RCA REVIEW

a technical journal

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

VOLUME IX SEPTEMBER 1948 NO. 3

RCA REVIEW

GEORGE M. K. BAKER
Manager

CHAS. C. FOSTER, JR.
Business Manager

SUBSCRIPTIONS:

United States, Canada and Postal Union: One Year $2.00, Two Years $3.50, Three Years $4.50
Single Copies: 75¢ each

Other Countries: One Year at $2.40, Two Years $4.30, Three Years $5.70
Single Copies: 85¢ each

Published quarterly in March, June, September, and December by Radio Corporation of America, RCA Laboratories Division, 30 Rockefeller Plaza, New York 20, N. Y.

Editorial and General Offices: RCA REVIEW, Radio Corporation of America, RCA Laboratories Division, Princeton, New Jersey

Entered as second class matter April 3, 1946, at the Post Office at New York, New York, under the act of March 3, 1879

RADIO CORPORATION OF AMERICA
DAVID SARNOFF, *President*

LEWIS MACCONNACH, *Secretary*

ARTHUR B. TUTTLE, *Treasurer*

PRINTED IN U.S.A.

RCA REVIEW

a technical journal

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

Published quarterly by
RADIO CORPORATION OF AMERICA
RCA LABORATORIES DIVISION
in cooperation with

RCA VICTOR DIVISION
RADIOMARINE CORPORATION OF AMERICA
RCA INTERNATIONAL DIVISION
RCA COMMUNICATIONS, INC.
NATIONAL BROADCASTING COMPANY, INC.
RCA INSTITUTES, INC.

VOLUME IX — SEPTEMBER 1948 — NUMBER 3

CONTENTS

PAGE

Automatic Gain Controls for Television Receivers 373
K. R. WENDT AND A. C. SCHROEDER

Details of the Simultaneous Equation Solver 394
E. A. GOLDBERG

Some Notes on Noise Theory and its Application to Input Circuit Design 406
W. A. HARRIS

The Brightness Intensifier 419
G. A. MORTON, J. E. RUEDY AND G. L. KRIEGER

Theoretical Analysis of Various Systems of Multiplex Transmission (Concluding Part) 433
V. D. LANDON

The Pack Transmitter 483
J. L. HATHAWAY AND W. HOTINE

Electro-Optical Characteristics of Television Systems; Part III—Electro-Optical Characteristics of Camera Systems 490
O. H. SCHADE

Some Applications of Frequency-Modulated Radar 531
I. WOLFF AND D. G. C. LUCK

RCA Technical Papers 556

Authors 558

RCA REVIEW

AUTOMATIC GAIN CONTROLS FOR TELEVISION RECEIVERS*

BY

K. R. WENDT AND A. C. SCHROEDER

Research Department, RCA Laboratories Division,
Princeton, N. Y.

Part I

GENERAL CONSIDERATIONS

BY

K. R. WENDT AND A. C. SCHROEDER

Summary—The general theory of automatic gain controls for television receivers is discussed. Several specific circuits are described in detail with the advantages and disadvantages of each.

INTRODUCTION

AUTOMATIC gain control (AGC) is as important in television receivers as in sound receivers, and actually serves more useful purposes in television than in sound. Manual gain adjustments in a television receiver are annoying, and a non-technical person experiences difficulty in learning to set the control properly, since the optimum level for limiting and sync separation is not easily judged from the picture contrast. Also, television signals may suffer from violent fading due to passing airplanes. This can be reduced or eliminated only by a fast AGC circuit. The use of an AGC not only provides easier adjustment, but also may allow the simplification of portions of the receiver, such as the sync separator, which would not be required to operate over wide ranges of amplitude. A properly designed and operating AGC makes the setting of the contrast and background controls simple and infrequent. AGC has, however, been little used because its design has not been sufficiently understood, and the early circuits have not performed satisfactorily.

In a sound receiver for amplitude modulation, the signal which is measured and which is held constant by the AGC, or AVC as it is called,

---

* Decimal Classification: R583.5

is the average carrier level. This is easily measured because the dc output of the detector is proportional to the average carrier level. In a television receiver it is the peak carrier level which must be held constant by the AGC. This peak carrier level may be obtained by measuring the voltage of the peaks of the synchronizing pulses at the output of the detector, provided that the load of the detector has the same dc as video-frequency impedance. The output of the measuring device is then fed back to the intermediate-frequency amplifier in such a way as to decrease the gain as the signal increases. Amplification of the AGC signal may be obtained by amplifying the signal before peak measurement or amplifying the dc output of the peak measuring device, or both. In either case, the amplification must be dc, and in the former case it must also amplify video frequency. The AGC measuring device consists of some detector, such as a diode feeding a capacitor. Since this circuit receives information only during the sync pulse, which is 8 per cent of the time, the capacitor must hold its charge between pulses.

NOISE CONSIDERATIONS

A simple peak detector, when used for television AGC, is quite susceptible to peaks of impulse noise. The noise is predominantly in the black or increasing signal direction, and may be quite high as compared with the signal. The peak detector then measures the noise height rather than the signal height, and may reduce the gain of the receiver to a small value, giving very unsatisfactory performance under noise conditions which may be encountered quite often in outlying areas. A poor AGC circuit may render unsatisfactory a picture which, with a good AGC or manual control, would contain nothing more than nearly unnoticeable short black streaks. Many devices, such as automobiles, buzzers, electric shavers, etc., produce such interference. It is therefore important that AGC circuits be designed with noise immunity as a primary consideration.

AGC SPEED

After a simple peak detector has responded to a noise peak, it can return to normal only as fast as the capacitor may be discharged through its associated resistor. The capacitor may be charged much more quickly than it is discharged, since it is charged through the low impedance video circuit and diode. The AGC thus normally is quicker in responding to than in recovering from a noise peak, the effect of noise peaks being thereby greatly extended.

It is desirable to have the AGC quite fast in order to have it follow fast fading, and to minimize the effect of noise. However, the vertical sync signal interferes with fast operation. If the AGC is made fast, it removes a considerable portion of the vertical sync signal and produces a transient following the vertical signal. The mechanism, briefly, is as follows: The detector is not a true peak device. It must operate with a certain area (voltage $\times$ time) of the sync extending into the conduction region of the diode. The horizontal pulses are 8 per cent of a line interval in width. The vertical pulses, however, are 84 per cent of a line interval in width, and for the same height pulses approximately ten times as much current will be drawn during the vertical pulse. To the AGC circuit this appears as an increased signal. In a fast circuit the intermediate-frequency gain is therefore reduced during the vertical until the average diode current approximates that drawn during the horizontals. The vertical signal is thus reduced, or "pushed into a hole" by an amount approximately equal to the height that the horizontal pulses extended into the conduction region. It might appear advisable to reduce this conduction area by increasing the diode resistor so that it would require less power. Such operation is impractical, however, due to noise. A small amount of noise power is then able to take the control away from the signal. Nor is a limiter of much help with such a peak-operated detector. If the limited noise extends beyond the sync—and it must, in order to keep the limiter from destroying the information as to the height of the signal—the gain of the receiver will be reduced until the noise no longer fills the limiter. If the received noise-to-signal ratio were 10 to 1, and this can easily be, the signal would be reduced to approximately 1/10 the desired value by the AGC. For noise immunity, therefore, the simple AGC detector must be of relatively low impedance, and be energy operated, thereby using large areas of the sync signal. This also means that the circuit can not be made fast, due to the consequent loss of the vertical sync signal.

Early AGC Circuits

The receivers used in the field tests conducted in the early 1930's contained a picture AGC circuit. It consisted of a diode peak detector coupled directly to the video output of the second detector. The output of this peak detector was amplified by a triode amplifier and applied to the intermediate-frequency amplifier grids. The proper dc voltages were obtained by operating the detector and the cathode of the dc amplifier at about —50 volts and the plate of the amplifier at —3 volts.

Figure 1 shows the AGC circuit in the field test sets. T_1 is the detector, T_2 is the peak detector, and T_3 is the dc amplifier.

The first TRK-12 had a similar AGC except that the peak detector

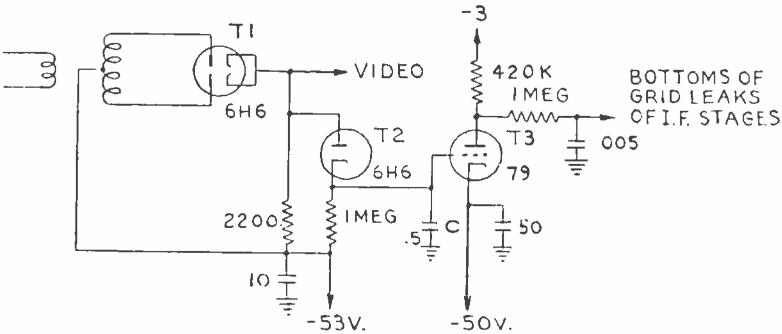


Fig. 1—Field test AGC (early 1930's).

was a triode in a cathode follower type circuit. Figure 2 shows the AGC circuit in the TRK-12 where T_1 is the detector, T_2 is the peak detector, and T_3 is the dc amplifier. It was found necessary to remove this AGC from these sets, partly to make room for a limiter which, under noisy conditions, was more necessary than an AGC, and partly because high peak noise caused the AGC to operate unsatisfactorily, as explained previously. A peak of noise would cause the gain to go down, and due to the long time constants in the circuits, appreciable time was required for the receiver to return to normal operation. This caused the picture to flash very badly in the presence of high peak noise, which would otherwise have been almost unnoticeable.

In both the field test set and the TRK-12 a high peak of noise charged C from a low impedance to nearly the height of the pulse; C then had to discharge through the high impedance of the cathode circuit of the peak detector.

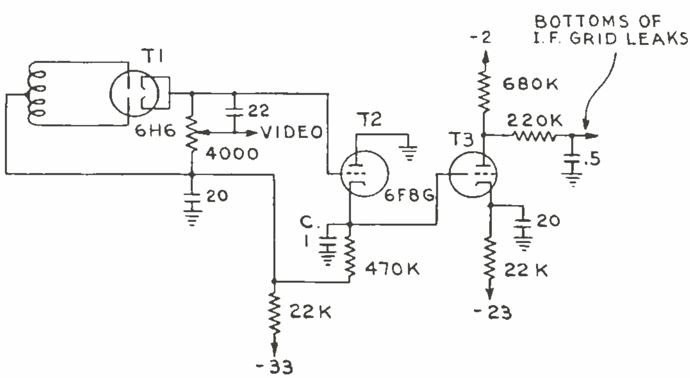


Fig. 2—TRK-12 AGC.

CIRCUITS WITH IMPROVED NOISE IMMUNITY

A satisfactory AGC must be immune to high peak noise. A limiter is of considerable help in this respect, if used with an energy-operated AGC detector. If the limiter cuts the noise down nearly to the level of sync and the noise is narrower than the sync, which it nearly always is, a detector which works on the energy in the sync peaks will be quite immune to the noise. It must be emphasized here that such a detector much be energy operated and use an appreciable area of the sync pulses.

Figure 3 shows a simple AGC circuit which has a limiter before the AGC detector and which operates quite satisfactorily. T_1 is the detector, T_2 is the limiter, T_3 is the AGC detector, and T_4 is a glow tube which brings the dc voltage down to the proper value to operate the intermediate-frequency grids. In operation, the limiter T_2 prevents

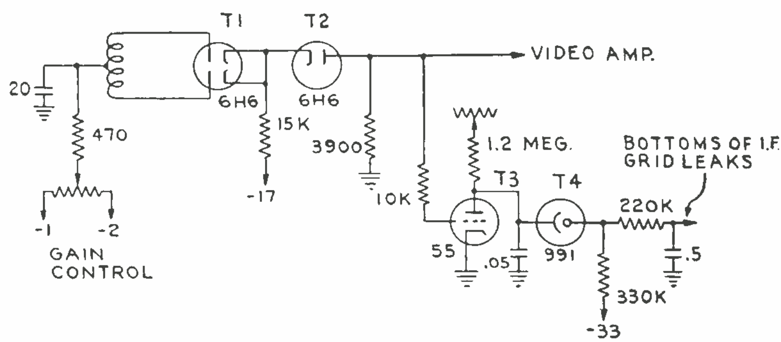


Fig. 3—Simple AGC with limiter.

currents flowing through the 3900-ohm load resistor of the detector if the detected voltage is higher than ground. If the AGC holds the peaks of sync at approximately ground, the gain can be varied if desired by changing the output voltage of the detector for zero carrier. This is accomplished by varying the potentiometer marked gain control. The signal after limiting is applied to the grid of T_3, which is cut off except when the signal approaches ground. In other words, T_3 conducts only if the peaks of sync are near ground potential. The peaks of sync then cause pulses of plate current to flow, which causes the dc voltage at the plate of T_3 to decrease. The voltage at the supply end of the plate resistor is adjusted so that with no plate current and just the current through the series combination of the 1.2-megohm plate resistor, the 991 glow tube, and the 330,000-ohm biasing resistor, the voltage at the low end of the glow tube (the voltage applied to the grids of the intermediate-frequency tubes) is zero. Then, when plate current flows through T_3, causing the plate voltage to decrease, the voltage fed to the intermediate-frequency grids is also decreased by the same amount, decreasing the intermediate-frequency gain. The glow tube need not be of the selected, voltage regulator type, such as the 991, but can be of the type ordinarily used for illumination.

Another means of improving the noise immunity is by means of a double time constant arrangement.

As may be seen from Figure 4, this circuit uses one diode fed from the last intermediate-frequency transformer. Bias is applied to its cathode to delay the application of AGC bias until the video output is sufficient for full contrast. Two time constants are used and account for the improved performance obtained with this circuit over that obtained from a long-time-constant peak detector as has sometimes been used. R_1C_1 form the first time constant which is relatively fast—of the order of one picture line. R_2C_2 forms the second time constant and is much longer. It should not be less than 1/20 second. Because of the short input time constant a relatively small amount of energy is

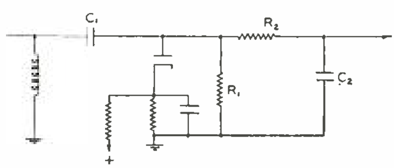


Fig. 4—Double time constant integration.

stored in C_1 and by the end of each line the voltage across it has dropped to approximately the black level at which time C_1 is again charged. R_2 and C_2 filter out both this ac component and the 60-cycle component caused by the vertical synchronizing pulses. In the presence of noise pulses C_1 is charged to approximately the peak amplitude of the pulses, but because its capacitance is low the energy stored is small. The ac component of this energy is removed by the subsequent filter, and the dc component is so small that it affects the AGC output only slightly. To operate in this ideal manner the R_1C_1 time constant should be of the order of four or five lines. However, it has been determined empirically that the AGC voltage is not greatly affected by the picture content if the R_1C_1 time constant is made as short as one line. Because the noise susceptibility of the system decreases as the input time constant is decreased, the shorter time constant is more desirable. This principle can be applied to advantage in many AGC systems.

Simplified Fast Circuit

Due to the desirability of a fast-acting AGC for the elimination of airplane fading, considerable work has been done in this direction. In the foregoing discussion of fast circuits, in which it was shown that the vertical sync signal was removed from the video, no mention was made of the fact that in such a case the vertical sync signal appears on the AGC control voltage. Attempts have been made to use this as the vertical sync signal. Such a vertical is unsatisfactory, however, since its amplitude varies radically with the incoming signal. For instance, a strong signal may provide very little vertical sync, since the intermediate-frequency tubes operate near cutoff, and require very little signal to change their gain appreciably. Also, the AGC voltage

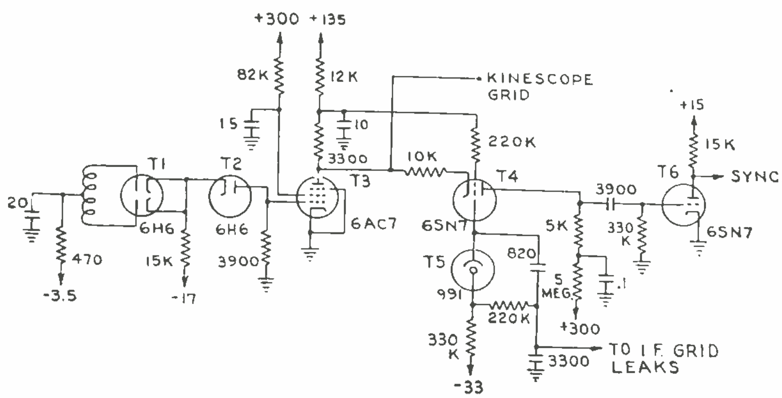


Fig. 5—Combined AGC and sync separator.

contains all the low frequency variations in the signal, such as that from fading caused by airplanes, which makes proper use of the AGC voltage as a vertical sync signal quite difficult.

It has been found possible to combine the operation of the AGC detector and the sync separator in such a manner that a satisfactory sync signal may be obtained with a fast circuit. The video signal still has the vertical "in a hole". However, the separated signal has in effect the AGC signal added to it, as far as vertical sync is concerned, and that which is lost from the video and appears on the AGC is added to the separated signal to produce a normal separated signal.

Figure 5 shows a circuit of this type. Up to the grid of the video amplifier the circuit is identical with the one previously described in Figure 3. In Figure 5, however, T_3 is the video amplifier which feeds the grid of the kinescope. Since the grid of T_3 is dc-connected to the detector, the correct second detector dc information is available from

T_3 by making its plate circuit flat to dc. This is done by simply choosing the ratio of the plate filter resistance to the screen filter resistance such that the ac amplification is equal to the dc amplification, and making the two time constants of the filters equal. The plate of T_3 is dc-connected to the grid of the kinescope so that if the AGC holds the peaks of sync at a given level it will hold not only the gain constant but also the black level on the kinescope constant so that there is no need for an additional dc setter. The AGC which is connected to the plate of T_3 is also the sync separator. The grid and cathode of T_4 act as a peak detector and the detected voltage is dropped to the proper level by the glow tube T_5 in a similar manner to the previous circuit. However, here the time constants are very much faster; in fact the glow tube and series resistance to the intermediate-frequency grids is by-passed for high frequencies by the 820-micromicrofarad condenser. The plate of T_4 is kept low with respect to its cathode by the large series resistance to +300 volts (5 megohms) so that it acts as an effective clipper. This allows only sync pulses to appear across the plate resistor and after further clipping by T_6 the sync is fed to the automatic-frequency control (AFC) sync circuits. The sync appearing at the plate of T_6 has full amplitude vertical even though the picture signal on the grid of the kinescope has its vertical "in a hole". This is because the grid of T_4, which is the AGC output, has the signal on it which puts the vertical "in the hole". Therefore the grid to cathode signal of T_4 has not lost the vertical sync.

This circuit, then, is a fast AGC circuit in which the disadvantage of losing vertical sync is overcome. Another incidental advantage is that, since there is no coupling condenser to the grid of the kinescope, capacity to ground on this lead is reduced, allowing greater gain or bandwidth in the amplifier. Also the three functions of sync separation, dc restoration, and AGC, are all obtained with a tube which usually only serves as sync separator.

Keyed AGC Circuits

Another method of obtaining improved AGC performance is by means of the keying principle. A keyed AGC system is one which is turned on, or made sensitive, for only small intervals of time, usually less than 8 per cent. A narrow pulse, occurring at horizontal frequency, is used to key the AGC. The pulse is usually obtained from the local horizontal oscillator, in which case synchronism must be established for proper operation. The pulse may also be obtained from separated sync, although this method is much inferior with respect to noise, since noise peaks become keying signals which measure themselves.

Keyed systems have several fundamental advantages. First, if the pulse is 5 per cent, a theoretical advantage of 20 times in noise immunity is obtained, since for 95 per cent of the time noise can not affect the AGC. Although this full gain is not obtained, since small noise pulses on a white signal do not affect a simple AGC, nevertheless a considerable gain is realized. Second, the vertical sync information is completely eliminated from the AGC, allowing it to be made as fast as desired without impairing vertical synchronization, and, third, since the speed of response may be made fast, the effects of rapid fading are not only reduced, but the receiver quickly recovers from any residual effects due to noise.

The basic balanced type of keyed AGC is shown in Figure 6.[1] Many arrangements are possible, but the simplest, providing proper

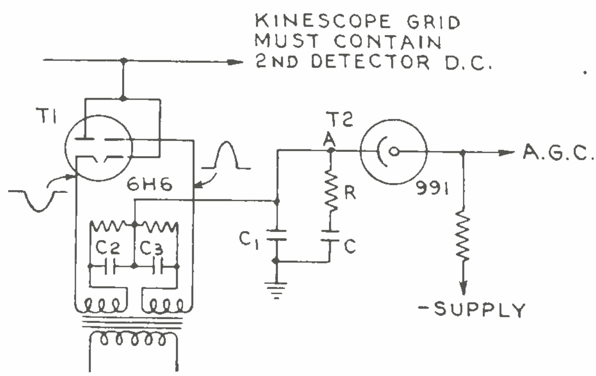


Fig. 6—Balanced keyed AGC detector.

2nd detector picture dc can be maintained to the kinescope, is the arrangement shown, in which the AGC detector operates from the maximum video signal. The keying pulses are supplied by the transformer shown, which may be driven by a tube, supplied with a pulse from the horizontal blocking oscillator, or the horizontal output stage. Alternatively, the two output windings shown may actually be wound on either the horizontal blocking, or output transformer, with the windings connected, of course, such that the pulses are of proper polarity to drive the diodes into conduction. As the pulses drive the diodes into conduction, capacitor C_1 is charged or discharged until point A is brought to the potential of the video circuit during the time of the pulse. The resistance-capacitance circuit across C_1 supplies damping and prevents overall oscillation. The dc voltage at A is positive with respect to ground, and must be reduced by means of one or

[1] K. R. Wendt, "Television DC Component", *RCA Review*, Vol. IX, No. 1, pp. 85-111, March, 1948.

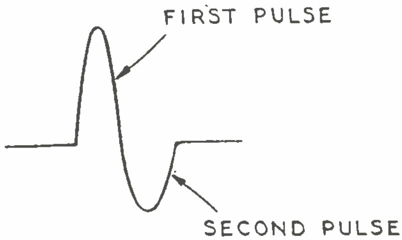


Fig. 7—Oscillator transformer wave form.

more glow tubes, to a negative potential suitable for the AGC control voltage. This AGC can be improved by following capacitor C_1 with a dc amplifier, or a cathode follower, in which case there is no dc load on C_1.

This keying circuit is independent of the height of the keying pulses, as long as they are sufficiently high to maintain the diodes open between pulses for all conditions of video signal.

The keying pulses should be appreciably narrower than the 8 per cent sync pulse in order to allow some variation in the phasing of synchronization. Also, for the same reason, the pulse should occur in the middle of the sync pulse when the synchronizing circuits are operating normally. For a triggered synchronizing system, this may be accomplished by using the first of the two oscillator pulses, as shown in Figure 7. For AFC synchronization, the second pulse may be used. The location of the pulse, with respect to the synchronizing pulse, may be varied in the AFC circuit by changing the delay in the sawtooth signal fed into the AFC, or by changing the total width of the pulse of Figure 7 by adding capacity across the oscillation transformer, or by altering the inductance of the transformer.

This balanced circuit can not be used to key on the "back porch," in order to improve the dc restoration on the kinescope. This is due to the fact that the vertical sync pulse occurs during the time normally occupied by the "back porch." Hence, wrong information would be supplied to the AGC, and the vertical sync pulse would be practically removed, with a severe transient occurring after it. This effect could be made small by making the circuit very slow, but thereby losing, at the same time, one of the advantages of keyed AGC.

Simplified Unbalanced Keyed AGC

In this arrangement a pulse in the black direction is added to the signal, and a normal AGC peak detector is used, or the detector is pulsed so that it "reaches down" to the sync signal only during the pulse. If the pulse is added to the video, the sync separation must be ahead of the AGC since the pulse would interfere with the separator. Such a system has the following advantages:

1. It has considerably increased noise immunity over a plain AGC, especially if a limiter is used on the signal, and the pulse is made large enough to maintain the AGC detector above the clipped noise peaks;

2. It may be made quite fast; and

3. It is somewhat simpler than the balanced type.

It also has these disadvantages:

1. It is very critical to pulse amplitude changes (any change in the pulse amplitude is interpreted the same as a signal amplitude change); and

2. It is faster in one direction than the other—that is, it is faster in the increased signal direction, as is a simple AGC. (This limits its noise immunity, since the effect of a noise pulse is prolonged, and, in order to make it fast enough in the decreased signal direction, it must be made very fast in the increased signal direction, and hence sensitive to noise pulses of very small energy or area that may occur during the keying pulse.)

Effect of Loss of Synchronism

The keyed AGC can not operate satisfactorily when synchronism has been lost. Depending upon the speed of the AGC, the pattern becomes darker or entirely black. This is due to the keying pulse attempting to hold at black the various portions of the picture upon which it falls. However, either sync or the limited edges of blanking are present to act as sync pulses. These tend to synchronize the oscillator frequency in the normal manner, and when the oscillator approaches synchronism or passes through the correct phase, normal conditions are restored by the AGC, and the oscillator can lock in properly. Although the pull-in range may be reduced somewhat by the AGC, synchronizing systems have in practice operated normally and satisfactorily in conjunction with keyed AGC systems.

Some General Considerations

The Video DC Amplifier. Any AGC system for television must maintain the 2nd detector dc information to the AGC detector point. The most economical point for the AGC detector is at the kinescope where the level is the highest. The easiest method of getting the 2nd detector dc to the kinescope is a dc amplifier. Usually, the video tube is a pentode, and some dc is unavoidably lost by the screen grid. As explained previously, it is possible to introduce into the plate circuit a low boost filter which can exactly make up this loss. The two filter time constants should match. However, if the by-pass capacitors are reasonably large, and the screen does not have too high a resistor, compensation becomes quite non-critical. Figure 8 shows such a circuit. A limiter may be installed before the video tube, but care must be taken to insure that the dc signal is really maintained through the limiter. A two-stage video amplifier may also be used, in which the first tube acts as a limiter by being driven beyond cutoff. Such a limiter is not as sharp as a diode, but is satisfactory for operation with AFC and keyed AGC.

Kinescope Operation with Second Detector dc. Operation in this manner is quite different from operation with a dc restorer. First, the screen, in the absence of a signal, is white, with noise showing,

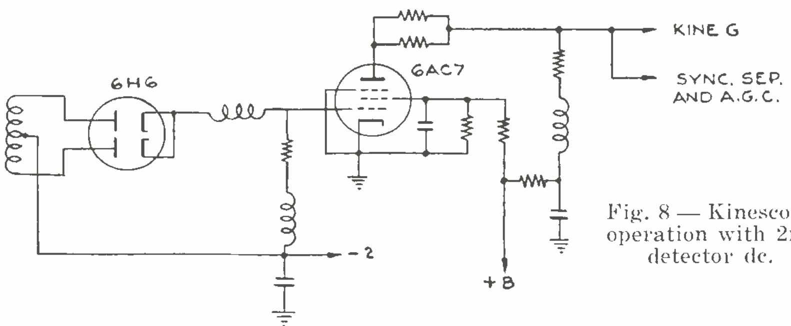


Fig. 8 — Kinescope operation with 2nd detector dc.

which is a definite indication that the station is not on. There is thus little danger of forgetting that the set is turned on after the station has signed off. If an unmodulated carrier is on, the screen is black, which is again a definite indication not obtainable with a separate dc restorer. This information is very helpful, especially if the contrast and brightness controls need not be operated. Its chief advantage is that the dc restoration for the kinescope can be made exceedingly accurate, yet no tube or equipment is required other than the AGC.

Gain and Filtering in AGC Circuits

In a sense, all television AGC circuits are keyed. In other words, information as to the output voltage is received only during short intervals of time. The control, of course, must be made to apply continuously.

The control voltage must therefore be integrated, or stored. This involves filtering and delay. If the ac gain of the system is too high it will oscillate. That is, the control voltage may be larger than needed, which will not be measured until the next pulse, at which time an opposite correcting voltage will be received, and the system will therefore oscillate. The gain of the system must be as low as possible for frequencies higher than approximately ⅓ the keying rate, or line frequency. The gain below this frequency may be allowed to rise as rapidly as possible; delays should be avoided.

The dc gain may be quite high. The flatness of the AGC is determined by the dc gain. If the dc gain is too high, however, any low frequency delay, such as power supply filtering, may cause low frequency oscillation or motor boating.

In general, for maximum AGC speed, as in all feedback circuits, the filtering should all be in one place, which should be the capacitor receiving the charge from the detecting device. This capacitor should be by-passed by another larger capacitor, in series with a resistor, called the damping resistor, which should be adjusted for maximum stability. All other by-passing should be in the nature of intermediate- or radio-frequency grounding, and should not cause appreciable delay. The dc gain should be adjusted in accordance with the flatness or economy considerations desired.

Part II

A NEW FAST NOISE-IMMUNE TELEVISION AGC CIRCUIT

By

K. R. Wendt

Summary—*A new inverted keyed AGC circuit has been developed, which is fast enough to remove airplane interference, possesses very high noise immunity, and is simple and non-critical. It requires 1 or 1½ tubes, and in addition restores the dc for the kinescope. The dc is automatically restored with the blanking level as the reference. The sync height thus becomes non-critical, and the noise limiter may actually remove a portion of sync without an undesirable result.*

Introduction

As EXPLAINED in Part I, television AGC rectifiers must be of the peak type, and since they operate on such a small area signal, they become excessively sensitive to noise. Usual peak rectifiers have two speeds: the response speed and the recovery speed. The response is the fast speed, and the recovery the slower. The response is caused by an increased signal drawing more rectifier current, and, of course, the larger the increase in the signal, the more the response. The slow speed of the circuit is its own recovery speed, and applies whenever the circuit is not drawing current. A sudden decrease in the signal will be followed only according to a resistance-capacitance time constant and not at the speed of decrease of the signal. Since the noise is usually in the increased signal direction, and the circuit is fast to respond to an increase in signal, the circuit is therefore fast to respond to noise and slow to recover from the effect of noise. Keyed circuits are an improvement in this respect in that they are completely inactive between keying pulses. Therefore noise pulses between the keying pulses have no effect upon the operation of the keyed circuit. Furthermore, keyed circuits are fast in both directions, so that although they may respond quickly to a noise pulse they also recover quickly, thus reducing the effect of noise.

The circuit described herein has been called the "inverted keyed circuit" because the response to signal is inverted over the previous rectifier circuits in that it *responds* to a *decreased* signal rather than

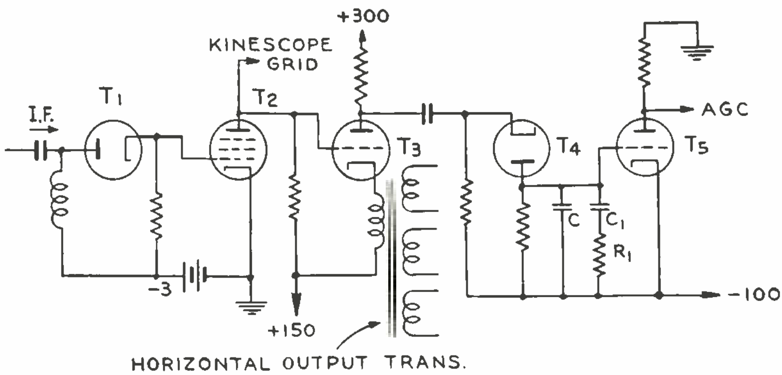


Fig. 9—Simplified schematic of inverted keyed AGC.

an *increased* signal. Furthermore, since it is keyed, no record remains of noise pulses occurring between keying pulses. A noise pulse which occurs during the time of the keying pulse usually appears as a sudden increase of signal. The circuit completely ignores such a sudden increase, and follows only in its slow speed. When the circuit returns to normal, that is, when a normal synchronizing pulse arrives, the circuit responds quickly in this decrease direction and promptly returns to normal operating conditions. The circuit therefore responds slowly to noise and recovers quickly from the effect of noise.

The New Circuit

The circuit which accomplishes this inverted keyed action is shown in Figure 9, in which T_1 and T_2 are respectively the second detector and the video amplifier. Two stages of video could be used, but, in any case, as explained in Part I, it is necessary for the amplifier to be dc-connected from the second detector to the AGC tube. For simplicity, no peaking coils are shown in this diagram. The video signal is connected directly to the grid of T_3. The signal has its sync pulse negative, and is on the order of 50 volts. A negative pulse obtained from a few turns on the horizontal output transformer is applied to the cathode of T_3. This pulse must be of greater amplitude than the video signal, and may be approximately 75 volts. The pulse causes current to flow in T_3, depending upon the actual amplitude of the video signal. A large video signal will be more negative during sync and blanking time, and will cut off T_3 and very little pulse will be produced. A smaller video signal will hold the grid of T_3 more positive and a larger pulse will be produced. This pulse is connected by way of a capacitor through an ac connection only, to T_4, which is a diode for rectifying and obtaining a dc corresponding to the amplitude of the pulse. On the grid of T_5, therefore, a dc exists which is negative for small video signals, and positive for large video signals. T_5 inverts the polarity and supplies an appropriate signal for an AGC voltage. The dc levels require that a negative supply be used, or at least that the intermediate-frequency tubes have their cathodes at some point more positive than the cathode of tube T_5. The voltage for the grid of T_5 is, of course, obtained by the way of an ac connection, and the rectifier T_4. The noise immunity of this circuit is seen, therefore, to come from two causes. First, the circuit is keyed and is completely immune to noise between keying pulses. Since the noise always extends negative on the grid of T_3, the noise pulses merely cut this tube off further than it has been cut off by the absence of the keying pulse and therefore no current can flow in its plate circuit. Second, due to the inverse action of this circuit, noise which produces a suddenly smaller pulse from T_3 does not cause current to flow in the diode T_4, and, hence, the noise is ignored by the circuit. Therefore, regardless of the amplitude of the noise, it is not measured by the circuit, which merely considers the noise as an increase in the signal which is followed slowly and from which it can recover quickly. If there is any noise in the white direction, however, the circuit is sensitive to it. Normally, however, the noise which occurs in this direction is a result of very infrequently encountered phase and amplitude conditions of the interfering carrier. In order for the detector output to be *reduced* by noise, and thus to give a white signal, the noise must be of approximately the same frequency and amplitude as the signal, and must remain in a phase opposite to that of the signal for an appreciable time. An appreciable area of noise which extends toward white is very seldom encountered. White noise may also be caused by intermediate-frequency overload. However, intermediate-frequency overload is minimized by the low impedance of the AGC driving circuit of tube T_5. This tube may be made to conduct very large currents necessary for a low impedance, since its current opposes the change in current in the intermediate-frequency and hence serves to stabilize the overall load of the receiver. Also, the circuit opposes excessive white noise generation due to overload in any tube which has AGC applied to its grid; i.e., the white noise is a result of bias produced by grid current drawn by the noise pulses. The AGC immediately responds to the reduced signal applied to it and *reduces* the bias, thus opposing the bias produced by the grid current.

