four fundamental methods of feedback applied simultaneously.

back proportional to the output voltage. In each case the added elements between the input terminals have the effect of increasing the input power required to correct for drift.

Silicon and Germanium with Extended Temperature Ranges

Results obtained with a germanium transistor (2N109) and with a silicon transistor (904-A) for different temperature ranges are plotted in Figures 11 and 12, respectively. The power to correct for drift increases rapidly as the temperature range is extended. The

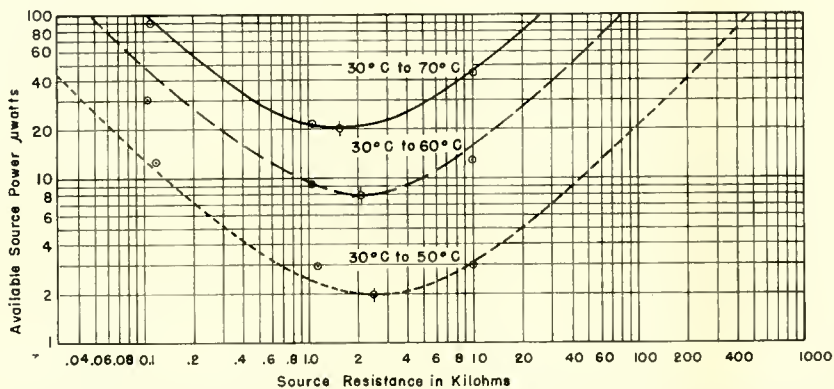


Fig. 11—Available power required to correct for drift due to temperature change plotted as a function of source resistance for a germanium transistor (2N109). The temperature ranges are 30°C to 50°C, 30°C to 60°C, 30°C to 70°C.

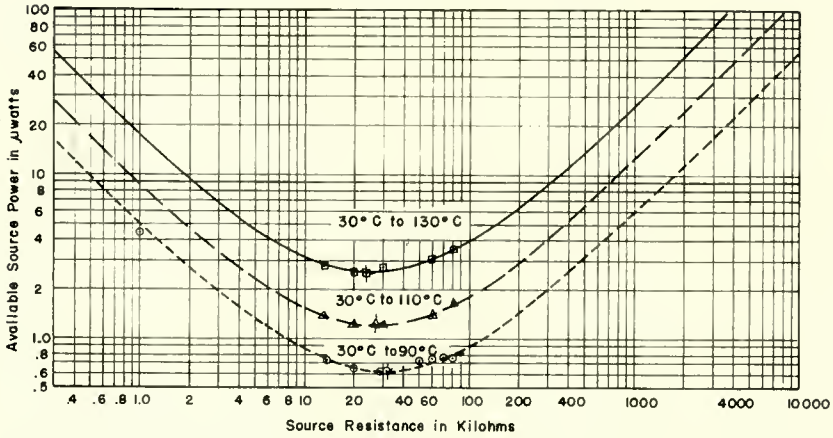


Fig. 12—Available power required to correct for drift due to temperature change plotted as a function of source resistance for a silicon transistor (904-A). The temperature ranges are 30°C to 90°C, 30°C to 110°C, 30°C to 130°C.

optimum source resistance decreases slightly. The results for the silicon transistor represent considerable improvement over that obtainable with germanium.

In Figure 13, also, the uncorrected collector current is plotted against the source resistance used. The magnitude of drift current decreases with increasing source resistance as previously reported.¹ The drift current magnitude is not a minimum for the optimum source

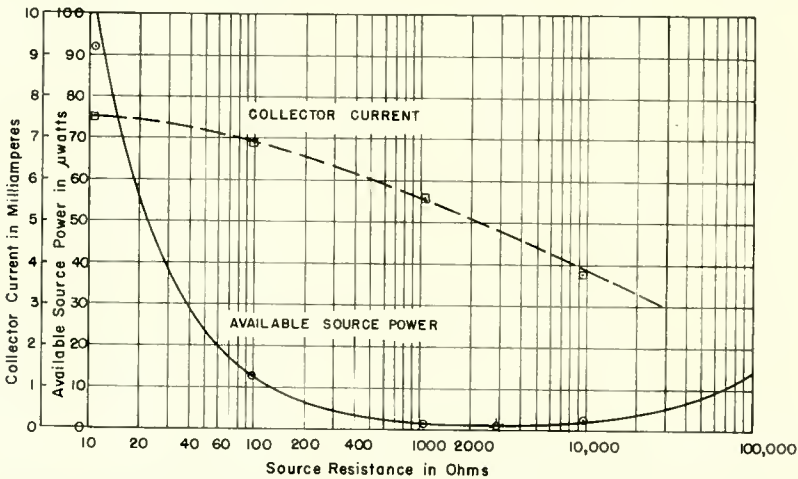


Fig. 13—Available power required to correct for drift due to temperature change and collector current plotted as a function of source resistance. The transistor is a 2N109. The temperature range is 30°C to 50°C.

resistance in this case. It is the available power needed to correct for drift that reaches a minimum for the optimum source resistance and this achieves the maximum obtainable signal-to-drift ratio. This is not a contradiction. Even though the drift magnitude is minimized by using a very large source resistance, at the same time the mismatch between the source and the amplifier becomes so great that the available power required to correct for drift is greatly increased. The drift represents the least error in terms of available power from the source when the optimum source resistance is used.

Operating Point

The obtainable signal-to-drift ratio can be improved by operating

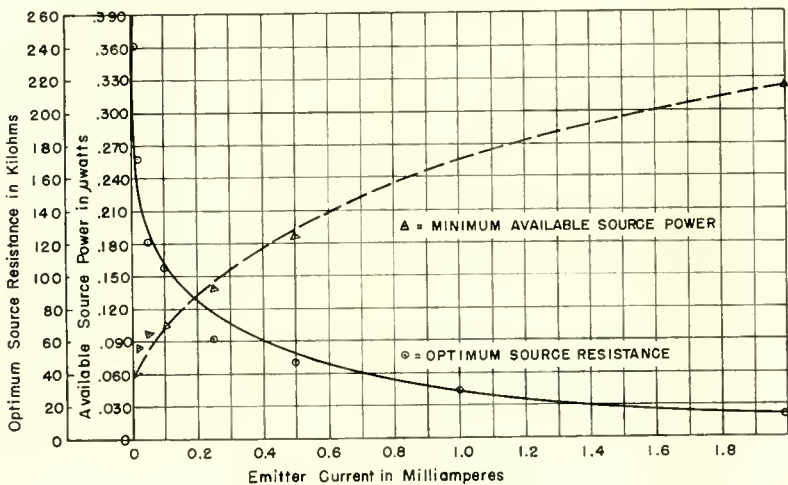


Fig. 14—The optimum source resistance and the minimum available power required to correct drift plotted as a function of emitter current for a silicon transistor (904-A). Collector-to-emitter voltage is 6 volts. The temperature range is 30°C to 70°C.

at lower emitter currents. Figure 14 is a plot of optimum source resistance and drift correction power at different emitter currents for a silicon transistor (904-A). The drift correction power decreases with lowered emitter currents while the optimum source resistance increases. Here the limit of usefulness is set by the decrease of transistor gain at low emitter currents.

Compensation

Compensation is useful for improving the signal-to-drift ratio of transistor direct-coupled amplifiers. Compensation can be used at the input of an amplifier so as to improve the signal-to-drift ratio and also

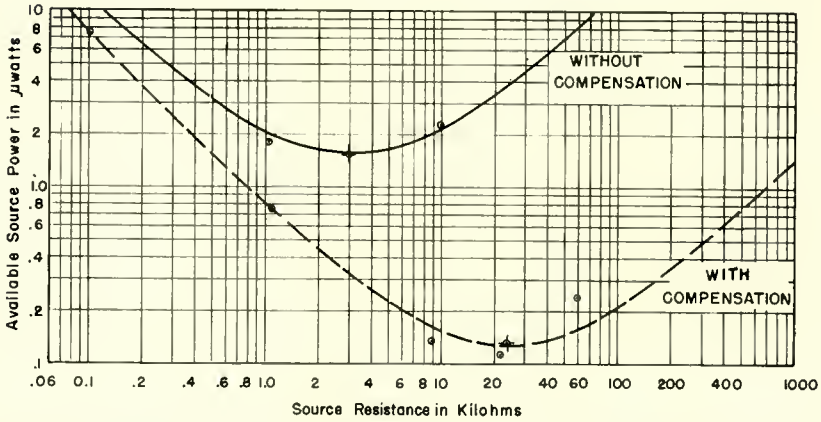


Fig. 15—Available power required to correct for drift due to temperature change plotted as a function source resistance for a germanium transistor (2N109) with and without compensation at the input.

to alter the optimum source resistance. This is important because it admits the possibility of designing the amplifier to match a fixed source. A simple scheme is to supply compensation current at the input of the amplifier by means of a reversed biased diode which is chosen to follow the temperature-produced changes of the transistor as closely as possible. Results obtained for a 2N109 with and without such diode compensation are plotted in Figure 15. The optimum source resistance has been increased and the obtainable signal-to-drift ratio improved.

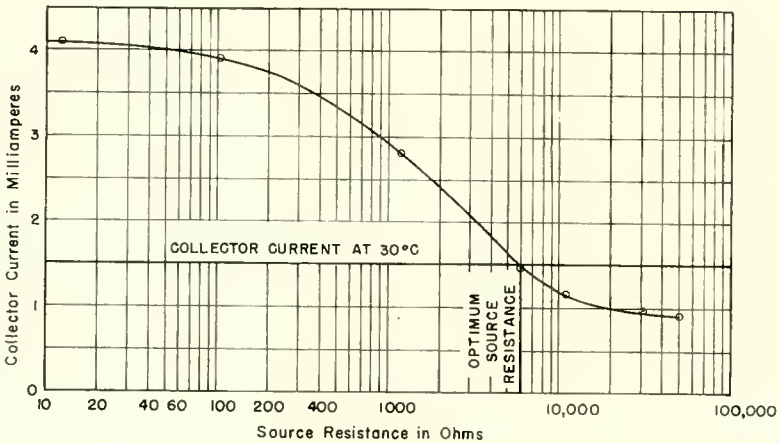


Fig. 16—Collector current at 50°C plotted as a function of source resistance for a germanium transistor with excess reverse current compensation at the input. The drift component of collector current is eliminated at the optimum source resistance.

It is possible, by supplying an excess of reverse current at the base of the transistor, to overcompensate so that the direction of drift due to temperature change becomes reversed. This yields the interesting result that drift in the output is eliminated if the optimum source resistance is used with the amplifier. In Figure 16 collector current is plotted against source resistance for such a case. The drift component of the collector current passes through zero and reverses at the optimum value of the source resistance.

The means by which this eliminates drift is illustrated by Figure 17. The reverse saturation current from the diode flows to the device and to the source. If the source resistance is so chosen that the resulting voltage drop is that required for compensation, drift is eliminated. Such a result is possible if the voltage and current change needed for compensation are opposite in sign.

Certain practical limits exist for this case. The optimum source resistance for an amplifier will be subject to some spread of tolerance

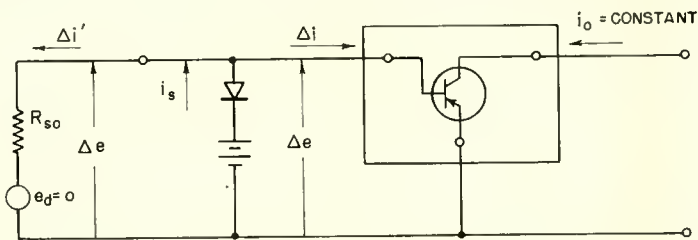


Fig. 17—Compensation reverses drift correction current.

limits; a like situation exists for the source resistance so that an exact match is improbable. Also the optimum source resistance changes somewhat with the temperature increment involved. Such factors combine so that drift will not be completely eliminated.

APPLICATION

The correct application of these principles to the design of a direct-coupled amplifier requires careful appraisal of design objectives. For example, the distinction between drift magnitude and signal-to-drift ratio is an important one. Minimum drift does not necessarily imply best signal-to-drift ratio. Negative feedback reduces drift but also decreases the signal-to-drift ratio. A common-collector stage drifts much less than a common-emitter stage, but the signal-to-drift capabilities are nearly equal. Performance requirements with respect to drift, gain, and signal-to-drift ratio are therefore separate factors to take into account.

The signal-to-drift ratio of an amplifier is improved by reducing available source power which will correct drift. With a particular amplifier this value is a minimum for the optimum source resistance. However, in many cases the source resistance is fixed and not equal to the optimum source resistance of the amplifier. This means only that if the source resistance could be altered without losing available power, then the maximum obtainable signal-to-drift ratio for that amplifier could be achieved.