The pulse for operating this AGC circuit, which is applied to the cathode of T_3, is, as has been explained, obtained from the horizontal output transformer. This gives a low impedance pulse, and since the load of T_3 is small, it has no adverse effect upon the horizontal output circuit. The pulse should be reasonably wide. It may occur during the synchronizing pulse, but it must occur during at least a portion of the "back porch," or blanking level, following the synchronizing pulse. The pulse must end before the end of this blanking level, or very wrong information will be obtained. As receivers are normally operated, this condition obtains because it is necessary for the return line to occur and be completed during the blanking time of the signal. Therefore this horizontal output pulse is admirably suited for operating this type of AGC.

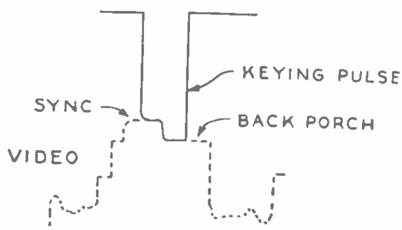


Fig. 10—The inverse modulated pulse.

Blanking Level Reference

This AGC actually operates by means of the blanking level, and not the synchronizing peak level. Therefore the actual sync height is unimportant, and the dc which is used on the kinescope is obtained from the blanking or true black level. This may be understood by noting that the signal on the grid of T_3 is at two levels during the keying pulse: first, at the synchronizing level, which is the more negative; and then at the blanking level, the more positive of the two levels. Therefore the larger pulse on the plate of T_3 is produced by the blanking level. The pulse obtained from the blanking level, therefore, is the pulse which actually operates the diode, T_4, and from which the AGC dc voltage is derived. Figure 10 shows in dotted form, the signal that would exist on the plate of T_3 from the video, and in solid form, the signal that does exist as the result of the video *and* the keying pulse. The maximum height of the pulse is seen to occur as a result of the blanking level of the video signal.

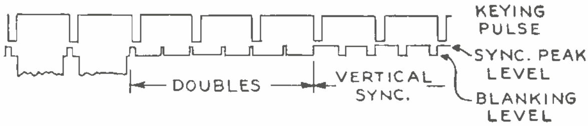


Fig. 11—Absence of blanking level during vertical sync and keying pulse.

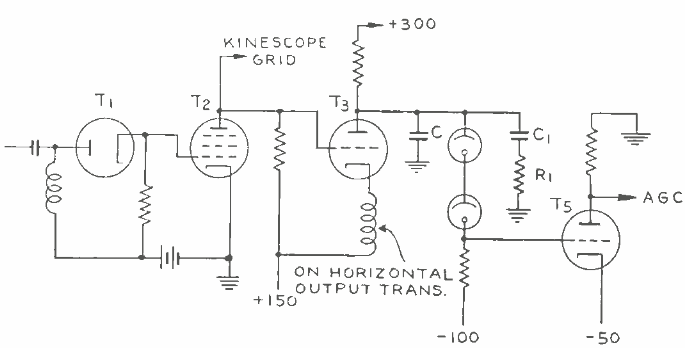
THE VERTICAL SIGNAL

During the vertical sync signal there is no blanking level during the time of the keying pulse. This can be seen by referring to Figure 11. The AGC detector is therefore not operated during the vertical sync signal and it interprets the vertical as a sudden increase. Therefore the gain of the amplifier, and the vertical sync signal are reduced according to the time constant speed of the detector. This fact limits the speed for this arrangement of the inverted keying system. It would be impossible to operate on the blanking level, or "back porch," except by virtue of the inverted system, wherein an increased signal is followed slowly, and a decreased signal quickly. The circuit speed is therefore limited by the response desired during the vertical. If the vertical signal is treated too severely, the speed must be decreased somewhat. However, when the vertical signal is reasonably unaffected, the AGC is still fast enough to remove most airplane interference.

A SIMPLER ARRANGEMENT

In Figure 12 is shown a simpler arrangement of the circuit of Figure 9. T_3 itself is made the detector. The plate resistor is made quite large, and is bypassed by a capacitor which then makes T_3 a peak plate rectifier. The dc is generated on the plate and is conducted to the grid of the AGC output tube, T_5, by means of glow tubes such as the 991. It is not necessary in this application, however, to use the more carefully selected variety of glow tube, such as the 991. The commercial ¼-watt neon lamp without the ballast resistor may be used. These tubes act essentially as batteries in series with the circuit from the plate of T_3 to the grid of T_5 and dc changes appearing on the plate

Fig. 12—T_3 operated as rectifier.



are therefore transmitted to the grid of T_5 except at a more negative dc level. This circuit eliminates one-half of a tube, but uses in its place the glow lamps. The operation of the circuit is relatively the same as the operation for the circuit using the diode, except that the gain of T_3 is reduced somewhat when used as a detector over the application in which it is used as an amplifier.

Filtering and Damping

Information for television AGC circuits, whether they be keyed or not, is obtained only over a small interval of time during the sync or blanking pulse. This information must therefore be integrated to control the signal until the next pulse is received. Such filtering is apt to introduce delays which may cause overswing and oscillation. If the filtering is accomplished entirely by one circuit, this tendency may be eliminated. The filtering in Figure 9 is accomplished by the capacitor C. Some bypass is necessary at each intermediate-frequency grid, but this filtering should be in the nature only of bypass for intermediate frequencies and should in no way affect the frequencies necessary for the AGC control. Since these frequencies are widely different it is possible to introduce sufficient intermediate-frequency bypass without introducing adverse delay for the AGC voltage. However, it is usually necessary to add some sort of damping. This damping should be introduced across the filter capacitor C. It is shown in Figure 9 as the circuit C_1R_1. C_1 is approximately 10 times C. R_1 is adjusted to give the minimum oscillation conditions. Providing there are no delays in other portions of the circuit, R_1 will be found to be non-critical. When R_1 is very small, oscillation may occur at a low frequency corresponding to the frequency of filtering of C_1. When R_1 is very large, oscillation may occur at the frequency caused by the filter capacity C. A point midway between these two values gives quite stable performance.

Circuit Adjustments

Figure 13 is Figure 9 redrawn with the required adjustments, screen supplies, etc. The most essential adjustment is that labeled "black level control". This control is not for the purpose of adjusting the black level on the kinescope, which should be accomplished in the normal manner, but is for setting the black level near the video noise clipping level. Clipping may be accomplished by plate current saturation on the video output tube, or by using a two-stage amplifier, and cutoff clipping in the first tube. In either case, the AGC equilibrium point must be adjusted by means of the black level control, until the sync pulses are held just inside the clipping level. The control may be

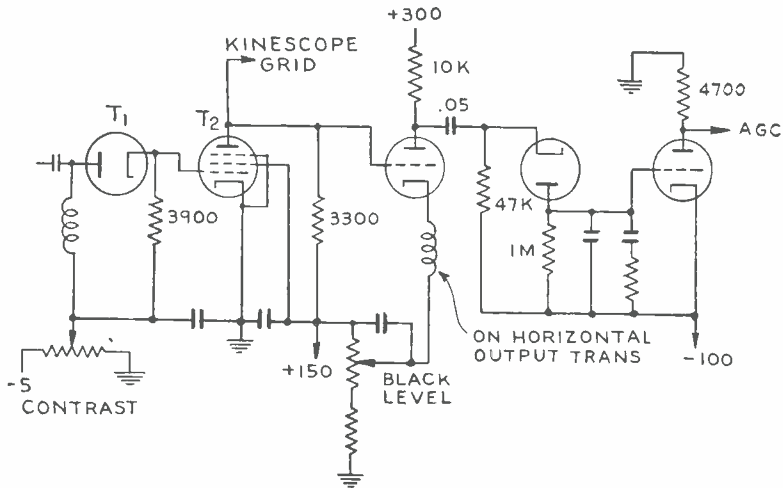


Fig. 13—Modified inverted keyed AGC circuit, showing screen supplies and controls.

a factory adjustment and need not appear on the front of the receiver. This control operates by varying the dc voltage on the cathode of T_3 and hence essentially adjusting the absolute level at which the peak of the measuring pulse occurs. Since it is the difference between this measuring pulse and the video which produces the AGC voltage, the video black level may be made to occur at any desired level by varying the dc in the pulse circuit as shown.

The other control shown is the video contrast control. This control adjusts the highlights of the picture without affecting the black level. As explained above, the black level is held at a fixed plate current level in T_2. If the bias on T_2 is changed by the dc control, the AGC circuit will bring the black level to the same value with reference to the cathode of T_2 as previously, and the difference between the zero output of T_1, which is the white level, and the black level, will have been changed, thus effecting a change in the overall signal swing or the white level on the plate of T_2 and the kinescope grid. This control may be included on the front panel, although the only important control of this nature is the background control for the kinescope, which, even so, should need adjusting only occasionally. A properly operating AGC will maintain the maximum whites at as near as possible the blooming level for the kinescope under all conditions except those of low modulation.

If the circuit of Figure 12 has been used, the black level control shown in Figure 13 is also able to correct for variation in the glow tubes, and is a satisfactory adjustment for tolerance in these tubes.

A Circuit for Maximum Speed

In cases where it is desired to increase the speed of operation of the inverted keyed AGC, a pulse must be used which occurs during the synchronizing pulse time. The keying pulse must be narrow and occur only during the synchronizing pulse time, which means that for correct operation, the keying pulse should occur wholly within the double frequency pulses, which are half as wide as the regular horizontal pulses. If the keying pulse should last over into the "back porch" time following the doubles, their height will be increased by the AGC, and difficulty will be experienced in properly separating the vertical pulse. Likewise, if the keying pulse lasts over into "back porch" time following the *regular* horizontal pulses, the AGC will operate as described previously,

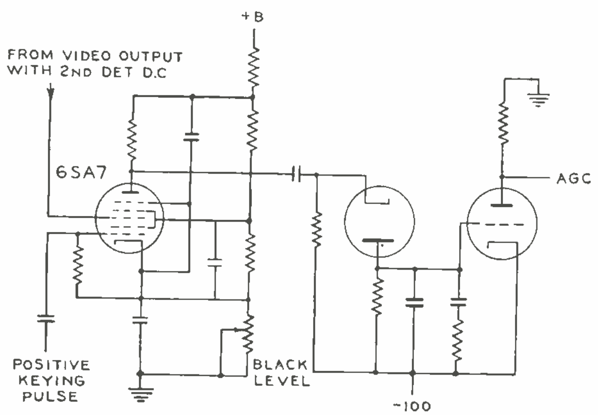


Fig. 14 — Circuit for high impedance or variable amplitude pulse.

and if too fast, will remove the vertical pulse. When a suitably narrow pulse is used, and it is made to occur on the sync pulse, extremely fast operation can be obtained without any errors being introduced.

A satisfactory keying pulse may be obtained from the blocking oscillator. However, difficulty will be encountered in properly applying such a pulse in the circuit of Figure 13. Figure 14 shows a circuit which is quite satisfactory. A 6SA7 has been substituted for the triode T_3 of Figure 13. The video is applied to the second control grid, and the keying pulse to the first grid. Black level control is effected by varying the dc potential of the cathode. The circuit otherwise operates essentially the same as the triode circuit of Figure 13, and is to be preferred over that circuit except for reasons of economy.

Use of Keying Pulse Which May Vary in Amplitude

In receivers which obtain the second anode voltage from the kickback pulse in the horizontal output stage, the pulse varies in ampli-

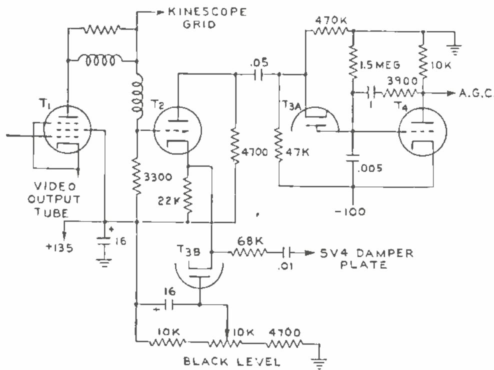


Fig. 15—Additional circuit for variable amplitude pulse.

tude with changes in kinescope current. The use of such a pulse for the AGC keying pulse results in an apparent loss of the dc component in the kinescope, and a limiting level which varies with the picture dc component, unless steps are taken to reference the pulse at its peak. This is accomplished satisfactorily by the circuit of Figure 14 which automatically references the pulse by means of grid current in the 6SA7. Another circuit for accomplishing this is shown in Figure 15. Actual values are given here for installing the AGC in a receiver of the 630TS type. It is necessary to dc connect the video amplifier, T_1 being the output stage. The video which appears across the 3300-ohm plate load is fed directly coupled to the grid of the AGC tube T_2. The cathode of this tube is driven negative by the pulse from the 5V4 damper tube plate. The pulse is reduced in amplitude by the 68K, 22K divider, and its peak referenced by the diode T_{3B} at a voltage determined by the setting of the 10K black level control. The resistance divider also serves as a series impedance between the pulse source and the diode T_{3B}, so that the loading of the diode widens the pulse to the desired width.

The modulated pulse from T_2, which appears across the 4700-ohm resistor, is applied through an .05 capacitor to T_{3A} with the bias resistor arranged as a divider which essentially applies a positive bias to the T_{3A} cathode. The pulse then need not become very small (with T_2 working near cut-off), when a small amount of rectified dc is required from the diode. The discharge resistor for the diode plate is returned to ground, which is positive, in order to make the discharge, or AGC speed, constant regardless of the signal level. The damping can be improved by connecting it as degenerative from the plate to the grid of T_4, as shown.

DETAILS OF THE SIMULTANEOUS EQUATION SOLVER

By

Edwin A. Goldberg

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—An electronic device for solving systems of linear simultaneous equations such as those encountered in circuit analysis work, quantitative chemical analysis, and a wide range of physical problems is described in this paper. Emphasis is placed on the actual electrical design employed in the execution of a practical model and the operation of the device is considered.

Principle of Operation

FIGURE 1 is a block diagram illustrating the principle of operation of a computer for solving a system of n linear equations of the form:

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots\cdots + a_{1n}X_n \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots\cdots + a_{2n}X_n \\ &\vdots \\ Y_n &= a_{n1}X_1 + a_{n2}X_2 + \cdots\cdots + a_{nn}X_n \end{aligned} \tag{1}$$

where the Y's and a's are known and it is desired to find the values of the X's. The values of all the known constant terms (Y's) and coefficients of the unknowns (a's) are set on potentiometers. Voltages proportional to the unknowns (X's) appear on the output terminals of the computing amplifiers and are measured by a potentiometer bridging method. The actual system of equations which the computing circuits solve is:

$$\begin{aligned} Y_1 &= \left(a_{11} + \frac{n+1}{G}\right) X_1 + a_{12}X_2 + \cdots\cdots + a_{1n}X_n \\ Y_2 &= a_{21}X_1 + \left(a_{22} + \frac{n+1}{G}\right) X_2 \cdots\cdots + a_{2n}X_n \\ &\vdots \\ Y_n &= a_{n1}X_1 + a_{n2}X_2 + \cdots\cdots + \left(a_{nn} + \frac{n+1}{G}\right) X_n, \end{aligned} \tag{2}$$

---

* Decimal Classification: R800(621.375.2).

where n is the number of equations and G is the gain of a computing amplifier. Note that this system differs from the system to be solved by the addition of the factor $\dfrac{n+1}{G}$ to the diagonal coefficients. This discrepancy may be taken care of either by changing the coefficients of the diagonal terms by the factor $\dfrac{n+1}{G}$, or by making this term so small, by making G large, that its effect may be neglected.

In order to handle positive and negative coefficients, constant terms, and unknowns, the outputs of the signal generator and each computing

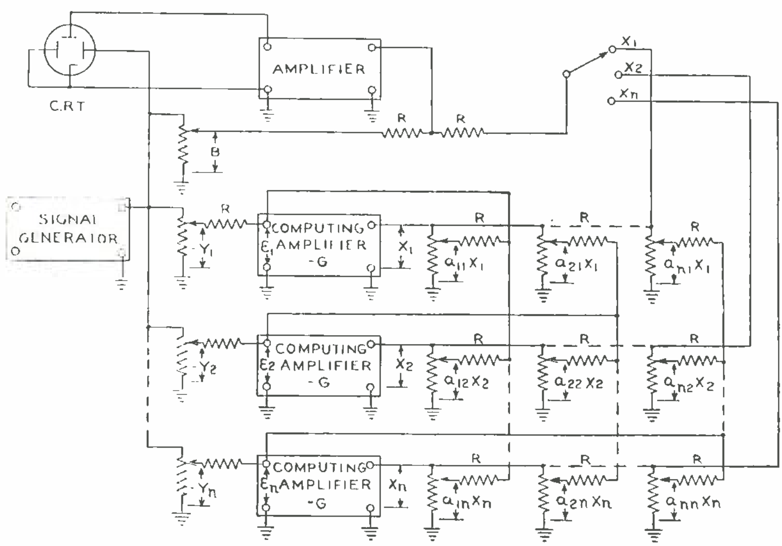


Fig. 1—Block diagram of simultaneous equation solver.

amplifier are push pull. The system uses an alternating-current signal generator, and the difference between positive and negative signals is a difference in phase of 180 degrees.

Computing Amplifiers

The computer circuits comprise a system of amplifiers with many feedback paths, the nature of these paths being a function of the system of equations being solved. Consequently, in order that the computer be useful in the practical sense, the phase and gain characteristics of the amplifiers must be so controlled that the computer will be stable (i.e., will not oscillate) for a wide range of equations which will normally be encountered. This problem has been thoroughly dis-

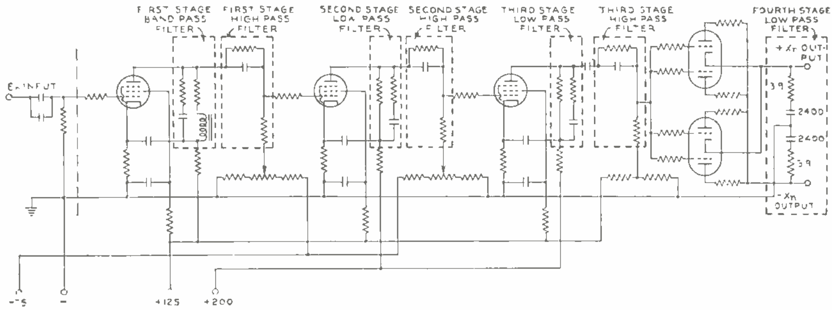


Fig. 2—Computing Amplifier.

cussed in another paper on this computer.[1] It is shown that the computer will be stable for any system of linear equations whose determinant has no characteristic roots with negative real parts. However, each amplifier and its associated networks must be so designed that their phase shift does not exceed ± 90 degrees in the frequency range in which the gain is greater than unity.

Figure 2 is a schematic diagram of a computing amplifier. It has a gain (G) of about 15000 at the operating frequency of the signal generator (1000 cycles), but its phase shift is controlled by networks so that it does not exceed ± 90 degrees from 0.2 cycle to 5 mega-

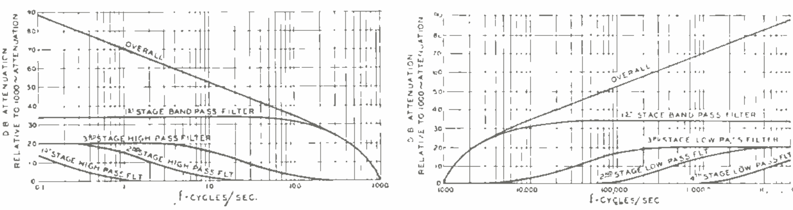
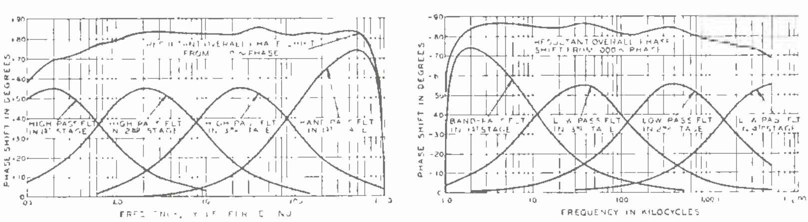


Fig. 3(a)—Attenuation vs. frequency characteristic of computing amplifier.

Fig. 3(b)—Phase shift vs. frequency characteristic of computing amplifier.



[1] E. A. Goldberg and G. W. Brown, "An Electronic Simultaneous Equation Solver," *Jour. Appl. Phys.*, Vol. 19, No. 4, pp. 339-345, April, 1948.

cycles, the frequencies at which the gain is reduced to unity. Figure 3(a) is a set of curves showing the attenuation vs. frequency for each stage, and also for the overall amplifier. Figure 3(b) is a set of curves showing phase shift vs. frequency.

Each computing amplifier drives a load consisting of ten "Micropot"* potentiometers. The settings of these potentiometers are the same as the coefficients of the unknowns in the system of equations being solved. In order to obtain either positive or negative values for the coefficients of an unknown, provisions are made for inserting any coefficient potentiometer either in the plate circuit or the cathode circuit of the output stage. The total impedances in the plate and cathode circuits are held constant by the use of a switch for each po-

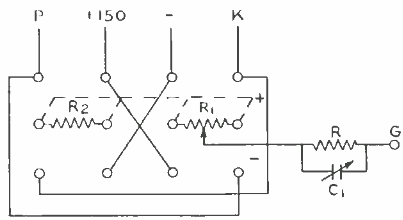


Fig. 4—Coefficient polarity reversing circuit.

tentiometer. The switch will insert a resistance equal to the potentiometer resistance in the plate circuit when the potentiometer is in the cathode circuit, and vice versa (Figure 4). This insures that the push-pull output of each amplifier is balanced at all times.

The arm of each potentiometer is connected to a summing resistor (R in Figure 1). Each summing impedance consists of a resistor shunted by a variable condenser for phasing. The resistance value for the summing resistors is made large relative to the resistance value for the potentiometers to minimize loading effects on potentiometer linearity. Each summing resistor is held within $\pm$ 0.1 per cent of its nominal value.

Compensation for Non-Uniformity in the Micropot Potentiometers

The linearity of the coefficient setting potentiometers was $\pm$ 0.1 per cent or better, but compensation was necessary to correct for other forms of non-uniformity between units.

Most of the potentiometers which were used could not be set to zero, their total resistance values were not all equal, and the percentage of total resistance included for a rotation of 3600 degrees was

* Helical type potentiometer manufactured by Thos. B. Gibbs and Co.

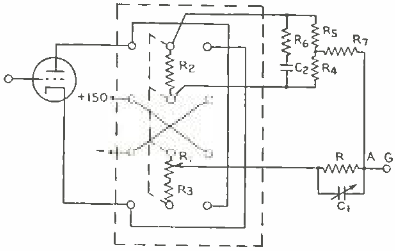


Fig. 5—Compensation circuits for micropots.

not the same for all potentiometers. These factors necessitated using circuits to make all potentiometers perform electrically alike. Figure 5 illustrates the circuits used.

The electrical rotation for each potentiometer was made equal for 3600 degrees mechanical rotation by adding a resistor R_3 in series with it. The value of R_3 is readily calculated as follows:

Let $R_1 =$ Total potentiometer resistance

$\Delta R =$ Amount of resistance traversed by the arm for 3600 degrees mechanical rotation

$k =$ Ratio of variation of voltage at arm of potentiometer to the total applied voltage of the output stage

$R_3 =$ Resistance to be inserted in series with the potentiometer to standardize the ratio of electrical to mechanical rotation.

$$R_3 = \frac{\Delta R}{k} - R_1 \qquad (3)$$

The value of R_2, the matching resistor, is given by the equation: $R_2 = R_1 + R_3$ (4)

Correction for the fact that most potentiometers could not be set to give zero output was made by the network of resistors R_4, R_5, and R_7 which supplies a current to the summing point A which is equal in magnitude to that fed to A by the potentiometer R_1 when it is set at zero, but 180 degrees out of phase. R_5 and R_7 are high value resistors, and R_4 is selected for each micropot. R_6 and C_2 are a network shunted

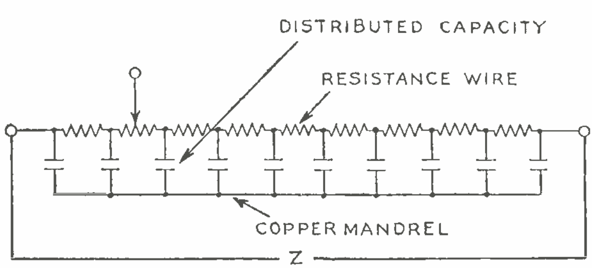


Fig. 6 — Equivalent circuit of micropots.

across R_2 in order to make the phase angle of the matching impedance the same as the phase angle of the potentiometer impedance at 1000 cycles. The construction of the micropot, is such that its impedance has a reactive component. The resistance wire is wound on an insulated mandrel of copper, which makes the impedance equivalent to that of a resistor with distributed capacity, as illustrated in Figure 6. The impedance looking into such a configuration has a phase angle which approaches 45 degrees for high frequencies and which decreases in value 1.41 times per octave. This characteristic was used to obtain high-frequency attenuation in the 6J6 output stages in lieu of the low-pass filter shown in Figure 2.

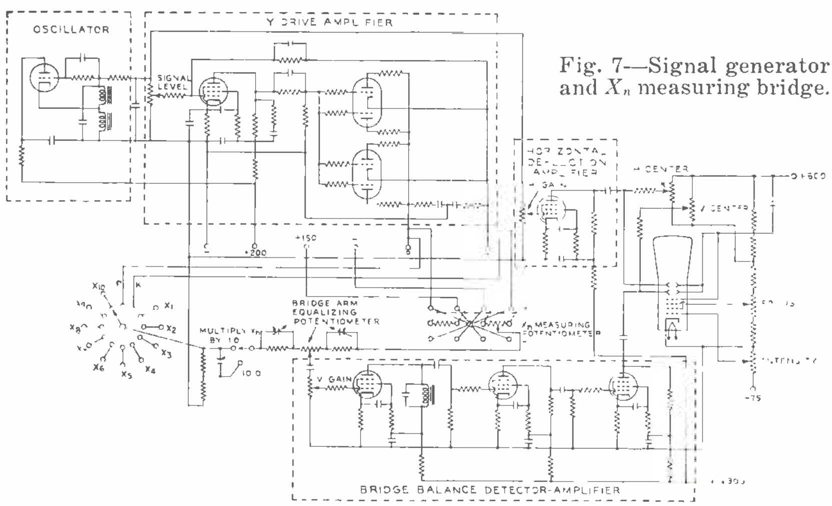


Fig. 7—Signal generator and X_n measuring bridge.

Signal Generator and Bridging Circuit

The signal generator (Figure 7) consists of a 1000-cycle oscillator followed by a two-stage amplifier, the output stage of which generates a push-pull signal to drive the Y potentiometers and the measuring bridge potentiometer. The amplifier has negative feedback so that it will not introduce distortion as the level is changed.

A potentiometer bridging method is used for measuring the output voltages of the computing amplifiers. The output of a computing amplifier is applied to a summing resistor through a selector switch. The output of the bridging potentiometer, whose input is connected across the signal generator, is tied to another summing resistor. The common ends of these resistors are connected to the input of a high-gain amplifier whose output is connected to the vertical deflection plates of the cathode-ray tube. When the potentiometer is set so that there is no vertical deflection on the oscilloscope, the voltage at the arm of the potentiometer is of the same magnitude as the voltage being measured, but 180 degrees out of phase. The unknown voltage, Xn, referred to the signal generator voltage can thus be read directly from the bridging potentiometer dial. A 10:1 attenuator position is provided so that values of Xn up to ten (represented by voltages up to ten times that of the signal generator) may be measured by the potentiometer. Phase adjustments by means of small adjustable air capacitors are provided in the bridging circuits to facilitate balancing the bridge when slight differences in phase (other than the normal 180-degree difference) exist between the signal generator output and the signal being measured. A simple network of this sort can shift the phase up to about 2.5 degrees without causing the magnitude to change more than 0.1 per cent.

Power Supplies

Voltages of + 200, + 150, + 125, and — 75 are required for the operation of the computer proper. In addition, voltages of + 300 and + 600 are used for operating the signal generator, bridging amplifiers, and cathode-ray tube. All voltages with the exception of the + 600 are regulated. Electronic regulators are used for the + 300, + 200, + 150, and + 125-volt supplies. The — 75-volt supply is glow tube regulated. Figure 8 is a schematic diagram of the power supplies.

Basically, all the electronic regulated power supply circuits are identical. The main purpose of using regulated voltages is to prevent interaction between amplifiers and circuits through common power supply impedance by making the power supply impedances extremely low. Since the absolute voltage stability required is not excessive, a glow tube standard is used.

All of the electronically regulated supplies contain two stages of amplification preceding the series regulator tubes. It was necessary to incorporate *R-C* (resistance-capacitance) networks in the amplifiers of these power supplies to eliminate circuit instability by controlling the high frequency cut-off characteristic. Extremely low dynamic impedance was required (at 1000 cycles particularly) for the + 150-volt power supply because of the nature of the push-pull output stage of the computing amplifiers. This was obtained by increasing the gain of the regulator amplifier by adding a 1000-cycle parallel resonant circuit in the plate circuit of the first stage, and also by inserting a small impedance in series with the load and taking a portion of the voltage drop across this impedance and adding it to the standard voltage applied to the grid of the first tube of the regulator.

This latter expedient is known as bridging, and the stability of the improvement gained therefrom is a function of the gain of the regulator. It is feasible to increase the effective gain and decrease the effective dynamic impedance from five to ten times by bridging. The resultant dynamic impedance of the + 150-volt supply was about 0.0001 ohm at the regulating points which are located at a level between the fifth and sixth computing amplifiers. The + 150-volt bus is a heavy copper strip so that there is very little voltage drop along it. Careful measurements indicated more voltage drop at 1000 cycles

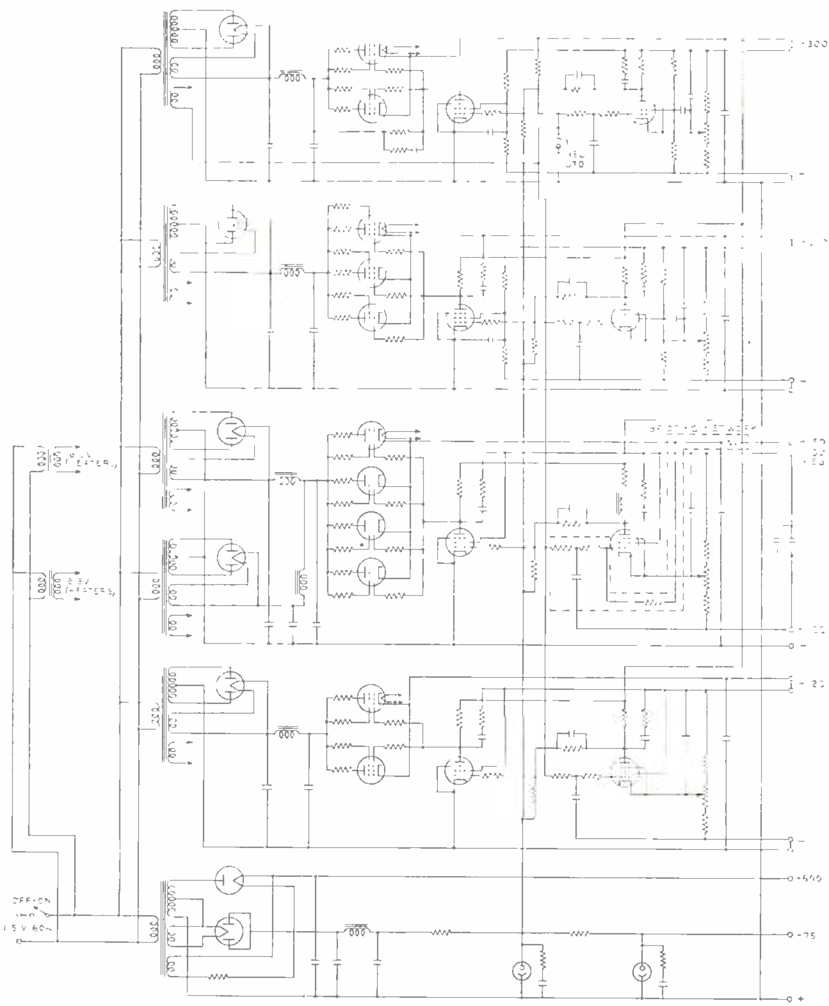


Fig. 8—Power supplies.

along one foot of this bus than across the regulating points of the power supply.

CONSTRUCTION

The whole computer is rack-mounted on a special rack. The panel at the top of the rack contains the signal generator and bridging circuits used for measuring the values of the X's. The next ten panels each contain eleven ten-turn potentiometers (the coefficient and constant potentiometers), the ± switches, and the summing resistors. Connections between each of these panels are made to the proper buses

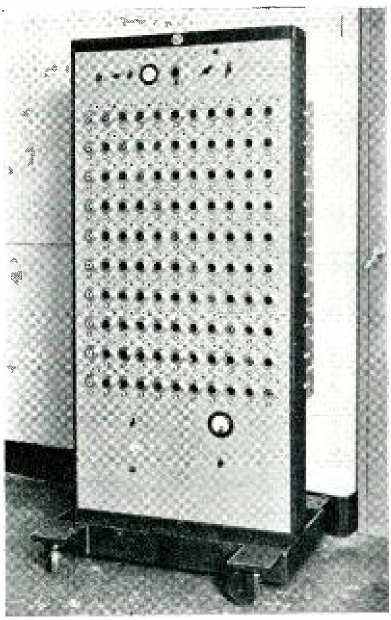


Fig. 9—Front view of equation solver.

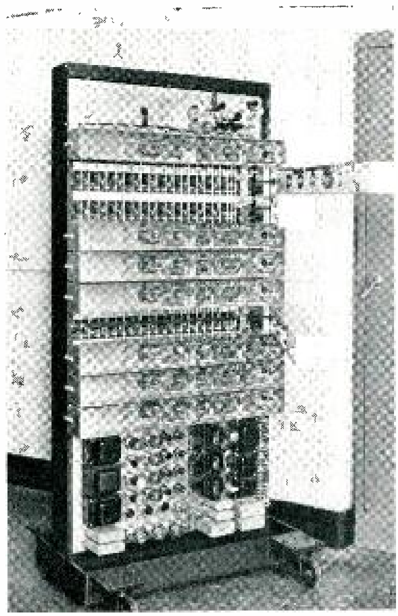


Fig. 10—Rear view of equation solver.

by means of clips so that any one panel may be easily removed for servicing on the bench the potentiometers, matching resistors, summing resistors, and phasing adjustments. The lower panel is occupied by the power supplies. The ten computing amplifiers are each hinged on the back of the rack so that they may be individually disconnected from the computing circuits. It is necessary to swing these amplifiers out in order to change tubes. When the amplifiers are swung into the closed position, the proper input and output connections are automatically made. Filament and plate supply connections are made through an octal socket and plug connection for each amplifier. Figures 9 and 10 are the front and rear views of the device.

Three of the amplifiers are swung out in Figure 10 to illustrate the constructional features referred to.

OPERATION

A system of equations must be reduced to a suitable form before being set up on the computer. All coefficients (a_{ij}) and constant terms (Y_i) must be reduced in such a manner that the maximum value for any one is unity or less. This may be done by (I) multiplying both sides of each equation by suitable constants, (II) multiplying the constants (Y_i) by a suitable constant, and (III) multiplying columns of coefficients (a_j) by suitable constants. Application of step I does not alter the numerical values of the unknowns. If step II is employed, the results obtained must be divided by the constant used in order to make them satisfy the original equations. If step III is employed, the values obtained for the unknowns must be multiplied by the same constants as those by which the coefficients of the respective unknowns were multiplied. The reduction should be made in such a manner that the largest constant term (Y_i) is near unity, and the largest coefficient (a_{ij}) in each row and column is near unity. This procedure will result in operation of the computer at the most favorable scale factors.

The coefficients (a_{ij}) are then set on their respective potentiometer dials, and the cathode-ray tube bridge balance indicator is observed. If there are no signs of oscillation, the system is stable, and the constant terms (Y_i) are set on their respective potentiometer dials. Next, the X selector switch on the top panel is set to X_j and the bridging potentiometer is adjusted for minimum vertical deflection on the balance indicator. The setting on this dial is the value of X_j. This procedure is repeated for each different X_j.

If the system is unstable, it is necessary to rearrange the equations. The equation solver is inherently stable for all systems of linear equations whose determinants have no characteristic roots with negative real parts. Normally, equations should be so arranged that the largest coefficient in each equation is a diagonal term and is positive if possible. This does not necessarily insure that the system will be stable. Any system with finite values for the unknowns may be made stable if the number of equations does not exceed half the number that the computer may normally handle. In such a case, the system

$$Y = AX \tag{5}$$

where Y is the matrix of constant terms, A is the matrix of the determinant, and X is the matrix of unknowns, may be solved by inserting the constant terms and coefficients as follows:

$$\begin{matrix} O & I & -A^{tr} \\ Y & A & O \end{matrix} \tag{6}$$

where I is a unit matrix (diagonal elements of unity and zeros elsewhere), $-A^{tr}$ is the negative transpose of A (rows and columns interchanged, and the signs reversed), and O is a zero matrix (all terms zero). The unknowns are measured in the normal manner.

The accuracy of the results obtained depends not only on the accuracy of the components, but also on the nature of the system of equations being solved. When the computer itself is the limiting factor, the results obtained will normally satisfy the equations within ± 0.1 per cent. If the determinant of the equations has one or more very small characteristic roots, the results will not be as accurate. Through the use of an iteration process, however, essentially unlimited accuracy may be obtained. Iteration consists of the following steps:

1. Set up the equations and solve for the unknowns (X_j) in the normal manner.
2. Substitute the values obtained for the X_j into the original equations. Take the differences between the values thus obtained and the values of (Y_i). Call these differences ΔY_i.
3. Change the Y_i dials only to $K\Delta Y_i$ (K generally being 100 or 1000 and chosen to utilize as much scale factor as possible) and measure $K\Delta X_j$.
4. Add ΔX_j to X_j to obtain corrected values for X_j.

If higher accuracy is desired, repeat the process.

The following system of equations was solved on the computer:

$$\begin{aligned} 1 &= 1\,X_1 + 0.4\,X_2 + 0.1\,X_3 \\ O &= -0.9\,X_1 + 1\,X_2 + 0.4\,X_3 \\ O &= -0.2\,X_1 - 0.4\,X_2 + 1.0\,X_3 \end{aligned} \tag{7}$$

The computer yielded the following results:

$$\begin{aligned} X_1 &= +0.753 \\ X_2 &= +0.531 \\ X_3 &= +0.363 \end{aligned} \tag{8}$$

Substituting these values into the original equations yielded the following errors:

$$\begin{aligned} \Delta Y_1 &= -0.0017 \\ \Delta Y_2 &= +.0015 \\ \Delta Y_3 &= .0000 \end{aligned} \tag{9}$$

These errors were multiplied by 100, inserted in the computer in place of Y_i, and 100 ΔX_j was measured. The corrected values for X_j were:

$$X_1 = +\,0.751290$$
$$X_2 = +\,0.531070 \qquad (10)$$
$$X_3 = +\,0.362690$$

These results, when checked, yielded the following errors:

$$\Delta Y_1 = +\,.000013$$
$$\Delta Y_2 = +\,.000015 \qquad (11)$$
$$\Delta Y_3 = -\,.000004$$

Circuit problems involve the solution of systems of equations with complex constants, coefficients, and unknowns. Such a system of n equations may be readily converted to a system of $2n$ equations with real quantities only by equating reals and imaginaries in the original equations. The resulting system is solved on the computer in the normal manner.

Acknowledgement

The author wishes to acknowledge the guidance of V. K. Zworykin in connection with this work, the useful contributions of J. M. Morgan concerning the mechanical design, and of G. W. Brown on the mathematical aspects of the problem.

SOME NOTES ON NOISE THEORY AND ITS APPLICATION TO INPUT CIRCUIT DESIGN*

BY

WM. A. HARRIS

Tube Department, RCA Victor Division,
Harrison, N. J.

Summary—*The mechanism by which noise is produced in an electron tube and the relation between induced grid noise and plate noise are discussed. An equivalent circuit with noise generators supplying voltages and currents to simulate noise derived from the plate current of a tube, from the grid by passage of this current, and from the input circuit is then analyzed to determine the optimum noise factor obtainable under various conditions. The frequency for which the quantity $R_{eq}\ g_{in}$ is unity is seen to be an appropriate figure of merit for the noise produced by an electron tube. The frequencies corresponding to chosen values for the noise factor are presented for several receiving tube types. The paper concludes with a discussion of the circuit requirements which must be satisfied in order to obtain noise factors approximating the theoretical values.*

INTRODUCTION

NOISE generated in the first tube of a receiving system is frequently the factor controlling the over-all sensitivity of the system. An understanding of the mechanism by which such noise is produced is helpful in the design of receiving equipment, particularly with respect to the choice of tube types. If the electrons in a tube were to leave the cathode at a perfectly uniform rate, there would be no noise, or at least, none in the frequency range in which a tube is useful. The rate of emission of electrons, however, is not uniform. In any given interval of time there are probably a few more or a few less electrons leaving the cathode than the average number for that amount of time. The classical shot-effect derivations predict the magnitudes of fluctuations of this sort. Furthermore, because theory and experimental data have revealed the extent to which space-charge effects can reduce these fluctuations in electron tubes, the noise components of the plate current of a tube can be computed in many instances.[1]

At high frequencies, the fluctuation current induced in the grid of

* Decimal Classification: R138 × R161.6 × R361.211.

[1] B. J. Thompson, D. O. North, and W. A. Harris, "Fluctuations in Space-Charge-Limited Currents at Moderately High Frequencies", *RCA Review,* Vol. IV: No. 3, pp. 69-285, January 1940; No. 4, pp. 443-473, April, 1940; Vol. V: No. 1, pp. 106-124, July, 1940; No. 2, pp. 244-260, October, 1940; No. 3, pp. 371-388, January, 1941; No. 4, pp. 505-524, April, 1941; Vol. VI, No. 1, pp. 114-124, July, 1941.

a tube by the passage of the fluctuating plate current through the grid is another noise source which must be considered. The magnitude of the mean square of this current is proportional to the component of input conductance due to transit time.[2]

Since the grid noise and plate noise are derived in part from the same current fluctuations, they cannot be treated as entirely independent noise sources. Nevertheless, valuable working formulas and principles have been derived for conditions under which coherence between grid noise and plate noise can be ignored.[3] It is possible, moreover, that the improvements obtainable by taking coherence into account are not very important for the majority of tubes and circuits in current use. Theoretical considerations, however, indicate that a substantial improvement in noise factor may be obtained by taking advantage of the coherence between grid noise and plate noise if the conditions assumed for the theory can be realized in actual tubes. Experimental evidence shows that at least part of this improvement can be obtained in a practical system.[4]

In this paper, the relation between induced grid noise and plate noise is illustrated by an examination of the result of the passage of a single electron through a tube. Then, the conditions giving optimum noise factors are derived, using the methods employed by Herold and others. Herold[3] showed that the noise factor for a tube is a function of the product $R_{eq}\, g_{in}$, where R_{eq} is the equivalent noise resistance and g_{in} is the input conductance. In this paper, the frequency for which the product $R_{eq}\, g_{in}$ is unity is recommended as an appropriate noise "figure of merit" for a tube. The ratios of the operating frequencies to this reference frequency therefore can be used as the abscissas for curves of optimum noise factor.

CURRENT IMPULSES FROM ONE ELECTRON

Figure 1 shows the distribution of potential in a parallel-plate triode. The potential curve is based on the assumption of a Maxwellian distribution of initial velocities, with a cathode temperature of approximately 1000 degrees Kelvin. The dotted curve represents the velocity of an electron with just enough initial velocity to allow it to pass the point of minimum potential and continue to the anode. The time of

[2] D. O. North and W. R. Ferris, "Fluctuations Induced in Vacuum-Tube Grids at High Freqencies", *Proc. I.R.E.*, Vol. 29, pp. 49-50, February 1941.

[3] E. W. Herold, "An Analysis of the Signal-to-Noise Ratio of Ultra-High-Frequency Receivers", *RCA Review*, Vol. VI, No. 3, pp. 302-331, January, 1942.

[4] M. J. O. Strutt and A. vanderZeil, "Signal-Noise Ratio at VHF", *Wireless Engineer*, Vol. 23, pp. 241-249, September, 1946.

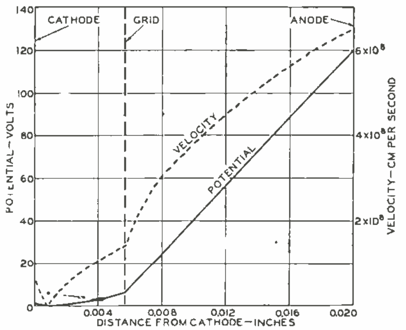


Fig. 1—Potentials and velocities in a parallel-plane triode.

transit for such an electron can be computed by a graphical integration process.

When a charge is in motion between two electrodes in a tube, current flows between the electrodes bounding the region containing the charge. For a parallel-plane structure this current is proportional to the velocity of the charge and inversely proportional to the distance between the two plane boundaries; it does not depend on the position of the charge relative to the boundaries.[5] The velocity curve of Figure 1 is applicable for a charge of the indicated initial velocity. If the velocities for various points in the cathode-grid space are divided by the distance between cathode and grid, and the velocities for various points in the grid-anode space are divided by the distance between grid and anode, quantities proportional to the current due to the motion of the charge for these various positions are obtained. Then, the relation between position and time obtained by integration may be used to obtain a current-time curve.

The curves of current versus time for the tube structure of Figure 1 are shown in Figure 2. The solid curve shows the current to the grid and the dotted curve the current to the plate which would result from the passage of one electron. The choice of the velocity of the slowest electrons which can reach the anode leads to a computation difficulty; the time required for such electrons to pass the potential-minimum region is theoretically infinite. Consequently, transit times are computed from the cathode to a point near the potential minimum on the cathode side, and from the grid back to a point near the potential minimum on the grid side. The three rectangles between the ends of the two curves show the times and currents for charges passing between the terminal points of these curves with velocities exceeded by 90, 50, or 10 per cent of the electrons reaching the anode. The use of one of these velocities would cause some change in the remainder of the curve, both in the current and the time scale, but the shape of the curve would be about as shown. The effect of a change in initial velocity on the current between grid and anode would be almost negligible. The cathode-to-grid transit time for an electron with an initial velocity

[5] S. Ramo, "Currents Induced by Electron Motion", *Proc. I.R.E.*, Vol. 27, pp. 584-585, September, 1939.

corresponding to one of the rectangles in Figure 2 is, therefore, approximately the sum of the transit times represented by the two curves and the appropriate rectangle. The indicated range is from 6 to 8×10^{-10} seconds. The transit time from grid to anode for the conditions of Figure 2 is about 1×10^{-10} seconds.

The curve of Figure 2 does not show the compensating effect which takes place when an extra charge passes through a tube. The potential minimum is depressed by an amount depending on the position of the added charge, during the whole time this charge is between cathode and grid. The result is a reduction of the current, which can be considered equivalent to the passage of a series of charges of opposite sign between potential minimum and grid, and the passage of charges of the same sign but of opposite direction between potential minimum and cathode. These effects account for the shot-effect reduction factor computed by North[1]. It appears, however, that between the initiating pulse and the compensating current there is some time delay which may be important in the determination of the noise at very high frequencies.

As soon as the extra noise-producing charge leaves the cathode a small effect on the minimum potential will be noted; some electrons which are reaching the potential minimum at this instant turn back instead of continuing toward the plate. The effect of the extra charge persists until it reaches the grid. The compensating effect cannot be completed until the time at which an electron, turned back because of the depression of the potential minimum when the extra charge was near the grid, would have reached the grid had it not been turned back.

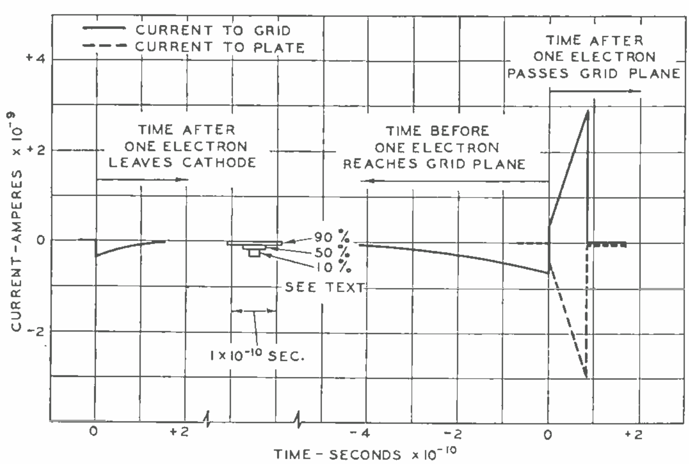


Fig. 2—Grid and plate currents due to passage of a single electron.

The compensating charges in motion after the passage of the causing charge, however, must themselves have an effect on the minimum potential. The complete result, consequently, is probably of the nature of a damped oscillation with a period related to the transit time of an electron from cathode to grid. Thus, the curves of Figure 2 do not present a complete picture of the generation of noise in a parallel-plane triode, but they do show some of the characteristics of the basic noise impulses.

It is pertinent at this point to discuss the extent to which grid and plate noise currents can be made to cancel each other. The pulse shapes are quite different, so it is evident that complete cancellation cannot be expected. Partial cancellation may be obtained if a voltage is developed at the grid by allowing the grid current to flow into a capacitor. This voltage is proportional to the integral of the grid current and has the effect of momentarily reducing the plate current. A suitable choice of capacitor value can give a plate-current pulse of zero net area for electrons of a particular initial velocity. However, there will always be some conductance in the grid circuit which will result in a grid-voltage component tending to increase the noise output. Moreover, electrons leaving the cathode with velocities too low to allow them to pass the point of minimum potential produce pulses of grid current without producing corresponding plate-current pulses.

Determination of Frequency Spectra

The curves of Figure 3 illustrate the method by which the frequency spectra corresponding to the grid-current and plate-current pulses may be obtained. The current which could be measured in the small frequency range represented by $d\omega$ at a frequency $\omega/2\pi$ is obtained from the Fourier integral:

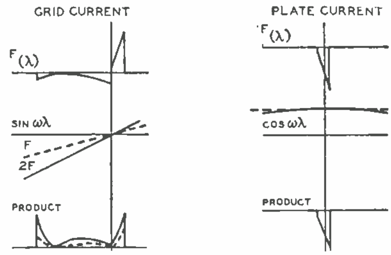
$$A d\omega = \frac{2}{\pi} \int F_{(\lambda)} \cos \omega\lambda \, d\lambda \cos \omega t \, d\omega + \frac{2}{\pi} \int F_{(\lambda)} \sin \omega\lambda \, d\lambda \sin \omega t \, d\omega. \tag{1}$$

The function $F_{(\lambda)}$ represents the pulse.

When the frequency is low in comparison with the reciprocal of the transit time, the value of the term $\cos \omega\lambda$ in Equation (1) is nearly constant over the region in which $F_{(\lambda)}$ has a value other than zero. In addition, the value of the term $\sin \omega\lambda$ in Equation (1) can be represented as a straight line with a slope directly proportional to the frequency over the same region.

Because the grid pulse has equal positive and negative areas, the

Fig. 3—Development of frequency spectra.



integral containing the cosine terms is zero for any frequency low enough so that cos $\omega\lambda$ can be considered constant. The integral containing the sine terms will have a value which is represented graphically by the area under the curve labeled "product" in the grid-current curves of Figure 3. This area will be directly proportional to the frequency because the slope of the sin $\omega\lambda$ line is proportional to frequency. It will also be proportional to transit time because, if the areas and shapes of the parts of the $F_{(\lambda)}$ curve are maintained constant and the base line is extended, the area under the "product" curve will increase in proportion to the increase in base-line length.

The plate pulse will give a zero value for the integral containing sine terms if a suitable point of origin is chosen. The integral containing cosine terms, then, gives the current, which is independent of the frequency when the frequency is low. Because the area of the plate-current pulse represents the amount of charge producing the pulse, the current in a small frequency band resulting from a given amount of charge is also independent of the transit time.

The mean-square noise current measurable in any frequency band results from large numbers of pulses distributed at random with respect to time. For the plate current, consequently, the mean-square current $d\overline{i^2}$ in a frequency band of width df can be represented by the equation

$$d\overline{i^2} = k_2\, df \tag{2}$$

and for the grid current, by

$$d\overline{i^2} = k_1\, \omega^2\, \tau^2\, df \tag{3}$$

where τ is the transit time, or

$$d\overline{i^2} = k_1\, \theta^2\, df \tag{3a}$$

where θ is the transit angle.

The electronic component of input conductance is proportional to the square of the transit angle[6], so a proportionality between the mean-square noise current and the input conductance is indicated, thus:

$$d\overline{i^2} = k_3\, g_1\, df. \tag{4}$$

North and Ferris[2] found that the complete relation for grid-current noise is

$$d\overline{i^2} = \theta_1\, g_1 \cdot 4k\, T_o\, df \tag{5}$$

[6] D. O. North, "Analysis of the Effects of Space Charge on Grid Impedance", *Proc. I.R.E.*, Vol. 24, pp. 108-136, January, 1936.

where θ_1 has a numerical value of approximately 5 when the cathode temperature is 1000 degrees Kelvin and the reference temperature T_o is approximately 300 degrees Kelvin. The dependence on temperature is discovered only when the analysis is extended to include the compensating currents.

The plate current noise from a tube may be represented as if it were derived from a noise voltage at the grid sufficient to produce the noise current. The appropriate equations[1] are $d\overline{e^2} = 4kT\,R_{eq}\,df$ (6)

where, for oxide-coated cathode tubes, theory indicates approximately that, for triodes,

$$R_{eq} = 2.5/g_m \tag{7}$$

and for pentodes

$$R_{eq} = \frac{I_b}{I_b + I_{c2}}\left(\frac{2.5}{g_m} + \frac{20\,I_{c2}}{g_m^2}\right) \tag{8}$$

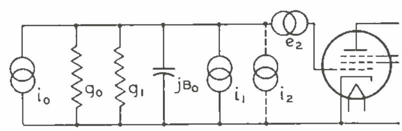


Fig. 4—Equivalent circuit.

Circuit Analysis

The circuit of Figure 4 represents the replacement of a real tube by a fictitious noise-free tube with zero input admittance and suitable noise generators and external circuit elements. The plate noise is introduced by a constant-voltage generator delivering a voltage e_2 in series with the grid. The noise current to the grid, i_1, is represented by a constant-current generator across the grid circuit. The noise from the input system, i_o, is represented by a second constant-current generator. The plate-noise generator can be replaced by another constant-current generator; the voltage output of the plate-noise generator is multiplied by the total admittance of the input circuit to give the required current i_2. The relations between the noise currents and the tube and circuit parameters are given by the equations:

$$e_2 = K\sqrt{R_{eq}} \quad (9) \qquad i_1 = (-j)\,K\sqrt{\theta_1 g_1} \quad (10)$$

$$i_2 = e_2(g_o + g_1 + j\,B_o) = K\sqrt{R_{eq}}\,(g_o + g_1 + j\,B_o) \tag{11}$$

$$i_o = K\sqrt{\theta_o g_o}. \tag{12}$$

where $K = \sqrt{4kT\Delta f}$

g_1 is the electronic component of input conductance;

R_{eq} is the resistance equivalent for the plate noise, referred to the grid;

θ_1 is a multiplier relating grid noise to input conductance; its value is approximately 5 for tubes with oxide-coated cathodes;

θ_o is a multiplier representing the ratio of antenna noise to the noise in a resistor at room temperature;

g_o is the antenna conductance, referred to the grid;

B_o is the net susceptance of the circuit at the operating frequency.

It is assumed that the conductance g_o can be varied arbitrarily by some such means as a variable-ratio transformer between antenna and grid. Also, it is assumed that means such as a tuning capacitor are provided so that B_o can be varied arbitrarily. Ohmic losses in the input circuit are neglected.

The quantity $(-j)$ in parenthesis in the expression for i_1 indicates that i_1 may be in quadrature with e_2 over a specified frequency range. The preceding discussion suggests that this assumption is legitimate in the case of a triode, when the frequency is not too high and the frequency band is not too wide. The assumption is not valid, however, for a pentode because in that case the larger part of the plate noise results from the division of current between plate and screen grid[1], and consequently it cannot be correlated with the grid noise.

The total mean-square current from the three generators of Figure 4 can be found as follows: Add i_1 and i_2, taking coherence, if assumed, into account. Then, determine the sum of the squares of i_o, the real part of $(i_1 + i_2)$, and the imaginary part of $(i_1 + i_2)$. When coherence is not assumed, simply add the mean-square values of i_o, i_1, *and* i_2. The results follow:

When a quadrature relation between grid and plate noise is assumed, the mean-square current is

$$\overline{i^2} = K^2\{g_o\theta_o + R_{eq}(g_1 + g_o)^2 + (B_o\sqrt{R_{eq}} - \sqrt{\theta_1 g_1})^2\} \tag{13}$$

When no coherence is assumed

$$\overline{i^2} = K^2\{g_o\theta_o + R_{eq}(g_1 + g_o)^2 + R_{eq}B_o^2 + \theta_1 g_1\} \tag{14}$$

Optimum Noise Factors

Optimum performance with respect to noise is obtained when the term $g_o\theta_o$ is as large as possible in comparison with the other terms, and, in fact, the noise factor as defined by North[7], Friis[8], and others is obtained by dividing Equation (13) or (14) by $K^2 g_o \theta_o$ and assuming $\theta_o = 1$. The first step in finding conditions for minimum noise is the adjustment of B_o to eliminate the term in which it appears in either equation.

Then, either equation can be differentiated with respect to the ratio g_1/g_o and an optimum value of noise factor can be obtained. The noise factors after adjustment of B_o are given by the equation

$$NF = 1 + R_{eq}\, g_1 \left(\frac{g_1}{g_o} + \frac{g_o}{g_1} + 2 \right) \qquad (15)$$

when coherence is assumed, and the equation

$$NF = 1 + R_{eq}\, g_1 \left(\frac{g_1}{g_o} + \frac{g_o}{g_1} + 2 \right) + \theta_1 \frac{g_1}{g_o} \qquad (16)$$

when coherence is not assumed.

The minimum noise factors, with the conditions for obtaining them are

$$NF = 1 + 4\, R_{eq}\, g_1 \quad (17) \qquad \frac{g_1}{g_o} = 1 \quad (18) \qquad B_o^2 = \theta_1 g_1^2 / R_{eq}\, g_1 \quad (19)$$

when coherence is assumed; and, when coherence not assumed

$$NF = 1 + 2\, R_{eq}\, g_1 + 2 \sqrt{\theta_1 R_{eq}\, g_1 + (R_{eq}\, g_1)^2} \qquad (20)$$

$$\frac{g_1}{g_o} = \sqrt{R_{eq}\, g_1 / (\theta_1 + R_{eq}\, g_1)} \quad (21) \qquad\qquad B_o = 0 \quad (22)$$

The quantities R_{eq} and g_1 are both tube parameters Since they appear as the product $R_{eq}\, g_1$ in Equations (17) and (20), the magnitude of this product indicates the noise performance obtainable from a tube. The quantity g_1, however, varies with the square of the frequency. For purposes of computation, it is preferable to use as a reference parameter the square root of the product $R_{eq}\, g_1$, which

---

[7] D. O. North, "The Absolute Sensitivity of Radio Receivers", *RCA Review*, Vol. VI, No. 3, pp. 332-343, January, 1942.

[8] H. T. Friis, "Noise Figures of Radio Receivers", *Proc. I.R.E.*, Vol. 32, pp. 419-422, July, 1944.

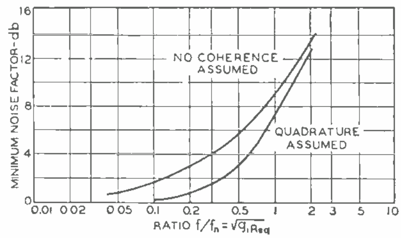


Fig. 5—Minimum noise factor.

varies with the first power of the frequency. The curves of Figure 5 show the optimum noise factors for the two cases considered, plotted against the quantity $\sqrt{R_{eq}\, g_1}$ for the condition $\theta_1 = 5$. If the frequency for which $\sqrt{R_{eq}\, g_1}$ is unity is designated as f_n, the quantity $\sqrt{R_{eq}\, g_1}$ for any frequency f is equal to the ratio f/f_n.

The method of analysis described above is essentially the same as that used by Herold[3]. The curve for the case of no coherence (Figure 5) can be identified with one of the curves (Figure 5) of Reference 3 when differences in the coordinates used are taken into account. Equation (21), giving the required ratio of tube input conductance to circuit conductance, is equivalent to Equation (7) of Reference 3.

The susceptance required for the case of quadrature, as found from Equation (19), is obtained by the same amount of capacitance at any frequency. Equation (19) can be rewritten

$$B_o^2 = \theta_1 g_1 / R_{eq}. \tag{19a}$$

Because g_1 is proportional to the square of the frequency and R_{eq} and θ_1 are independent of frequency, it is evident that the susceptance B_o is directly proportional to the frequency and, consequently, can be produced by a fixed capacitance.

Comparison of Tubes

The data for Tables I and II were obtained by calculating values for the equivalent noise resistance and using measured values for input conductance for the tube types listed. Table I gives the reference frequency for noise, f_n, and the frequencies for which noise factors of 1, 3, and 10 decibels are calculated for a number of pentode types. No coherence is assumed between plate noise and grid noise for this case. Table II gives similar data for two triodes and for several pentodes, connected as triodes, under the alternate assumptions of no coherence between plate and grid noise, and a quadrature relation between plate and grid noise. The 10-decibel column for the quadrature case is omitted because the indicated frequencies are too high to make the assumption appear reasonable.

Table I—Pentodes

Type	R_{eq} ohms	g_{in} (100 mcs.) micromhos	f_n mcs.	Frequency for Noise Factor (1 db.) mcs.	(3 db.) mcs.	(10 db.) mcs.
6SK7	11,600	440	45	2.5	8.9	56
6AC7	650	1,730	94	5.3	19	119
6BA6	3,800	580	67	3.8	13.4	85
6AG5	1,900	300	133	7.5	27	169
6AK5	1,900	125	208	11.6	41	262
6BH6	2,360	340	122	6.8	24	154
6BJ6	3,800	275	98	5.5	19.3	124

The input conductance values used in Tables I and II were measured by the susceptance-variation method[9] and include the effects of lead inductance. For pentodes, the predominant lead effect is that of the cathode-lead inductance, which tends to increase the input conductance. For triodes, inductance in the plate lead tends to reduce the input conductance and this effect may be equal or greater than the effect of cathode-lead inductance. For the triode-connected pentodes, the input-conductance data obtained with the tubes connected as pentodes are used.

Triode "A" in Table II is a developmental triode, designed primarily for use as a high-frequency oscillator. The low input conductance and the consequent high "f_n" value recorded for this type

[9] "Input Admittance of Receiving Tubes", RCA Application Note AN-118, RCA Tube Department, Harrison, N. J., April, 1947.

Table II—Triodes and Triode-Connected Pentodes

Type	R_{eq} (Triode) ohms	g_{in} (100 mcs.) micromhos	f_n (Triode) mcs.	Frequency for Indicated Noise Factor — No Coherence Assumed (1 db.) mcs.	(3 db.) mcs.	(10 db.) mcs.	Quadrature Assumed (1 db.) mcs.	(3 db.) mcs.
6SK7.....	970	440	72	4.0	14	91	18	36
6AC7.....	214	1,730	164	9.2	33	207	41	82
6BA6.....	410	580	204	11.5	41	258	51	102
6AG5.....	380	300	294	17	59	374	74	148
6AK5.....	380	125	476	26	92	580	116	230
6BH6.....	390	340	274	15.4	54	345	68	137
6BJ6......	485	275	274	15.4	54	345	68	137
6J6.......	470	195	320	18.0	63	410	80	160
"A"*......	360	50	747	42	147	940	186	373

* Developmental triode.

is probably accounted for by close spacing, high current density, and a symmetrical cylindrical structure which contributes to uniformity in the cathode-to-grid and grid-to-plate transit times.

Relation of Reference Frequency to Transit Time

The reference frequency for noise for a triode depends primarily on the electron transit time between cathode and grid. The noise equivalent resistance for a triode is approximately $R_{eq} = 2.5/g_m$ (7)

and the electronic component of the input conductance[6] is approximately

$$g_1 = g_m (\omega\tau_1)^2/20. \tag{23}$$

The product, therefore, is

$$R_{eq}\, g_1 = (\omega\tau_1)^2/8. \tag{24}$$

This product is equal to unity when

$$\omega\tau_1 = 2.83 \tag{25}$$

$$\text{so} \quad f_n = 0.45/\tau_1. \tag{26}$$

The values of f_n obtained from Equation (26) are even higher than the values given in Table II. The cathode-to-grid transit time for a tube such as Type 6AK5 is of the order of 7×10^{-10} seconds, so the value of f_n from the above equation is

$$f_n = 0.064 \times 10^{-10} \text{ cycles} = 640 \text{ megacycles.}$$

The value obtained for Type 6AK5 from input conductance data (Table II) is 476 megacycles.

It appears that the only way to increase the frequency for a given noise factor with electron tubes of conventional design is to reduce the transit time. Triode types such as the 6J6, 6J4, and 2C43 are designed with close enough spacings and, consequently, short enough transit times to give promise of good results in equipment designed for minimum noise.

Effect of Circuit Losses

An important question with reference to the application of the curves and tables presented is the attainability of the circuit conditions assumed. The conditions are not hard to realize in practice, as the following examples illustrate:

1. Consider the use of Type 6AK5 as a pentode amplifier at 40 megacycles. The reference frequency f_n is 208 megacycles, so the ratio f/f_n is 0.192; the product $R_{eq}\, g_1$ is 0.037. The calculated noise factor is 3 decibels. The required ratio g_1/g_a is 0.046. Because the tube input conductance for 40 megacycles is 19.7 micromhos, the required antenna loading is 230 micromhos. For a tube input capacitance of 6 micromicrofarads, the quantity ωC is 1500 micromhos; because

the total conductance at the grid is 250 micromhos, the minimum value of Q is 6. Higher Q values may be obtained by adding more capacitance with appropriate inductance values. It is evident that there will be no serious increase in the noise factor until the conductance of the added elements becomes appreciable in comparison with 250 micromhos. If the Q is improved to 50 by addition of a resonant circuit with a Q of 200, the added conductance is approximately 60 micromhos. The noise factor would be increased from 3 to 3.5 decibels by the added circuit losses.

2. Consider the 6AK5 or an equivalent tube connected as a triode used at 200 megacycles. Neutralization may be used to avoid feedback, but feedback generally does not have an important effect on the question of obtainable noise factors. The reference frequency f_n is 476 megacycles; the ratio f/f_n is 0.42; the expected noise factors, from the two curves of Figure 5, are 5.1 decibels for no coherence, 2.3 decibels if the quadrature relation holds. In the first case, the required antenna loading is 2600 micromhos and the resulting Q for the input circuit is only 2. Adjustment of Q to any moderate desired value can be made by the addition of circuit elements as before without materially affecting the noise factor. In the second case, the antenna loading would be adjusted to equality with the tube conductance, which is 500 micromhos for this frequency. Then, the susceptance which must be added is 5700 micromhos, corresponding to a capacitance of 4.6 micromicrofarads.