No increase of signal-to-drift ratio is realized by attempting to improve the match with added resistance between the signal source and the input of an amplifier. If such added resistance is regarded as part of the signal source, the resulting available signal power is lower. If the added resistance is regarded as part of the amplifier it absorbs

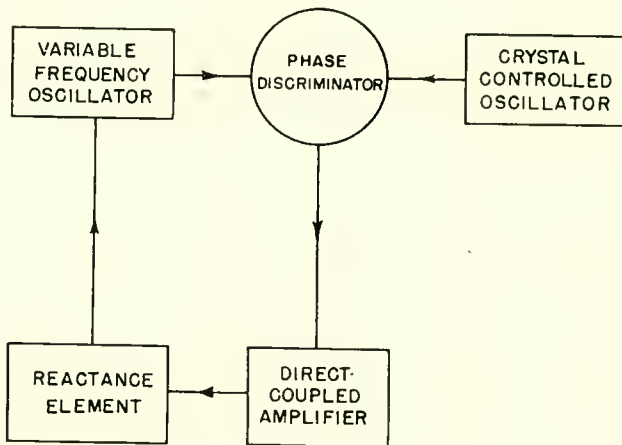


Fig. 18—Direct-coupled amplifier for automatic frequency control.

power from the source, reducing the power delivered as effective drive to the amplifier. In any case, that amplifier which provides the best signal-to-drift ratio requires the least available power from the given source to correct drift.

There are widespread applications of direct-coupled amplifiers for which the source resistance can be altered without loss of available source power. An example is the use of a direct-coupled amplifier in the a-f-c loop of a controlled oscillator as shown in Figure 18. In this application the direct-coupled amplifier is used to drive a frequency-control element which varies the frequency of the variable frequency oscillator. The d-c source for this amplifier is a phase discriminator which provides an output if there is a frequency shift of the VFO. The output of the phase discriminator represents a comparison between the

VFO and the crystal-controlled oscillator, which provides a signal for automatic frequency control.

In this application, drift of the d-c amplifier results in an associated drift of the VFO. Consequently, the frequency stability of the VFO depends upon the signal-to-drift ratio of the amplifier. In such a case the obtainable signal-to-drift ratio of the amplifier can be achieved if the source resistance represented by the phase discriminator is set equal to the optimum source resistance of the amplifier.

MEASUREMENT OF AVAILABLE POWER REQUIRED TO CORRECT FOR DRIFT

The circuit of Figure 19 was used to determine the available source power required to correct for drift due to temperature change. In this circuit E_{st} is a standard voltage source and E_v is a variable voltage source. The important measured quantities are E , the voltage differ-

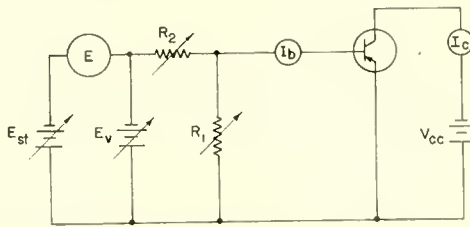


Fig. 19—Measurement of available source power required to correct for drift.

ence between the sources, I_b , the base current and I_c , the collector current. The procedure is to set the variable voltage equal to the standard at the initial temperature and then to adjust R_2 and R_1 for the desired source resistance and operating point. Then the variable voltage source, E_v , is used to correct for drift due to the temperature change of interest; i.e., the collector current is maintained constant. The available source power required to do this is

$$\frac{\left(\frac{E R_1}{R_1 + R_2}\right)^2}{4 \left(\frac{R_1 R_2}{R_1 + R_2}\right)} = \frac{E^2 R_1}{4 R_2 (R_2 + R_1)}$$

Values of this available power obtained over a range of source resistance show a minimum for the optimum source resistance.

It is possible to determine the optimum source resistance directly from one such measurement since $R_{so} = \Delta e / \Delta i$.

The current change at the input, Δi , is the difference between base currents at the initial and final temperatures for constant collector current. The corresponding voltage change at the input, Δe , can be determined since the equivalent source voltage and resistance are known. The relationship in terms of measured quantities is

$$\Delta e = \frac{E R_1}{R_1 + R_2} - \frac{R_1 R_2}{R_1 + R_2} (I_{bf} - I_{bi}),$$

where I_{bi} and I_{bf} are the initial and final base-current readings, respectively. This permits a check of each determination and fixes the optimum value more precisely than would be possible with plotted curve alone.

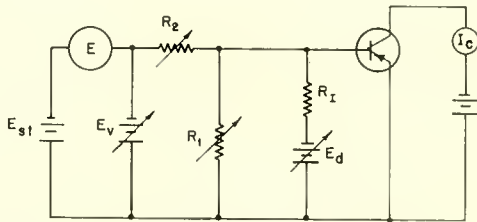


Fig. 20—Measurement of small currents required to correct for drift.

If the input-current changes are very small, some difficulty is encountered with this procedure of direct measurement. Very small currents can be introduced readily by the method shown in Figure 20. The operating point and source resistance are established as before by manipulating R_1 and R_2 . Small input-current changes are introduced by means of the large resistor R_1 and voltage source E_d . For example, if R_1 is 10 megohms and E_d is one volt, an input current of 1/10 microampere is obtained. Two measurements are required to fix the values of current and voltage change at the input of the transistor. Thus

$$\Delta e = (i_1 - \Delta i) R_{s1},$$

$$\Delta e = (i_2 - \Delta i) R_{s2},$$

where i_1 and i_2 are the small-current increments introduced and R_{s1} and R_{s2} are the different equivalent source resistances chosen for the determination. Then

$$\Delta i = \frac{R_{s1} i_1 - R_{s2} i_2}{R_{s1} - R_{s2}},$$
$$\Delta e = \frac{R_{s1} R_{s2} (i_1 - i_2)}{R_{s2} - R_{s1}}.$$

ACKNOWLEDGMENT

Most of the measurements described were made by Sheldon Edelman and William Nolan.

HARMONIC GENERATION WITH NONLINEAR REACTANCES*

BY

KERN K. N. CHANG

RCA Laboratories,
Princeton, N. J.

Summary—A simple theory of frequency multiplication using nonlinear reactances is derived. The conversion efficiency is shown to increase with the circuit Q 's, the degree of nonlinearity, and the input power. The theory is compared with experiments performed with the nonlinear capacitance of a germanium point-contact diode. An efficiency of 22 per cent has been attained on a 3,300 to 6,600 megacycle doubler. It is found that experimental results agree in general with the theory at low input power levels. At high powers, the efficiency is found to drop below the theoretical value. The drop in efficiency may be caused by the power lost into higher harmonics and by the intermittent voltage breakdown of the diode.

INTRODUCTION

HARMONIC generation through tube techniques has been no problem at low and moderately high frequencies. In the centimeter or millimeter wave region where traveling-wave multipliers are used, however, the r-f fields shrink closely around the r-f traveling-wave structures thus weakening the interaction between field and electrons and causing the multiplication process in the tube to become extremely inefficient. Manley and Rowe¹ have pointed out the possibility of efficient frequency multiplication using nonlinear reactances. Since the nonlinear reactance is essentially lossless, the conversion efficiency should be high compared to efficiencies achieved with nonlinear resistances. Higa² has suggested using nonlinear ferroelectric reactances for frequency multiplication. Melchor et al³ have made an efficient microwave doubler through use of a ferrite which is essentially a nonlinear inductive reactance. Kita⁴ has built a frequency

* Manuscript received July 11, 1958.

¹ J. M. Manley and H. E. Rowe, "Some General Properties of Nonlinear Elements," *Proc. IRE*, Vol. 44, p. 904, July, 1956.

² W. H. Higa, "Theory of Nonlinear Coupling in a Novel Ferroelectric Device," *Jour. Appl. Phys.*, Vol. 27, p. 775, July, 1956.

³ J. L. Melchor, W. P. Ayres, and P. H. Vartanian, "Microwave Frequency Doubling from 9 to 18 KMC in Ferrites," *Proc. IRE*, Vol. 45, p. 643, May, 1957.

⁴ Shaichi Kita, "A Harmonic Generator by use of the Nonlinear Capacitance of Germanium Diode," *Proc. IRE*, Vol. 46, p. 1307, June, 1958.

doubler (24,000 to 48,000 megacycles) using the nonlinear capacitance of a germanium diode and has achieved an efficiency of 2.6 per cent. Uhlir⁵ has also used a semiconductor diode and has produced a second harmonic (430 to 860 megacycles; efficiency, 29 per cent), a third harmonic (330 to 990 megacycles; efficiency, 16 per cent) and a seventh harmonic (1,200 to 8,400 megacycles; efficiency .125 per cent).

This paper, too, deals with experimental results on frequency multiplication using a semiconductor diode. In addition, it derives a simple theory of frequency multiplication and compares it with the experimental data.

The experiments were performed on a 3,300-to-6,600 megacycles frequency doubler and a 2,200-to-6,600 megacycles frequency tripler using a germanium point-contact diode.* The conversion efficiency obtained in the doubler case was found to be comparable with that which was attained with a ferrite.³

THEORY

Manley and Rowe¹ have derived a power relation connecting three linear circuits which are resonant individually at frequencies ω_1 , ω_2 , and $\omega_1 + \omega_2$ and which are coupled together through a nonlinear reactance. This power relation states that

$$\frac{W_{12}}{\omega_1 + \omega_2} = -\frac{W_1}{\omega_1} = -\frac{W_2}{\omega_2}. \quad (1)$$

Here W_{12} , W_1 , W_2 are the average powers flowing into the nonlinear reactors at the respective frequencies $\omega_1 + \omega_2$, ω_1 , ω_2 . The minus signs indicate that when high-frequency pumping power W_{12} flows into the reactor, lower frequency powers, W_1 and W_2 flow out of the reactor. However, the above power relation can also be written as

$$-\frac{W_{12}}{\omega_1 + \omega_2} = \frac{W_1}{\omega_1} = \frac{W_2}{\omega_2}. \quad (2)$$

This suggests that if the reactor is pumped at the two lower frequencies, ω_1 and ω_2 , then power at the high frequency, $\omega_1 + \omega_2$, will flow out of the reactor. As a corollary, the low-frequency pumps can have

⁵ A. Uhlir, Jr., "The Potential of Semiconductor Diodes in High-Frequency Communications," *Proc. IRE*, Vol. 46, p. 947, June, 1958.

* The diodes used in these investigations were of an experimental design provided by C. Stocker of RCA Laboratories.

equal frequencies; also there can be more than two pumps. This will lead to the following multiplier system.

Consider the two resonant circuits shown in Figure 1. The signal circuit contains a current generator $I_1 \cos \omega_1 t$, and a tank conductance G_1 . The output circuit, being resonant at a frequency n times the signal frequency, n being an integer, has an output tank, a tank conductance G_t , and a load conductance G_L . These two tank circuits are capacitively coupled through a nonlinear capacitor, which has a loss conductance G_c . The electric charge q on the nonlinear capacitance is taken as n^{th} power of the voltage v ; e.g.

$$q = c_0 v - C_n v^n \tag{3}$$

where c_0 is the linear capacitance, C_n is the nonlinearity coefficient, and $C_n v^n$ is presumed small compared to $c_0 v$.

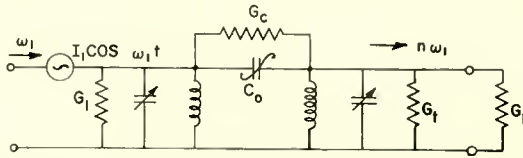


Fig. 1—Equivalent circuit of a frequency multiplier using a nonlinear capacitor.

The general charge-voltage equations for the signal circuit and the multiplier circuit are similar to those which were derived in a previous article⁶ by replacing flux for charge, current for voltage, and impedance for admittance. These equations are

$$I_1 = V_1 Y_1 - j \omega_1 C_n \left(\frac{n}{2^{n-1}} \right) (V_1^*)^{n-1} V_n, \tag{4}$$

$$0 = V_n Y_n - j \omega_1 C_n \left(\frac{n}{2^{n-1}} \right) V_1^n.$$

At resonance, Equation (4) becomes

$$I_1 = V_1 G_1 + \omega_1 C_n \left(\frac{n}{2^{n-1}} \right) V_1^{n-1} V_n, \tag{5}$$

⁶ S. Blom and K. K. N. Chang, "Theory of Parametric Amplification Using Nonlinear Reactances," *RCA Review*, Vol. 18, p. 578, December, 1957.

$$0 = V_n G_n - \omega_1 C_n \left(\frac{n}{2^{n-1}} \right) V_1^n,$$

where $G_n = G_L + G_t$.