CONCLUSIONS

The conclusions which may be drawn from this discussion may be summarized as a set of principles to be followed in the design of amplifiers for low noise.

1. Choose an input tube with low transit time. For frequencies above 30 megacycles, use a triode or a triode-connected pentode.

2. Adjust the input circuit with signal-to-noise ratio as the criterion. This adjustment is most readily made by using a noise generator, such as a diode, as a signal source.

3. Try the effect of detuning the input from resonance and the effect of increasing the coupling to the antenna beyond the value for maximum gain.

When theoretical considerations indicate a very low noise factor, it may be necessary to pay considerable attention to the design of the load circuit of the first tube and the input circuit for the second tube to obtain optimum results.

THE BRIGHTNESS INTENSIFIER*

By

G. A. Morton, J. E. Ruedy, and G. L. Krieger

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—By certain combinations of optical and electronic means it is possible to reproduce the image of a scene in greater brightness than that of the original. A general discussion is given as to the possible improvement in seeing under low light conditions by the use of such brightness intensifiers, and the fundamental limit to such improvement is shown to be the statistical fluctuation in the number of photons entering the eye. When the object viewed is a small luminous area or point of light, a gain in brightness may be obtained by optical means alone and the electronic intensification adds no further useful information. When a constant large angular field of view must be maintained, an optical system consisting of an objective of large diameter forming an enlarged image on the primary photosensitive screen of a brightness intensifier can be used to advantage, since the brightness intensifier permits passing along to the eye the gain in seeing resulting from the greater photon gathering power of the large objective.

An interesting application of the image intensifier would be in rendering more information available from medical fluorescent X-ray screens. It is shown that the information available is limited by the randomness in the X-ray beam itself, and the only significant gains would be in ease of viewing or in a more convenient size of the final image.

Introduction

MODERN electronic techniques have reached a point where it is possible to design equipment which will reproduce the image of a scene projected onto its pickup element in greater brilliancy than the original subject. Two examples of such equipment might be cited. The first and most obvious is a television pickup unit employing an image orthicon and reproducing a scene on the bright screen of a projection kinescope. The second is the brightness intensifier image tube as illustrated in Figure 1. Here the image is focused on the photocathode and the electrons released from it are imaged on a fluorescent screen as in an ordinary image tube. The fluorescent screen, however, is mounted on a thin, transparent membrane the back of which has been photosensitized. The light from the fluorescent screen excites photoelectrons from its photosensitized side. These electrons are accelerated and again imaged on a second fluorescent screen. This second fluorescent screen may be the viewing screen or

* Decimal Classification: R800 (621.375 × 535)

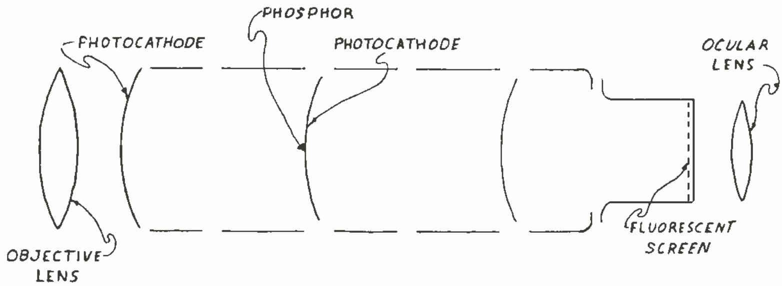


Fig. 1—Two stage brightness intensifier image tube.

may be another intensifier unit similar to the one just described. This procedure can be carried out for as many stages as desired, a brightness gain being obtained at each stage.

It is not the intention of the present discussion to consider in detail the design of any particular type of brightness intensifier, or to discuss the technical problems involved in its construction or operation. It is rather to determine what can be gained in the way of vision at lower light levels and under unfavorable seeing conditions by the use of a brightness intensifier. In order to do this, it will be necessary to examine the fundamental limits of the eye under low-light conditions and to compare them with similar fundamental limits which must necessarily exist in any type of brightness intensifier that might be built.

It has been shown that a fairly satisfactory agreement with the observed performance of the eye at low-light levels can be obtained by postulating a relatively simple model for the eye, namely, that it consists of an optical system forming an image of the scene before it upon the retina of the eye; that the latter consists of a photoactive surface with a quantum efficiency in the neighborhood of 10 per cent; that this surface is connected to the optic nerve leading to the brain; and that the information transmitted from this photoactive surface is limited only by the statistical fluctuations of the photo-effect existing at each of its elements. Such a model allows rather a direct comparison with an electronic brightness intensifier system consisting of an optical system forming an image on a photocathode (which, on the average, releases one electron for every λ photons falling on it) and some means for producing visible luminescence on the final viewing screen for each electron released at the photocathode.

It will be shown that if the optical system used with the brightness intensifier has the same numerical aperture and focal length as that of the eye, no brightness intensifier can make it possible to see at appreciably lower light levels than can be done with the unaided eye. However, if optical systems are used, taking advantage of the magnification properties, etc. which are possible in the brightness intensifier, very considerable gains may be expected and it is possible to see under conditions of illumination which are totally impossible with the unaided eye. The nature of these optical systems and their expected performance in conjunction with a brightness intensifier will be considered in some detail below.

Mechanism of Seeing

In order to obtain a quantitative understanding of the mechanism of seeing, consider the problem of visually differentiating a small element of area having a brightness B from its background which has a brightness B_o. Any luminous surface emits photons distributed in time in such a way that the average number of photons per unit area per unit time in a given solid angle is proportional to the brightness B of this surface. Thus, if the element under consideration has a size of h^2 square feet, the number of photons emitted by it per second per unit solid angle will be given by $n = \mathcal{K}h^2B$ where n is the number of photons per second and $\mathcal{K}$ the constant of proportionality connecting brightness and number of photons emitted ($\mathcal{K} = 1.3 \times 10^{16}$ photons per second per lumen.) An element of the background having the same area will correspondingly emit $n_o = \mathcal{K}h^2B_o$ photons per second.

The seeing device collects a certain fraction of these photons by means of its optical system and receives them on its photosensitive surface. The problem of seeing the element in question thus resolves itself into the ability of the seeing device to differentiate between n_o photons per second from an area the same size as the element under examination and the n photons per second from the test element. For simplicity, suppose that the seeing device is that illustrated in Figure 2. Here the element h is imaged on to the sensitive detector by a lens. This lens has a focal length f and subtends an angle α at the element l which is at a distance f from the lens. Conventional optics can provide the following relations between the brightness of the element and the number of lumens L falling on the image of this element formed on the detecting device.

$$\alpha \cong \frac{D}{f}; \quad l = \frac{f}{d}h; \quad L = Bl^2 \sin^2 \alpha \cong Bl^2\alpha^2; \quad L = Bl^2\frac{D^2}{f^2} = B \cdot \frac{h^2}{d^2}D^2. \tag{1}$$

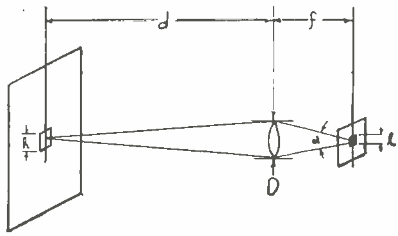


Fig. 2—Generalized seeing device.

Consequently, the number of photons received per second on the image of the element is given by

$$n = \mathcal{GH}L = \mathcal{GH}B \frac{h^2}{d^2} D^2. \tag{2}$$

Before proceeding further it is necessary to assign some properties to the sensitive element. For simplicity first the unrealistic assumption is made that the device is sufficiently sensitive to detect each separate photon. Furthermore, it will be assumed that the sensitive device integrates the effect of the photons over a time t corresponding to a persistence of vision. Based on this model, the image of the test element will receive

$$N = \mathcal{GH}B \frac{h^2}{d^2} D^2 t \tag{3}$$

units of signal during the period of integration. An element of equal area adjoining it will receive

$$N_o = \mathcal{GH}B_o \frac{h^2}{d^2} D^2 t \tag{3a}$$

units of information. The difference signal will, of course, be $N - N_o$. If the arrival of photons were not a random process $(N - N_o)$ might be a very small quantity and still be distinguishable. However, the statistics of random events indicates that the average N arriving at the image of the test element per integration interval will have superimposed on it a fluctuation whose root-mean-square value is proportional to the square root of the number of photons reaching this point. Unless the difference $N - N_o$ is larger than the fluctuation in N, it will be fundamentally impossible to detect the element in question. The extent to which $N - N_o$ must be greater than the square root of N depends upon the degree of certainty required in the determination that the difference actually exists. This conclusion can be easily modified to take into account the more realistic detector or seeing device whose quantum efficiency is not unity but has some fractional value λ. A quantum efficiency λ means that only one out of every $1/\lambda$ photons will produce a useful signal. Therefore, the relation indicating the possibility of distinguishing the test element now becomes

$$\lambda(N - N_o) \geqq \gamma(\lambda N)^{\frac{1}{2}} \tag{4}$$

where γ is the number of times the difference signal must exceed the root-mean-square fluctuation to insure detection. Referring back to Equations (3), (3a) and (4), the condition required for detection can be written as follows:

$$\Delta B \geqq \gamma B^{\frac{1}{2}} \left(\lambda \mathcal{H} \frac{h^2}{d^2} D^2 t\right)^{-\frac{1}{2}}.$$

Furthermore, if contrast be defined in the usual way as $C = \dfrac{\Delta B}{B}$

the following relation must hold $C \geqq \gamma \left(\lambda \mathcal{H} \dfrac{h^2}{d^2} D^2 t\right)^{-\frac{1}{2}} B^{-\frac{1}{2}}$

or $$C^2 \left(\frac{h}{d}\right)^2 \geqq K/B \tag{5}$$ where $K = \gamma^2 (\lambda \mathcal{H} D^2 t)^{-1}$.

K is a quantity which depends only on the character of the optical system and the properties of the photosensitive detector. This derivation shows that for a given element size and brightness level there is a fundamental lower limit to the contrast step that can be detected. Furthermore, for optimum performance the seeing device should integrate or count the whole area being resolved rather than attempting to discriminate the difference in brightness by evaluating the number of photons arriving on some constant small area, e.g. an area corresponding to the limit of its maximum resolving power. It is interesting to note that the eye is found to have just this property as will be discussed later.

Equation (5) suggests that it might be convenient to name as a figure of merit a small element defined as follows:

$$M' = \frac{1}{C^2 \left(\dfrac{h}{d}\right)^2} = \frac{1}{C^2 \Omega}$$

where Ω is the solid angle subtended at the eye by the test element. With the aid of Equation (5), M' can be expressed as follows:*

$$M' = B \times \frac{\lambda \mathcal{H} D^2 t}{\gamma^2}. \tag{6}$$

* The symbol M will be reserved for the efficiency of seeing in the case of an extended image where angle of view cannot be sacrificed.

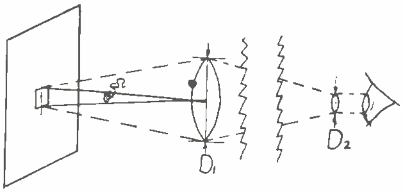


Fig. 3—Optical means for improving seeing.

Thus, it is evident that in addition to brightness, quantum efficiency, and γ, the figure of merit M' depends upon the diameter of the viewing lens or objective. Where the seeing device is the unaided eye, this diameter is simply that of the pupil. It is obviously possible to increase M', the figure of merit by the purely optical means illustrated in Figure 3. The improvement in the value of M' can be calculated as follows:

$$\frac{C_1^2 \dfrac{h_1^2}{d^2} D_1^2 t}{C_2^2 \dfrac{h_2^2}{d^2} D_2^2 t} = 1; \qquad \frac{M_2'}{M_1'} = \frac{D_2^2}{D_1^2}. \tag{7}$$

From Equation (7) it will be seen that M' is increased by a factor equal to the square of the ratio of the diameter of the new objective to that of the pupil of the eye. This is the principle upon which night glasses are based. Further, it is evident from the above that no form of brightness intensifier can improve the value of M' over that obtained by the eye using an optical system having the same absolute diameter (and some absorption losses) unless the quantum efficiency of the photosensitive detector of the brightness intensifier exceeds that of the retina of the eye.

Performance of Various Seeing Devices

A great deal of work on the part of many scientists[1] over a period of almost half a century has been done in the field of vision and the human eye in order to evaluate the signal-to-noise ratio and quantum efficiency of this very efficient seeing device. The quantum efficiency on the basis of these investigations has been estimated to lie between 1 and 10 per cent with 5 per cent as a good working value.

On the other hand for a photosensitive surface such as might be used in a brightness intensifier, a fair value of the photoelectric yield is 50 to 100 microamperes per lumen. Inasmuch as 1 lumen is constituted of some 10^{16} photons per second, the quantum efficiency of the photoelectric surface in question is of the order of 5 per cent. Since

[1] A. Rose, "The Sensitivity Performance of the Human Eye on an Absolute Scale", *Jour. Opt. Soc. Amer.*, Vol. 38, No. 2, pp. 196-208, February, 1948.

the quantum efficiency of the present known photoelectric surfaces does not exceed that of the eye, the type of brightness intensifier under consideration has no advantage over the unaided eye working with a suitable optical system for viewing a small element. Thus, if one wishes to see a star which is invisible to the unaided eye, one uses a telescope of large aperture (e.g. a 200-inch telescope) and views the object directly.

Figure 4 is a curve[2] indicating the performance of the eye relative to the performance of an ideal viewing device (quantum efficiency 1) together with the performance of an image orthicon television pickup tube, super XX pan photographic film, and a hypothetical brightness intensifier. The ordinate of this curve is the figure of merit for a single element while the abscissa is the scene brightness. It will be noted that the performance of the brightness intensifier and the eye parallel one another over a considerable range, but that there is a threshold to the brightness intensifier. This threshold is due to the background illumination of the viewing screen of the brightness intensifier. Some form of background is rather fundamental to all types of brightness intensifier. For the intensifier illustrated in Figure 1 for example, it is the result of thermionic emission from the first photoelectric surface.

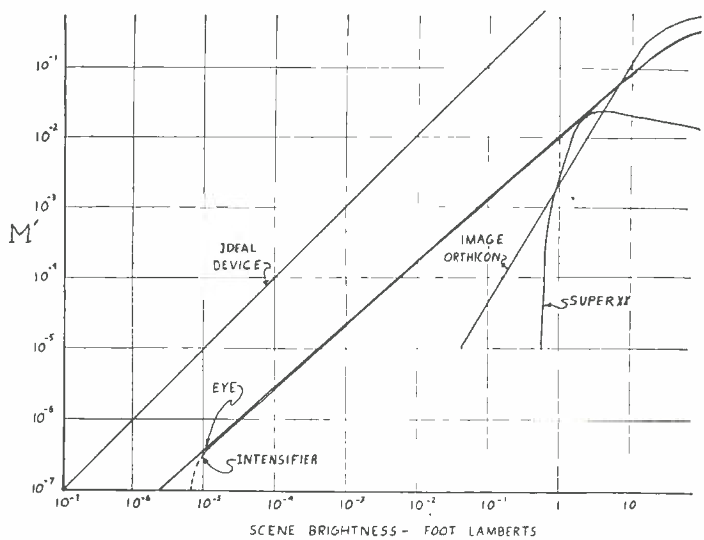


Fig. 4—Performance curves of various seeing devices.

[2] A. Rose, "Television Pickup Tubes and the Problem of Vision", ADVANCES IN ELECTRONICS, Academic Press Inc., New York, 1948 (in press).

The thermionic background can be reduced to a minimum (and possibly well below the background of the eye) by properly choosing the long wavelength threshold of the photoemitter, by cooling the photocathodes, etc., but it cannot be reduced to zero in a practical device.

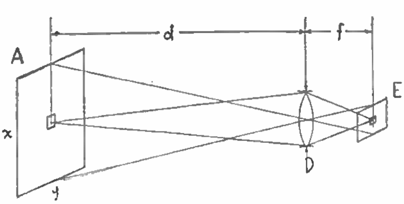
In the case of a brightness intensifier in the form of a television system, the d-c or average value of the background can be reduced by properly biasing the viewing tube, but the fluctuations in this background cannot be eliminated and, therefore, represent a threshold.

Wide-Angle Vision

When considering the question of seeing an extended image where it is important to retain a large angular field of vision, the problem is quite different. It is evident from the above that the optical methods used to permit seeing a small element under condition of illumination such that it cannot be seen by the unaided eye required an optical system which enlarged the image of the element on the sensitive area and, therefore, decreased the angular field of vision. Since this conclusion is fundamental for any type of optical system, it is not possible to lower the threshold for viewing an extended object, where the angle of vision must be kept constant, by purely optical means.

In considering the case of viewing an extended object, a line of reasoning is used similar to that followed above. A schematic diagram of a generalized seeing device is illustrated in Figure 5. Where this seeing device is the eye, the sensitive element is the retina while the lens system is the pupil and iris. This assumes that an extended object viewed is at a distance d from the device and covers an area A which completely fills the angle of view of the eye. It can be assumed to be made up of small elements of area h^2, each varying in brightness and thus making up the picture elements of the scene. From the optical arrangement of the elements of this system, the following relations are obvious:

$$L = B\left(\frac{hf}{d}\right)^2 \sin^2\alpha \cong B\left(\frac{hf}{d}\right)^2 \frac{D^2}{f^2}$$



but $E = A\dfrac{f^2}{d^2}$ and

$$L = B\frac{h^2}{A}E\frac{D^2}{f^2} \qquad (8)$$

Fig. 5—Seeing device for wide-angle vision.

where A is the area of the scene and E the area of the sensitive surface

$$\text{or} \qquad L = B \frac{\Omega}{\phi} E \frac{D^2}{f^2} \tag{8a}$$

where Ω is the solid angle of a test element and ϕ the solid angle of view.

Proceeding as in the case of viewing a single element, the condition which must be satisfied in order to insure the differentiation of a particular picture element of the scene from its neighbors is found to be:

$$\frac{1}{C^2} \geqq \mathscr{H}\gamma\lambda \frac{\Omega}{\phi} E \frac{D^2}{f^2} t\, B \tag{9}$$

$$\text{or} \quad \frac{\phi}{C^2\Omega} = KB \qquad \text{where} \quad K = \mathscr{H}\gamma\lambda E \frac{D^2}{f^2} t.$$

This leads logically to a figure of merit M for viewing an extended image related to the scene in question and the seeing device as follows:

$$M = \frac{\phi}{C^2\Omega} = \frac{A}{C^2h^2}.$$

Thus from Equation (9):

$$M = \frac{D^2}{f^2} \times Et \times \mathscr{H}\gamma\lambda. \tag{10}$$

Clearly M is a measure of the number of image elements that can be seen and the contrast between the elements.

Improvement in Seeing Using Brightness Intensifier

Considering again the case where the seeing device is the human eye, the first question to be asked is can the figure of merit for viewing an extended image, wherein ϕ is fixed, be improved simply by optical aids. It will be noted from Equation (10) that the only two optical factors which enter into the expression for the figure of merit M are the F number (i.e., $F = D/f$) of the lens and the area E of the viewing device. Since in the case where the viewing device is the eye, the value of E is fixed by the size of the useful area of the retina, the optical system associated with the eye as an aid to seeing may not produce, in combination with the lens of the eye, a focal length longer than that for the normal eye. Also, since the iris of the eye will be the aperture stop of the system, the F number or numerical aperture of the system cannot be greater than that of the unaided eye. Consequently, no purely optical device can be devised which will lower the brightness threshold for seeing without decreasing the effective angle of vision.

Where the primary surface upon which the optical system forms the image is the photocathode of a brightness intensifier, the situation is quite different. No longer is there a limit to the area of this sensitive surface nor is there any a priori limitation on the numerical aperture of the optical system. Therefore, it is possible with the brightness intensifier and optical system illustrated schematically in Figure 6 to surpass greatly the unaided eye in its ability to see under low levels of brightness.

With this arrangement, the larger diameter of the lens as compared to that in the eye permits the collection of a greater number of photons from the element being viewed. At the same time the larger area of the sensitive surface does not require any reduction in the angle of vision in spite of the greater focal length of the lens. The enhanced image appearing as output of the image intensifier is just bright enough so that the number of photons entering the pupil of the eye from an image element is equal to the number of photons entering the larger aperture of the objective of the system from the corresponding element in the object being viewed. The observer's eye is placed at such a distance from the final image on the intensifier screen that it subtends the same angle of view as would have the original scene when viewed with the unaided eye. The amount of useful image brightness intensification which should be produced by the brightness intensifier is equal to the ratio of the area of the objective of the device to the area of the pupil of the eye.

In the foregoing it was assumed that the size of the image reproduced by the brightness intensifier was equal to that of the image focused on its sensitive surface. Under these circumstances, the total number of lumens from the reproduced image must be equal to the product of the number of lumens falling on the photocathode and the ratio of areas of the objective lens and pupil of the eye.

However, the brightness intensifier system can be arranged in such a way that the reproduced image is smaller than the image formed on its photocathode, the reproduced image being viewed through a magnifier. This arrangement is shown schematically in Figure 7. When a system such as this is used, the same brightness relationships hold between the image on the photocathode and the reproduced image. However, since the area of the latter is now smaller, the required number of lumens put out by the device per lumen incident is smaller by the ratio of the area of the primary image to that of the reduced final image. Indeed it can be shown that even with an intensifier which

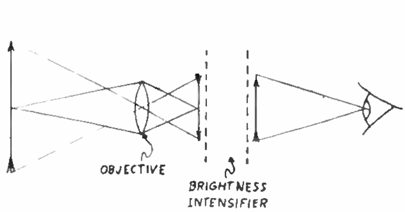


Fig. 6—Brightness intensifier used to improve seeing.

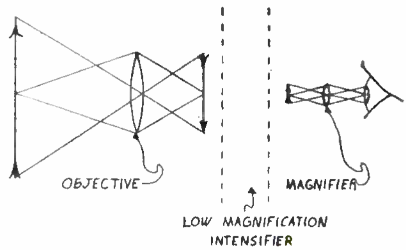


Fig. 7—Low magnification brightness intensifier.

simply produces one lumen of light per lumen of light incident can give any degree of lowering of the brightness threshold of vision limited only by the size of the objective forming the primary image. There are also certain practical limitations encountered in the design of the magnifying ocular for viewing the reproduced image.

In practice it is found most expedient to obtain most of the threshold lowering through the lumen gain of the brightness intensifier, but at the same time to use some reduction in size of the reproduced image as compared to that of the original image.

To summarize the foregoing, it is possible to lower the limit of brightness for threshold vision in the case of an extended object, where the angular field of view is predetermined, by the use of an image intensifier. (It is not possible to accomplish this by means of a passive optical system alone.)

The extent of the lowering of the threshold for vision is determined by the ratio of the area of the viewing lens to the area of the pupil of the dark-adapted eye.

Limitations in Use of Brightness Intensifier

The preceding discussion raises a number of points which require further consideration. First, the lowering of the threshold of vision by the use of the brightness intensifier as outlined above is not obtained without sacrificing some of the properties of the eye. Here the point of performance that is lost is depth of field. Simple optical considerations show that the depth of field is determined by the diameter of objective forming the image on the photocathode of the brightness intensifier, just as in the case of a photographic camera it is determined by the absolute diameter of the objective of the camera. Since the improved "signal-to-noise ratio" of seeing is only obtained through the use of a larger diameter objective, depth of focus is necessarily lost.

A second point of interest is the question of whether or not it is undesirable to use greater brightness intensification than is warranted by the "signal-to-noise" considerations of the over-all device. As was pointed out earlier in the discussion, one of the properties of the eye is that of integrating picture elements in such a way that the area over which photons are integrated increases as the number of photons per unit area decreases. In other words, as the "signal-to-noise" ratio gets poorer at the retina, its resolution decreases in such a way as to maintain a balance between contrast perception and definition. If the brightness intensification of the device is too great, the resolution of the eye will not adjust itself to match the information available in the image end, therefore, the eye will not be able to see as well as though a smaller brightness intensification had been used. This effect is not altogether uncommon. For example, one place where it may be encountered is where a halftone picture is too greatly enlarged. Such a picture when held at the normal viewing distance, so that every detail that is in the reproduction can be resolved by the retina, is almost unintelligible. However, when viewed from across the room, or at such a distance that the individual dots making up the halftone screen are not resolved, the picture becomes intelligible and gives the information that it was intended to give. Similar effects can be obtained when an ordinary photographic negative is over-enlarged or when the gain control on a television receiver is set too high. However, it might be mentioned that the deleterious effect of over-brightness intensification increases slowly with brightness and it is probably desirable to operate a practical instrument with some over-brightness intensification inasmuch as it tends to increase the comfort of the observer.

APPLICATIONS

Among a number of applications of the brightness intensifier, the use of the device to enhance the brightness of an image from the fluorescent screen used for X-rays warrants specific consideration. Modifications of the brightness intensifier for this purpose are shown in Figures 8(a) and 8(b). In Figure 8(a) the image of the fluorescent screen is reproduced on the sensitive element of the brightness intensifier by a fast objective. With this arrangement, the image on the intensifier sensitive surface may be reduced or magnified as desired. The second form shown in Figure 8(b) is that where the fluorescent screen and photocathode are essentially the same surface. This gives optimum optical coupling between the photocathode and fluorescent screen, but obviously permits only unity magnification.

Considerations relating to the signal-to-noise ratio that can be obtained from a fluorescent image and reproduced in the final image of the intensifier are exactly the same as the general relations given

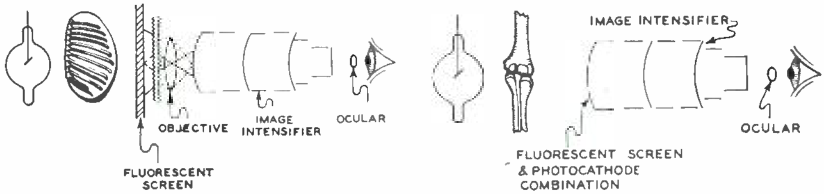


Fig. 8(a) Brightness intensifier used with x-ray screen. Fig. 8(b)

above. However, another aspect must be considered in this connection. This is that the X-ray photons each possess considerable energy and a relatively small number per unit area per second is required to produce visible luminosity in the fluorescent screen. For example, a single photon of X-rays corresponding to 60 kilovolts releases several thousand photons of visible light from an efficient fluorescent screen. Such a flash from a point source should be visible to the eye when the point is viewed with a 10-power magnifier. Under these circumstances, the amount of information seen on the fluorescent screen used with an X-ray tube is not necessarily limited by the low-level of luminosity of the screen itself but rather by the small number of X-ray photons reaching a unit area of the screen per second. Where the information is thus limited, a brightness intensifier cannot increase the information obtained from the screen. When it is desired to observe the image at the normal viewing distance an image intensifier may be used to maintain the brightness level at such a value that the eye usefully absorbs approximately one light photon for each X-ray photon. A rough estimate indicates that the useful brightness intensification that can be used with an X-ray screen is ten times or less. Some further intensification may increase convenience of viewing but intensification beyond this will cause an actual loss in information that can be seen.

It is, however, possible to use the brightness intensifier to advantage in connection with X-rays. This is when the arrangement shown in Figure 8(a) is used and the optical system produces a considerably reduced image on the sensitive area of the brightness intensifier. With such arrangement, a reduced fairly bright image of, for example, the entire chest and abdomen of the subject examined can be produced. This reduced image can be made bright enough so that the eyes of the observer function under conditions of photopic vision so that visual acuity is high. However, if the size of the image is correctly adjusted, the normal limitations of the perception will be just sufficient so that the eye could not attempt to glean more information from the image than is present in the original image on the X-ray fluorescent screen.

Where the brightness intensifier is used simply to reduce the brightness required for seeing, very real gains can be obtained if appropriate optical systems are used. For example, it is quite feasible to make the optical objective used with the brightness intensifier a Schmidt system with an aperture 10 inches in diameter. Since the diameter of the dark adapted eye is about ¼ inch, this means that only 1/1600 of the amount of light required for the unaided eye will give equivalent vision when the brightness intensifier is employed. This figure is probably optimistic because of inevitable optical inefficiencies and certain losses in the brightness intensifier itself. However, it can be predicted with a good deal of assurance that such a system would result in the lowering of the brightness required for equivalent vision by two or perhaps three orders of magnitude.

CONCLUSION

Experimental studies of the performance of the eye show that the limit of contrast and detail perception of the eye, working under conditions of low-level illumination, is determined by the limiting signal-to-noise ratio resulting from the fundamental statistical fluctuations in the photo effect produced by photons entering the optical system of the eye. No system of brightness intensification using an optical system equivalent to that of the eye and having a photosensitive primary element with a quantum efficiency no greater than that of the eye can hope to enhance the performance of the eye. Where the object viewed is a small luminous element and the angle of view is not important, considerable gain in sensitivty is possible through optical means. However, for this purpose a magnifying system is used having an objective which is larger in diameter than the pupil of the eye. Such a system produces an enlarged image of the object being seen on the retina of the eye and at the same time decreases the angle of vision. Where it is necessary to retain a constant large angular field of view, no passive optical system can give improved seeing. Here, however, an optical system consisting of an objective of large diameter forming an enlarged image on the primary photosensitive screen of a brightness intensifier can be used to advantage since the brightness intensifier permits passing along to the eye the gain in seeing resulting from the greater photon gathering power of the large objective system used. It might, therefore, be said in conclusion that the gain in seeing which can be achieved by a brightness intensifier is not the result of the action of the brightness intensifier itself, but actually the result of the greater absolute aperture of the objective which can be used when a brightness intensifier is intermediate between the objective and the observer's eye.

THEORETICAL ANALYSIS OF VARIOUS SYSTEMS OF MULTIPLEX TRANSMISSION*†

BY

VERNON D. LANDON

Research Department, RCA Laboratories Division,
Princeton, N. J.

NOTE: The first portion of this paper was published in the June 1948 issue of *RCA Review* (Vol. IX, No. 2) and contained: Complete Summary; Introduction (including Outline, Specification of Systems, Definitions and List of Symbols); Signal-Noise Analysis. (A limited number of copies of the June 1948 issue of *RCA Review* is still available for those desiring a complete file on this multiplex transmission analysis.) The remainder of this paper which is published herewith includes:

Minimum Signal Requirements; Bandwidth Economy; Impulse Noise, Cross Modulation and Interference Analysis; Propagation Study; Experimental Results.

The Manager, RCA Review

MINIMUM SIGNAL REQUIREMENTS

Thresholds of the Various Systems

The threshold of any system is the received power required to realize the wideband improvement and is equal to,

$$P_t = KTBF_nC_f \tag{57}$$

where K = Boltzmanns constant

$= 1.37 \times 10^{-23}$ joules per degree Kelvin

T = absolute temperature

= 293 degrees approximately

B = the effective noise bandwidth

F_n = the noise factor[7] of the receiver, usually about 12 decibels or 16 to 1 power ratio for frequencies above 500 megacycles

C_f = the ratio of signal power to noise power at threshold for a particular system.

= 8 for AM, AM-AM, FM-AM, and SS-AM

---

* Decimal Classification: R460 × R148.

† This paper summarizes work done under contract with the U. S. Army Signal Corps.

[7] For a definition and general discussion of noise factor (or noise figure) see H. T. Friis, "Noise Figures of Radio Receivers", *Proc. I.R.E.*, Vol. 32, No. 7, page 419, July, 1944.

$= 8$ for AM-PM, FM-PM, and SS-PM

$= 8Dn$ for PAM-AM

$= 32Dn$ for PPM-AM, PNM-AM, PWM-AM

$= 8$ for PAM($\pm$)FM, PPM-FM, PNM-FM, PWM-FM*. (58)

$nD =$ duty factor (of a pulse system)

$KTB =$ the available noise power from the antenna

$= 0.40 \times 10^{-20}$ watts per cycle of bandwidth times B

$KTBF_n =$ the noise from the whole receiver expressed as the equivalent noise applied through the antenna.

$= 6.40 \times 10^{-20}$ watts per cycle times B (59)

$P_t = 0.512 \times 10^{-18}$ watts per cycle times B (60)
(for AM, AM-AM, FM-AM, and SS-AM, also for AM-PM, FM-PM, SS-PM and SS-PM-FM)

$P_t = 0.512 \times 10^{-18}$ watts per cycle times nDB for PAM-AM

$P_t = 2.05 \times 10^{-18}\ nDB$ for PPM-AM, PNM-AM, PAM-PCM-AM and PWM-AM

$P_t = 0.512 \times 10^{-18}\,B$ for PPM-FM, PNM-FM, PAM-PCM-FM and PWM-FM

$P_t = 0.512 \times 10^{-18}\,B$ for PAM ($\pm$)-FM

In PAM-AM and PPM-AM, $nD = \dfrac{3n\,f_m}{B}$ while for PNM-AM and PAM-PCM-AM, $nD = \dfrac{21\,nf_m}{B}$. Thus P_t is independent of B for these systems, providing B exceeds the minimum value for system operation.

Comparison of Systems on the Basis of the Received Signal Power Required for a Given Signal Noise Ratio in the Output Circuit

The previous analysis of S/N ratios was made to compare the sys-

---

* This is on the basis of threshold being when the highest noise peaks equal carrier amplitude for both AM and FM. Some engineers use a 2 to 1 factor for FM, stating that threshold is where noise peaks have an amplitude ½ that of the carrier. Others make use of a 1 to 1 ratio as used here. It is true that for a 1 to 1 ratio of signal to noise some very high modulation frequencies are generated in the limiter. However, these are mostly outside the pass band, and the time duration of the disturbance from a single pulse is so short that the response in the audio-frequency band is minute. The balanced discriminator also helps to minimize such disturbances except when they occur on modulation peaks. The 1 to 1 peak factor, or 8 to 1 on a power basis will be used in this report.

tems for signal noise ratio with a given bandwidth, assuming that all systems are operated above the threshold level. This is satisfactory if the maximum bandwidth is given and the best signal noise ratio at a given range is desired.

On the other hand if it is desired to obtain maximum range with a specified signal noise ratio, then the comparison must be set up a little differently. Most systems have the characteristic that the signal noise ratio may be improved, at the expense of a poorer threshold, by widening the bandwidth. To obtain maximum range with a given signal noise ratio or better, the bandwidth of the system must be adjusted so that the signal noise ratio at threshold is equal to the specified minimum value. If the bandwidth is wider than this value the range is reduced because the threshold has a higher value. If the bandwidth is narrower, the range is reduced because the signal noise ratio is lower at all points.

To make the comparison on this basis it is necessary to,

(1) Obtain the formula for the signal noise ratio at threshold.

(2) Set this equal to the desired minimum signal noise ratio.

(3) Solve for the bandwidth.

(4) Solve for the threshold at that bandwidth for the particular System.

For the pulse systems (PAM-AM), (PPM-AM) and (PNM-AM), the threshold is independent of the bandwidth. In these cases the only reason for adjusting the signal noise ratio to the minimum acceptable value at threshold is to economize on bandwidth.

In a large number of cases, the bandwidth required to obtain the desired minimum signal noise ratio at threshold, is so great as to be impractical. In view of this, it seems reasonable to set a maximum value of bandwidth. The figure chosen is 10 megacycles. A somewhat larger figure could be used but the gain per stage would be quite low in the radio-frequency and intermediate-frequency amplifiers. The powers required for other bandwidths can be readily obtained as explained later.

System Characteristics Required

In order to obtain a comparison under widely varying conditions, the system types will be compared on the basis of four different sets of specifications.

1. A 4 channel system

Signal to noise ratio required (including cross talk) (peak)	40 decibels
Assumed signal to noise ratio required on random noise (peak)	46 decibels
Equivalent signal noise ratio on an r-m-s basis	55 decibels

2. A 12 channel system

Signal to noise ratio required (including cross talk) (peak)	50 decibels
Assumed signal to noise ratio required on random *noise* (peak)	56 decibels
Equivalent signal noise ratio on an r-m-s basis	65 decibels

3. A 48 channel system

Signal to noise ratio required (including cross talk) (peak)	60 decibels
Assumed signal to noise ratio required on random noise (peak)	66 decibels
Equivalent signal noise ratio on an r-m-s basis	75 decibels

4. A 72 channel system

Signal to noise ratio required (including cross talk)	60 decibels
Assumed signal to noise ratio required on random noise	66 decibels
Equivalent signal noise ratio on an r-m-s basis	75 decibels

Type of Summation to Be Used

In all of the frequency division systems two equations are presented, one for linear summation and one for root-mean-square (r-m-s) summation. For AM-XX, and FM-XX where the sub-carriers are not suppressed the linear summation equations should be used if n is small, and the r-m-s summation equation when n is large. For the following minimum signal calculations the linear equations will be used for $n = 4$ and the r-m-s for $n = 12$, 48, and 72.

While it seems probable that it is slightly optimistic to use r-m-s summation for $n = 12$, the error cannot be great. For FM-AM the r-m-s, and linear summation equations give the same result for $n = 8$. For $n = 12$ the indicated value of S/N is greater for r-m-s summation than for linear in the ratio of 1.2 to 1. The true answer must lie between these two values. A greater possibility for error lies in the two AM-XX systems where allowance is made for channel modulation in the linear summation equations and not in the r-m-s summation equations. This probably makes the results slightly optimistic for these two systems for $n=12$, and somewhat pessimistic for $n=4$ where the linear summation equations are used.

For SS-XX, the sub-carriers are suppressed and the proper type of summation depends on the type of service being transmitted. In particular it depends on the crest factor of the modulation of each channel. For 2-way telephone service the crest factor in each channel is greater than for a large number of channels. Hence the signal noise ratio is always somewhat better than the r-m-s summation equations indicate and the linear equations should never be used. The amount that the signal noise ratio is better than indicated by the r-m-s equations has been calculated from data available.[8]

The values are as follows:

n	*improvement factor*		
4	2 decibels or	1.26	voltage ratio
12	4	1.58	
48	7	2.24	
72	8	2.5	

The value approaches 12 decibels or 4 to 1 for very large values of n. For these values the assumption is made that the volume is adjusted to the same average level in each channel. If this is not done the improvement factor is slightly larger especially for large values of n. The improvement factor is not included in later calculations because the type of service is not specified.

The Improvement Factors on SS-XX Systems Derived

The above factors were obtained from Reference 8 in the following manner.

Curve A of the figure is a plot of the r-m-s test tone capacity required, plotted as a function of the number of channels. Volume in each channel is assumed controlled to the level —12.1 decibels. Curve D is linear at 3 decibels per octave and is asymptotic to Curve A, and hence D is a plot of the test tone capacity which would be required if the complex signal (all channels added) had a crest factor equal to that for random noise. (This is true because we know that the crest factor is the same as for random noise when n is very large.)

Extrapolating curves A and D to $n=1$, it can be seen that the two intercepts are at +6 decibels and —6 decibels, indicating that the crest

[8] J. T. Dixon, "Load Rating Theory for Multichannel Amplifiers," *Bell Sys. Tech. Jour.*, Vol. 18, p. 638, October, 1939 (Figure 7).

factor for a single channel is 12 decibels greater than for random noise. Let a line E be drawn parallel to D through the point +6 decibels for 1 channel. This line E represents the test tone capacity that would be required if the crest factor on the multichannel signals were the same as for a single channel.

As stated above the values of curves A and D for 1 channel shows that the crest factor for a single average channel is 12 decibels greater than the crest factor of random noise. (Presumably this includes 6 decibels for the difference between an active channel and an average channel, since this latter is active 25 per cent of the time. Then the remaining 6 decibels is the increase in crest factor of an active channel over that of random noise.)

Since the crest factor of random noise is 12 decibels, a transmitter will be modulated 100 per cent by the highest peaks of this noise if the r-m-s noise voltage is 12 decibels below the *peak* value of a voltage giving 100 per cent modulation. Then the r-m-s voltage of a single average signal that is to modulate a subcarrier must be kept 24 decibels below the peak voltage giving 100 per cent modulation.

Thus if there were only a single channel and it were modulated by this hypothetical average signal, the level of this latter must be set to the —24-decibel value just given. One might assume that for n channels each must be set to this same level. In fact, Equations (18) and (20) assume that the crest factor for n channels is the same as for one channel. Actually the crest factor decreases from 24 decibels for 1 channel to 12 decibels for a very large number. The difference between curves E and A represents the magnitude of the decrease for any number of channels. This difference is the source of the factors quoted. The factors represent the amount by which the modulator volume control can be advanced over that allowable when a constant crest factor (equal to that for 1 channel) is assumed.

The above applies accurately to SS-AM systems. For SS-PM, the analysis does not apply exactly because the frequency deviation caused by the various channels is different. This means that, for the purpose of solving for the improvement factor, the effective value of n is less than the actual value, so that the improvement factors tabulated above may be slightly too large.

For one-way telephone, the improvement factor is reduced by not more than 3 decibels, and less for small values of n.

Multiplexing Telegraph Signals

It is a general practice to utilize (under some circumstances) each audio channel of a multiplex system for multiplexing many telegraph channels. These telegraph channels may be either AM or FM on audio frequency sub-carriers. It is believed that the number of such channels will never be few enough to warrant the use of linear summation in the SS-PM system. For the AM case the r-m-s value of a single telegraph channel is its peak value times $\frac{D_1}{2}$ where D_1 is the duty factor of the telegraph signal. For the FM case the r-m-s value of a single channel is the peak value divided by 2.

In case a noise measuring set such as the Western Electric Type 2B is used, the appropriate weighting factor (for the changed fidelity) should be used.

Calculation of Minimum Signal Power Required

The value of the minimum signal power required to produce the desired signal noise ratio R_r may be calculated from the formula,

$$P_r = P_t \frac{R_r^2}{R_t^2}. \tag{61}$$

Where $P_t =$ the power required in the signal wave at threshold. (See Equation (60).)

$R_r =$ the desired signal-noise voltage ratio (peak).

$R_t =$ the signal noise voltage ratio in each channel of the output circuit of the system under discussion when operated at threshold $= R_o R_a$.

$$R_o = \frac{\text{S/N (XX-XX)}}{\text{S/N (AM)}}$$

$R_a =$ S/N (AM) (peak) at the signal strength required for threshold in the XX-XX system.

$$= \left(\frac{B}{2f_m}\right)^{1/2} \left(\frac{C_f\ \text{(XX-XX)}}{C_f\ \text{(AM)}}\right)^{1/2} \text{(See Equation (58).)}$$

As pointed out in the last section, for systems in which the value of S/N increases with B, the lowest value of P_r is obtained when B is adjusted to the point where $R_t = R_r$. Then $P_r = P_t$.

Minimum Signal Required for (AM-AM)

For this system the signal noise ratio is independent of the bandwidth when above threshold. Hence there is no use in adjusting the threshold to occur at the point where minimum acceptable signal noise ratio is obtained. If the bandwidth has the critical value or less, then the signal power required to produce the desired minimum signal noise voltage ratio R_r is,

$$P_r = P_t \frac{R_r^2}{R_o^2 R_a^2}.$$

Where $P_t = 0.512 \times 10^{-18} 4n f_m$ (From Equation (60))

$R_r =$ the required signal noise ratio

$$R_o = \frac{\sqrt{2}}{4n} \quad \text{(For linear summation, see Table III)*}$$

$R_a =$ S/N (AM) at the signal strength required for threshold in the AM-AM system.

$$= \left(\frac{B}{2f_m}\right)^{1/2} \left(\frac{C_f \text{ (AM-AM)}}{C_f \text{ (AM)}}\right)$$

$$= \left(\frac{B}{2f_m}\right)^{1/2} = (2n)^{1/2} \text{ (See Equation (58))}; \quad \frac{1}{R_o^2 R_a^2} = 4n.$$

Then $\quad P_r = 0.512 \times 10^{-18}\, 4n\, f_m\, R_r^2\, 4n = 8n^2 R_r^2\, 2f_m\, 0.512 \times 10^{-18}$ watts

For a ratio of signal to combined noise and cross talk of 40 decibels, it seems reasonable to set 46 decibels as the limit for signal to random noise, or $R_r = 200$, $R_r^2 = 40{,}000$. The value of f_m is 3500 cycles per second and $n = 4$. Then P_r (for $n = 4$) $= 40{,}000 \times 7{,}000 \times 0.512 \times 10^{-18} \times 128 = 18{,}300 \times 10^{-12}$ watts.

For $n = 12$, the r-m-s summation equation is used so that,

$$R_o = \frac{1}{4n^{\frac{1}{2}}}; \quad R_o R_a = \frac{\sqrt{2}}{4}; \quad \frac{1}{R_o^2 R_a^2} = 8.$$

Substituting this value for $4n$ in Equation (61) gives,

$$P_r \text{ (for } n = 12) = R_r^2\, KT 2f_m\, F_n C_f \times 16n.$$

The new value of R_r^2 corresponding to 56 decibels is 400,000. Then,

$$P_r \text{ (for } n = 12) = 400{,}000 \times 7000 \times 0.512 \times 10^{-18} \times 192$$
$$= 275{,}000 \times 10^{-12} \text{ watts.}$$

For larger values of n and R_r, P_r varies directly with n and R_r^2, so that

* Part I, p. 350—*RCA Review*, Vol. IX, No. 2, June, 1948.

$$P_r \text{ (for } n = 48) = \frac{48}{12} \times \frac{4{,}000{,}000}{400{,}000} \times 275{,}000 \times 10^{-12}$$

$$= 40 \times 275{,}000 \times 10^{-12} = 11{,}000{,}000 \times 10^{-12} \text{ watts}$$

$$P_r \text{ (for } n = 72) = 1.5 \times 11{,}000{,}000 \times 10^{-12} = 16{,}500{,}000 \times 10^{-12} \text{ watts.}$$

Minimum Signal for (AM-PM)

For this system the required minimum signal noise ratio R_r may be set equal to the signal noise ratio at threshold,

$$R_t = R_o R_a = R_r \tag{62}$$

where $R_o = \dfrac{\text{S/N (AM-PM)}}{\text{S/N (AM)}}$ (See Equation (6) or (9))*

R_a = S/N (AM) at signal strength required for threshold in the AM-PM system =

$$\left(\frac{B}{2f_m}\right)^{1/2} \left(\frac{C_f \text{ (AM-PM)}}{C_f \text{ (AM)}}\right)^{1/2} = \left(\frac{B}{2f_m}\right)^{1/2} \text{(See Equation (58))}$$

Using the linear summation formula for R_o, (See Table III)

$$R_t = \frac{\sqrt{2}}{8} \frac{B}{n^2 f_c} \times \left(\frac{B}{2f_m}\right)^{1/2} = 200$$

$$B_t = (R_t\, 8n^2 f_c f_m^{\frac{1}{2}})^{2/3} = \text{bandwidth for (S/N} = R_r) \text{ at threshold} \tag{63}$$

$$B_t \text{ (for } n = 4) = (200 \times 8 \times 16 \times 5000 \times 59)^{2/3}$$

$$= 3.85 \times 10^6 \text{ cycles per second.}$$

For this value of the bandwidth the power P_r required for the required signal noise ratio (R_r) is equal to the power at threshold (P_t) and P_t (for $n = 4$) $= 0.512 \times 10^{-18} \times B_t = 1.97 \times 10^{-12}$ watts.

For $n = 12$ the r-m-s summation equation for R_o is used so that, $R_t = R_o R_a$ (See Table III)

$$= \frac{\sqrt{3}}{16} \frac{B}{n^{3/2} f_c} \left(\frac{B}{2f_m}\right)^{1/2}; \qquad B_t = \left(16 \sqrt{\frac{2}{3}} R_t f_c f_m^{\frac{1}{2}}\right)^{2/3} n \tag{64}$$

$$B_t \ (n = 12) = (16 \times .815 \times 630 \times 5000 \times 59)^{2/3} \times 12$$

$$= 21.8 \times 10^6 \text{ cycles per second.}$$

* Part I, pp. 310-311.

P_t (for $n = 12$) $= 0.512 \times 10^{-18} B_t = 11.2 \times 10^{-12}$ watts.

For larger values of n and R_t, P_t and B_t vary directly with $R_t^{2/3}$ and n, (See Equation (64)) so that,

B_t ($n = 48$) $= 21.8 \times 10^6 \times 4 \times 10^{1/3} = 188 \times 10^6$ cycles per second

B_t ($n = 72$) $= 188 \times 10^6 \times 1.5 = 282 \times 10^6$ cycles per second

P_t ($n = 48$) $= 0.512 \times 10^{-18} \times 188 \times 10^6 = 96 \times 10^{-12}$ watts

P_t ($n = 72$) $= 0.512 \times 10^{-18} \times 282 \times 10^6 = 145 \times 10^{-12}$ watts.

The bandwidths given above for $n = 12$, 48 and 72 are probably impractically great. For such cases it seems reasonable to assign an upper limit to the band width and solve for the received power required for the system at that bandwidth. Threshold will not then be involved. This bandwidth limit will be arbitrarily chosen as 10 megacycles, but the rate of variation of the required power with bandwidth will be specified in each case to make the calculation easy for other bandwidths. For the AM-PM system, the signal noise ratio varies directly with B. So at a given signal noise ratio, B and the carrier voltage are inversely proportional. Then the power required (for $n = 12$, $R_t = 630$, $B = 10$) is

$P_{10}(n = 12) = 11.2 \times 10^{-12} \times 2.18^2 = 53 \times 10^{-12}$ watts

$P_{10}(n = 48) = 96 \times 10^{-12} \times 18.8^2 = 34{,}000 \times 10^{-12}$ watts

$P_{10}(n = 72) = 145 \times 10^{-12} \times 28.2^2 = 116{,}000 \times 10^{-12}$ watts.

Minimum Signal for FM-AM

For this system, the S/N ratio compared to AM, for linear summation is,

$$R_o = \frac{\sqrt{6}}{8}\,\frac{B}{n^2 f_m} \quad \text{(See Table III)}; \quad R_a = \left(\frac{B}{2f_m}\right)^{1/2}$$

$$R_t = \frac{0.217\, B^{3/2}}{n^2 f_m^{3/2}} = R_o R_a \quad (65); \qquad B_t = \left(\frac{R_t\, n^2}{0.217}\right)^{2/3} \times f_m \quad (66)$$

$$B_t(n = 4) = \left(\frac{3200}{0.217}\right)^{2/3} \times 3500 = 2{,}100{,}000.$$

For this bandwidth $P_r = P_t$, and

$P_t(n = 4) = 0.512 \times 10^{-18} B_t = 1.08 \times 10^{-12}$ watts.

For $n = 12$ and larger the r-m-s summation equation is used for R_o.

The equation is the same as that for AM-PM except for the factor $\frac{f_m}{f_c}$. As a result of this factor all values of B_t and P_t (for $n > 12$) are smaller by the factor $\left(\frac{f_m}{f_c}\right)^{2/3}$ or 0.79. Then,

$$B_t \text{ (for } n = 12) = 0.79 \times 21.8 \times 10^6 = 17.2 \times 10^6 \text{ cycles per second}$$
$$B_t \text{ (for } n = 48) = 0.79 \times 188 \times 10^6 = 148 \times 10^6 \text{ cycles per second}$$
$$B_t \text{ (for } n = 72) = 0.79 \times 282 \times 10^6 = 222 \times 10^6 \text{ cycles per second}$$
$$P_t \text{ (for } n = 12) = 0.79 \times 11.2 \times 10^{-12} = 8.85 \times 10^{-12} \text{ watts}$$
$$P_t \text{ (for } n = 48) = 0.79 \times 96 \times 10^{-12} = 76 \times 10^{-12} \text{ watts}$$
$$P_t \text{ (for } n = 72) = 0.79 \times 145 \times 10^{-12} = 114.2 \times 10^{-12} \text{ watts.}$$

If a maximum bandwidth of 10 megacycles is set, the required received power can be calculated from the above using the fact that a given signal noise ratio, B and the carrier voltage are inversely proportional. Then,

$$P_{10}\ (n = 12) = 8.85 \times 10^{-12} \times 1.72^2 = 26.0 \times 10^{-12} \text{ watts}$$
$$P_{10}\ (n = 48) = 76 \times 10^{-12} \times 14.8^2 = 16600 \times 10^{-12} \text{ watts}$$
$$P_{10}\ (n = 72) = 114.2 \times 10^{-12} \times 22.2^2 = 56{,}300 \times 10^{-12} \text{ watts.}$$

Minimum Signal for (FM-PM)

$$R_o = \sqrt{\frac{3}{2}\,\frac{B}{2n2f_m}} \quad \text{for linear summation;} \qquad R_a = \left(\frac{B}{2f_m}\right)^{1/2}$$

$$R_t = \frac{\sqrt{3}}{4}\,\frac{B^{3/2}}{n^2 f_m^{3/2}}\,; \qquad B_t = \left(\frac{R_t\, n^2}{0.433}\right)^{2/3} f_m \qquad (67)$$

$$B_t(n = 4) = \left(\frac{3200}{0.433}\right)^{2/3} \times 3500 = 1.33 \times 10^6 \text{ cycles per second.}$$

For this bandwidth $P_r = P_t$ and,

$$P_t(n = 4) = 0.512 \times 10^{-18}\, B_t = 0.68 \times 10^{-12} \text{ watts}$$

For $n = 12$ or larger the r-m-s summation equation is used so that,

$R_t = R_o R_a$ (See Table III)

$$= \frac{3}{16} \frac{B}{n^{3/2} f_m} \times \left(\frac{B}{2f_m}\right)^{1/2} = \frac{3\sqrt{2}}{32} \frac{B^{3/2}}{n^{3/2} f_m{}^{3/2}}$$

$$B_t = (7.53\, R_t)^{2/3}\, n\, f_m \tag{68}$$

$B_t(n = 12) = (7.53 \times 630)^{2/3} \times 12 \times 3500 = 11.8 \times 10^6$ cycles per second.

Equation (68) shows that B_t varies as n and $R_t^{2/3}$. Then

$B_t(n = 48) = 11.8 \times 10^6 \times 10^{1/3} \times 4 = 102 \times 10^6$ cycles per second

$B_t(n = 72) = 102 \times 10^6 \times 1.5 = 153 \times 10^6$ cycles per second

$P_t(n = 12) = 11.8 \times 10^6 \times 0.512 \times 10^{-18} = 6.05 \times 10^{-12}$ watts

$P_t(n = 48) = 52.3 \times 10^{-12}$ watts

$P_t(n = 72) = 78.5 \times 10^{-12}$ watts.

If a maximum bandwidth of 10 megacycles is set, the required received power can be calculated from the above using the fact that at a given signal noise ratio, B and the carrier voltage are inversely proportional. Then

$P_{10}(n = 12) = 6.05 \times 10^{-12} \times 1.18^2 = 8.42 \times 10^{-12}$ watts

$P_{10}(n = 48) = 52.3 \times 10^{-12} \times 10.2^2 = 5{,}450 \times 10^{-12}$ watts

$P_{10}(n = 72) = 78.5 \times 10^{-12} \times 15.3^2 = 18{,}300 \times 10^{-12}$ watts.

Minimum Signal for SS-AM

In this system the signal noise ratio is not a function of bandwidth as long as the signal is above threshold. The minimum signal for the desired signal noise ratio R_t, (assuming r-m-s- summation) is,

$$P_r = P_t \frac{R_r^2}{R_t^2} = 0.512 \times 10^{-18}\ 2nf_c \frac{R_r^2}{R_o^2 R_a^2} \qquad \text{(See Equation (61))}$$

$$R_o = \frac{1}{n^{\frac{1}{2}}}; \qquad R_a = \left(\frac{B}{2f_m}\right)^{1/2} = \left(\frac{nf_c}{f_m}\right)^{1/2}$$

$$\frac{1}{R_o^2 R_a^2} = \frac{f_m}{f_c}; \qquad P_r = 0.512 \times 10^{-18}\, 2nf_m\, R_r^2$$

$P_r(n = 4) = 0.512 \times 10^{-18} \times 8 \times 3500 \times 200^2 = 574 \times 10^{-12}$ watts.

Since the r-m-s summation equation for R_o is used for all values of n, the values of P_t may be computed from the value for $n = 4$ assuming that it varies directly with n and R_t^2.

$P_r(\text{for } n = 12) = 574 \times 10^{-12} \times 3 \times 10 = 17220 \times 10^{-12}$ watts

$P_r(\text{for } n = 48) = 690{,}000 \times 10^{-12}$ watts

$P_r(\text{for } n = 72) = 1{,}040{,}000 \times 10^{-12}$ watts.

The bandwidth required to transmit the signal is $2nf_c$ in each case or,

$B_t(\text{for } n = 4) = 40{,}000$ cycles per second

$B_t(\text{for } n = 12) = 120{,}000$

$B_t(\text{for } n = 48) = 480{,}000$

$B_t(\text{for } n = 72) = 720{,}000$.

Minimum Signal for (SS-PM)

$$R_o = \frac{\sqrt{3}}{2} \frac{B}{n^{3/2}f_c}; \quad R_a = \left(\frac{B}{2f_m}\right)^{\frac{1}{2}}$$

$$R_t = \frac{\sqrt{3}\ B^{3/2}}{2\sqrt{2}\, n^{3/2} f_c f_m^{\frac{1}{2}}} = 200; \quad B_t = \left(\frac{400\sqrt{2}}{\sqrt{3}} n^{3/2} f_c f_m^{\frac{1}{2}}\right)^{2/3} \qquad (69)$$

$$B_t(n = 4) = \left(\frac{3200\sqrt{2}}{\sqrt{3}} \times 5000 \times 59\right)^{2/3} = 0.84 \times 10^6.$$

For this bandwidth $P_r = P_t$, and

$$P_t = 0.512 \times 10^{-18} B_t = 0.512 \times 10^{-18} \left(\frac{2\sqrt{2}\, R_t\, n^{3/2} f_c f_m^{\frac{1}{2}}}{\sqrt{3}}\right)^{2/3}$$

$P_t(\text{for } n = 4) = 0.512 \times 10^{-18} \times 0.84 \times 10^6 = 0.430 \times 10^{-12}$ watts.

The values for larger values of n and R_t may be calculated directly remembering that B_t and P_t vary directly with n and $R_t^{2/3}$.

$B_t(\text{for } n = 12) = 0.84 \times 10^6 \times 3 \times 10^{1/3} = 5.4 \times 10^6$ cycles per second

$P_t(\text{for } n = 12) = 2.77 \times 10^{-12}$ watts

$B_t(\text{for } n = 48) = 46.5 \times 10^6$ cycles per second

$P_t(\text{for } n = 48) = 23.8 \times 10^{-12}$ watts

$B_t(\text{for } n = 72) = 70 \times 10^6$ cycles per second

$P_t(\text{for } n = 72) = 35.9 \times 10^{-12}$ watts.

If a maximum bandwidth of 10 megacycles is set, the required received powers for $n = 48$ and $n = 72$ can be calculated from the fact that at a given signal noise ratio the carrier voltage is inversely proportional to B. Then,

$$P_{10}(n = 48) = 23.8 \times 10^{-12} \times 4.65^2 = 515 \times 10^{-12} \text{ watts}$$

$$P_{10}(n = 72) = 35.9 \times 10^{-12} \times 7^2 = 1760 \times 10^{-12} \text{ watts.}$$

Minimum Signal for (PAM-AM)

The signal noise ratio is independent of bandwidth providing the bandwidth exceeds a minimum value. The minimum received signal power for 200 to 1 signal noise ratio is,

$$P_r(n = 4) = 0.512 \times 10^{-18} \times 7000 \times 200^2 \times 4 = 574 \times 10^{-12}.$$

The minimum value of $B = 4nf_p = 4 \cdot 3 \cdot 4 \cdot 3500 = 0.168 \times 10^6$ cycles per second.

For larger values of n and R_t, P_t is proportional to n and R_t^2 so that,

$$P_r(\text{for } n = 12) = 574 \times 10^{-12} \times 3 \times 10 = 17{,}220 \times 10^{-12} \text{ watts}$$

$$P_r(\text{for } n = 48) = 690{,}000 \times 10^{-12} \text{ watts}$$

$$P_r(\text{for } n = 72) = 1{,}040{,}000 \times 10^{-12} \text{ watts.}$$

The minimum values of B for system operation are as follows:

$$B_t(\text{for } n = 4) = 0.168 \times 10^6 \text{ cycles per second}$$

$$B_t(\text{for } n = 12) = 0.505 \times 10^6 \text{ cycles per second}$$

$$B_t(\text{for } n = 48) = 2.02 \times 10^6 \text{ cycles per second}$$

$$B_t(\text{for } n = 72) = 3.03 \times 10^6 \text{ cycles per second.}$$

Minimum Signal for (PAM ± FM)

$$R_o = \frac{\pi}{12} \frac{B}{n^{3/2} f_m}; \qquad R_a = \left(\frac{B}{2f_m}\right)^{1/2}; \qquad R_t = R_o R_a = 0.184 \frac{B^{3/2}}{n^{3/2} f_m^{3/2}}.$$

The bandwidth for $(R_r = R_t)$ at threshold is,

$$B_t = \frac{R_t^{2/3} n f_m}{0.184^{2/3}} \qquad (69a). \qquad \text{For } R_t = 200,\ n = 4,\ f_m = 3500$$

$$B_t\ (n = 4) = \left(\frac{200}{.184}\right)^{2/3} \times 4 \times 3500 = 1.5 \times 10^6 \text{ cycles per second.}$$

For this bandwidth $P_r = P_t$ and $P_t = 0.512 \times 10^{-18} B_t$

$$P_t \text{ (for } n = 4) = 0.512 \times 10^{-18} \times 1.5 \times 10^6 = 0.77 \times 10^{-12} \text{ watts.}$$

B_t and P_t vary as n and $R_t^{2/3}$ so that (See Equation (69a))

$$B_t \ (n = 12) = 1.5 \times 10^6 \times 3 \times 10^{1/3} = 9.7 \times 10^6 \text{ cycles per second}$$

$$P_t \ (n = 12) = 0.77 \times 10^{-12} \times 3 \times 10^{1/3} = 4.97 \times 10^{-12} \text{ watts}$$

$$B_t \ (n = 48) = 1.5 \times 10^6 \times 12 \times 10^{2/3} = 84 \times 10^6 \text{ cycles per second}$$

$$P_t \ (n = 48) = 0.77 \times 10^{-12} \times 12 \times 10^{2/3} = 43 \times 10^{-12} \text{ watts}$$

$$B_t \ (n = 72) = 84 \times 10^6 \times 1.5 = 126 \times 10^6 \text{ cycles per second}$$

$$P_t \ (n = 72) = 43 \times 10^{-12} \times 1.5 = 64.5 \times 10^{-12} \text{ watts.}$$

If a maximum bandwidth of 10 megacycles is set, the required received powers for $n = 48$ and $n = 72$ can be calculated from the above using the fact that at a given signal noise ratio the required input voltage is inversely proportional to the bandwidth. Then,

$$P_{10} \ (n = 48) = 43 \times 10^{-12} \times 8.4^2 = 3050 \times 10^{-12} \text{ watts}$$

$$P_{10} \ (n = 72) = 64.5 \times 10^{-12} \times 12.6^2 = 10{,}250 \times 10^{-12} \text{ watts.}$$

If cross talk balancing is used, then

$$R_o(\text{PAM} \pm \text{FM}) = 0.555\, B\, f_m^{-1}\, n^{-3/2} = 0.555\, n^{-\frac{1}{2}}\, r.$$

This value of R_o is better than that obtained using Equation (6a) by the factor 2.12. Then the values of B_t and P_t are reduced by the factor $(2.12)^{2/3}$ or 1.65, so that,

$$B_t(n = 4) = 0.91 \times 10^6 \text{ cycles per second}$$

$$P_t(n = 4) = 0.465 \times 10^{-12} \text{ watts}$$

$$B_t(n = 12) = 5.9 \times 10^6 \text{ cycles per second}$$

$$P_t(n = 12) = 3.0 \times 10^{-12} \text{ watts}$$

$$B_t(n = 48) = 51 \times 10^6 \text{ cycles per second}$$

$$P_t(n = 48) = 26 \times 10^{-12} \text{ watts}$$

$$B_t(n = 72) = 76 \times 10^6 \text{ cycles per second}$$

$$P_t(n = 72) = 38.7 \times 10^{-12} \text{ watts}$$

$$P_{10}(n = 48) = 26 \times 10^{-12} \times 5.1^2 = 675 \times 10^{-12} \text{ watts}$$

$$P_{10}(n = 72) = 38.7 \times 10^{-12} \times 7.6^2 = 2{,}250 \times 10^{-12} \text{ watts.}$$

Minimum Signal for PPM-AM

The signal power required at threshold is independent of bandwidth and is equal to,

$$P_t = 2.05 \times 10^{-18} \times 3n\,f_m \quad \text{(see Equation (60))}$$

$$P_t(n = 4) = 2.05 \times 3 \cdot 4 \cdot 3500 \times 10^{-18} = .086 \times 10^{-12}.$$

For larger values of n and R_t, P_t varies directly with n, but not at all with R_t so that,

$$P_t(\text{for } n = 12) = 0.086 \times 10^{-12} \times 3 = 0.258 \times 10^{-12} \text{ watts}$$

$$P_t(\text{for } n = 48) = 1.03 \times 10^{-12} \text{ watts}$$

$$P_t(\text{for } n = 72) = 1.55 \times 10^{-12} \text{ watts.}$$

The output signal noise ratio improves with increased bandwidth. To obtain $R_t = 200$,

$$R_t = R_o R_a; \quad R_o = \frac{2}{9} \frac{B}{n^{3/2} f_m} \quad \text{(See Table III)}$$

$$R_a = \left(\frac{B}{2f_m}\right)^{1/2} \left(\frac{C_f \text{ (PPM-AM)}}{C_f \text{ (AM)}}\right)^{1/2} = \left(\frac{B}{2f_m}\right)^{1/2} \frac{32Dn}{8}$$

(See Equation (58))

$$= \left(\frac{B}{2f_m}\right)^{1/2} \left(4\frac{3n\,f_m}{B}\right)^{1/2} = 2\sqrt{\frac{3}{2}}\,n^{\frac{1}{2}} = 4\sqrt{\frac{3}{2}} \text{ (for } n = 4)$$

$$R_t(\text{for } n = 4) = \frac{2}{9} \frac{B}{n^{3/2} f_m} 4\sqrt{\frac{3}{2}}; \quad B_t = \frac{9}{8} R_t\, n^{3/2} f_m \sqrt{\frac{2}{3}} \qquad (70)$$

$$B_t(n = 4) = 1.125 \times 200 \times 8 \times 3500 \times 0.815 = 5.15 \times 10^6 \text{ cycles/second.}$$

This is a minimum value. Larger values would do as well since the threshold would be the same and S/N would be better.