These give

$$I_1 = V_1 G_1 + \left[\omega_1 C_n \left(\frac{n}{2^{n-1}} \right) \right]^2 \frac{V_1^{2n-1}}{G_n}. \quad (6)$$

Equation (5) reveals that due to the nonlinear interaction a power $W_1 = n\omega_1 C_n (n/2^{n-1}) V_1^n V_n$ in the signal circuit is transformed to a power $W_n = n\omega_1 C_n (n/2^{n-1}) V_1^n V_n$ in the output circuit. Thus the power relation of Equation (2) is satisfied.

CONVERSION EFFICIENCY

The conversion efficiency of a frequency multiplier is defined as the ratio of power output at the multiplied frequency to power input at the signal frequency. The power output at the multiplied frequency, according to Equation (5), is

$$P_n = V_n^2 G_L = \left[\omega_1 C_n \left(\frac{n}{2^{n-1}} \right) \right]^2 V_1^{2n} \frac{G_L}{G_n^2}. \quad (7)$$

The power input at signal frequency is

$$P_1 = V_1 I_1 = V_1^2 G_1 + \left[\omega_1 C_n \left(\frac{n}{2^{n-1}} \right) \right]^2 \frac{V_1^{2n}}{G_n}. \quad (8)$$

Therefore, the efficiency η_n is

$$\eta_n = \frac{P_n}{P_1}. \quad (9)$$

In Equation (9), the efficiency can be computed if V_1 is known. V_1 can be determined from Equation (6), an equation of degree $(2n-1)$ if I_1 is known.

It is possible to give general solutions to equations of the third and fourth degree, but it is impossible to obtain general algebraic solutions of equations higher than the fourth degree. Nevertheless, since I_1 , G_1 and $[\omega_1 C_n (n/2^{n-1})]^2 / G_n$ are all positive values in Equation (6), by Descartes' Rule Equation (6) has one and only one positive real

root and no negative real roots. Positive and real values of V_1 are, of course, consistent with physical realization.

FREQUENCY DOUBLER

For the case of frequency doubling, $n = 2$. In this case Equation (4) is a cubic of which the general solution is

$$\frac{V_1}{V_0} = \left[\frac{1}{Q_{02}} + \sqrt{\left(\frac{1}{Q_{02}}\right)^2 + \frac{1}{27} \left(\frac{1}{Q_{12}}\right)^3} \right]^{1/3} + \left[\frac{1}{Q_{02}} - \sqrt{\left(\frac{1}{Q_{02}}\right)^2 + \frac{1}{27} \left(\frac{1}{Q_{12}}\right)^3} \right]^{1/3} \quad (10)$$

where

$$Q_{12} = \frac{1}{4} Q_1 Q_2, \quad (11)$$

$$Q_{02} = \frac{1}{2} Q_0 Q_2, \quad (12)$$

$$Q_1 = \frac{\omega c_0}{G_1}, \quad (13)$$

$$Q_2 = \frac{\omega c_0}{G_2}, \quad (14)$$

$$Q_0 = \frac{\omega c_0}{G_0}, \quad (15)$$

$$G_0 = \frac{I_1}{V_0}, \quad (16)$$

$$V_0 = \frac{c_0}{2C_2}. \quad (17)$$

In Equations (10), (16), and (17), V_0 is the normalization voltage. It has the physical significance of a driving voltage across the non-linear capacitor when a maximum charge q is reached. This voltage is a theoretical one. In practical cases, the driving voltage is always less than the normalized voltage.

The efficiency for the doubler is then, according to Equation (9),

$$\eta_2 = \frac{G_L Q_{12} \left(\frac{V_1}{V_0} \right)^2}{G_2 \left[1 + Q_{12} \left(\frac{V_1}{V_0} \right)^2 \right]}, \quad (18)$$

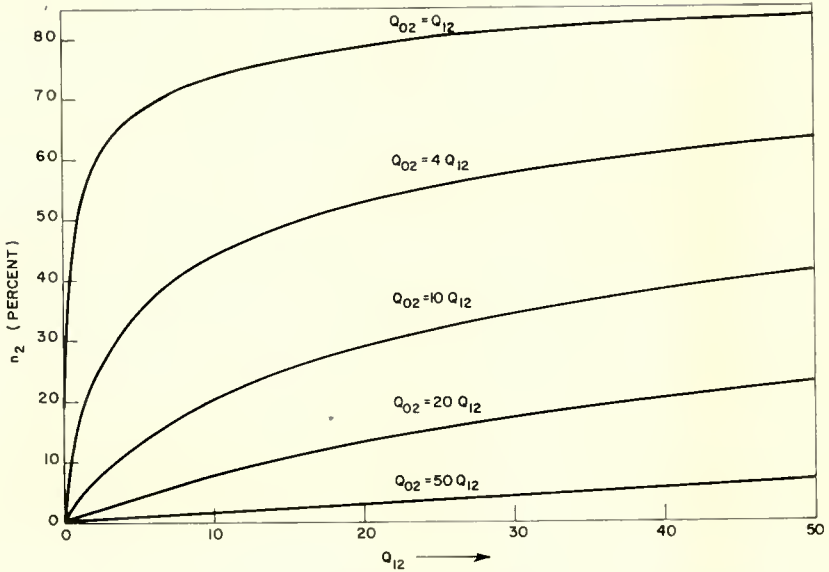


Fig. 2—Variation of efficiency with circuit Q_{12} and driving parameter Q_{02} for a doubler.

or, in terms of a normalized input power \bar{P}_1 where $\bar{P}_1 = P_1/V_0^2 G_1$,

$$\eta_2 = \frac{(\sqrt{1 + 4Q_{12}\bar{P}_1} - 1)^2}{4Q_{12}\bar{P}_1} \frac{G_L}{G_2}. \quad (19)$$

η_2 is plotted versus Q_{12} in Figure 2. The curves shown indicate that higher efficiencies occur at high Q_{12} 's and lower Q_{20} 's. High Q_{12} 's and Q_{20} 's mean respectively low-loss circuits and high pumping powers.

FREQUENCY TRIPLER

In the tripler case ($n = 3$), the normalization voltage V_0' is taken as

$$V_0' = \sqrt{\frac{c_0}{3\epsilon_s}} \quad (20)$$

which again is the theoretical driving voltage for a maximum charge q . After normalization, Equation (6) can be written as

$$\left(\frac{V_1}{V_0'}\right)^5 + \frac{4}{Q_{13}} \left(\frac{V_1}{V_0'}\right) - \frac{8}{Q_{03}} = 0, \quad (21)$$

where

$$Q_{13} = \frac{1}{4} Q_1 Q_3,$$

$$Q_{03} = \frac{1}{2} Q_0 Q_3,$$

$$Q_1 = \frac{\omega C_0}{G_1},$$

$$Q_3 = \frac{\omega C_0}{G_3},$$

$$Q_0 = \frac{\omega C_0}{G_0},$$

$$G_0 = \frac{I_1}{V_0'}. \quad (22)$$

Equation (9) then becomes

$$\eta_3 = \frac{\frac{1}{4} Q_{13} \left(\frac{V_1}{V_0'}\right)^4}{1 + \frac{1}{4} Q_{13} \left(\frac{V_1}{V_0'}\right)^4} \left(\frac{G_L}{G_3}\right), \quad (23)$$

or, in terms of a normalized input power \bar{P}_1

$$\eta_3 = \frac{G_L}{G_3} \left[1 - \frac{\left[\frac{2\bar{P}_1}{Q_{13}} + \sqrt{\frac{4\bar{P}_1^2}{Q_{13}^2} + \frac{64}{27Q_{13}^3}} \right]^{1/3} + \left[\frac{2\bar{P}_1}{Q_{13}} - \sqrt{\frac{4\bar{P}_1^2}{Q_{13}^2} + \frac{64}{27Q_{13}^3}} \right]^{1/3}}{\bar{P}_1} \right] \quad (24)$$

The fifth-degree Equation (21) can be solved by the aid of a computer. With known values of V_1/V_0' , the plot of efficiency, η_3 , versus Q_{13} is shown in Figure 3. As shown in the figure, the efficiency rapidly drops with decrease of the pumping power compared with the doubler case.

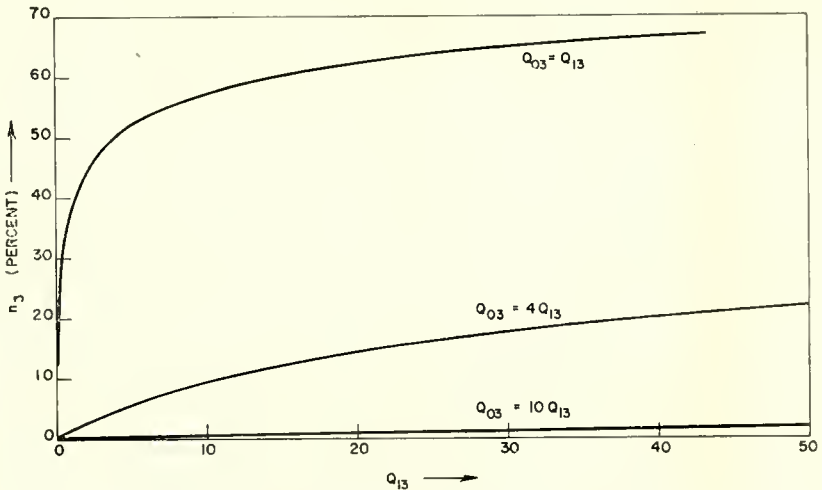


Fig. 3—Variation of efficiency with circuit Q_{13} and driving parameter Q_{03} for a tripler.

This is not surprising since in the tripler case the efficiency is a function of the 4th power of the driving voltage (see Equation (21)) instead of the 2nd power.

EXPERIMENTAL

A schematic diagram of an experimental multiplier is shown in Figure 4. A pump signal of a frequency f_1 is sent through a bandpass filter to a cavity C_1 which is resonant at the frequency f_1 . The cavity C_1 is coupled to a cavity C_2 by a semiconductor diode which has a non-linear capacitance. The cavity C_2 is tuned at nf_1 , which is the desired harmonic of the pump frequency. The produced harmonic signal is then delivered to a matched load through a bandpass filter.

For convenience, coaxial wave guides have been chosen as the resonant cavities. These wave guides are shown in the photograph of Figure 5. Two tunable wave guide cavities, made of standard coaxial lines, are coupled together by a "tee" junction. A point contact germanium diode mounted in a standard capsule is installed in one

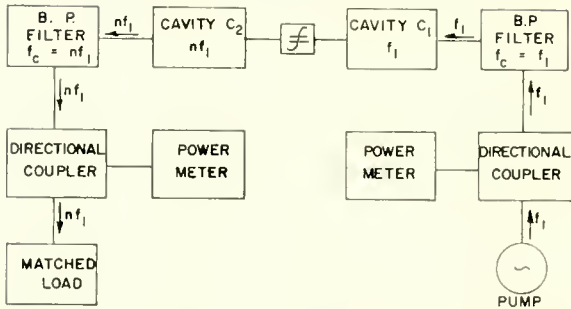


Fig. 4—Schematic of a frequency multiplier using a nonlinear capacitor.

branch of the tee junction and is put in direct contact with the center conductor of the coaxial line.

The diode has a zero-bias capacitance in the order of 1 micro-microfarad and a Q of 2. It has a breakdown reversed voltage of 6 volts.

Measurements were made on frequency doubling with a pump at 3,300 megacycles, and on frequency tripling with a pump frequency of 2,200 megacycles. The 6,600 megacycles multiplied signal is picked up through a directional coupler (Figure 4).

The measured efficiency versus the input power is plotted in Figure 6. By assuming $Q_{12} = Q_{13} = 1.6$, $V_0 = .5$, $V_0' = .47$ and $G_1 = .2$, which are reasonable values for the present circuit, the theoretical electronic efficiencies (i.e., η_2 , η_3) based on Equations (19) and (24) are plotted as dotted lines in the same figure. The figure shows that at low input

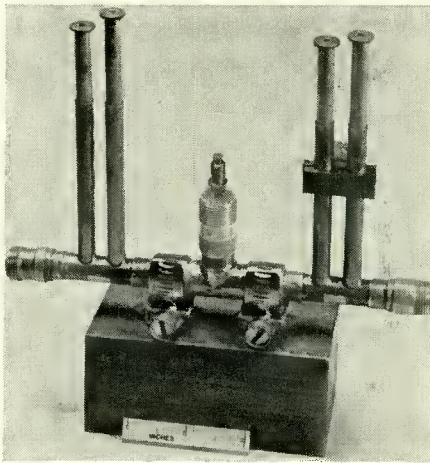


Fig. 5—Resonant cavities of a frequency multiplier.

powers the measured results, in general, follow the theoretical ones. As the input power becomes larger, however, the measured efficiency deviates more and more from the theoretical value and tends to drop at very high input power levels. This drop in efficiency at high powers may be due to two causes: first, power is lost in higher harmonics and second, at high input powers the diode breaks down over part of the cycle of the driving voltage.