To obtain the desired values of signal noise ratio with larger values of n, the minimum value of B varies directly with n and R_t so that,

$$B_t(\text{for } n = 12) = 3 \times 3.16 \times 5.15 \times 10^6 = 49 \times 10^6 \text{ cycles per second}$$

$$B_t(\text{for } n = 48) = 620 \times 10^6 \text{ cycles per second}$$

$$B_t(\text{for } n = 72) = 930 \times 10^6 \text{ cycles per second.}$$

If a maximum bandwidth of 10 megacycles is set, the required received powers for $n = 12$, $n = 48$ and $n = 72$ can be calculated from the above using the fact that at a given signal noise ratio the required signal voltage is inversely proportional to the bandwidth. Then,

$$P_{10}(n=12) = 0.258 \times 10^{-12} \times 4.9^2 = 6.2 \times 10^{-12} \text{ watts}$$

$$P_{10}(n=48) = 1.03 \times 10^{-12} \times 62^2 = 3950 \times 10^{-12} \text{ watts}$$

$$P_{10}(n=72) = 1.55 \times 10^{-12} \times 93^2 = 13500 \times 10^{-12} \text{ watts.}$$

Minimum Signal for PPM-FM

$$R_o = \frac{2}{3}\,\frac{B^{\frac{1}{2}}}{f_m^{\frac{1}{2}}\,n}; \qquad R_a = \left(\frac{B}{2f_m}\right)^{\frac{1}{2}}$$

$$R_t = R_oR_a = \frac{\sqrt{2}}{3}\,\frac{B}{f_m n}; \qquad B_t = \frac{3}{\sqrt{2}}\,R_t\,f_m n \tag{71}$$

$$B_t(n=4) = \frac{3 \times 200 \times 3500 \times 4}{1.414} = 5.92 \times 10^6 \text{ cycles per second}$$

$$P_t(n=\ \ 4) = 0.512 \times 10^{-18} \times 5.92 \times 10^6 = 3.03 \times 10^{-12} \text{ watts}$$

$$B_t(n=12) = 5.92 \times 10^6 \times 3 \times 3.16 = 56.2 \times 10^6 \text{ cycles per second}$$

$$B_t(n=48) = 56.2 \times 10^6 \times 4 \times 3.16 = 712 \times 10^6 \text{ cycles per second}$$

$$B_t(n=72) = 712 \times 10^6 \times 1.5 = 1070 \times 10^6 \text{ cycles per second}$$

$$P_t(n=12) = 0.512 \times 10^{-18} \times 56.2 \times 10^6 = 28.9 \times 10^{-12} \text{ watts}$$

$$P_t(n=48) = 0.512 \times 10^{-18} \times 712 \times 10^6 = 364 \times 10^{-12} \text{ watts}$$

$$P_t(n=72) = 0.512 \times 10^{-18} \times 1070 \times 10^6 = 546 \times 10^{-18} \text{ watts.}$$

If a maximum bandwidth of 10 megacycles is set, the required received power may be calculated from the above using the fact that at a given signal noise ratio, the required carrier voltage is inversely proportional to $B^{\frac{1}{2}}$. Then,

$$P_{10}(n=12) = 28.9 \times 10^{-12} \times 5.62 = 163.0 \times 10^{-12} \text{ watts}$$

$$P_{10}(n=48) = 364 \times 10^{-12} \times 71.2 = 26{,}000 \times 10^{-12} \text{watts}$$

$$P_{10}(n=72) = 546 \times 10^{-12} \times 107.0 = 58{,}500 \times 10^{-12} \text{ watts.}$$

Minimum Signal for PNM-AM

Here again threshold is independent of B.

$$P_t(\text{for } n=4) = 2.05 \times 10^{-18} \times 21\,f_m \ \left(\text{Assuming } \frac{nf_p}{f_m} = 21\,n\right)$$

$$= 2.05 \times 84 \times 3500 \times 10^{-18} = 0.60 \times 10^{-12} \text{ watts.}$$

The bandwidth required to obtain proper operation is about

$$B_{\text{min.}} = 3 \times 21n\, f_m = 252 \times 3500 = 0.88 \times 10^6 \text{ cycles per second.}$$

The factor* 3 is probably sufficient because of the inherent noise and cross talk eliminating properties of the system.

The signal noise ratio is theoretically infinite anywhere above threshold.

For larger values of n, P_t is larger in direct proportion but R_t has no effect, so that,

$$P_t (\text{for } n = 12) = 0.60 \times 10^{-12} \times 3 = 1.80 \times 10^{-12} \text{ watts}$$

$$P_t (\text{for } n = 48) = 7.4 \times 10^{-12} \text{ watts}$$

$$P_t (\text{for } n = 72) = 10.8 \times 10^{-12} \text{ watts.}$$

The bandwidth to obtain proper operation is also proportional to n.

$$B_t (\text{for } n = 12) = 2.65 \times 10^6 \text{ cycles per second}$$

$$B_t (\text{for } n = 48) = 10.6 \times 10^6 \text{ cycles per second}$$

$$B_t (\text{for } n = 72) = 16.0 \times 10^6 \text{ cycles per second.}$$

If a maximum bandwidth of 10 megacycles is specified, the 72 channel system is inoperative, unless changes are made in the system parameters.

Minimum Signal for PNM-FM

The threshold varies with bandwidth but the S/N ratio is infinite, above threshold. Assume a value of $B = 3 \times 21n\, f_m$ as above, to be the minimum for proper operation. Then,

$$P_t (\text{for } n = 4) = 0.52 \times 10^{-18} \times 0.88 \times 10^6 = 0.46 \times 10^{-12} \text{ watts.}$$

The value of P_t is proportional to n. $P_t (n = 12) = 1.36 \times 10^{-12}$ watts

$P_t (n = 48) = 5.42 \times 10^{-12}$ watts $P_t (n = 72) = 8.1 \times 10^{-12}$ watts.

Minimum Signal for PAM-PCM-AM

Same as PNM-AM for equal values of f_p.

Minimum Signal for PWM-AM

$$R_o = \frac{1}{\sqrt{6}\, n} \left(\frac{B}{f_m} \right)^{1/2} \text{(See Table III)}; \quad R_a = \left(\frac{B}{2f_m} \right)^{1/2} \left(\frac{32\, Dn}{8} \right)^{1/2} = \left(\frac{B}{2f_m} \right)^{1/2} \sqrt{2}$$

* Since the above was written, it has been pointed out that 3 is unnecessarily large. It has been argued that operation should be possible with the value unity for PNM and PCM systems.

$$R_t = \frac{1}{\sqrt{6}\,n}\;\frac{B}{f_m} = 200; \qquad B_t = \sqrt{6}\,R_t\,n\,f_m \tag{72}$$

$B_t(n=4) = 6.9 \times 10^6$ cycles per second.

For this bandwidth $P_r = P_t$ and

$$P_t(\text{for } n=4) = 2.05 \times 10^{-18}\,DnB_t$$
$$= 2.05 \times 10^{-18} \times \tfrac{1}{2} \times 6.9 \times 10^6 = 7.1 \times 10^{-12} \text{ watts.}$$

Both B_t and P_t are proportional to n and R_t (See Equation (72))

$B_t(n=12) = 6.9 \times 10^6 \times 3 \times 3.16 = 65 \times 10^6$ cycles per second

$B_t(n=48) = 820 \times 10^6$ cycles per second

$B_t(n=72) = 1230 \times 10^6$ cycles per second

$P_t(n=12) = 67 \times 10^{-12}$ watts

$P_t(n=48) = 850 \times 10^{-12}$ watts

$P_t(n=72) = 1270 \times 10^{-12}$ watts.

If a maximum bandwidth of 10 megacycles is set, the required received powers for $n=12$, $n=48$ and $n=72$ can be calculated from the above using the fact that at a given signal noise ratio the required signal voltage is inversely proportional to $B^{\frac{1}{2}}$. Then,

$P_{10}(n=12) = 67 \times 10^{-12} \times 9.2 = 615 \times 10^{-12}$ watts

$P_{10}(n=48) = 850 \times 10^{-12} \times 116. = 98{,}500 \times 10^{-12}$ watts

$P_{10}(n=72) = 1270 \times 10^{-12} \times 174. = 222{,}000 \times 10^{-12}$ watts.

Minimum Signal for PWM-FM

$$R_o = \frac{\pi}{\sqrt{3}\,n}\left(\frac{B}{f_m}\right)^{\frac{1}{2}} \qquad \text{(See Table III)}$$

$$R_a = \left(\frac{B}{2f_m}\right)^{\frac{1}{2}}; \quad R_t = \frac{\pi}{\sqrt{6}\,n}\;\frac{B}{f_m}\;; \quad B_t = R_t\,f_m\,n\,\frac{\sqrt{6}}{\pi} \tag{73}$$

$B_t(n=4) = 200 \times 3500 \times 3.12 = 2.18 \times 10^6$ cycles per second.

For this bandwidth $P_r = P_t$, and,

$P_t(n=4) = 0.512 \times 10^{-18} \times 2.18 \times 10^6 = 1.12 \times 10^{-12}$ watts.

P_t and B_t are directly proportional to n and R_t. (See Equation (73))

$$B_t(n = 12) = 2.18 \times 10^6 \times 3 \times 3.16 = 20.7 \times 10^6 \text{ cycles per second}$$

$$B_t(n = 48) = 262 \times 10^6 \text{ cycles per second}$$

$$B_t(n = 72) = 292 \times 10^6 \text{ cycles per second}$$

$$P_t(n = 12) = 10.6 \times 10^{-12} \text{ watts}$$

$$P_t(n = 48) = 134 \times 10^{-12} \text{ watts}$$

$$P_t(n = 72) = 201 \times 10^{-12} \text{ watts.}$$

If a maximum bandwidth of 10 megacycles is set, the required signal powers for $n = 12$, $n = 48$ and $n = 72$ may be calculated from the above, using the fact that for a given signal noise ratio the required carrier voltage is inversely proportional to $B^{\frac{1}{2}}$.

$$P_{10}(n = 12) = 10.6 \times 10^{-12} \times 2.07 = 22 \times 10^{-12} \text{ watts}$$

$$P_{10}(n = 48) = 134 \times 10^{-12} \times 26.2 = 3500 \times 10^{-12} \text{ watts}$$

$$P_{10}(n = 72) = 201 \times 10^{-12} \times 29.2 = 5870 \times 10^{-12} \text{ watts.}$$

Minimum Signal for (SS-PM-FM)

B_t and P_t are larger than for SS-PM by $5^{2/3}$ or 2.92. Then,

$$B_t(n = 4) = 2.92 \times 0.84 \times 10^6 = 2.46 \times 10^6 \text{ cycles per second}$$

$$B_t(n = 12) = 2.92 \times 5.4 \times 10^6 = 15.8 \times 10^6 \text{ cycles per second}$$

$$B_t(n = 48) = 2.92 \times 46.5 \times 10^6 = 136 \times 10^6 \text{ cycles per second}$$

$$B_t(n = 72) = 2.92 \times 70 \times 10^6 = 205 \times 10^6 \text{ cycles per second}$$

$$P_t(n = 4) = 2.92 \times 0.43 \times 10^{-12} = 1.26 \times 10^{-12} \text{ watts}$$

$$P_t(n = 12) = 2.92 \times 2.77 \times 10^{-12} = 8.0 \times 10^{-12} \text{ watts}$$

$$P_t(n = 48) = 2.92 \times 23.8 \times 10^{-12} = 70 \times 10^{-12} \text{ watts}$$

$$P_t(n = 72) = 2.92 \times 35.9 \times 10^{-12} = 105 \times 10^{-12} \text{ watts.}$$

If a maximum bandwidth of 10 megacycles is set the required received powers can be calculated from the fact that at a given signal noise ratio the carrier voltage is inversely proportional to B.

$$P_{10}(n = 12) = 8.0 \times 10^{-12} \times 1.58^2 = 20.0 \times 10^{-12} \text{ watts}$$

$$P_{10}(n = 48) = 70 \times 10^{-12} \times 13.6^2 = 13{,}000 \times 10^{-12} \text{ watts}$$

$$P_{10}(n = 72) = 105 \times 10^{-12} \times 20.5^2 = 44{,}100 \times 10^{-12} \text{ watts.}$$

Minimum Signal for (XX-XX-AM)

Adding an extra-sub-carrier using AM makes P_t larger by 8 to 1, since the signal noise voltage ratio is poorer by $\sqrt{8}$. Threshold on the second demodulation will be at the same signal noise ratio as though

(XX-XX) were the whole system but operated at 8 times lower signal power level. Threshold on the first demodulation will be at a lower level since the bandwidth is only changed by 2 to 1, but S/N is reduced by $\sqrt{8}$.

Minimum Signal for PAM-FM (Slow)

$$R_o = \frac{\sqrt{3}\,B}{2nf_m}\,; \quad R_n = \left(\frac{B}{2f_m}\right)^{1/2}; \quad R_t = \frac{\sqrt{3}}{\sqrt{2}\,2n}\left(\frac{B}{f_m}\right)^{3/2} = 200$$

$$B_t(n=4) = \left(\frac{R_t\sqrt{2}\,2n}{\sqrt{3}}\right)^{2/3} f_m \tag{74}$$

$$= \left(\frac{200 \times 1.41 \times 8}{1.73}\right)^{2/3} \times 3500 = 0.412 \times 10^6.$$

For this bandwidth $P_r = P_t$ and,

$P_t(n=4) = 0.512 \times 10^{-18}\,B_t = 0.211 \times 10^{-12}$ watts.

Both B_t and P_t vary directly with $n^{2/3}$ and $R_t^{2/3}$ (See Equation (74))
$B_t(n=12) = 0.412 \times 10^6 \times 3^{2/3} \times 10^{1/3} = 1.84 \times 10^6$ cycles per second
$B_t(n=48) = 1.84 \times 10^6 \times 4^{2/3} \times 2.15 = 10.0 \times 10^6$ cycles per second
$B_t(n=72) = 10.0 \times 10^6 \times 1.31 = 13.1 \times 10^6$ cycles per second
$P_t(n=12) = 0.211 \times 10^{-12} \times 4.46 = 0.94 \times 10^{-12}$ watts
$P_t(n=48) = 5.12 \times 10^{-12}$ watts
$P_t(n=72) = 6.70 \times 10^{-12}$ watts.

The bandwidths all lie within the 10-megacycle maximum except for $n = 72$.

$$P_{10}(n=72) = 6.70 \times 10^{-12} \times \left(\frac{13.1}{10}\right)^2 = 11.5 \times 10^{-12} \text{ watts.}$$

Direct Formula For P_{10}

In any system for which B_t exceeds 10 megacycles, the value of P_{10} (the received power required for a 10-megacycle setup) can be solved for directly by using the following:

$$P_{10} = \frac{8KT\,F_n\,2f_m\,R_r^2}{R_o^2} \quad (75)\,; \qquad P_{10}(n=72) = \frac{14.4 \times 10^{-9}}{R_o^2}.$$

This can be used as a check on the method used above. It is not used to replace the first method because the values of B_t and P_t have some interest. In particular it is necessary to know whether or not B_t exceeds 10 megacycles. If not, then P_{10} is below threshold.

The above Equation (75) applies for the value of P_B where P_B is the received power required for a setup having the bandwidth B. The value of R_o corresponding to the value B must be used. It is necessary to know that the specified bandwidth B is less than B_t. Otherwise P_B will be below threshold.

Summary of the Received Power Requirement

The minimum signal powers required to meet the various specifications for each system are tabulated below, assuming a 10-megacycle maximum bandwidth. The systems are arranged in the order of merit for a 48-channel set up.

Required Signal Power in $\mu\mu$ Watts

System	$R_r = 200$ $n = 4$		$R_r = 632$ $n = 12$	$R_r = 2000$ $n = 48$	$R_r = 2000$ $n = 72$
PAM-FM (slow)	0.211	(2)	0.94	5.12	11.5
PNM-FM } PAM-PCM-FM }	0.46	(4)	1.36	5.42	——
PNM-AM } PAM-PCM-AM }	0.60	(5)	1.8	7.4	——
SS-PM	0.43	(3)	2.77	*515.*	*1,760.*
PAM(±)-FM	0.77	(7)	4.97	*3,050.*	*10,250.*
FWM-FM	1.12	(9)	*22.*	*3,500.*	*5,870.*
PPM-AM	0.086	(1)	*6.2*	*3,950.*	*13,500.*
FM-PM	0.68	(6)	*8.42*	*5,450.*	*18,300.*
SS-PM-FM	1.27	(10)	*20.0*	*13,000.*	*44.100.*
FM-AM	1.08	(8)	*26.*	*16.600.*	*56,300.*
PPM-FM	3.03	(12)	*163.*	*26,000.*	*58,500.*
AM-PM	1.97	(11)	*53.*	*34,000.*	*116,000.*
PWM-AM	7.1	(13)	*615.*	*98,500.*	*222,000.*
SS-AM	574.	(14)	17,200.	690,000.	1,040,000.
PAM-AM	574.	(15)	17,200.	690,000.	1,040,000.
AM-AM	18,300.	(16)	275,000.	11,000,000.	16,500,000.

Setups in which the minimization of required received power is limited by the 10-megacycle bandwidth have the power figure *italicized.* Thus if the power figure is not *italicized*, the setup did not utilize the full 10-megacycle available bandwidth, but only such a bandwidth that the required value of signal noise ratio is obtained at threshold.

Note that the order of merit is somewhat different in the ($n = 4$) column as indicated by the number in parenthesis. However, the PAM-FM (Slow) system shows up the best by quite a margin in every column except for $n = 4$ where PPM-AM is better. PPM-AM would take first place in the other three columns except for the bandwidth limitation. PNM-FM and PNM-AM are almost equal and are assumed to be equivalent to PAM-PCM-FM and PAM-PCM-AM in noise reducing performance. These systems would always lead if the signal noise ratio requirement was raised because it is assumed that when operated above threshold their signal noise ratio is unlimited. Actually this is unfair because the system noise generated because of quantization has not been considered. These systems with $n = 72$ are inoperative in a 10-megacycle bandwidth without parameter changes.

If judged by the ($n = 4$) and ($n = 12$) columns any of the first twelve systems would be fairly satisfactory. The small differences in power required might easily be offset by other considerations. For the $n = 48$ column, the list should probably be narrowed to the first 5 or 6 systems. For $n = 72$, the system PAM-FM (slow) has no near competitors. It is unfortunate that more practical work has not been done with this system.

Bandwidth Economy

The comparisons of required received power were all made on the basis of a 10-megacycle available bandwidth. For all systems in which best results were obtained by fully occupying this band, no further comments are necessary. For $n = 48$ or 72 all of the better systems fully utilized the band, with the exception of PAM-FM (slow). Thus this system not only provide the best signal noise ratio, but does so with a reduced bandwidth requirement. For $n = 48$ or less PNM-XX and PAM-PCM-XX are included in the systems requiring less than 10 megacycles. For $n = 12$ or less SS-PM and PAM(±)FM are included. For $n = 4$ none of the systems utilized the full band for best results. The actual bands required for each system in the $n = 4$ column of the minimum signal requirements table are given below:

PAM(±)-FM	910,000	PPM-FM	5,920,000
PAM-FM (Slow)	412,000	SS-PM-FM	2,460,000
PNM-FM / PAM-PCM-FM	880,000	PWM-FM	2,180,000
		FM-AM	2,100,000
PNM-AM / PAM-PCM-AM	880,000	AM-PM	3,850,000
		PWM-AM	6,900,000
SS-PM	840,000	SS-AM	40,000
PPM-AM	5,150,000	PAM-AM	168,000
FM-PM	1,330,000	AM-AM	56,000

Following is a listing of the formulas for the minimum bandwidth required for operation without cross modulation:

AM-AM	$4nf_m y^{-1}$	PPM-AM	$36nf_m$
AM-PM	$4nf_m y^{-1}$	PAM(±)-FM	$6nf_m$
FM-AM	$4nf_m y^{-1}$	PPM-FM	$36nf_m$
FM-PM	$4nf_m y^{-1}$	PNM-XX	$63nf_m$
SS-AM	$2nf_m y^{-1}$	PAM-PCM-XX	$63nf_m$
SS-PM	$2nf_m y^{-1}$	PWM-XX	$24\ nf_m$
PAM-AM	$12nf_m$		

The above figures for pulse systems represent conservative design. Somewhat smaller values could be used if necessary.

A complete discussion of the factors affecting bandwidth economy should give consideration to transmitter and receiver oscillator drift, frequency calibration error, undesired sidebands produced outside the nominal frequency band, and guard bands.

Impulse Noise Analysis

Signal Noise Ratio on Impulse Noise

Impulse noise from automobiles, electric motors, etc. becomes progressively of less importance as we go to higher and higher frequencies. One reason for this is that the initial wave form of such disturbances is essentially that of the unit step function which has a distribution of energy versus frequency such that the voltage amplitude per cycle of bandwidth is inversely proportional to the frequency of measurement.

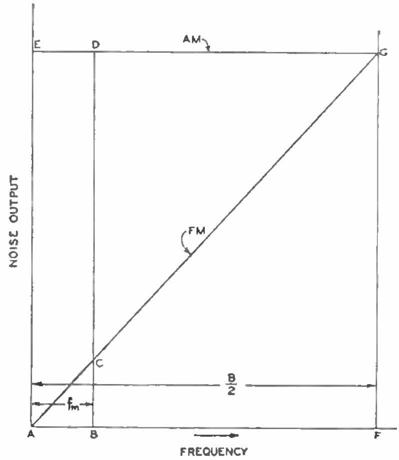
The frequency spectrum usually considered for this type of transmission is about 30 megacycles to 30,000 megacycles. At the low end of this spectrum impulse noise may assume major importance for some antenna locations. However, over most of this frequency spectrum at most locations, impulse noise is of minor importance. Nevertheless it is occasionally important and hence merits further discussion.

The Nature of Impulse Noise

"Pure" impulse noise is the disturbance resulting from an input voltage wave form consisting of isolated or recurrent pulses, the time duration of the pulses being less than the time of a half wave at the mean frequency of the pass band of the receiver. Under these conditions a wave train is applied to the detector of the receiver for each input pulse. The duration of the wave train is about $1/B$ (at 50 per cent amplitude). The amplitude is proportional to the area under the input pulse.

Except for the last statement the above paragraph applies for

Fig. 13—Noise frequency spectrum for AM and FM.



somewhat longer duration input pulses. An essential limitation is that the input pulses should be far enough apart so that there is no overlapping of wave trains of different pulses. A complicated input wave of long time duration may be considered as pure impulse noise if it has only a few points where the slope of the wave changes with sufficient abruptness to appreciably excite the radio-frequency circuits and providing these points are sufficiently separated so that the wave trains do not overlap. Ignition interference from automobiles frequently does not meet this requirement. A single ignition spark may consist of as many as several hundred separate discharges very closely spaced so that the wave trains overlap in the output of the radio-frequency amplifier. The disturbance created by this type of interference may be a great deal worse than would be predicted by theory if the assumption is made that there is only one pulse per spark. This last fact may change the ratio of S/N ratios for the various systems for this type of noise but it will not be considered in the following because of difficulty in specifying the noise wave form.

Comparing random noise and "pure" impulse noise, both have a uniform distribution of energy over the pass band. The essential difference is in the relative phases of the various frequency components. For random noise this phase is random but for impulse noise the phases are so related that at the instant of the peak amplitude of the wave train, the components are all in phase (or nearly so).

Weak Impulse Noise in Frequency Division Systems

The term weak impulse noise is taken to mean impulse noise of less than signal amplitude. For weak noise of any type, a given noise frequency component has an effect on the output which is independent of the presence of the other noise components. It follows that for all frequency division systems the signal noise ratio on an r-m-s basis compared to AM has the same formula for impulse noise as for random noise.

On the basis of the ratio of peak signal amplitude to peak noise amplitude for impulse noise the numerical coefficient will be slightly different from that for random noise if FM is involved. This can be more clearly understood by referring to Figure 13 which shows a comparison of the noise spectra for FM and AM. In this figure frequency is plotted along the line ABF and AF = B/2. Then the noise in a broadband amplitude modulation case is represented by the area AFGE. Reducing the frequency band to f_m = AB, the noise is represented by the area ABDE. In the FM case the broad band noise is represented by the triangle AFG. Reducing the frequency band to f_m = AB, the noise is represented by triangle ABC. In the case of random noise the areas are summed up r-m-s and $\dfrac{\mathrm{ABC}}{\mathrm{ABDE}} = \sqrt{3}\,\dfrac{B}{2f_m}$ (r-m-s summation of noise).

In the case of impulse noise, if it is desired to compare the peak value of the noise in the two cases the areas are summed up linearly (making the assumption[9] that all components are in phase at the peak) in which case, $\dfrac{\mathrm{ABC}}{\mathrm{ABDE}} = 2\,\dfrac{B}{2f_m}$ (linear summation of noise).

Since the noise is distributed differently with frequency for AM and FM, neither formula is perfectly accurate from an annoyance standpoint. Also the difference between the formulas is rather slight namely 13 per cent. It is believed that the following approximation is justified.

The signal noise ratio on weak impulse noise for all frequency division systems compared to AM is the same as for random noise.

Because of its discontinuous nature impulse noise can be tolerated at a somewhat higher peak amplitude than random noise. The wideband gain of the better systems will be found sufficient to make the interference of weak impulse noise negligible.

Weak Impulse Noise in a Time Division System

On an r-m-s basis, the wide band gain of a time division system on weak impulse noise should be the same as for random noise, just as it is in a frequency division system. There is, however, one important difference. In the time division system only $1/n$ of the individual impulses are heard at all. But the wideband gain is normal on an r-m-s basis, so the pulses that do come through are stronger in amplitude (than in a frequency division system having the same wide band gain on random noise) by a factor $\sqrt{n}$ (more accurately factor is $1/\sqrt{D}$).

---

[9] This method of attacking the problem was first used by M. G. Crosby; see "Frequency Modulation Noise Characteristics", *Proc. I.R.E.*, Vol. 25, No. 4, pp. 483-485, April, 1937.

Strong Impulse Noise in Frequency Division Systems

The term strong impulse noise is taken to mean impulse noise of greater than signal amplitude. In XX-PM systems the use of a limiter sets a maximum on impulse noise at a value corresponding roughly to operating at threshold. In a similar manner a maximum can be set for AM or XX-AM systems by the use of a limiter adjusted to saturate at the amplitude of 100 per cent modulation. Here also the limit is at a level corresponding roughly to operating at threshold.

If in the AM case, the limiter is used after a broad band radio-frequency (or intermediate-frequency) amplifier, the noise is limited to a level corresponding to operating at threshold with that wide amplifier. (This gives a value of S/N in the output better than unity by the divisor $\frac{B}{2f_m}$). Then to compare the signal noise ratio on AM and FM, reference may be made again to Figure 13 where AFGE is the broad band AM noise and ABDE the narrow band noise, while AFG is the broadband FM noise and ABC the narrow band noise. It can be seen that as long as the limiter is used as stated in the AM cases as well as in the FM or PM cases, the S/N ratio compared to AM is roughly the same for strong impulse noise as for random noise for all frequency division systems.

The Wave Form of Impulse Noise

It may seem peculiar that for an AM system, impulse noise has a uniform distribution versus audio frequency while for an FM system the amplitude of the audio frequency components is proportional to the frequency. There is a corresponding difference in the wave form of the output noise. For an AM case with nearly ideal band pass characteristics the wave form is an impulse of approximately the shape of $\sin x/x$ or roughly a half cycle of a sine wave. Maximum response occurs when the radio-frequency phase of the noise wave train is the same as that of the carrier. For an FM system, maximum response (on weak impulse noise and on strong impulse noise when the carrier is not deviated from the center of the band) occurs when the phase of the noise radio-frequency wave train is 90 degrees from the phase of the carrier. For this condition (accurately for very weak impulse noise and approximately for strong impulse noise) the phase displacement has the same wave form as for the AM case at zero phase angle. The frequency deviation is, however, the derivative of the phase displacement and is hence a bi-directional pulse, having as much area above the axis as below. This is in agreement with the fact that the low-frequency components are attenuated relative to the highs in the noise output of an FM system.

Strong Impulse Noise in an FM System Under Conditions of Peak Modulation

When an FM system is modulated (or detuned) to the edge of the pass band at the instant a noise impulse occurs a phenomenon[10] may take place which greatly increases the noise. The noise wave train is at the frequency of the center of the pass band (i.e. it is synchronous with the unmodulated carrier). When the carrier is modulated to the edge of the pass band at the instant the pulse occurs, the pulse wave train beats with the carrier. The time of one cycle of the beat note is $2/B$. The time duration of the pulse (at its base) is also $2/B$. Thus if the relative phase of the pulse is 0 degrees at its start, it is 180 degrees at its peak and 0 degrees again at the end of the pulse. This is the condition for maximum noise response. If the noise radio-frequency wave train happens to be of twice signal amplitude, the resultant of signal plus noise has almost a constant amplitude even without a limiter. In the output circuit of the limiter the output amplitude is almost perfectly flat (even if the pulse is many times signal amplitude) but there is a unidirectional pulse of frequency modulation, in that one complete radio-frequency cycle has been subtracted (or added) in a time $2/B$ so that the frequency is deviated $B/2$. In other words the frequency is deviated back to the center of the pass band at the peak of the pulse.

It will be remembered that when the carrier was unmodulated the noise output had both a positive and a negative peak and hence a triangular frequency spectrum. For the case discussed in the last paragraph the noise pulse is unidirectional and hence has an approximately uniform frequency spectrum. That is the low frequency components have been pushed up to the level of the high frequency components. The signal noise ratio (wide band or narrow band) is the same as for the AM case (with limiter and wideband ahead of limiter).

Thus the gain of frequency modulation is completely lost if a strong noise impulse (of proper phase) occurs at the instant of peak frequency deviation. It should be noted, however, that the crest factor (in multi-channel operation or one channel on voice) of the signal is large which means that for a great majority of the time the instantaneous value of the frequency deviation is quite small.

[10] For a more detailed discussion with oscillograms see V. D. Landon, "Impulse Noise in FM Reception", *Electronics*, Vol. 14, No. 2, pp. 26-31, February, 1941.

If the deviation of the carrier from the center of the band is quite small at the instant the pulse starts then no complete cycles are lost or gained. If the overall number of cycles is constant then the frequency must be high just as much as it is low. In other words the noise wave form must be above the axis as much as it is below the axis. This means that the low frequency audio components are attenuated. Thus the ratio $R_o = \sqrt{3}\dfrac{B}{2f_m}$ is much more nearly accurate than the value $R_o = 1$ on strong impulse noise.

Impulse Noise in Time Division Systems

For time division systems, on the contrary, the value of R_o is widely different for impulse noise from the value for random noise. On the average only $1/n$ of the noise pulses are effective at all in a given channel because of time division. This advantage is more than compensated for, by the fact that the noise pulses which do come through produce interference much stronger than the value of R_o for random noise would indicate. This is because the noise and signal have essentially the same wave form. Thus the signal noise ratio for PAM-AM is unity for strong impulse noise. The bandwidth reduction (of a low pass filter with cut off at f_m) produces no improvement since noise and signal amplitude are affected in the same way. The bandwidth reduction from f_p to f_m would be effective if the ratio were larger. Being only 2 or 3 it is inappreciable. (At 2, the effect would be unity.)

For PPM-AM the same argument is true since strong impulse noise can deflect the phase at least as much as the signal can.

The same is true for PWM-AM.

For PNM-AM the ratio f_p/f_m is larger and there is a corresponding effective bandwidth reduction factor of about $\sqrt{\dfrac{20}{2}} = \sqrt{10}.$

For PAM-PCM-AM the increased bandwidth reduction factor is lost because in the presence of strong impulse noise the wrong code will be transmitted which can give unity signal noise ratio since the code for the highest amplitude might be sent when the lowest was desired and vice versa.

Even more serious than the poor signal noise ratio is the fact that often synchronism will be lost in the presence of strong impulse noise in a PXX-XX system, thus completely blocking communication. The use of "flywheel effect" circuits (which reduce the effectiveness of isolated pulses on the synchronizing) such as are used in television receivers would be a help in this respect. Using synchronizing pulses considerably longer than the minimum the circuit will pass, should also be a help.

For PXX-FM, wideband gain is largely lost in the presence of strong impulse noise because the carrier is frequency modulated 100 per cent a high percentage of the time. (This is not true for PAM(±)-FM which retains wideband gain.)

The system PAM(±)-FM warrants separate consideration. Since the degree of FM is proportional to the signal in each channel the average modulation remains low and the wideband gain is not lost. Not only are all but $1/n^{th}$ of the impulses unheard, in a given channel, but the pulses which do come through are suppressed by the same ratio as that for random noise.

Summary—Impulse Noise

On strong impulse noise the signal noise ratio compared to AM (wide band with a limiter) is roughly the same as R_0 for random noise, in systems XX-XX and PAM(±)-FM. For PXX-XX, not including PAM(±)-FM, however, the signal noise ratio on strong impulse noise is unity or worse as there is no bandwidth reduction factor except in the case of PNM-XX where it is about $\sqrt{10}$.

Cross Modulation Analysis

Interchannel Cross Modulation in Frequency Division Systems

If the modulator of the transmitter or the demodulator of the receiver is non-linear, cross modulation may occur between sub-carriers. At a relay station it is necessary to receive the intelligence, amplify it and retransmit it on a new or the original frequency. Cross modulation may easily occur in this process and the effect might be accumulative when the signals are passed through many successive relay stations. Hence it is necessary at this point to discuss methods by which the modulation of the incoming wave can be transferred to the outgoing wave.

An obvious possibility is to amplify and demodulate the incoming wave, recovering the various sub-carriers or single sideband signals (as the case may be) passing these waves through a common wide band amplifier and using them to modulate the outgoing carrier. A defect of this scheme is the fact that the repeated demodulation and modulation provide the possibility of accumulative cross modulation if a slight non-linearity is consistently present in either the modulator or demodulator or in the sub-carrier amplifier. This type of cross modulation can be minimized in the amplifier (and possibly in the modulator) by using negative feedback to improve the linearity.

Another possibility is to heterodyne the incoming signal to an intermediate frequency, amplify it, and re-heterodyne it to the outgoing frequency. This scheme seems to offer much less possibility for accumulative cross modulation since the carrier is never demodulated. A difficulty is present because of the necessity for providing a power amplifier at the output frequency. The technique for using power amplifiers in transmitters has not been fully developed above about 500 megacycles. Some successful amplifiers have been built however and no doubt the problem will be completely solved in the near future.

Another inherent defect of this scheme for relaying by a double heterodyne at each relay station is the fact that any frequency drift of the heterodyning oscillators is accumulative and might be quite serious if the number of relay stations is large. This defect might be remedied by stabilizing the frequency of the intermediate-frequency at each station by an automatic-frequency-control circuit. Other methods of reducing the accumulative drift may also be used.

Another solution of the relaying problem is to go to a triple modulation system similar to that used in the microwave radio relay equipment types CW1a and CW2a.[11] This equipment uses the system SS-PM-FM, frequency modulating the carrier with a single intermediate sub-carrier which is modulated as an SS-PM system. At a relay station a demodulator is employed to recover the main sub-carrier which is not further demodulated but is amplified and used to frequency modulate the carrier of the outgoing wave. It is felt that this arrangement is unusually free of any tendency for accumulative cross modulation. As pointed out previously it involves a sacrifice of about 5 to 1 in signal noise ratio.