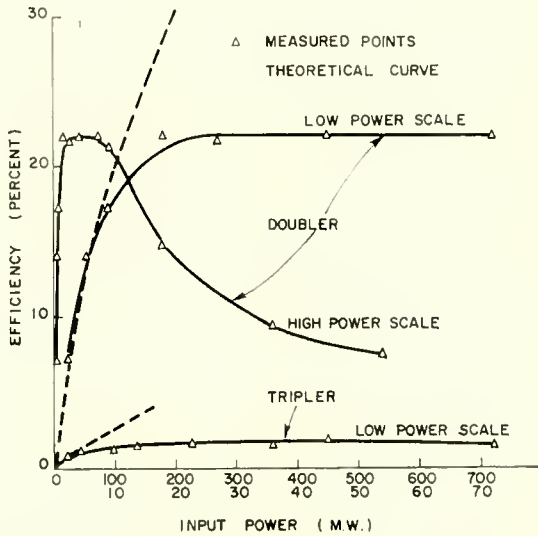


Fig. 6—Experimental and theoretical results of efficiency variation with input power.

CONCLUSION

Frequency multiplication can be achieved in a very efficient way through use of a nonlinear reactance. The efficiency increases as the circuit Q 's, the degree of nonlinearity, and the input power increase. By a simple demonstration, a germanium point-contact diode is shown to give a conversion efficiency of 22 per cent for doubling and 2 per cent for tripling at 6,600 megacycles. The efficiency drops at high input power levels. If the diode had a higher breakdown voltage and higher harmonics rejection, the efficiency would have been much higher than that observed. It is thus concluded that efficient harmonic generation is possible in the centimeter-wave region or even in the millimeter-wave region through nonlinear semiconductor diodes provided that the diodes have a high value of Q and a high breakdown voltage.

A COLOR "ELECTROFAX" PROCESS

BY

J. S. RYDZ AND S. W. JOHNSON

RCA Defense Electronic Products,
Camden, N. J.

Summary—Recent developments with Electrofax toners have resulted in color Electrofax prints. These prints, while similar in appearance to photographic results, do not follow the subtractive color mixture laws which describe dye images. In the present paper, the basic principles of pigment mixtures and the application of the Kubelka analysis to the color Electrofax process are discussed. Colorimetric data for the printing of Electrofax particles is presented. A method of separating color for producing line work is discussed. The method differs from the conventional techniques of using tri-color filters in that the separation of color is achieved by controlling the amount of exposure on a scene. The mixing of colors which occurs due to the fusing operation in the Electrofax process provides the potential means for a color print system based on particle mixing rather than over-laying dyes as in a halftone printing ink process.

INTRODUCTION

THE DIRECT electrophotographic printing process known as "Electrofax" and described by C. J. Young and H. G. Greig¹ has led to many interesting new concepts. Among the newly developed techniques is an Electrofax color print process. The development of color toners and the characteristics of electrostatic printing results in several techniques for producing color prints.

This paper presents some of the techniques for achieving electrophotographic color prints and discusses the colorimetric laws describing the behavior of Electrofax particles.

Initially, the basic concepts of the Electrofax process are reviewed. The basic types of color mixture laws and their application to the Electrofax process are discussed. A unique method for achieving color separation without the use of tricolor filters is presented. Finally, experimental data is analyzed and basic color mixture laws applied to the color Electrofax process.

THE ELECTROFAX PROCESS

Before describing the colorimetric aspects of the color Electrofax process, the basic concepts as described by Young and Greig will be reviewed. A simple 4-step procedure describes the basic process.

¹ C. J. Young and H. G. Greig, "Electrofax—Direct Electrophotographic Printing on Paper," *RCA Review*, Vol. XV, p. 469, December, 1954.

1. An Electrofax surface which is a thin flexible white coating of special zinc oxide in a resin binder is coated on ordinary paper. The coating is charged to a negative potential rendering it sensitive to light.
2. The light sensitive sheet is exposed by any of the usual photographic techniques resulting in a reduction of charge in the exposed areas. This produces a latent image in the form of electrostatic charge which remained in the areas receiving little or no light.
3. The latent image is developed by applying a pigmented resin powder carrying a positive electrostatic charge. The powder is attracted electrostatically and held by the negatively charged image areas.
4. A durable light-fast image is then produced by applying heat to melt the resin powders so as to fuse it to the paper surface.

Carrying out these steps by hand results in a finished copy in less than a minute.

To implement the 4 basic steps described above, methods presented by Young and Greig were utilized in the preparation of color prints. Initially, the Electrofax surface is placed on a ground plate and exposed to a corona discharge from a 6-kilovolt negative d-c source. This results in a charge density distribution on the coated surface. The voltage, which can be measured by electrometer techniques, is initially between 300 and 600 volts at the surface.

Exposing the charged electrofax surface to light decreases the charge in the illuminated areas. Furthermore, the charges in these areas leave in proportion to the amount of incident light. This results in a latent electrostatic image on the top of the surface of the coating.

Developing consists in bringing fine positively charged particles of developer powder or toner in contact with the surface so that they will adhere to those areas which remained charged. The technique of developing involves use of a magnetic brush which is a mass of iron filings mixed with toner and picked up on the end of a permanent magnet. The iron filings take on a negative charge and the toner particles become positive due to the triboelectric effect. Sweeping the brush over the latent image results in the formation of an image of full density on a clear background.

The development of color Electrofax particles results in the ability to produce color prints by these techniques. An analysis of the printed results reveals that the colorimetric laws describing present-day printing or photographic processes are not directly applicable to color

Electrofax prints. Before discussing the color Electrofax process, the basic color mixture laws will be reviewed.

TYPES OF COLOR MIXTURES

Generally speaking, color mixture phenomena are classified into two types, namely, additive and subtractive. This classification is, however, an oversimplification. In practice, many substances follow laws which are neither purely additive nor purely subtractive. For purposes of analysis, color mixtures can be classified as follows:

1. average additive mixture,
2. integrated additive mixture,
3. true subtractive mixture,
4. halftone mixture,
5. dye mixture,
6. pigment mixture.

An average additive mixture occurs when

1. a rapid succession of different colored light flashes occur,
2. a multicolored sector disc is rapidly revolved,
3. a composite of different color patches is viewed from a distance.

In the average additive mixture, the tristimulus values of the result are the weighted means of the tristimulus values of the components. That is, if X_a, Y_a, Z_a are the tristimulus values of one component in a two-component mixture and X_b, Y_b, Z_b are the tristimulus values of the other component, then the resultant tristimulus values of the mixture will be

$$X_R = fX_a + (1 - f)X_b, \quad (1)$$

$$Y_R = fY_a + (1 - f)Y_b, \quad (2)$$

$$Z_R = fZ_a + (1 - f)Z_b, \quad (3)$$

where the subscript R denotes the resultant mixture and f is the fractional part of the area occupied by component a . It is noted that a mixture of yellow and blue will appear gray in the case of an average additive mixture. This is due to the lower luminance resulting from f in Equation (2).

An integrated additive mixture occurs when colored lights are projected simultaneously on a white screen. In this case

$$X_R = X_a + X_b, \quad (4)$$

$$Y_R = Y_a + Y_b, \quad (5)$$

$$Z_R = Z_a + Z_b. \quad (6)$$

It is noted that in an integrated additive system, a mixture of complementary yellow and blue produce white.

A true subtractive mixture occurs when

1. transparent fluids are mixed (assuming no chemical reaction takes place),
2. light passes successively through colored transparent layers,
3. a transparent colored layer is placed over a colored surface.

In the case of the subtractive type of mixture, no simple relationship exists between the tristimulus values of the components and the resultant mixture. According to Beers Law

$$T_i = e^{-cK_1X}, \quad (7)$$

where T_i is the transmittance, X is the thickness of the layer, c is the concentration, and K_1 is the absorption coefficient for unit concentration ($c = 1$). By applying Equation (7) at a sufficient number of wavelengths, the intensity of light as a function of wavelength for a resultant mixture can be determined.

Halftone mixtures occur in the color printing process when dots of different colors are printed on a reflecting surface such as paper. The familiar Neugebauer equations govern the colors produced by this resultant combination of subtractive and additive effects which occur in halftone printing.² In the case of a three-color process, these equations are written

$$\begin{aligned} X_R = & (1 - c)(1 - m)(1 - y)X_w + cX_c(1 - m)(1 - y) \\ & + mX_m(1 - c)(1 - y) + yX_y(1 - m)(1 - c) + cmX_{cm}(1 - y) \\ & + cyX_{cy}(1 - m) + myX_{my}(1 - c) + cmyX_{cm y}, \end{aligned} \quad (8)$$

$$\begin{aligned} Y_R = & (1 - c)(1 - m)(1 - y)Y_w + cY_c(1 - m)(1 - y) \\ & + mY_m(1 - c)(1 - y) + yY_y(1 - m)(1 - c) + cmY_{cm}(1 - y) \\ & + cyY_{cy}(1 - m) + myY_{my}(1 - c) + cmyY_{cm y}, \end{aligned} \quad (9)$$

² A. C. Hardy and F. L. Wurzburg, "Color Correction in Color Printing," *Jour. Opt. Soc. Amer.*, Vol. 38, p. 300, April, 1948.

$$\begin{aligned}
 Z_R = & (1 - c)(1 - m)(1 - y) Z_w + cZ_c(1 - m)(1 - y) \\
 & + mZ_m(1 - c)(1 - y) + yZ_y(1 - m)(1 - c) + cmZ_{cm}(1 - y) \\
 & + cyZ_{cy}(1 - m) + myZ_{my}(1 - c) + cmyZ_{cmy}, \quad (10)
 \end{aligned}$$

where c , m , y are the percentages of cyan, magenta, and yellow ink dots in any area; the subscript w denotes the component for a white area.

Dye mixtures occur when two or more dyes are applied either simultaneously or successively to a material. The retention of dye by textile fibers is an example of this type of mixture. An analysis of a system of dyed textile fibers is complicated by many variables. Generally, a Kubelka-Munk analysis describes this type of mixture in an approximate manner. Empirical equations have been developed which can be applied to this class of mixture for special cases.³

Pigment mixtures occur when materials are mixed which scatter light and are either translucent or opaque. If the particles are nearly transparent, an approximation to a subtractive mixture occurs. A mosaic pattern composed of completely opaque particles results in colors predicted by averaged additive mixture laws. In practice, neither of these extremes is evidenced. Resultant colors cannot be predicted by subtractive or additive mixture laws. In pigment mixtures, it is necessary to take into account the coefficients of scatter of the components as a function of wavelength in addition to their coefficients of absorption. Experimental evidence exists which demonstrates that a mixture of present color Electrofax particles follow pigment mixture laws.

PIGMENT MIXTURE LAWS AND THEIR APPLICATION TO ELECTROFAX TONERS

Laws governing the optical properties of pigmented or other light-scattering materials have been mathematically expressed in many forms. These formulas are equivalent or related to one another since they are all derived from the same basic formula by assuming different approximations.⁴ The fundamental laws start from the theory of Kubelka and Munk which states that a colored material laid down on a background of any reflectance R_g as a function of the thickness X is given by

³ Deane B. Judd, *Color in Business, Science and Industry*, John Wiley & Sons, Inc., New York, N. Y., 1952, p. 347.

⁴ P. Kubelka, "New Contributions to the Optics of Intensely Light-scattering Materials," *Jour. Opt. Soc. Amer.*, Vol. 38, p. 448, May, 1948.

$$R = \frac{\frac{(R_g - R_\infty)}{R_\infty} - R_x \left(R_g - \frac{1}{R_x} \right) e^{Sx \left(\frac{1}{R_\infty} - R_\infty \right)}}{R_g - R_\infty - \left(R_g - \frac{1}{R_x} \right) e^{Sx \left(\frac{1}{R_\infty} - R_\infty \right)}}, \quad (11)$$

where

S represents the spectral response characteristics of the emulsion. R_x is defined as the reflectance as a function of wavelength of a colorant layer sufficiently thick that a further increase in its thickness does not significantly change its reflectance. Furthermore, R_x depends upon the ratio of the absorption and scattering coefficients K/S as follows:

$$\left(\frac{K}{S} \right)_\lambda = \frac{(1 - R_x)^2}{2R_x}. \quad (12)$$

Appendix A contains a more detailed discussion of these formulas as well as a derivation of Equation (12). A summary of Kubelka's results has been presented by Judd⁵ and has proven to be very useful in the application of these laws to practical colorant analysis. Results of experimental tests with pure Electrofax color toners has revealed that mixtures follow Equation (12). Pure toners in this case refers to the particles of developer powder, free of any iron particles and not subjected to fusing.

Color Electrofax toners were mixed thoroughly and packed in a glass cell. Spectrophotometric curves were obtained for the various mixtures of two-component systems. Using a highly absorbent black surface or diffused white surface as a backing for the cell revealed no significant difference in the reflectivity of the sample. Thus, a spectrophotometric curve obtained in this manner represents R_x as a function of wavelength.

If R_∞ at selected wavelengths is plotted against the fractional amounts of a two-component Electrofax particle mixture, then the shape of the resulting curve will be an indication of the departure from additivity. Judd has illustrated this effect by plotting systems which obey or depart from additive mixture laws.⁵

⁵ Deane B. Judd, *Color in Business, Science and Industry*, John Wiley & Sons, Inc., New York, N. Y., 1952, p. 343.

Figure 1 represents plots of a two-component Electrofax mixture. In the case of thoroughly mixing a two-color Electrofax component system, the resultant curves depart from the straight line as predicted by additive color mixture laws. Furthermore, the curves are analogous to the family of curves illustrated by Judd as representative of systems which obey the Kubelka-Munk analysis.

The application of Equation (12) provides an excellent means for predicting the spectral reflectance curve of an Electrofax particle mixture. By obtaining spectrophotometric curves of the toners in pure

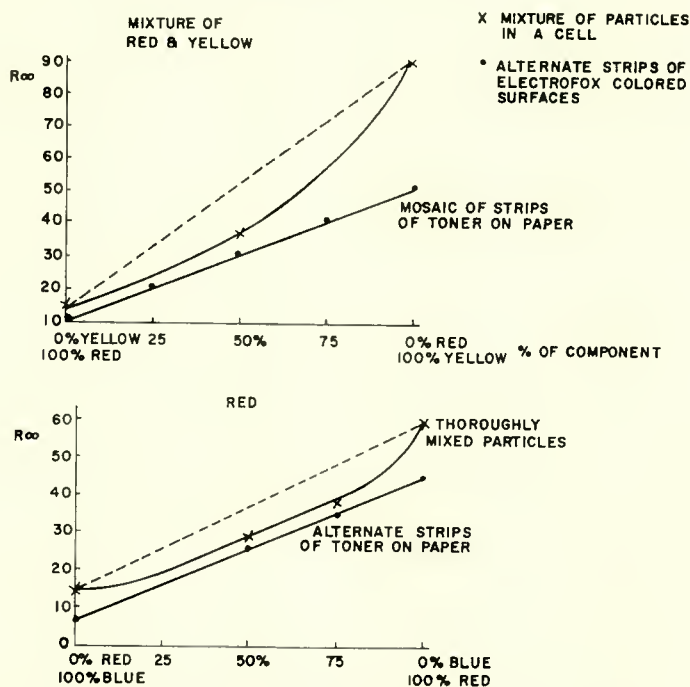


Fig. 1— R_{∞} versus percentage of components in a two-component mixture (R_{∞} taken at $560 \text{ m}\mu$ wavelength).

form and applying Equation (12) wave length by wavelength, resultant mixture curves can be predicted.

In detail, the application of Equation (12) to Electrofax color systems can be accomplished as follows:

1. Pure toners are packed in a cell and spectrophotometric curves obtained of each component.
2. Equation 12 is applied to the spectrophotometric curves at different wavelength ordinates. That is, R_{∞} can be read directly from the curves obtained and $(K/S)_{\lambda}$ solved at the selected wavelength for each of the components.

3. A sum of the absorption scattering ratios $(K/S)_\lambda$ for the components at the selected wavelengths results in predicted values of R_∞ for the *mixture*.

Figure 2 illustrates the curve predicted by this method compared with observed data. It can be seen that the application of Equation (12) predicts rather closely the behavior of thoroughly mixed Electrofax

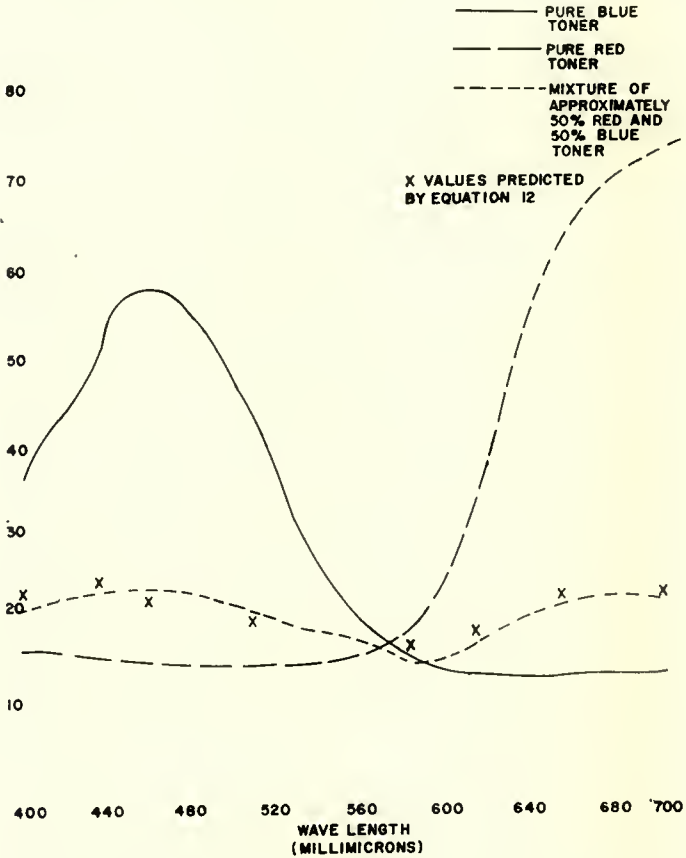


Fig. 2.—Spectrophotometric curves of pure toners and mixtures.

toners in various proportions. Thus, the application of the complex pigment behavior laws derived by Kubelka-Munk are applicable to the prediction of mixing of Electrofax particles in their pure form.

In practice, toners are not in pure form. Also, the particles cover a surface whose spectral reflectance characteristics differ from the toner. Thus, highlight areas will depend on the color of the paper as well as

the particles. In general, the effects of developing, fusing, and paper surface reflections are complex and not easily predicted by theoretical methods. However, empirical data can be developed, and with the aid of the Kubelka-Munk equations, color Electrofax results can be predicted with a fair degree of accuracy. The application of Equation (12) to toner mixtures as described in this section results in the following:

1. A determination can be made without trial mixes as to the color which will be obtained from combining color Electrofax toners.
2. A theoretical color gamut from tristimulus values obtained with predicted curves can be established for which the color bounds of a process can be hypothesized.
3. Relative comparisons can be made between toners in establishing the relative suitability for applications requiring mixing.
4. It is conceivable that the application of these techniques will aid in the converse problem; that is, given a certain desired result, what components are needed to fulfill these requirements.
5. Materials can be evaluated for desired results and factors of availability, economics, and other considerations can be determined in the selection of optimum components for desired applications.

COLOR SEPARATION TECHNIQUES

It has been shown in the previous section that mixtures of present color Electrofax particles follow the Kubelka-Munk analysis. The obtaining of information for the degree of mixing and the amounts of the components in a color process depends upon the color separation method utilized. Through the color separation process, color information for every area in a scene is obtained, usually in the form of tristimulus values. The relative amounts of the tristimulus values determines the color properties of every area in a scene being reproduced.

Two basic requirements which must be considered in a color separation process are

1. The recording of tristimulus color information in the media.
2. The obtaining of tonal values.

The first consideration is related to the spectral response characteristics of the recording media, while the latter consideration is related to the gradient or characteristic curve of the material.

All practical methods of 3-color photography are based on the

trichromatic mechanism of color vision. In the case of photography, color television, and color printing, the color information is obtained by "separating" the scene through red, green, and blue filters. In the photographic case, the exposure values, E_n , impressed on the three emulsions can be written

$$E_n = \int_{\lambda_1}^{\lambda_2} I_\lambda R_\lambda S_\lambda T_n d\lambda. \quad (13)$$

where the subscript n denotes R , G , or B ,

I_λ = spectral distribution of the illumination,

R_λ = spectral reflectance characteristics of the scene,

S_λ = spectral sensitivity of the recording medium,

T_n = spectral characteristics of the red, green or blue filter,

$\lambda_2 - \lambda_1$ = wavelength region of interest.

Analogous expressions can be applied to an Electrofax system. In the analysis of a color Electrofax process, the spectral response characteristic, S_λ , is of prime consideration.

The coating generally associated with Electrofax has a spectral sensitivity response to the near ultraviolet, that is, below 4,200 angstroms. It has been found that quite a number of organic dyes can impart a wider spectral response to the white coating. With the selection of dye, special papers may be designed for nearly any desired spectral response with broad or narrow sensitivity bands. Thus, the addition of minute amounts of certain dyes results in Electrofax coatings usable for color applications.

Figure 3 illustrates the relative reflectance values for two types of dye-sensitized surfaces at different regions of the spectrum. Both surfaces exhibit characteristics similar to panchromatic photographic materials. (It should be noted that the reflectance values illustrated in Figure 3 do not necessarily correspond to the spectral sensitivity of the papers for the Electrofax process. The correlation between the color peaks of the sensitizing dyes with the sensitization peaks is approximate and for purposes of analysis, it will be assumed that these peaks correspond.) Thus, with these surfaces, a color process based on photographing a scene through filters and developing the image after each exposure with the appropriate color toner will result in a direct color print. Of course, the surface must be fused and recharged after each exposure and development.

The wide range of sensitivity at the different portions of the spectrum for the coating illustrated in Figure 3-b indicates a unique

potential method for separating color. The exposure range of this surface has approximately a two-photographic-aperture stop between the spectral color regions. The separation of color can be achieved, by the controlling of exposure rather than with filters as normally used for the making of separations. For example, initially exposing the Electrofax surface with a controlled amount of illumination can result in producing only a blue image. Developing with a conventional blue toner results in rendering the blue-developed areas insensitive to any further charging. Then, reducing the level of light by a controlled amount results in only a red latent image on the Electrofax surface. As in the case of the blue image, red toner is applied, the surface recharged and the exposure time again reduced. Development with a

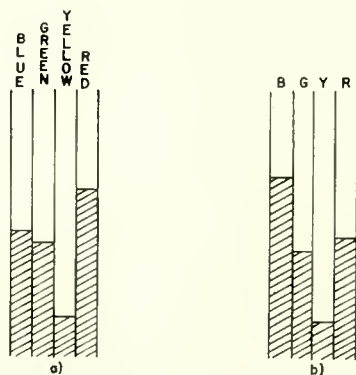


Fig. 3—Relative reflectance values of dye-sensitized surfaces at different regions of the spectrum.

green toner then produces a green image. This method provides a potential means for separating color without the use of tri-color filters.

The technique of separating by exposure is unique to an electrostatic printing process. Inspection of Equation (13) without the filter terms T_n and a familiarity with the conventional Electrofax process reveals that this technique merely utilizes the electrostatic printing characteristics. Consider the effects of increasing or decreasing the exposure, E_n , in Equation (13). Since the conventional process is a direct positive, an increase in E_n will result in a decrease of the resulting image density. Thus, in order to produce the same reflectance values throughout the spectrum with the surface illustrated in Figure 3-b, a value of E_n for green and red exposures must be less than the value of E_n for blue.

If the illumination on a scene is white light, then the color as imaged on an Electrofax surface will be only characteristic of the

reflectance values, R_λ , from the original scene. With a given value of S_λ for a certain portion of the spectrum, the amount of white-light illumination can be controlled to produce an exposure on the Electrofax surface so as to record only those areas with high reflectance values, R_λ , in desired spectral bands.

$$E_n = \int_{\lambda_1}^{\lambda_2} R_\lambda S_\lambda I_\lambda d\lambda. \quad (14)$$

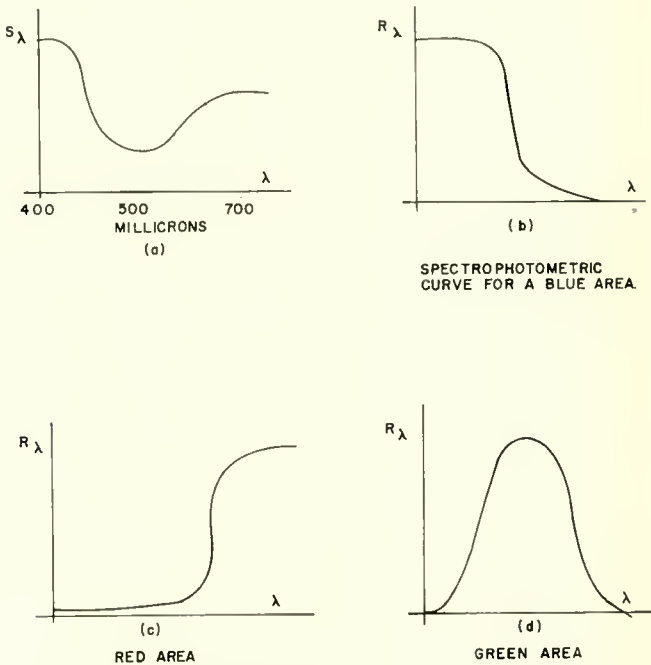


Fig. 4—Hypothetical S_λ for a dye-sensitized surface.