Interchannel Cross-Modulation in Time Division Systems

In time division systems cross-modulation will occur if insufficient time separation is provided between pulses of adjacent channels. This type of cross talk can always be reduced by decreasing the duty factor of the pulses. A decrease of duty factor in a PAM-AM system requires an increase in bandwidth. Thus there is a minimum bandwidth to avoid cross modulation. In calculating this minimum (See page 456) it was assumed that $B = 4nf_p$ would be a reasonable figure. A more accurate calculation will now be made.

Up to now no definite statement has been made about the shape

---

[11] Leland E. Thompson, "A Microwave Relay System", *Proc. I.R.E.*, Vol. 34, No. 12, pp. 936-942, December, 1946.
G. G. Gerlach, "A Microwave Relay Communication System", *RCA REVIEW*, Vol. VII, No. 4, pp. 576-600, December, 1946.

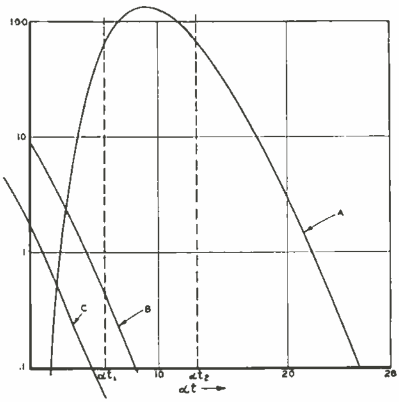


Fig. 14—Response to unit impulse for ten tuned circuits.

of the radio-frequency selectivity curve. It has been assumed that "a", the pulse duration (above 50 per cent of peak amplitude), is equal to $1/B$, where B is the effective bandwidth. To know how rapidly a pulse decays toward zero it is necessary to know something more definite about the frequency characteristic of the amplifier.

A type of amplifier which has easily calculated performance (and which has been used a great deal in radar work) is one consisting of a number of amplifier stages, all alike, with one tuned circuit per stage. With unit impulse excitation, an amplifier has an output wave form which is a pulse of the minimum duration that can be passed by the amplifier.

1. Shape of Minimum Duration Pulse

An amplifier of the type described above has a response to unit impulse of the following form,

$$e = A\alpha \frac{(\alpha t)^{m_1-1}}{m_1-1} e^{-\alpha t} \sin \omega t \tag{76}$$

where $A =$ the amplification at resonance

$\alpha = R/2L$ for each tuned circuit

$m_1 =$ the number of tuned circuits

$\omega =$ the resonant frequency.

The expression $\dfrac{(\alpha t)^{m_1-1}}{m_1-1} e^{-\alpha t}$ gives the form of the envelope of the pulse. A graph of this expression as a function of (αt) with $m_1 = 10$ is given in Figure 14.

It can be seen from the graph that the time duration (above 50 per cent of peak amplitude) is $\alpha(t_2 - t_1) = 7$.

The value of α is $R/2L$ where R and L are the effective series resistance and inductance of one of the tuned circuits. The bandwidth of a single circuit is $B_2 = R/2\pi L$ (equals the bandwidth of one tuned circuit at 70 per cent, equals the bandwidth of two tuned circuits at 50 per cent). The shape of the resonance curves (very near the peak) is the same as a parabola so that the band width at 50 per cent amplitude for "m" tuned circuits is $B = \dfrac{B_2}{\sqrt{m_1/2}} = \dfrac{B}{\sqrt{5}} = \dfrac{B}{2.24}$

for $m_1 = 10$. However, $B_2 = \dfrac{R}{2\pi L}$ so that $B = \dfrac{R}{2L \times 2.24\pi}$

$$= \frac{\alpha}{2.24\pi} = \frac{\alpha}{7.03}.$$

The value of $\dfrac{1}{B}$ is $\dfrac{7.03}{\alpha}$ which is seen to equal the time duration of the pulse quite accurately.

2. Pulse Spacing Required

If, in a PAM-AM system, the gate for each channel is turned on and off at the 50 per cent amplitude points, the interference from the decay train of the preceding pulse need be considered only during that interval, shown between the dotted lines. The curve B represents the decay envelope of a preceding pulse spaced by a time interval $\dfrac{2.5}{B}$. The interference can be seen to be down at least 43 decibels or 140 to 1 at the highest point and is considerably better when averaged over the gated interval. The line C represents a pulse spaced $3/B$ ahead. For this condition the cross modulation is down by more than 60 decibels. It appears that Roberts and Simonds[12] used a less advantageous arrangement as they only obtained a signal-to-cross talk ratio of 40 decibels with a pulse spacing of $\dfrac{3.4}{B}$. In any case $\dfrac{4}{B}$ (the value assumed for PAM-AM) seems to be more than adequate. An increment of time spacing equal to $\dfrac{1}{B}$ increases the attenuation about 35 decibels.

Systems PAM($\pm$)-FM, PPM-XX, and PWM-XX will usually be designed so that undistorted cross-modulation will not be found. Interference in the form of distorted undesired modulation will occur only when the pulses are modulated to too great a degree.

[12] F. F. Roberts and J. C. Simmonds, Multichannel Communication Systems", *Wireless Engineer*, Vol. 22, page 548, November and December, 1945.

Thus the *minimum* spacing between pulses in a PPM-AM system should be about $\frac{3}{B}$ for peak deviation of the pulses toward each other. If (as assumed previously*) this is ⅓ of the spacing when unmodulated, then $\frac{9}{B} = \frac{1}{n f_p}$ and B (min.) $= 9n f_p$ or $27nf_m = 6\frac{1}{4}$ megacycles for 72 channels. Then the 10 megacycles used (in calculating minimum required signal power) is adequate.

When operating long chains of relay stations using a time division system it may be found that the pulses grow continuously in effective width with an increase in the number of relay stations, eventually reaching a point where serious cross modulation results. For all such systems except PAM-XX, and PWM-XX this increase in pulse width may be corrected by demodulating at one or more relay stations to obtain the pulse sub-carriers. The pulses may then be reshaped by a process of amplifying, limiting, differentiating, clipping and re-amplifying. For PWM-XX systems the trouble manifests itself as a more gradual slope of the rise and decay of the pulses. This may be corrected by passing the pulses through a clipper-limiter. (See definition 34)† For PAM-AM neither a clipper nor a limiter may be used on the pulses but the width of the pulses may be reduced by the use of a time gate. (See definition 36)† For PAM(±)-FM the rise and decay of the pulses may be steepened by passing the pulses through a time gate.

3. *Pulse Shape for Coupled Circuits*

If a pair of coupled circuits is used in each stage of the amplifier with very weak coupling, the transient response is the same as for one tuned circuit per stage, for the same total number of tuned circuits. If the coupling is gradually increased it will be found that the decay to zero, is of shorter duration, but that the response returns (with reversed radio-frequency phase) after the zero. The tighter the coupling, the nearer the zero approaches the main part of the pulse, and the higher the response after the zero. A proper choice of the coupling coefficient will place the zero at the most advantageous position near the center of the following pulse, and help to minimize the cross talk. Because of the phase reversal at the zero, the portions of the interfering wave before and after the zero have opposing effects in a PAM-AM system and may buck each other out. For any PXX-XX system using a limiter, minimum cross modulation will occur when the

* Part I—pp. 327-334.
† Part I—page 296.

zero occurs at the instant the limiter saturates on the following pulse. For PWM and PPM the advantage of this is largely lost when the degree of modulation is large.

The mathematics of the transient response of coupled circuits for a large number of circuits is too complex to make detailed analysis worth while.

4. Cross Modulation With Coupled Circuits

The above indicates that by using coupled circuits a lower cross modulation may be achieved in a given bandwidth (or a lesser bandwidth may be used without making the cross modulation worse). The gain in this respect may be quite appreciable especially for PAM-AM, if a critical balance is obtained. Such a balance should not be depended on for a large portion of the desired signal to cross-modulation ratio, because a critical balance is difficult to maintain.

5. Cross Modulation in PNM-XX and PAM-PCM-XX

It will bear repeating, that the above systems have theoretically perfect signal to cross modulation ratios, for the same reason that they have theoretically perfect signal noise ratios.

6. Relaying Time Division Multiplex Signals

In relaying PXX-XX systems, (as well as XX-XX systems) the scheme may be employed wherein the incoming wave is heterodyned to a suitable intermediate frequency, amplified and re-heterodyned to the frequency of the outgoing wave. This requires a class B power amplifier at transmitter frequency.

Another method of transferring the modulation to the output wave is to amplify and demodulate the incoming wave. The resulting series of pulses may be amplified (and limited except for PAM-XX cases) and used to modulate the transmitted wave. It appears that whatever cross modulation is present as a result of overlapping pulses, may be accumulative as the signals go from one relay station to another. However, the desired attenuation of the cross modulation can always be obtained by reducing the duty factor. (In PWM-XX, the duty factor with no modulation should be kept 50 per cent. Cross talk is reduced by using a wider band so as to obtain more nearly vertical sides to the pulses. The degree of width modulation must be limited to avoid overlapping of the up stroke and decay of adjacent pulses.)

Summary of Interchannel Cross-Modulation:

The actual signal-to-cross modulation ratio obtained in each system is more a function of the skill and techniques used in making a setup than a function of the type of system. The governing factors have been discussed. A discussion of the relative feasibility of the various systems in this respect must depend on an engineering and experimental study. It may be found that for long relay chains, a triple modulation system will be required to avoid accumulative cross modulation.

Interchannel Cross Talk:

Cross talk other than that caused by cross modulation should never be a problem in time division systems if proper shielding is used. In frequency division systems such cross talk is governed by sub-carrier spacing and the design of the frequency selective filters. The design of such filters is a well known art.

INTERFERENCE ANALYSIS

Mutual Interference Between Multiplex Systems Operating on the Same Assigned Frequency

1. Neither Signal Pulsed

First assume the interfering wave to be a frequency modulated carrier wave. Then for a XX-PM desired signal the interfering wave beats continually with the carrier, but the beat is audible in a given channel only when the beat frequency lies between the values f_{sc} and $f_{sc} \pm f_m$. Then the presence or absence of interference in a given channel at a given instant depends on the vagaries of the modulation of all the channels of the desired and undesired transmissions. This complex function is near enough to random that the interference should sound very similar to random noise. *If it could be established* that the instantaneous beat frequency between the two waves was equally apt to take any value from 0 to $\frac{B}{2}$, then the signal to noise ratio on an r-m-s basis would be,

$$S/N = R_0 R_a R_I K_5 \qquad (77)$$

where $R_0 =$ the system noise improvement ratio over AM

$R_a =$ S/N (AM) at signal strength required for threshold on random noise in the system under discussion.

$R_I =$ The ratio of the desired to the undesired carrier amplitude on an r-m-s basis.

$K_5 =$ The ratio of crest factors for the interfering signal and for random noise.

Equation 77 was arrived at by making the assumption that the interfering wave produces the same amount of interference as that which would be produced by random noise of the same r-m-s value. This assumption is obviously untrue for an AM desired signal. Hence the multiplex systems cannot be compared to AM in this characteristic. However, the comparison of one XX-PM system with another may be made using the ratio of signal noise ratios R_o as indicated in Equation (77) providing the instantaneous beat frequency is kept random by the vagaries of the two multiplex modulating signals.

The requirement of the beat frequency being random is seldom fulfilled. If this requirement is not met, the formula still holds for the average channel but some channels are worse than others. If the crest factor of the modulation is about 4 as in random noise, and if the interfering carrier is centered in the pass band the lower frequency beats (producing interference in the first one or two channels) will occur more often than for a uniform probability distribution by the ratio $4/(\sqrt{2}\,\pi) = 1.6$. Thus it seems likely that the noise in no channel will be worse than about twice that indicated by Equation (77). If the interfering carrier is at mid-band frequency, the interference will be greatest on the channels having the lower frequency sub-carriers. If the interfering carrier is near the edge of the pass band, the channels having higher frequency sub-carriers will receive the greatest interference. Of course R_I must exceed unity, or the wide band gain is lost.

The above discussion holds for any XX-PM system where the phase modulation keeps the noise well scattered in frequency. The same formula also applies for XX-AM, providing the interfering carrier has a continuous high percentage of frequency modulation. If the interfering wave is unmodulated however, the beat might fall near one sub-carrier sideband of an AM-AM or SS-AM system causing an audio beat of fixed pitch in that one channel and much more serious interference. If the system is FM-AM the beat will be scattered in frequency only when the channel in which the beat occurs, is modulated. The interference level in that channel would remain above the formula value by a factor of about $\dfrac{B}{4nf_m}$.

Equation (77) for the signal to interference ratio is also believed to be approximately correct for the systems PAM(±)-FM, PWM-FM and PPM-FM.

2. *Desired Signal Wave Unpulsed, Interfering Wave Pulsed*

If the interfering wave is pulsed, it is believed that the same formula (Equation (77)) still applies (for the same systems for which the above applies). R_I is still interpreted as the ratio of the desired to the undesired signal on an r-m-s basis.

However, the proviso must now be made that nDR_I must exceed unity where Dn is the overall duty factor of the interfering pulsed wave. If this proviso is not met the interference will be so severe as to render the XX-XX system inoperative. If the interfering pulsed carrier has a very small duty factor, conditions may be met where the pulsed carrier produces serious interference (because the above proviso is not met) while an XX-XX interfering system of the same average power might produce negligible interference. This condition will be most apt to occur if the interfering system is PPM-AM as this system has the smallest duty factor of any.

3. Desired Signal Pulsed and Interfering Wave Unpulsed

For PPM-AM, PAM-AM, or PWM-AM as a desired signal and with an FM wave as interference, the ratio of signal to interference should follow the formula, Equation (77). If the interfering wave is not continuously modulated, beats of fixed pitch might occur making the interference much worse.

For PPM-AM and PWM-AM, broadband gain in signal to interference ratio is maintained up to the point where the interference is equal to one half the pulsed wave in amplitude. This is of particular advantage for PPM-AM, especially if $\frac{B}{nf_m}$ is large, as this results in a very small duty factor for this system.

With PNM-AM, or PAM-PCM-AM as desired signal, no interference should be produced if the interfering wave is less than half of the pulsed wave in amplitude.

4. Desired Signal and Undesired Signal Both Pulsed

While the formula of Equation (77), should hold on the average for this case, the interference might be somewhat worse than this would indicate because of the following:

If the pulse rates of the two systems are nearly the same but they are not synchronized, there will be a slow beat as the pulses go in and out of phase. The interference will be missing (or slight) when the pulses are interleaved and will be worst when they are overlapping. If the condition corresponds to threshold, (interfering pulses one half the amplitude of the desired pulses) then the noise will, at that instant, have a level corresponding to threshold even though the received power of the interfering wave may be much less than the noise power that would be required to produce peaks of one half signal amplitude. This is especially true if both duty factors are small.

5. Summary on Common Channel Interference

In general the ratios of signal-to-interference ratios for the various systems correspond roughly to the ratios of signal-to-noise ratios for random noise. An exception is the fact that XX-AM and PXX-AM systems are more subject to fixed pitch beat note interference which is more serious. Also threshold conditions are more apt to be exceeded if the interfering wave is pulsed and less apt if the desired signal is pulsed, especially if the duty factor is small.

Mutual Interference Between Multiplex Systems Operating on Adjacent Frequency Assignments:

It would perhaps be ideal if a table could be constructed listing the various systems and giving for each, a numerical measure of first the side band splash or the power radiated outside its assigned frequency band, second the ability to produce interference neglecting sideband splash, third the susceptibility to interference from sideband splash from the adjacent multiplex channel, and fourth the susceptibility to interference from an adjacent channel wave which does not spread outside its own assigned band.

The construction of such a table cannot be attempted because the relative merit of setups of the various systems in the four respects is not so much a fundamental property of the systems as it is of the skill and technics used in making the setups. Hence, a discussion will be given of the parameters controlling these factors with only a few general summarizing remarks.

1. Radiation of Power Outside the Assigned Frequency Band

An XX-PM wave at a point immediately following the final modulator will always have some energy outside the assigned band as do all FM waves. When the sub-carrier (or SS signal) of a single channel is applied to the main carrier as phase modulation, one or more side frequencies of considerable amplitude are produced on each side of the carrier, spaced apart by the modulation frequency. Still more side frequencies of lesser (and eventually inappreciable) amplitude occur at greater distances from the carrier, according to the well known Bessel function formulas. When a second modulating frequency is applied, each of the above frequency components may be considered to be frequency modulated in the above manner. A new set of side frequencies is produced for each frequency mentioned above thus squaring the number of components having appreciable amplitude. Each time a new modulation frequency is introduced the number of frequency components is multiplied by a factor m so that for n channels the number of frequency components is m^n. If for example $m = 5$ and $n = 12$, the number of components is $5^{12} = 250 \times 10^6$. The average component amplitude is less than the unmodulated carrier amplitude by the square root of this factor or 16,000. Thus the individual components taken singly are of extremely small amplitude. If when modulating with a given frequency the second harmonic sidebands (or the highest considered to be appreciable) are a good deal smaller than the fundamental side bands, then, as more and more modulating frequencies are added, the outer frequency components rapidly become smaller and smaller in amplitude compared to the average component. Thus if m and n are both large, a large group of components may be dropped from the theoretical number m^n. The information transmitted on any given channel is partly contained on every single component of the 250 million. It is not surprising that a few of the weaker ones near the edges of the spectrum may be omitted without losing any appreciable information.

The exact extent to which the band can be cut down, can only be found by experiment. It is believed that trouble would first manifest itself in the form of noise. When the bandwidth is cut to a certain point, occasional peaks of the complex modulating wave will cause the instantaneous frequency to move outside the pass band for an instant. The result should be similar in effect to impulse noise.

When the degree of modulation is adjusted (as it should be) to the greatest value possible for a given bandwidth receiver without excessive noise or distortion, there will be considerable energy present outside the nominal band. This may be reduced by means of a filter located between the transmitter and the transmitting antenna. Antenna directivity, distance and the broad band gain of the receiver are the only other variables that may be used to help keep this interference out of a receiver on an adjacent multiplex channel.

The XX-AM systems have little or no spread of the sidebands outside the nominal band.

The PXX-AM systems may or may not have considerable energy outside the nominal frequency band depending on the wave form of the modulating pulses. It is assumed that the pulse duration (or else the time of rise) is the minimum transmissible in the assigned bandwidth as this gives the greatest signal noise ratio. The simplest way to obtain this condition is to start with the time still shorter. When such a wave is applied to circuits having the desired bandwidth the time duration and time of rise are automatically adjusted to the desired value. These filter circuits should preferably lie ahead of the final power amplifier to avoid a loss in efficiency. Theoretically the circuits determining the wave form may be ahead of the modulator operating at "video" frequencies. Practically, side frequencies may be reintroduced in the modulation process because of non-linearity.

2. The Ability to Produce Adjacent Channel Interference Aside from Sideband Splash

In common channel interference, the very serious interference which occurs when the interfering wave is above threshold, is more apt to occur when the interfering wave is a PXX-AM system (especially a PPM-AM system) because of the large crest factor. This effect is largely lost in adjacent channel interference, because of the slope of the selectivity curve in the adjacent channel. For PXX-AM interfering waves, a portion of the power of each pulse is received. The effective crest factor is reduced by the slope of the selectivity curve. For XX-PM interfering waves, the received power comes in pulses because of the frequency modulation and the slope of the selectivity curve. As a result PXX-AM systems and XX-PM and PAM(±)-FM systems should produce interference of the same order of magnitude. PXX-FM (not including PAM(±)-FM) should be worse because of the continuous high modulation index.

3. Susceptibility to Side Band Splash from the Adjacent Multiplex Channel

Since interference of this nature comes from energy in the same assigned frequency band as the desired signal wave, the subject is covered by the section on common channel interference.

4. Susceptibility to Interference from an Adjusted Multiplex Channel Which Does Not Radiate Outside Its Assigned Band

At first sight it appears that interference of this sort should be less if the desired signal wave is SS-PM or AM-PM, the idea being that appreciable attenuation (say 10 decibels) might be used at nominal band limits since the limiter effectively flattens the response in the limiter output circuit. The use of such a filter would make it possible to obtain higher attenuation in the adjacent multiplex channel than could be obtained in a filter having a flat response over the desired band. It has been pointed out however that the phase shift is not a linear function of the frequency deviation with such an arrangement. Hence it is feared that cross modulation would occur.

It is believed that for any of the better systems, a nearly flat response is required over the complete band if the best signal-noise ratio is to be attained and cross modulation avoided. Thus all of the systems (which have some wide band gain) are about of equal merit as regards the shape of the selectivity curve which may be used. Some of the components of the interfering signal and of the desired signal are separated in frequency by less than $B/2$ so that the nature of the interference produced is not greatly different from that of the common channel case.

5. Summary on Adjacent Channel Interference

The XX-AM systems produce little sideband splash in the adjacent channel. The PXX-FM systems (not counting PAM(±)-FM) are probably worst in this respect. The remainder of the systems are roughly equal in potential performance, but may vary greatly in practice according to the skill and techniques used in making the setups. Not counting side band splash, but simply the ability to produce interference because of the imperfect selectivity of the receiver on the adjacent multiplex channel, the order of merit of the interfering systems is about the same as above.

The susceptibility to sideband splash places the systems in the same order of merit as signal-to-noise ratio considerations, except for cases where the interference may exceed threshold. In these cases the PXX-AM systems have the advantage insofar as they have a high crest factor.

Propagation Analysis; Susceptibility to Selective Fading

To produce selective fading it is necessary to have two or more propagation paths of different length between the transmitting and receiving antennas. In order to simplify the problem only two paths will be considered; the direct and indirect paths differing in length by a distance d meters. In a C.W. system it is well known that

$$d = 3 \times 10^8/(f_2 - f_1)$$

where $f_2 - f_1$ is the frequency change of the transmitted signal which causes the relative phase between the direct and indirect waves to progress through 360 degrees at the receiving antenna. This relation is independent of the total length of path and the carrier frequency. If we assume the two waves are in phase at the carrier frequency f_o, then they will be in phase again at $f_o \pm (f_2 - f_1)$ or $f_o \pm \dfrac{3 \times 10^{-8}}{d}$. The

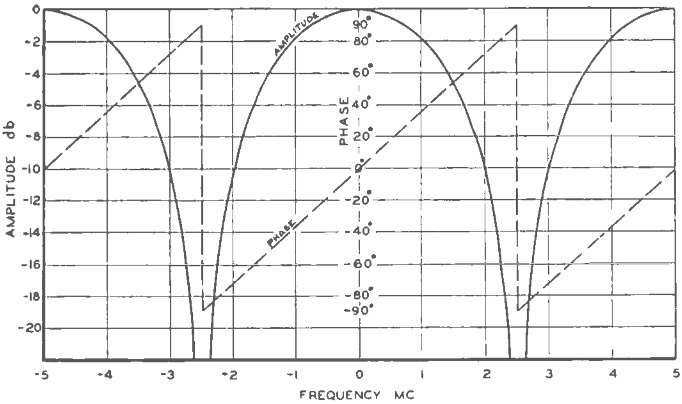


Fig. 15—Phase and amplitude versus frequency for two path transmission (amplitude ratio $= 1$; $d = 60$ meters).

waves will be out of phase at $f_o \pm \dfrac{f_2 - f_1}{2}$ or $f_o \pm \dfrac{3 \times 10^8}{2d}$.

In this paper it has been assumed that a 10-megacycle radio-frequency band is the widest to be considered. On this basis the curves of Figures 15-22 have been drawn showing the resultant amplitude and phase characteristic vs. frequency for the combination of two waves arriving over paths differing in length by the values of d shown for each figure. Figures 15-18 show the results for equal amplitudes of the two waves and Figures 19-22 the results for relative amplitudes of 1 to 2. The phase curves show the phase displacement of the resultant wave from that of the direct wave component. Path length differences

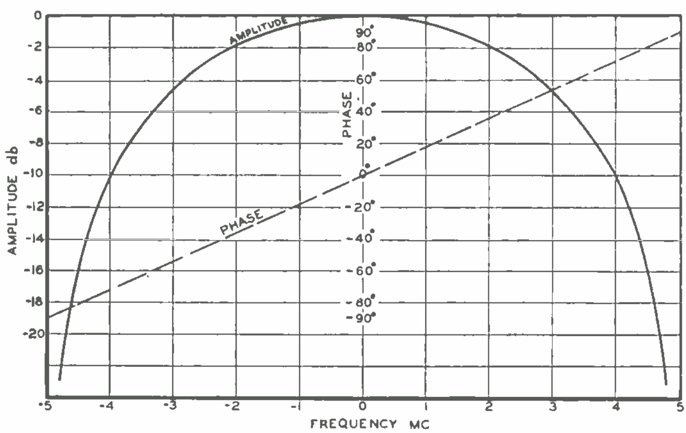


Fig. 16—Phase and amplitude versus frequency for two path transmission (amplitude ratio $= 1$; $d = 30$ meters).

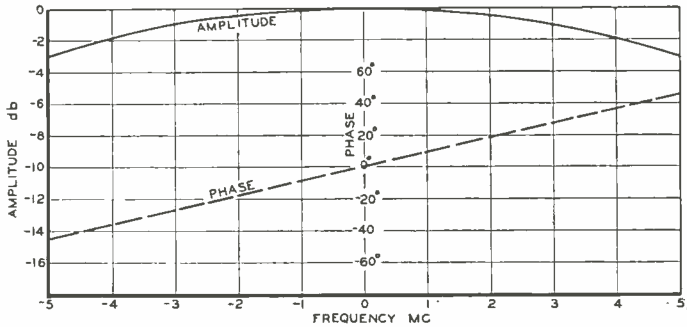


Fig. 17—Phase and amplitude versus frequency for two path transmission (amplitude ratio = 1; $d = 15$ meters).

d, of 60, 30, 15 and 7.5 meters have been used in these curves and it will be observed that a value of d equal to 7.5 meters (Figures 18 and 22) can be tolerated without difficulty as the phase curve is straight over the 10-megacycle band and the amplitude variation is less than one decibel. Of course smaller values of d would give even less trouble.

In Figure 23 is shown a plot of the required height of reflection point above the direct ray path, vs. distance, to give various constant path length differences. Thus for a path length of 100 miles and a path length difference of 7.5 meters the reflecting layer would have to be 2500 feet above the direct ray at the middle of the path.

Figure 24 is a plot of the angle between the direct and reflected rays at the receiving antenna, vs. distance for fixed values of d. This angle is also the grazing angle of the reflected ray at the point of reflection. In the previous example for a distance of 100 miles and a value of $d = 7.5$ meters, the angle is equal to 0.55 degree. Normally this is too small to allow the directivity of the receiving antenna to separate the two rays.

On the basis of present knowledge in the frequency range 1000-4000 megacycles we are aware of no atmospheric discontinuity occurring which is capable of reflecting or refracting a large percentage of

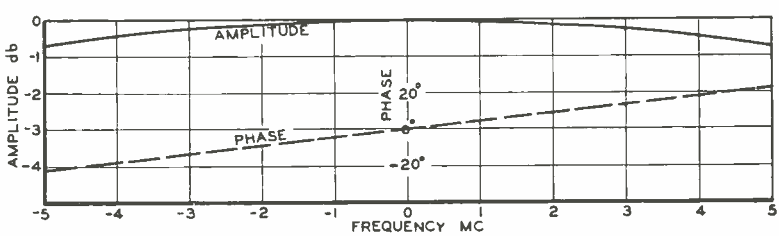


Fig. 18—Phase and amplitude versus frequency for two path transmission (amplitude ratio = 1; $d = 7.5$ meters).

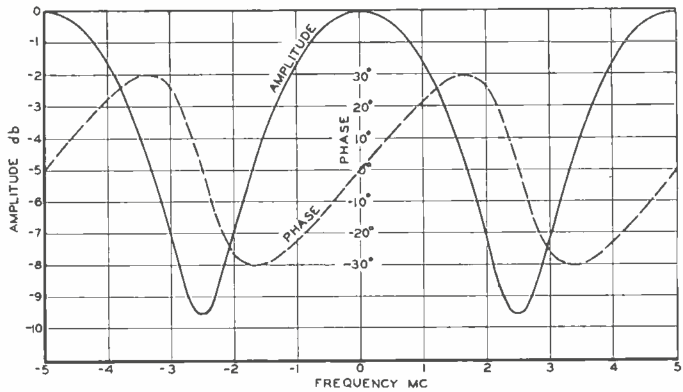


Fig. 19—Phase and amplitude versus frequency for two path transmission (amplitude ratio = ½; $d = 60$ meters).

the energy through as large an angle as 0.5 degree. It is known that very small percentages of the energy are turned through angles of this magnitude at times.

Recent experience with microwave radio relay equipment[11] operating in the neighborhood of 4000 megacycles has indicated very small path length differences due to discontinuities in the atmosphere. These path length differences appear to be much less than one meter. If this conclusion is correct the problem of selective fading disappears almost completely with respect to atmospheric discontinuities.

It should be born in mind that difficulties can arise due to too high locations of the ends of a radio link whereby a large component of ground reflected ray is present with a large path length difference as compared with the direct ray. Under these conditions a value of d of

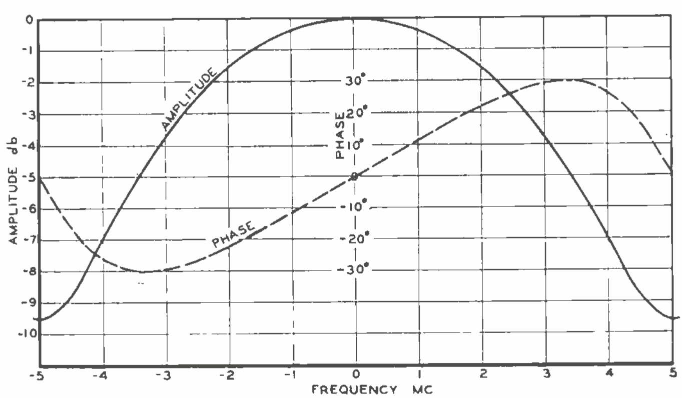


Fig. 20—Phase and amplitude versus frequency for two path transmission (amplitude ratio = ½; $d = 30$ meters).

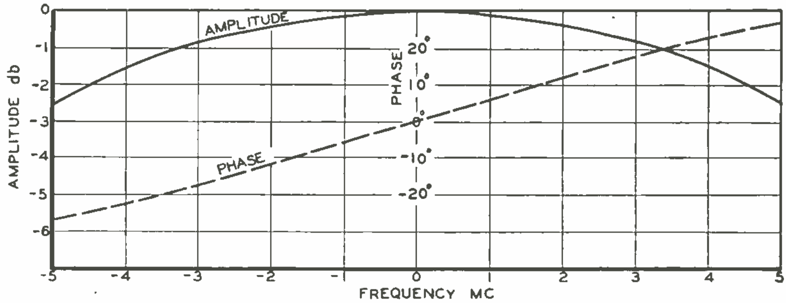


Fig. 21—Phase and amplitude versus frequency for two path transmission (amplitude ratio = ½; $d = 15$ meters).

30 meters would be obtained for a 100 mile path if the terminals were 5100 feet above the center ground level of the path. Here the angle between the direct and reflected rays would be 1.1 degrees which would not allow good separation of the rays by virtue of receiving antenna directivity in most cases. A value of $d = 30$ meters would give bad distortion. (See Figures 18 and 20).

The above discussion applies directly to C.W. systems but a similar analysis for a pulse system shows comparable results. As an example a 30-meter path length difference corresponds with a time delay of 0.1 microsecond. In a pulse system utilizing a 10-megacycle radio-frequency band the pulse width at half amplitude would be about 0.1 microsecond. Thus the combination of the pulse over the direct path along with the pulse over the indirect path gives rise to a new pulse of about double normal width or one having two peaks. In a PNM, PAM, or PPN system the widened pulse would give rise to interchannel cross modulation and in a PCM system the wrong code might be registered giving bad distortion.

It is to be expected that path length differences of less than one

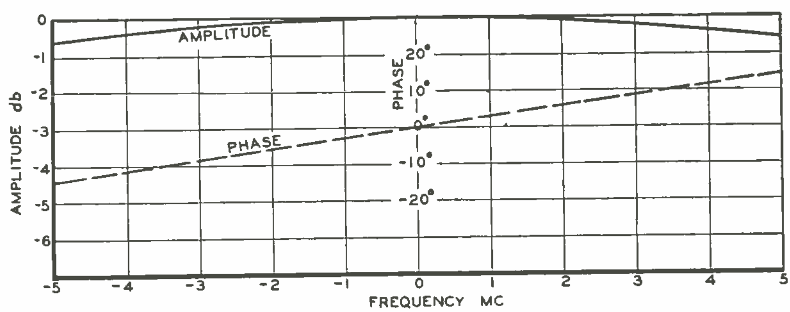


Fig. 22—Phase and amplitude versus frequency for two path transmission (amplitude ratio = ½; $d = 7.5$ meters).

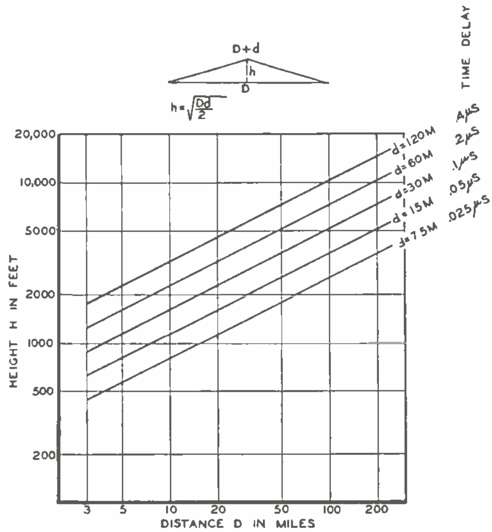


Fig. 23—Required height versus distance for various constant path differences.

meter would give no trouble from selective fading on any of the systems discussed in this paper. It is felt that sufficient data is not available on the subject of path length differences actually encountered and accordingly an experimental study of this phenomenon might prove fruitful.