In Equation (14), it is assumed that I_λ is white-light illumination. Furthermore, Equation (14) differs from Equation (13) in that the filter terms T_n have been eliminated. Assume S_λ has a spectral characteristic analogous to that illustrated in Figure 4-a. Blue areas which have R_λ characteristics similar to those illustrated in Figure 4-b can be recorded on an Electrofax surface by controlling the amount of white light illumination incident on the original scene. Establishing the proper level of illumination will result in the recording of blue areas alone since other colored areas will discharge the Electrofax

surface due to the value of R_λ for these areas being relatively greater than S_λ in other regions of the spectrum.

If the electrostatic latent image is developed with conventional blue toner, then S_λ for the blue portion of the spectrum becomes 0; hence, blue areas can no longer be recorded. Decreasing the amount of white-light illumination results in the recording of red areas which have spectral reflectance characteristics similar to those illustrated in Figure 4-c. As in the previous case, green areas with spectral characteristics illustrated in Figure 4-d will have R_λ relatively greater than S_λ in the corresponding regions of the spectrum. These areas will not be recorded until the amount of illumination is again decreased. After development, clear areas of an Electrofax print can again be rendered light sensitive, thus providing the basic mechanism for this color separation technique.

This method is applicable to producing results with a narrow tonal range. Hence, it is applicable only to line-type reproduction. Color prints which have been produced by this technique exhibit the separation between the three colors. While these results are not directly applicable to continuous-tone prints, the method offers the possibility of coding color in line drawings, maps, formats, and other line copy. In these applications, the color information can be in the form of different densities for each color in one photograph, analogous to microfilm. Upon exposure to the Electrofax surface, the different density levels can be transformed to appropriate color by means of the separation technique described in this section.

To summarize, the separation of color for producing Electrofax prints can be accomplished by either of the following methods:

1. Exposing the Electrofax surface through tristimulus color filters. This method is applicable to copying transparencies, art copy or an original scene.
2. Exposing the Electrofax surface at different illumination levels. This method is restricted in use, at present, to line-type copy.
3. Exposing the Electrofax surface to photographic separations made by conventional graphic arts methods.

APPLICATION OF COLOR MIXTURE LAWS TO THE ELECTROFAX PROCESSES

In the production of color images, the color of an area is the result of either areas printed singly, in juxtaposition, superimposed, or mixed thoroughly. In practice, most processes utilize techniques which obtain color by printing areas singly, in juxtaposition or superimposed. Printing by a thorough mixing technique appears to have practical limita-

tions. Colors printed singly, such as line work, usually present no severe colorimetric requirement.

Printing by juxtaposition alone imposes severe registration requirements. Furthermore, such a process obeys the average additive mixture laws described earlier. Considering Equation (2), regardless of how great the luminance values (Y) of the individual components, the resultant luminance of a mixture will be reduced below that of the luminance of the individual components. This is especially true for areas of medium saturation. This factor is not desirable and is probably the most important consideration for not basing a color process on the juxtaposition principle alone.

Tonal values can be produced by varying the density of the colorant, varying the size of dots in a structured image, or controlling the thickness of superimposed layers. There are other methods by which color

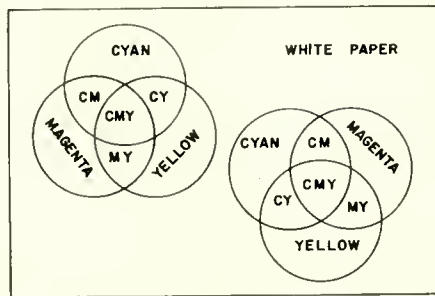


Fig. 5—Half-tone mixture of printing inks.

images can be produced; however, the methods discussed here are directly applicable to color Electrofax prints as well as being the most practical techniques in present day use.

Consider the half-tone dot process shown in Figure 5. The color of an area as seen by the eye from a distance is the integrated result of seeing 8 areas (including the white paper) in juxtaposition. Furthermore, these areas are the result of the inks printed singly or superimposed. In the case of transparent inks, the resulting color tristimulus values can be predicted by the application of the Neugebauer equations.

It was noted previously that the Neugebauer equations are the sum of the tristimulus values of the individual components weighted according to the amount of over-printing in any area of a color print. The application of this equation or any other similar analysis to the color Electrofax process requires that the colorimetric aspects of toner mixtures be considered.

The prime consideration in determining tonal rendition characteristics for the Electrofax process is the familiar H-D curve. Figure 6 illustrates the characteristic curve for typical Electrofax surfaces. It is noted that the gamma of the process is quite high. Thus, the present Electrofax method requires that tonal values be produced by structured images. It has been found that by utilizing conventional half-tone screening techniques, tonal values in color Electrofax prints can be obtained.

In the application of halftone mixture laws to the Electrofax process, several effects require consideration prior to utilizing any existing formulas. Consider, for example, the effects of a halftone image. In

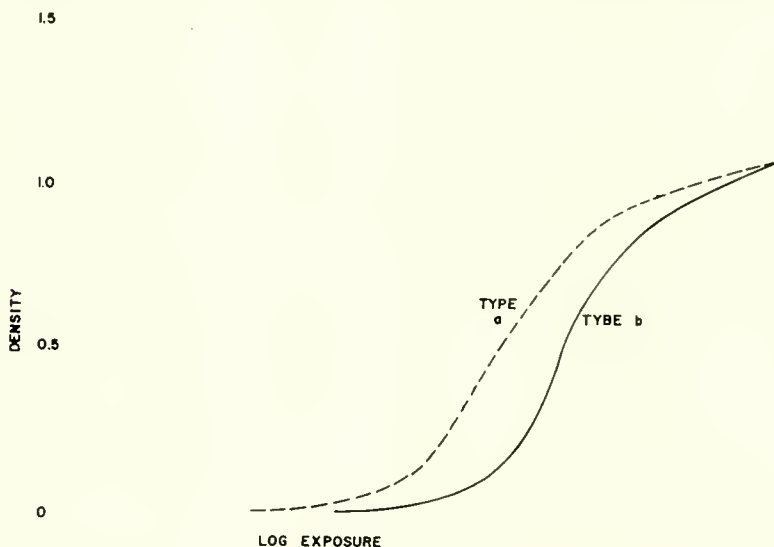


Fig. 6—Characteristic curve for Electrofax surfaces.

any area of variable-size dots, there will be particles adjacent, mixed, and over-printed. Thus, it is desirable to consider the behavior of each of these effects singly.

In the case of areas in juxtaposition, a simple test can be conducted to illustrate the effects which occur in practice. Alternate strips of toner particles printed on paper are arranged in a mosaic pattern. A spectrophotometric curve of such configuration represents the integrated result of viewing colored areas in juxtaposition. If the size of the individual strips are known, then applying Equations (1), (2), and (3) tristimulus values for this type of configuration can be determined.

In place of a strip mosaic type pattern, a series of halftone screened plates were also utilized to expose Electrofax surfaces. In this case,

color areas were printed in juxtaposition. As in the case of the strips, by measuring the size of the areas occupied by the individual colors and substituting these values in Equations (1), (2), and (3), the tristimulus values can be found. A comparison of the data obtained through this method with that obtained from the strips will reveal any departure from additive laws for areas printed in juxtaposition.

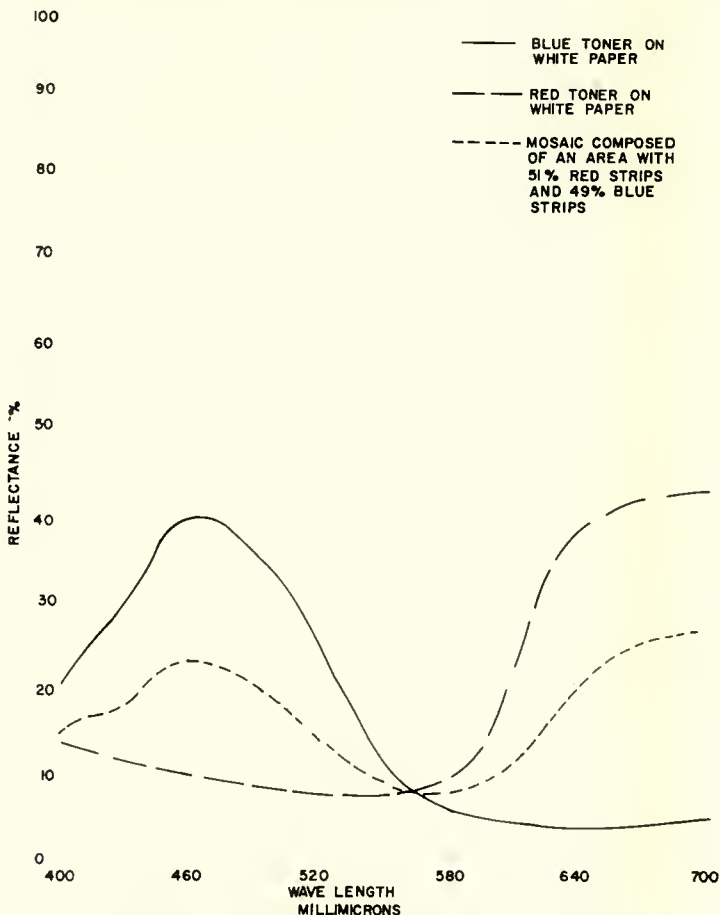


Fig. 7—Spectrophotometric data for toner strips in juxtaposition.

Figure 7 illustrates a spectrophotometric curve of a mosaic comprising strips of approximately 51 per cent red toner and 49 per cent blue toner. Since the mosaic is composed of alternate strips of red and blue, the curve follows the average additive color mixture laws.

A chromaticity diagram for printing in juxtaposition is illustrated in Figure 8. For all practical purposes, it can be seen that additive

laws are nearly followed by this configuration. That is, mixtures lie along a straight line connecting the components.

In the case of printing an Electrofax layer over another toner layer, limitations occur due to the characteristics of electrostatic printing. In particular, the printing of a conventional toner makes the surface unsuitable for further Electrofax use.

Figure 9 illustrates a chromaticity plot of toner layers overprinted. It can be seen that the color of an overprint area is primarily a single

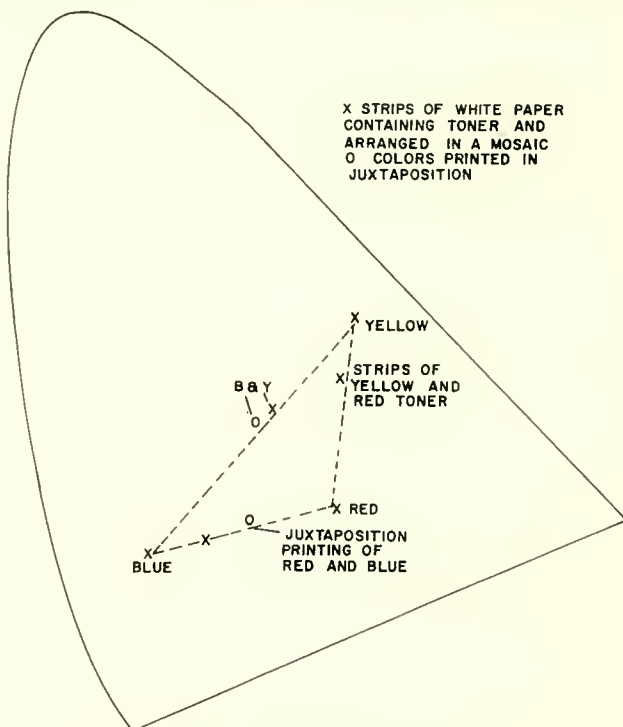


Fig. 8—Chromaticity diagram illustrating additivity effects for juxtaposition printing.

color rather than a mixture. Since the layer is an insulator, the Electrofax properties of the surface are no longer present in these printed areas. Thus, no mixing will occur in areas where two layers overlap.

The color of overlapping layers of non-overprint Electrofax toners is largely determined by the thickness of the applied layer. In the case of a thin layer, overprinting will occur due to an induced image. With a thick layer, the insulating toner will hold a charge, hence, resulting in a solid colored area of subsequent applied toner layers. In areas requiring a sparsely deposited toner distribution, the color

of the area below will be visible through the applied layers. Color prints made with non-overprint particles exhibit no purples, oranges, or browns; in other words, no mixing of the primary colors occurs.

To eliminate this undesirable characteristic, a special overprinting type of toner has been developed. That is, the toners themselves possess electrostatic characteristics; hence, overprinting can be achieved. Figure 9 also illustrates the effect of printing with this special overprinting type of toner. It can be seen that superimposed areas now

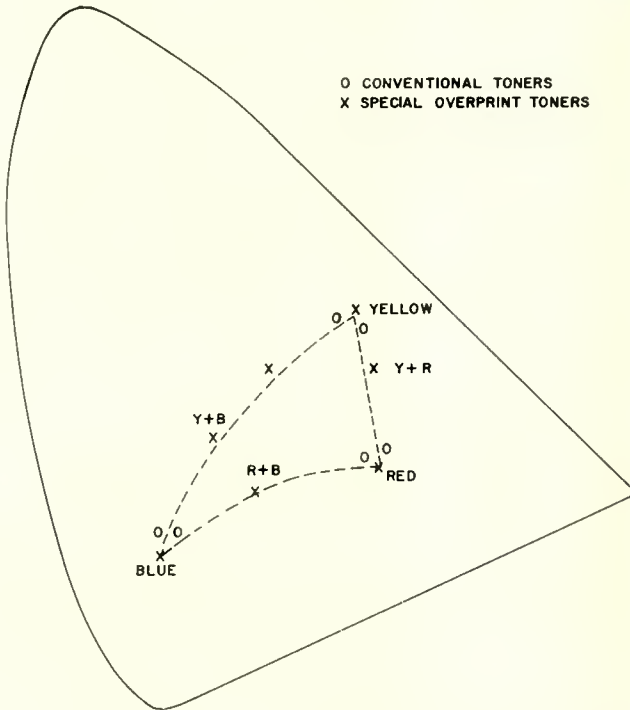


Fig. 9—Chromaticity diagram illustrating overprint effects.

result in mixtures of the color primaries. Prints made with these toners exhibit a remarkable improvement over those made with previous materials.

As illustrated in Figure 5, the overprint areas are those indicated as CM, CY, MY, and CMY. In the case of images produced by a color printing process, the ink layers overlap in these areas; hence, a true overprint results. In the case of printing by Electrofax techniques, the particles in these so-called overprint areas actually mix rather than overlay. Thus, the color resulting from the overlapping of the individual dots in a halftone Electrofax process is due to the mixing which takes place.

Several factors affect the degree of mixing which occurs in the overprint areas. The electrostatic characteristics which have been discussed in this section determine whether succeeding layers of toner material will print over the initial layer. The fusing operation determines the degree of mixing which takes place. This is due to the heat applied during the fusing operation which not only fixes the image, but also causes a melting, hence, particle mixing.

In addition to the mixing which takes place, an additional effect has been observed. The dye contained in the toner particles flows due to the heat of fusing. This results in the penetration of the Electrofax surface by the dye, resulting in further mixing in the overprint areas.

Materials and techniques have been developed which enable the control of these combined effects. Complete color mixing has been observed in prints which have been fused under controlled temperature conditions.

CONCLUSION

The basic colorimetric considerations of a newly developed color Electrofax process are presented. It is shown that laws describing the behavior of pigment mixtures are applicable to describing the factors affecting Electrofax prints. A potential new method of obtaining color separation is presented. The effects of printing by an Electrofax half-tone process are analyzed.

Prints which have been produced utilizing the techniques described in this paper are similar in appearance to a color photograph. To date, the major effort has been in the area of optimizing electrostatic characteristics of the particles with their printing characteristics. It is anticipated that considerable improvement will result from further considerations of the colorimetric properties of the color Electrofax process, as well as in the application of new process techniques.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to C. J. Young, H. G. Greig, and J. Amick, who developed the basic materials for the color Electrofax process, and who have provided valuable assistance in the analysis of the experimental data.

Thanks are also due to J. P. Lauriello and members of the M. I. T. Color Measurement Laboratory who assisted in the obtaining of spectrophotometric data.

APPENDIX—DERIVATION OF THE KUBELKA—MUNK EQUATIONS

For many practical applications, rather than consider in detail the absorption encountered by individual light rays in undergoing complex

paths due to Fresnel reflection and light scattering from opaque surfaces, the net effect can be determined by assuming essentially perfect diffusion of the light flux in every direction.

Referring to Figure 10, assume two diffuse light fluxes, a downward-proceeding flux, I , and an upward-proceeding flux, J . Consider an elementary layer of thickness dx . The thickness dx is small compared to the total thickness X , but large compared to the diameters of the Electrofax particles.

If K is the coefficient of absorption and S the coefficient of scattering, then I will be decreased by a reversal of direction by an amount $SIdx$.

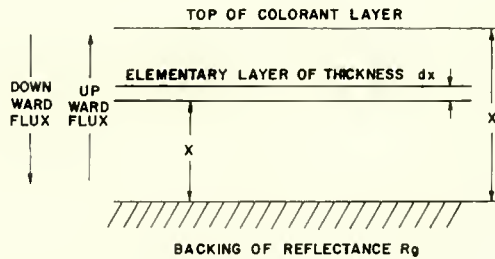


Fig. 10—Derivation of pigment mixture laws.

The flux J , during its passage through the elementary layer dx , is reduced by the amounts $KJdx$ and $SJdx$. $SIdx$ reversed in direction by scattering from I is added to J . Thus

$$dJ = -(S + K) Jdx + SIdx, \quad (15)$$

$$dI = -(S + K) Idx + SJdx. \quad (16)$$

Let

$$a = \frac{S + K}{S}.$$

Then, Equations (15) and (16) can be written

$$-\frac{dI}{dx} = -aI + J, \quad (17)$$

$$\frac{dJ}{dx} = -aJ + I. \quad (18)$$

In determinant form, these are written

$$0 = \begin{vmatrix} \left(-\frac{\Delta}{S} + a \right) & -1 & I \\ -1 & \left(\frac{\Delta}{S} + a \right) & J \end{vmatrix} \quad (19)$$

for which solutions exist in the form

$$I = \sum c_n x \exp \{r_n\}, \quad (20)$$

and

$$J = \sum d_n x \exp \{r_n\}, \quad (21)$$

where the c 's and d 's are arbitrary constants.

Solution of Equation (19) results in

$$I = c_1 x \exp \{s(a^2 - 1)^{1/2}\} + c_2 x \exp \{-s(a^2 - 1)^{1/2}\}, \quad (22)$$

$$J = d_1 x \exp \{s(a^2 - 1)^{1/2}\} + d_2 x \exp \{-s(a^2 - 1)^{1/2}\}. \quad (23)$$

Substituting Equations (22) and (23) in (17) and (18) results in the following relationship between the c 's and d 's:

$$c_1 = d_1 (a + [a^2 - 1]^{1/2}), \quad (24)$$

$$c_2 = d_2 (a - [a^2 - 1]^{1/2}). \quad (25)$$

Consider the case for an infinite layer of colorant material. Let

$$R_x = \left(\frac{I}{J} \right)_{x \rightarrow \infty}. \quad (26)$$

Integrating Equations (17) and (18) from $x = 0$ to $x = \infty$ results in

$$I = \frac{c_2}{s(a^2 - 1)}, \quad (27)$$

$$J = \frac{d_2}{s(a^2 - 1)}. \quad (28)$$

Substituting Equations (25) and (26) into (21) and imposing the conditions of the coefficients from Equation (23) results in

$$R_{\infty} = [a - (a^2 - 1)^{1/2}]. \quad (29)$$

Since

$$a = \frac{S + K}{S},$$

then

$$R_{\infty} = 1 + \frac{K}{S} - \left[\frac{K^2}{S^2} + \frac{2K}{S} \right]^{1/2}, \quad (30)$$

which can be written

$$(1 - R_{\infty})^2 = \left[\left(\frac{K^2}{S^2} + \frac{2K}{S} \right)^{1/2} - \frac{K}{S} \right]^2, \quad (31)$$

which can be transposed to the more familiar form

$$\frac{(1 - R_{\infty})^2}{2R_{\infty}} = \frac{K}{S}. \quad (32)$$

It should be noted that R_{∞} and K/S are functions of wavelength. The scattering coefficient and the absorption coefficient produced by a mixture of components is merely the sum of the coefficients of the constituents.⁶ This simple additivity of the coefficients is what makes the Kubelka-Munk analysis applicable to colorant-layer formulation. Equation (32) can be written for a mixture of particles in the form

$$\frac{(1 - R_{\infty})^2}{2R_{\infty}} = \sum_n c_n \frac{K}{S}, \quad (33)$$

where c_n refers to the proportional amount of the mixture constituents. Equation (33) is a useful expression since R_{∞} of the mixture components can be found directly by measurement. Knowing R_{∞} , one can determine K/S . A summation of the K/S ratios weighted according to the concentrations then enables one to predict the spectral distribution curves for the mixture. Equation (33) is similar to Equation (12), and it is this equation which was utilized in the analysis of data presented in Figure 4.

⁶ B. Persoz, "La détermination des couleurs," *Peintures, pigments, vernis*, Vol. 21, p. 329, 1945.

RCA TECHNICAL PAPERS†

Second Quarter, 1958

Any request for copies of papers listed herein should be
addressed to the publication to which credited.

"Analog Solution of Space-Charge Regions in Semiconductors," L. J. Giacoletto, <i>Proc. I.R.E.</i> (June)	1958
"Application of Transistors in Communications Equipment," D. D. Holmes, <i>Proc. I.R.E.</i> (June)	1958
"The Application of Very Precise Frequency Control to Minimize Television Cochannel Interference," Part I, W. C. Morrison, <i>Broadcast News</i> (April)	1958
"Commercial Airborne Weather Radar," A. W. Vose and F. V. Wilson, <i>RCA Review</i> (June)	1958
"Common Emitter Transistor Amplifiers," D. F. Dion, <i>Proc. I.R.E.</i> (May) (Letter to the Editor)	1958
"Controlled Thermonuclear Fusion—Its Meaning to the Radio and Electronic Engineer," E. W. Herold, <i>RCA Review</i> (June) ..	1958
"Controlled Thermonuclear Fusion—Promise of the Future," G. Warfield, <i>RCA Review</i> (June)	1958
"Design Considerations for Transistor Reflex Receivers," R. V. Fournier, <i>Semiconductor Products</i> (May/June)	1958
"Designing Transistor A-F Power Amplifiers," M. B. Herscher, <i>Electronics</i> (April 11)	1958
"Dry Cells Containing Various Aromatic Nitro Compounds as Cathode Materials," C. K. Morehouse and R. Glicksman, <i>Jour. Electrochem. Soc.</i> (June)	1958
"Effect of Collector Potential on the Efficiency of Traveling-Wave Tubes," H. J. Wolkstein, <i>RCA Review</i> (June)	1958
"Effect of Electron Lenses on Beam Noise," R. C. Knechtli, <i>Trans. I.R.E. PGED</i> (April)	1958
"Fix Your Scope," R. Samuel, <i>Radio-Electronics</i> (April)	1958
"A Gallium Arsenide Microwave Diode," D. A. Jenny, <i>Proc. I.R.E.</i> (April)	1958
"Graphical Method Simplifies Wien Bridge Oscillator Design," E. D. Grim, <i>Electronic Design</i> (May 28)	1958
"Ground Weather Radar for Broadcast Stations," R. C. Strang, <i>Broadcast News</i> (April)	1958
"High Electric Field Effects in n-Indium Antimonide," M. Glicksman and M. C. Steele, <i>Phys. Rev.</i> (June 1) (Letter to the Editor)	1958
"How the RCA Video Tape Recorder Works," J. L. Grever, <i>Broadcast News</i> (April)	1958
"How to Design the Combination Control Room-Announce Booth to Give Good Acoustical Performance," J. F. Palmquist, <i>Broadcast News</i> (April)	1958
"Investigation of the Electrochemical Properties of Organic Compounds. I—Aromatic Nitro Compounds," R. Glicksman and C. K. Morehouse, <i>Jour. Electrochem. Soc.</i> (June)	1958
"A Low-Cost Hi-Fi Amplifier," L. Kaplan, <i>Radio & TV News</i> (May)	1958
"On the Measurement of Transit Time Dispersion in Multiplier Phototubes," M. H. Greenblatt, <i>Trans. I.R.E. PGNS</i> (June)	1958

† Report all corrections or additions to *RCA Review*, RCA Laboratories, Princeton, N. J.

- "Minimize Local Oscillator Drift," Part 1, W. Y. Pan and D. J. Carlson, *Electronic Design* (June 25) 1958
- "The New BK-10A Bigradient Uniaxial Microphone," J. W. O'Neill, *Broadcast News* (April) 1958
- "New 500 and 1,000 Watt AM Transmitters," J. Novik, *Broadcast News* (April) 1958
- "A New High-Transconductance Electron Gun for Kinescopes," J. W. Schwartz, *RCA Review* (June) 1958
- "New 2-KW VHF Television Transmitter Type TT-2BH," H. E. Small, *Broadcast News* (April) 1958
- "Photoconductivity of Zinc Selenide Crystals and a Correlation of Donor and Acceptor Levels in II-VI Photoconductors," R. H. Bube and E. L. Lind, *Phys. Rev.* (June 1) 1958
- "Power Transistor Test Set," W. Hasenberg, *Electronic Industries* (May) 1958
- "The Preparation of Semiconductor Devices by Lapping and Diffusion Techniques," H. Nelson, *Proc. I.R.E.* (June) 1958
- "Progressive Steps Toward Automation in Television Programming," A. H. Lind, *Broadcast News* (April) 1958
- "Propagation Characteristics of Slow-Wave Structures Derived from Coupled Resonators," E. Belohoubek, *RCA Review* (June) ... 1958
- "A Proposed 'Automatic' Teaching Device," P. K. Weimer, *Trans. I.R.E. PGE* (June) 1958
- "Quick Design of Thermistor Compensation Networks," B. R. Schwartz, *Electronic Design* (April 30) 1958
- "Re-Invention by Young Engineers," E. W. Herold, *Proc. I.R.E.* (May) (Letter to the Editor) 1958
- "Reliability and Economics of Telecommunication Systems," Z. Prihar, *Trans. I.R.E. PGCS* (June) 1958
- "Science and Technology in a Changing Environment," E. W. Engstrom, *Elec. Eng.* (April) 1958
- "A Simplified Approach to Optimum Design of Pin Keys," Jack Paull, *Machine Design* (April) 1958
- "The Sixth Scintillation Counter Symposium," G. A. Morton, *Physics Today* (May) 1958
- "A Solid-State Amplifying Fluoroscope Screen," B. Kazan, *American Journal of Roentgenology Radium Therapy and Nuclear Medicine* (April) 1958
- "Specifying Military Thermistors," B. R. Schwartz, *Electronic Design* (April 30) 1958
- "Spectrographic Porous Cup Technique for the Analysis of Indium Alloys," H. M. Hyman, *Applied Spectroscopy* 1958
- "The Status of Transistor Research in Compound Semiconductors," D. A. Jenny, *Proc. I.R.E.* (June) 1958
- "Thermionic Energy Converter," K. G. Hernqvist, M. Kanefsky, and F. H. Norman, *RCA Review* (June) 1958
- "Tracing Distortion in Stereophonic Disc Recording," M. S. Corrington and T. Murakami, *RCA Review* (June) 1958
- "A Transistorized Grid-Dip Meter," C. A. West, *RCA Ham Tips*, Part I (April), Part II (June) 1958
- "The Transition from Engineer to Supervisor," H. M. Elliott, *Trans. I.R.E. PGEM* (June) 1958
- "Universal Coaxial Transmission Line," W. N. Mcule, *Broadcast News* (April) 1958
- "The Use of Vertical Polarization to Solve UHF Television 'Ghosting' Problems in a Shadowed Valley," D. W. Peterson, *RCA Review* (June) 1958
- "Very Low-Noise Traveling-Wave Amplifier," E. W. Kinaman and M. Magid, *Proc. I.R.E.* (May) 1958
- "WFII-TV and WRCV-TV Multiple Antenna System," *Broadcast News* (April) 1958
- "Worksheet Gives Optimum Conditions," C. H. Li, *Chemical Engineering* (April) 1958

AUTHORS



JOHN B. ATWOOD attended Washington and Jefferson College for one year and then transferred to Cornell University, where he received the M.E. degree in 1930 and the M.S.E. degree in 1937. He was a member of the technical staff of the Bell Telephone Laboratories, Inc. from 1930 to 1932. In 1937, he joined the Research and Advanced Development Group of RCA Communications, Inc. and transferred to the RCA Laboratories Division upon its formation in 1942. Mr. Atwood is a Senior Member of the Institute of Radio Engineers, and a member of the American Association for the Advancement of Science and Tau Beta Pi.

K. K. N. CHANG (see *RCA Review*, Vol. XIX, No. 1, March 1958, p. 86.)

ROBERT C. DEMMY received the B.S. degree in Physics from Franklin and Marshall College in 1951. Before entering college, he served for three years with the U. S. Marine Corps in Field Communications work. He was employed by the Hamilton Watch Company from 1951 to 1954 in their watch research group. He joined the RCA Electron Tube Division at Lancaster, Pa. in 1954, where he is engaged in product design and development of color kinescopes.



S. W. JOHNSON attended Illinois State Teachers College and the National Institute of Electronics. He was employed by the Leich Electric Co. of Genoa, Ill., from 1939 to 1945, during which time he worked as detailer and designer in the drafting department, and later as assistant engineer and design engineer on telephone equipment. He joined the Advanced Development Section of RCA in Camden, N. J. in 1945, where he has been associated with a variety of projects. Mr. Johnson is a member of S.M.P.T.E., George Eastman House Associates, and S.P.S.E.

J. E. LINDSAY received the B.S. degree in Engineering Physics in 1947 from the University of Colorado and the M.S. degree from the University of Illinois in 1948. During 1949-1950 he was an Assistant in Radiology at the Medical College of the University of Illinois in Chicago. He taught electronics at the Officer Training School, Sectt Air Force Base, Belleville, Illinois in 1951 and 1952. In 1952 he joined the RCA Service Company, teaching principles of radar at the Signal School, Fort Monmouth, New Jersey. In 1953 he joined the sound, optics and special engineering section, where he engaged in transistor circuit development. In 1958 he became associated with the Cornell Aeronautical Laboratory of Buffalo, N. Y. Mr. Lindsay is a member of Sigma Pi Sigma, Gamma Alpha and the Institute of Radio Engineers.





G. B. MACKIMMIE received the B. Eng. (Communications) degree from McGill University in 1943. He then spent two years with the Royal Canadian Navy as an electrical lieutenant. He joined RCA Victor Co., Ltd., as an engineer in 1945, where he was engaged for several years in the field of antenna design. During the past two years, he has been manager of the Broadcast Engineering group working on antenna design in the Engineering Department of the Technical Products Division. Mr. MacKimmie is a member of the Institute of Radio Engineers and the Corporation of Professional Engineers of the Province of Quebec.

ALBERT M. MORRELL received the B.S. degree in Electrical Engineering from Iowa State College in 1950, and is currently doing graduate work at Franklin and Marshall College. He served with the U. S. Army Air Force for three years before entering college. He joined the RCA Electron Tube Division in Lancaster, Pa. in 1950 as a design engineer on color kinescopes. At present, he is engineering leader of the color kinescope design unit. Mr. Morrell is a member of Phi Kappa Phi, Eta Kappa Nu, and the Institute of Radio Engineers.



T. MURAKAMI (see *RCA Review*, Vol. XIX, No. 2, June 1958, p. 216.)



FRITZ PASCHKE received the degree of Diplom-Ingenieur in 1953 and the Dr. techn. sc. in 1955 from the Technical University Vienna. Between 1953 and 1955 he was a research assistant at the Institute of High-Frequency Techniques in Vienna. Since April 1956, Dr. Paschke has been a member of the technical staff at RCA Laboratories, Princeton, N. J.

ARNOLD S. ROSE received the B.S. degree in 1935 and the M.S. degree in Chemical Engineering in 1939 from New York University. He has worked in various phases of metallurgical research; from 1939 to 1940 at the National Bureau of Standards, from 1940 to 1942 at the William's Gold Refining Company, and from 1942 to 1950 at the Lancaster plant of the RCA Electron Tube Division. In 1950, he joined the Special Products Division of the I-T-E Circuit Breaker Company as Supervisor of the Research and Development Laboratory. In 1955 he returned to RCA, joining the Semiconductor Division. Since 1957 he has been Manager of the Materials Service Laboratory in Somerville. Mr. Rose is a member of the American Society for Metals and the American Institute of Mining and Metallurgical Engineers.





FRED D. ROSI received the B.E. degree in Metallurgy from Yale University in 1942, the M.E. degree in 1947, and Ph.D. degree in 1949 from the same institution. From 1943-46 he served on the minesweeper, *USS Sprig*, as Engineering Officer in the Atlantic and Pacific theatres. From 1949 to 1954 he was an Engineering Specialist in the theoretical metallurgy group at Sylvania Electric Products, Inc., where he was engaged in fundamental research on the problems of plastic flow and recrystallization in various metals and alloys. Since March 1954 he has been with RCA Laboratories, where he is now

head of the Semiconductor and Electrochemical Research Group engaged in materials research on compound semiconductors for transistor and thermoelectric device applications. He is a member of the American Institute of Metallurgical Engineers and Sigma Xi. Dr. Rosi also holds the rank of Adjunct Professor at New York University, where he lectures in graduate courses on Crystal Mechanics and Advanced Physical Metallurgy.

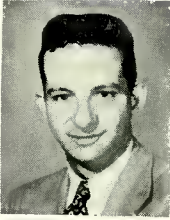
JOHN S. RYDZ received the B.S. degree in Physics from the Massachusetts Institute of Technology in 1952. From 1943 to 1946 he served as an electronic technician in the U. S. Navy. From 1946 to 1949 he worked in the field of industrial instrumentation for chemical process measurement and control with Merck & Co. While at M.I.T. he was a student engineer for Raytheon. He joined RCA in 1952 as a member of the Industrial Electronic Products activity at Camden, N. J. In 1956, he received the M.S. degree in Physics from the University of Pennsylvania. At present he is the Engineering Group Leader of the Applied Physics activity of RCA Defense Electronic Products. Mr. Rydz is a member of the Optical Society of America, the Institute of Radio Engineers, and a committee member to the Inter Society Color Council for the Technical Association of the Graphic Arts.



DON G. SHIPLEY graduated from RCA Institutes Inc. in 1952 after completing the Advanced Technology Course. He then joined the staff of the Radio Reception Laboratories, RCA Laboratories Division, Riverhead, Long Island, N. Y., where he has worked primarily on propagation programs. Mr. Shipley is an Associate Member of the Institute of Radio Engineers.

C. P. SMITH received the B.S. degree in Electrical Engineering from the University of Missouri in 1942, and joined the Lancaster Plant of the RCA Electron Tube Division the same year as a production engineer working on small power and gas tubes. In 1945, he became a design engineer responsible for work on the 6.4-ampere xenon thyratron. He was appointed Manager of the Tube Development Shop in 1948, and became Manager of Color Kinescope Engineering in 1954. At present, he is Manager of Engineering of the combined black-and-white and color kinescope operations. Mr. Smith is a member of the American Association for the Advancement of Science and a Senior Member of the Institute of Radio Engineers.





RICHARD W. SONNENFELDT received the B.S. degree in Electrical Engineering in 1949 at Johns Hopkins University. From 1941-43 and 1946-49 he designed control circuits and systems for the Charles Electric Company of Baltimore. In 1949 he joined RCA, where he was engaged in advanced development work on monochrome and color television receivers. He is now with Advanced Development, Industrial Electronic Products. Mr. Sonnenfeldt is a member of Tau Beta Pi, Omicron Delta Kappa, and a Senior Member of the Institute of Radio Engineers.

GILBERT S. WICKIZER received the B.S. degree in electrical engineering from the Pennsylvania State College in 1926, and after graduation he was employed by the Radio Corporation of America in the operating division at Rocky Point, N. Y. Since 1927, he has been with RCA at Riverhead, N. Y., engaged in communications research on problems of reception, including propagation studies in the VHF and UHF regions and development work on direction-finding, frequency measurement, receiving antennas, and field-strength measurement. Mr. Wickizer is a Senior Member of the Institute of Radio Engineers, and a member of Eta Kappa Nu, the American Meteorological Society, and an associate member of the Arctic Institute of North America.



H. J. WOLL received the B.S. degree in Electrical Engineering from North Dakota State College in 1940. He was a graduate assistant at the Illinois Institute of Technology in 1940-1941. He was the first recipient of an RCA employee fellowship in 1947, receiving the Ph.D. from the University of Pennsylvania in 1953. He has been engaged in advanced development in Indianapolis and in Camden since joining RCA in 1941. He is presently Manager of Development Engineering, Defense Electronic Products. Dr. Woll is a member of the Institute of Radio Engineers, the American Institute of Electrical Engineers, and Phi Kappa Phi.