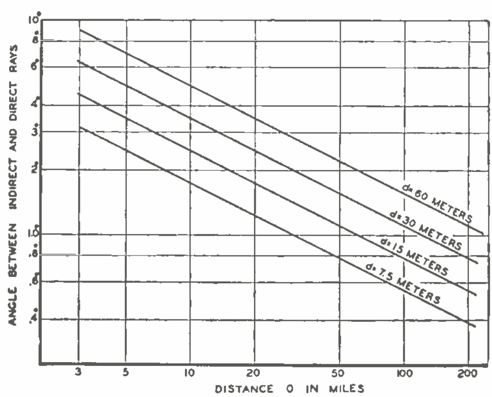


Fig. 24—Angle between direct and reflected rays versus distance for fixed values of *d*.

Experimental Confirmation

It was desirable to obtain experimental confirmation of the above theory for the major systems. Experimental setups were made of PPM-AM, PPM-FM, SS-PM, SS-PM-FM, PAM±FM, PAM±FM-FM and PNM-FM-FM.

The comparison was made with $B = 4$ megacycles, $n = 30$ channels, and $f_m = 3000$ cycles.

In checking for the signal-noise ratio compared to AM, it was found possible to check the formula within one or two decibels in each case, but certain detail considerations were brought out by the tests that warrant further discussion.

PPM-AM

1. The theory assumes that minimum width pulses are used. If the pulse width is greater than the minimum which the receiver will pass, the signal-noise ratio is poorer.

2. The theory assumes that both leading and trailing edges of the pulses are demodulated separately and the outputs added. Usually only one side is used and this involves a sacrifice of 3 decibels.

3. A synchronizing pulse must be used, sacrificing the equivalent of one channel for this purpose. Then $n + 1$ should be substituted for n in the formula for R_o.

4. The theory assumes $f_p = 3 f_m$. If a different ratio is used the formula must be corrected.

5. Sometimes greater power is used in the synchronizing pulse than in the others. If so, this power does not contribute to signal strength and a correction is necessary in the value of R_o.

6. The theory assumes that the phase modulation of each pulse can modulate it a peak distance of ⅓ of the space between adjacent pulses. A change in bandwidth changes the value of the usable fraction of this period. For best accuracy the true value should be substituted in each case.

When a correction was made for each of the above points the experimental setup checked the theoretical formula quite accurately.

PPM-FM

Of the corrections for PPM-AM mentioned above, points 2, 3 and 4 apply with equal force to PPM-FM. With these corrections applied the experimental setup checked the formula quite well.

SS-PM

The formula for SS-PM is derived on the assumption that the peak allowable frequency deviation is $B/2$. This seems to be a practical assumption if B/nf_c is of the order of 10 or larger. For smaller values of this ratio it may be found that the peak frequency deviation is limited to about half the theoretical figure of $B/2$ if cross modulation is to be avoided.

In one experimental setup this was found to be the case. The correction in the R_o formula required by this reduction in modulation is the same order of magnitude as the correction for the high crest factor of telephone signals. (See page 437.) Since these two corrections approximately cancel each other, it may be considered that the formula for R_o (SS-PM) is still that given, after both corrections are applied.

SS-PM-FM

As in the SS-PM system above, the modulation of the intermediate sub-carrier had to be reduced from the theoretical value about 3 decibels to reduce the cross modulation. In spite of this fact, the formula

$$R_o(\text{SS-PM-FM}) = 0.17 \frac{B}{n^{3/2} f_c}$$

checked very closely.

This formula was actually developed for SS-PM-PM and it was mentioned that if the final modulation was FM instead of PM it is possible that a 3-decibel improvement would result. Apparently the assumption is correct and this improvement just balances the modulation reduction.

PAM ± FM

The most important practical consideration in this system is whether or not cross talk balancing is employed. This is discussed in the theoretical work. The measured signal noise ratio checked the theoretical quite closely for both conditions.

PAM ± FM-FM

This system checked about 6 decibels below theoretical in the value of R_o. It is believed that this may have been due to one of the transmitter modulator circuits which was somewhat narrower in band width than desirable.

PNM-FM-FM

A pulse numbers system having only one channel was set up and operated. The circuit required is novel but will not be described in detail here. Since the signal is supposed to be proportional to the averaged density of the pulse sequence over an appreciable interval of time, it is necessary that a memory or storage circuit be used with a comparator circuit to determine whether each successive pulse should be transmitted and omitted.

In the setup the ratio of the fundamental pulse rate to the maximum audio frequency was 33 instead of 21 used in the theoretical work.

Voice transmission and other types of modulation were tried. Distortion was appreciable but not bad. The noise characteristics seemed to be about that predicted by theory. When operating above threshold the noise was below the hum level of 60 decibels.

Pulse Interference

The susceptibility of the various systems to pulse interference was also tested. The check with the theoretical predictions was only fair. In particular the performance of SS-PM on weak impulse noise was very poor for unknown reasons. On strong impulse noise it was best of all.

C.W. Interference

This test checked the theory quite well showing PPM-AM to be best for working through C.W. interference on the same frequency.

Conclusion

The various multiplex systems can now be placed in the order of merit from a signal noise ratio standpoint. The order of merit is somewhat different for various other characteristics. A final choice of the most desirable system cannot be made without weighting factors to determine the relative importance of various system characteristics.

ACKNOWLEDGMENT

Valuable contributions to the theoretical work were made by B. Trevor, D. Bond, H. O. Peterson and O. E. Dow of RCA Laboratories Division and by L. E. Thompson of RCA Victor Division. The experimental work was undertaken by D. R. Crosby, T. H. Pi, T. E. Jacobi and L. E. Thompson of RCA Victor Division, and O. E. Dow and W. D. Houghton of RCA Laboratories Division.

THE PACK TRANSMITTER*

By

J. L. Hathaway and W. Hotine

Engineering Development, National Broadcasting Co., Inc.,
New York, N. Y.

Summary—*A portable, low-powered, high-quality, high-frequency transmitter is described, which has its application in the remote pickup of sound broadcast programs when extreme mobility and freedom of action are required. This new model pack transmitter is approximately one third smaller in size and weight than previous models, yet is capable of improved performance. The savings in space and weight are made possible by the type of construction employed, together with the use of miniature type tubes. Simplified tuning controls, high-level modulation, automatic audio gain control, high-frequency pre-emphasis and high-quality monitoring are provided in the design of this unit.*

Introduction

PORTABLE, low-powered, high-frequency transmitters have been employed extensively for sound broadcast relay service and have been found useful particularly in such applications as the remote pickup of golf matches, parades, and other events requiring freedom of action during the broadcast. This type of transmitter has been termed a pack transmitter because it is a completely self-contained portable unit which can be operated and transported simultaneously by one man, and is usually carried on the back like a knapsack or pack.

In 1929 the first known successful broadcast using such a pack transmitter was made by a parachute jumper who described the sensations of his descent to earth on a network program. The transmitter in use on that occasion was quite heavy, had only fair audio quality, was somewhat unstable, and was very limited in its operating time due to short battery life. Since that time many improvements have been incorporated in succeeding transmitter models of this general type. This progress in development has culminated in the type of pack transmitter described in this paper, in which improvements have been made possible mainly by taking advantage of tube development in the miniature and subminiature class.

General Characteristics

The new pack transmitter has been designed for use within the

* Decimal Classification: R350 × R549

25-32 megacycle band. The operating frequency is determined by a quartz crystal within 0.01 per cent of the assigned frequency over a wide range of ambient temperatures.

The factor deemed most important in considering the design of this transmitter was optimum performance at the minimum possible weight and space, consistent with a reasonably long battery life. Additional desirable features included a rugged construction capable of withstanding abuse, and a simple tuning procedure with a minimum number of controls to facilitate operation and reduce errors.

Considerable savings in weight and space were made possible by

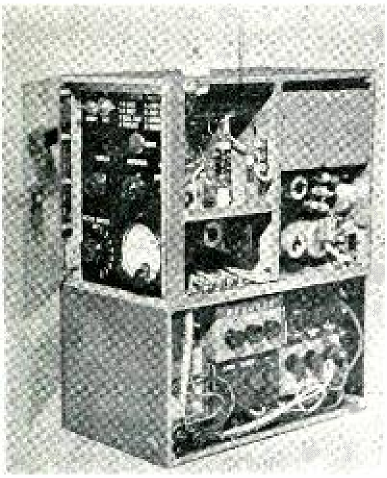


Fig. 1—Pack transmitter—front view (cover plates removed).

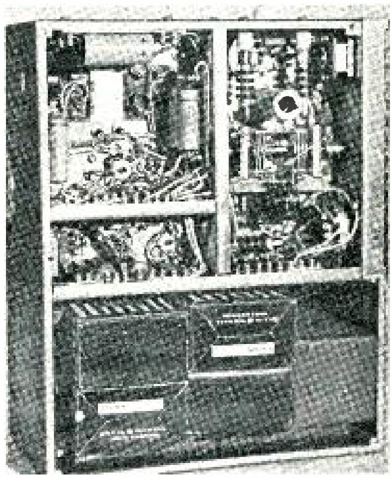


Fig. 2—Pack transmitter—back view (cover plates removed).

resorting to a type of construction which may be described as an "integral" chassis and case. This construction is shown in Figures 1 and 2. Figure 1 shows the front side of the transmitter, with the front cover plate removed, while Figure 2 shows the back of the transmitter, with the back cover plate removed. In this type of construction, the sides, top and bottom of the case enclose a vertical dividing wall which is used as a chassis. Shield partitions subdivide the chassis space on each side, and provide additional mounting space for components, all of which are easily accessible for servicing when the cover plates are removed. This construction provides interior bracing to the case, thus increasing its strength, and yet conserving space and weight.

The external housing dimensions of the pack transmitter are only 9¾ inches wide by 12¼ inches high by 5⅜ inches deep. The dimensions are such that carrying is extremely easy, since the unit does not protrude too far from the back and the height dimension is small enough to avoid bumping of the lower edge against the hip of the operator. The illustration of Figure 3 shows the transmitter in operation, strapped to the operator's back.

The overall weight, including batteries and protective cover, is 24 pounds. In order to withstand adverse weather conditions, the protective cover is made of treated canvas and is highly water repellent. The transmitter in its canvas case is pictured in Figure 4.

Batteries, of available standard types to simplify battery replacement, provide about 6 hours of continuous operation, which corresponds

Fig. 3—Pack transmitter in portable operating position.

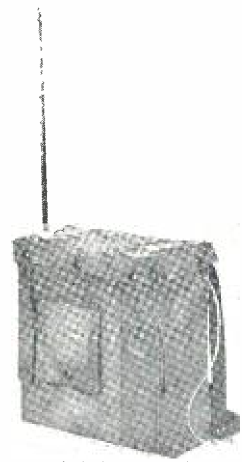


Fig. 4—Pack transmitter in waterproof case.

approximately to 15 hours of operation at a rate of about 1 hour per day.

Any high-quality, low-impedance microphone may be plugged in by way of a 3-pin Cannon receptacle, as shown in Figure 3. Polarization of the microphone should be such as to produce predominantly upward modulation rather than downward, inasmuch as more undistorted output power is available if polarization is properly maintained.[1]

Radio Frequency Portion

A standard type RC2, BT-cut quartz crystal is used in a simple Pierce oscillator circuit, as shown in the schematic diagram of Figure 5, utilizing one section of a Type 3A5 miniature dual triode envelope.

[1] J. L. Hathaway, "Effect of Microphone Polarity on Percentage Modulation", *Electronics*, October, 1939.

The other section of this tube is used as a frequency tripler, which is somewhat overneutralized in order to increase the harmonic amplitude. Output from the tripler stage is balanced to ground and is applied to the grids of the push-pull output stage which is a second Type 3A5 tube, operated as a Class-C neutralized power amplifier. Exact neutralization is achieved by adjustment of a single 5-micromicrofarad variable condenser, a fixed 5-micromicrofarad condenser "cross-neutralizing" the opposite portion of the dual triode.

The antenna is energized from a split output coil arranged on each side of the power amplifier tank inductance. It is closely coupled to the tank and is of sufficient inductance to load correctly a short length of vertical antenna. The four section telescopic antenna is adjusted in length during the tuning procedure for proper loading of the power amplifier.

As shown in the photograph of Figure 1 and as indicated in the schematic diagram (Figure 5), the only tuning adjustments required in normal operation are those of the tripler tank and power amplifier tank. Neither of these is extremely critical, and highly-skilled personnel are not required for their proper alignment.

Neutralizing may be accomplished very simply at the front panel, holes being provided for screwdriver adjustment of the variable neutralizing condenser and for screwdriver operation of a power amplifier plate supply on-off switch.

AUDIO-FREQUENCY PORTION

The audio-frequency section of the transmitter consists of a four-stage amplifier, the final stage of which is a push-pull class-B plate modulator for the radio-frequency power amplifier.

Some degree of high-frequency peaking is included at the input transformer to gain the desirable effects of pre-emphasis at the transmitter with subsequent de-emphasis in a receiver. The degree of pre-emphasis is equivalent to that of a 10-microsecond time constant circuit.

Automatic audio gain control is employed in order to eliminate the necessity for manual control, and to permit a high percentage of modulation at all times without appreciable overmodulation. The automatic control action is achieved by rectifying a portion of the output energy of the third audio amplifier stage and applying the rectified voltage to the grid circuits of the first two stages. Time constants employed are such that the second stage acts appreciably faster than the first, in order to reduce objectionable "thumps" and "plops". A germanium crystal rectifier is shunted by a fixed resistor

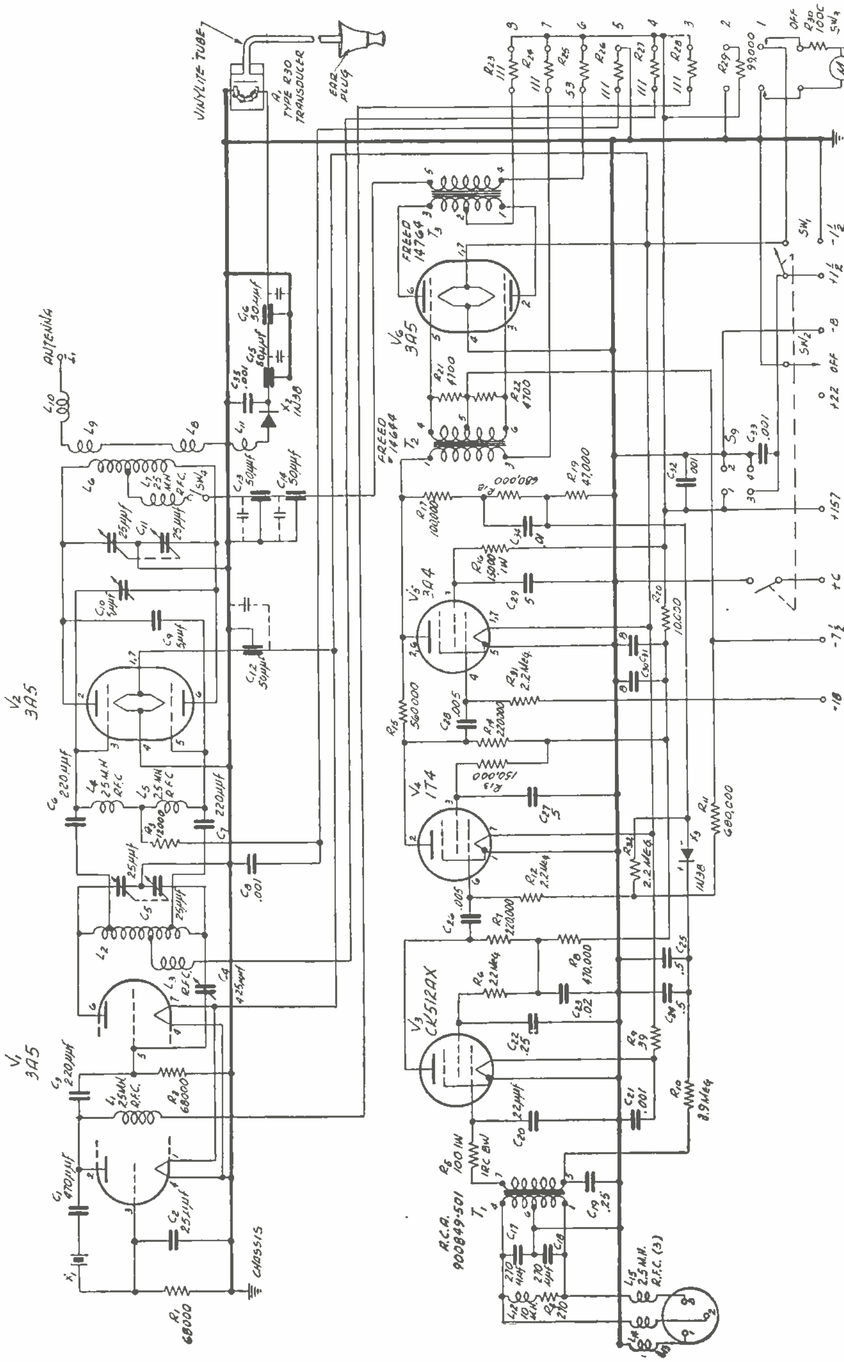


Fig. 5—Pack transmitter—schematic diagram.

so that the overall back resistance in the circuit remains as constant as possible with variations of ambient temperature and differences between individual crystal units. The "back resistance" must remain substantially constant in order to obtain the correct no-signal bias voltage on the grids of the first two stages. So called "delay bias" voltage is applied to the rectifier from a d-c voltage divider circuit which is bled off the plate of the third stage. This "delay bias" prevents rectification at extremely low input signal levels and thereby straightens the control characteristic of the system.

The Class-B driver and output transformers are physically rather large in order to prevent core saturation and are designed to provide optimum impedance transformations together with good frequency response characteristics.

Features and Facilities

A small milliameter may be utilized for measurement of any of the desired plate currents or battery voltages by means of a multi-position switch, as shown in Figure 1.

A toggle switch is provided for using either 157½ or 180 volt B supply. The prime purpose of this switch is to permit a shift to the higher voltage after the B batteries have become partially exhausted. However, the full 180 volts from new batteries may be applied with a consequent increase in power output, as described in the concluding section of this paper.

For applications where the transmitter need not be carried, it is possible to employ either large external batteries or a rectifier type of power supply. Plate and filament voltages may be applied to the unit by connection to the receptacle provided for this purpose.

High-quality monitoring of the transmitter is made available to the operator. A small coil is coupled to the output stage and the output, rectified by a germanium crystal, is then applied to an electro-acoustic transducer which is in the form of a miniature magnetic telephone receiver unit located within the transmitter case. The sound output from the receiver unit is conducted to the ear by way of a flexible vinylite tube of ⅛-inch outside diameter and equipped with a rubber ear plug. Figure 3 illustrates the use of the monitor, showing the tube plugged into the operator's ear.

Circuit Precautions

It will be noted in the schematic that the first audio amplifier tube is of the subminiature type, whereas all other tubes are miniatures.

Subminiatures would not be satisfactory for use throughout, since their output power ratings are too low. The use of a miniature tube is not satisfactory in the first audio stage because of excessive microphonics. The particular type subminiature employed in this transmitter has been found to be excellent in its lack of microphonics and is used for this reason, rather than for any saving of space or filament power.

Previously-designed pack transmitters have been subject to severe detuning with changes in the position of the microphone cord. Tuning adjustments of the final tank circuit and antenna length for one position of the microphone cord were altered when the position of the microphone cord was changed relative to ground or to the transmitter case. In order to avoid this detuning effect as much as possible, radio-frequency chokes were applied in the three microphone leads at the microphone receptacle.

RESULTS

Measurements of output power have been made using a 5-ohm dummy antenna. With the full 180 volt B supply the output power measures 3.5 watts over the tuning range; with 157½ volts it decreases to 2.5 watts. Harmonic distortion, as measured by feeding a sine wave into the audio system and rectifying a portion of the radiated energy at 90 per cent modulation, was about 4 per cent root-mean-square total.

The useful transmission distance is subject to wide variation, depending upon noise level conditions at the receiver, receiver sensitivity and selectivity, and the propagation path. In general, it is felt that the useful range lies within the limits of from 1 to 20 miles. However, in New York City, for example, over a reasonably good path but with very severe diathermy interference, it was found that the useful range was about 3 miles.

A number of pack transmitters of this type have been constructed and are now in use throughout the country. Their performance has been highly satisfactory. At the Philadelphia national political conventions they were found to out-perform the earlier models of pack transmitters by a substantial margin, at the same time being appreciably smaller, lighter, easier to operate, and in general more reliable than the old models.

ELECTRO-OPTICAL CHARACTERISTICS OF TELEVISION SYSTEMS*

By

Otto H. Schade

Tube Department, RCA Victor Division,
Harrison, N. J.

NOTE: This paper consists of an Introduction and four parts: Part I—Characteristics of Vision and Visual Systems; Part II—Electro-Optical Specifications for Television Systems; Part III—Electro-Optical Characteristics of Camera Systems; Part IV—Correlation and Evaluation of Electro-Optical Characteristics of Imaging Systems. The Introduction and Part I appeared in the March 1948 issue of *RCA Review* and Part II in the June 1948 issue; summaries of these parts are reprinted herewith for reference purposes. Part III is included in this issue. Part IV is scheduled for publication in the December issue.

INTRODUCTION; PART I—CHARACTERISTICS OF VISION AND VISUAL SYSTEMS

(Reprinted from RCA REVIEW, March, 1948)

Summary—*The optical and electro-optical conversion processes in television systems are examined as intermediate stages of a multi-stage process by which optical information at the real object is "transduced" into sensory "response" at the brain. The characteristics of the human eye and vision in the final stage of the process determine the requirements and standards for preceding stages. When expressed on a unified basis by "transfer" and "aperture response" characteristics, the properties of the process of vision can be correlated with those of external imaging and transducing processes. It is shown that image definition, or the corresponding information from optical or electrical image-transducing stages, can be specified by the characteristics of an equivalent "resolving aperture". These characteristics may be computed and measured for all components of the system.*

Quantitative data from measurements permit definite quality ratings of optical and electrical components with respect to theoretical values. A subjective rating of the resolution in an imaging process external to the eye such as a television system is derived by establishing a characteristic curve for the relative "sharpness" of vision as affected by the "aperture response" of the external imaging process.

A general review of the material and the broad methods of analysis employed are given in the Introduction. Following this, Part I treats characteristics of vision and visual systems. In this part, viewing angle, sensation characteristics, color response, persistence of vision, flicker, re-

* Decimal Classification: **R138.3 × R583.12.**

solving power, response characteristics, and steady and fluctuating brightness distortions are discussed and related to the characteristics of external imaging systems and the television process.

PART II—ELECTRO-OPTICAL SPECIFICATIONS FOR TELEVISION SYSTEMS

(Reprinted from RCA REVIEW, June, 1948)

Summary—The ability of an image-forming device to reproduce fine detail can be specified basically by the size and flux distribution of the small light spot formed as the image of a point source of light. It is shown that the defining ability of practical image transducers is specified more accurately by response characteristics obtained by scanning a test object with the elemental point image which represents the "resolving aperture" of the imaging device. Methods of computing and measuring the "aperture flux response" of practical image transducers are developed for correlation of optical and electrical system components.

The television raster is treated as a sampling process and its effect on the system resolution is evaluated as an aperture process. Brightness, repetition rate, and flicker in television images are treated in relation to the screen materials and performance of practical kinescopes.

* * *

There follows the third paper in the series: Part III—Electro-Optical Characteristics of Camera Systems.

The Manager, RCA Review

PART III—ELECTRO-OPTICAL CHARACTERISTICS OF CAMERA SYSTEMS

Summary—Fundamental relations between energy and signal quality are derived by the concepts of discrete energy "samples" to establish the principal parameters controlling the energy transfer in television camera systems. Signal-to-fluctuation ratios and the transfer and aperture-response characteristics of practical camera components are evaluated subsequently. The optical characteristics of camera elements are discussed with particular emphasis on the aperture response of lenses as measured by a television microphotometer.

A. Fundamental Relations Between Energy and Quality in Image Signals

The quality of an imaging process may be judged by three properties of the final image: graininess, sharpness, and tone scale. These properties are measured and specified by the signal-to-"noise" ratio $|R|$, the aperture response $r\Delta\overline{\psi} = f(N)$, and the light-transfer characteristic (ψ image $= g\psi$ object) of the system, and are controlled, to a large extent if not entirely, by the characteristics of the transducing processes in the camera. The three characteristics are not independent of each other because the energy required by each one increases with the respective figure of merit.

An imaging process may be regarded fundamentally as a sampling process. (See Part II, A.5 (page 257 et seq.)) With a given number of energy "samples", such as a certain number of silver grains in photographic film, graininess can be decreased in exchange for resolution by integration of samples (defocusing) without effect on the tone scale in large areas. The transfer characteristic may be altered by contracting samples into one level by eliminating half tones to increase sharpness (as in line copying). The resolution may be increased in exchange for graininess by increasing the relative energy in small areas by an "aperture correction" process (masking).

The fundamental relations of signal properties can be readily established by the concepts of discrete energy samples which can be applied to all forms of energy or energy flux used in television and photographic processes.

1. Picture Element Number, Resolution, and Frequency Channel

The smallest detail area a which can be resolved by an imaging process is a small equivalent square the linear dimension of which is specified by the limiting resolution line number N_c or the balanced cutoff line number $\overline{N}_{co}$ in the image. This area will be defined as a "picture element". The number of picture elements is, therefore, given by

$$n = \overline{N}_{co}^{\,2} \times (H/V) \tag{38}$$

where H/V is the aspect ratio of the image.

In a television channel, the total number of picture elements is the number of half-cycles at cutoff frequency Δf in the active scanning time for one frame. Because of "blanking" periods, the number of active picture elements, n_c, for the camera image or n_k for the kinescope image, are

$$n_c = 1.7\, T_f\, \Delta f \qquad \text{and} \qquad n_k = 1.58\, T_f\, \Delta f \tag{39}$$

where $T_f =$ frame repetition time.

The relation between line number $\overline{N}_{co}$ and frequency channel Δf

given by Equation (1), (Part I, page 9) is obtained by combining Equation (38) and n_k from Equation (39). These equations furnish $\overline{N}_{co} = 410$ and $n_k = 224{,}000$ for a standard kinescope image ($\Delta f = 4.25$ megacycles, $T_f = 30$ cycles) and $n_c = 240{,}000$ for the camera image.

2. Sample Number and Signal-to-Fluctuation Ratio

A critical comparison of picture elements representing a constant level discloses deviations of intensity from a mean value. To find the cause of these deviations, the element is inspected with high magnification by removing electrical or optical channel limitations (at least theoretically). Each element is found to have a structure consisting of a certain number of "sub-elements" which are groups or aggregates of still smaller "quanta": atom groups forming the silver grains in photographic film, quanta of light in optical images, electron groups in television images. The size or density of the sub-element can be grown from extremely small aggregates or even single quantum units by multiplication or "development" processes, but the number of sub-elements or "samples" remains substantially constant. By counting the number of samples, the cause for the observed deviations from an average intensity value is found to be a variation in the number of energy samples within the picture-element dimension. When the distribution of samples is random (following the Gaussian error function), the average number of samples n' in the selected area (N_c) multiplied by the sample energy q is the signal energy $n'q$; the root-mean-square value $q\sqrt{n'}$ is the root-mean-square fluctuation or "noise" energy. The signal-to-"noise" ratio, or using the more adequate term the "*signal-to-fluctuation*" ratio $|R|$ is, therefore,

$$|R| = n'q/q\sqrt{n'} = \sqrt{n'}. \quad (40)$$

The total energy Q in an image representing a constant level is obviously the number of samples per element multiplied by the element number (n) of the image

$$Q = nn'q. \quad (41)$$

Combining Equation (41), (40), and (38) yields the relations

$$Q = |R|^2 nq \quad (42a) \qquad Q = |R|^2 \overline{N}_{co}^2 (H/V) q. \quad (42b)$$

The image or signal energy increases as the square of both signal-to-fluctuation ratio and resolution, and in proportion to the sample energy (q).

The physical size, i.e., the effective diameter δ_q of the sample, imposes a restriction on the maximum obtainable value of the balanced line number $\overline{N}_{co}$ because of the aperture effect of the sample (compare Part II, A.6). A picture element is not resolved by *random* samples unless the spatial or time dimension of the sample is smaller than the picture element. The aperture effect of the samples in grain structures can be measured by the same methods used for normal defining apertures and will be discussed in Part IV.

The reader is cautioned that the picture element area, in which a subelement count is made without channel restrictions, *does not* represent a scanning or resolving aperture with uniform transmission; such apertures resolve elements (N_c) smaller than their diameter. Equation (42b) is valid only for the particular "aperture" represented by an electrical frequency channel of constant sine-wave-amplitude response and sharp cutoff, as specified by the symbol $\overline{N}_{co}$. Determination of the size of physical or optical "apertures" and response characteristics yielding the same root-mean-square (r-m-s) fluctuation value as a specified electrical channel ($\overline{N}_{co}$), requires evaluation of the "filter factor" m (see Part I, page 32) of the aperture. The equivalent scanning aperture size will be evaluated in Part IV in the discussion of motion picture film characteristics.

3. *Frame Number, Storage Factor, and Energy Flux*

The number of images which must be generated per unit time by a television or motion picture camera is determined by the property of the visual process to store impressions for a certain length of time. Incomplete storage causes flicker which has, therefore, determined the number of image fields per second in television and motion picture systems. An estimate of the storage time in the visual process may be obtained from the following observations.

A square-wave brightness change with 100 per cent amplitude (black-to-white) occurring at a rate of 24 cycles per second causes a just perceptible flicker (good storage) at a critical brightness $B_c = 0.1$ foot-lambert (Part I, Figure 6, page 22). The flicker amplitude, however, increases rapidly when the brightness is increased, indicating a likely decrease of storage.

At an average field brightness $\overline{B} = 1$ foot-lambert, the amplitude of the 24-cycle fluctuation must be decreased by a factor of 16 to reduce flicker again to a just perceptible value. The effective storage time, hence, has decreased to less than 1/24 second assuming 2 per cent as the minimum perceptible brightness difference. Observations made by the author on interlaced television fields repeating at 60 cycles per second have shown,* similarly, that flicker becomes noticeable at $\overline{B} = 10$ foot-lamberts when consecutive fields are made to differ only 8 per cent in intensity when a 30-cycle fluctuation of this small amplitude

* The measurements were made on (standard) sulfide screen materials.

is introduced. It may be concluded that motion picture or television images shown at a rate of 48 and 60 fields per second, respectively, are stored and blended by the eye at "normal" brightness levels of 1 and 10 foot-lamberts, respectively, over a period hardly exceeding two fields. The motion picture shows a complete image in every field, the image content, i.e., the sample distribution, being changed every two fields. The number of simultaneous sample impressions alternates, therefore, from Q when the same frame repeats to not more than 2 Q when successive fields are different frames. The average number of sample impressions is, therefore, increased by a storage factor $s' \simeq 1.5$ at $\overline{B} = 1$ foot-lambert.

The situation is different in television. Because of the interlacing process, new samples are shown in the same image area only at a constant 30-cycle rate. The number of simultaneous impressions is not increased by storage beyond that of a single frame and the storage factor is unity ($s' = 1$ at $\overline{B} = 10$ foot-lamberts). Changes in the decay characteristics of kinescope screen materials or of the interruption periods in motion picture shutters may increase these storage factors somewhat. Increases in brightness levels above these (normal) values decreases s', while reductions of $\overline{B}$ can effect considerable increases of the effective storage factor of the visual process.

The energy flux ($\mathcal{F}$) which must be generated by a television or motion picture camera system is, therefore, $\mathcal{F} = Q/(T_f \times s')$ (43)

when $1/T_f$ = number of complete images or frames in one second and s' = storage factor of the eye with the tentative values $s' = 1$ at $\overline{B} = 10$ foot-lamberts for 30-frame television and $s' = 1.5$ at $\overline{B} = 1$ foot-lambert for 24-frame motion pictures shown with twice the field number. Combining Equations (43) and (42a) expresses the signal flux in terms of the average number of samples per unit time

$$\mathcal{F} = |R|^2 nq/(s'T_f) \qquad (44a)$$

or with Equation (42b) $\mathcal{F} = |R|^2 \overline{N}_{co}^2 (H/V) q/(s'T_f)$. (44b)

In optical units, the energy flux is the light flux $\mathcal{F} = \psi$ lumen; the smallest sample energy being one quantum of light. For white light, $q_o = 7.7 \times 10^{-17}$ lumen.

In electrical units, the energy flux is the current $\mathcal{F} = \mathrm{I}$ ampere; the smallest sample being the energy of one electron $q_e = 1.6 \times 10^{-19}$ ampere.

In motion picture film, the sample q represents one average or equivalent grain.

The signal current in a television channel is expressed by combining Equations (44a) and (39) $\mathcal{F} = |R|^2\, 1.7 \Delta f q_e$;

$$I = 2.72\, |R|^2 \Delta f \times 10^{-19} \text{ ampere} \qquad (45)$$

This current is, for example, the photocathode current in a light-spot scanner using a kinescope phosphor with negligible time delay. It is also the signal current of an image dissector tube divided by the gain of its electron multiplier. For a frequency channel $\Delta f = 4.25$ megacycles and $|R| = 100$, $I = 0.0116$ microamps.

It is apparent that the signal flux (or current) which is required for a given signal quality should be determined at a point where succeeding processes do not decrease the sample number. Processes which increase the sample energy (q) by (electron) multiplication or amplification do not affect the value $\mathcal{F}$ (or I) although they may change the balance between $|R|$ and N_c by an aperture effect which can be evaluated separately.

4. Additive Fluctuation Energies (Q_f)

Many conversion or amplifying processes introduce additional fluctuation energy which contributes nothing to the desired signal but increases the fluctuation level. The resulting change of the ratio $|R|$ can be determined as follows.

The normal signal-to-fluctuation ratio in a process is given by Equation (40).

Additional random samples representing the fluctuation energy $n_f q$ per element do not increase the desired signal n', but by being added to the total number increase the fluctuation value to $q(n' + n_f)^{1/2}$, with the result:

$$|R'| = n'/(n' + n_f)^{1/2} \quad (46) \qquad \text{or} \qquad |R'|^2 = |R|^2/(1 + n_f/n'). \quad (47)$$

By substituting this expression into Equation (42) or (44), it is readily seen that the flux $\mathcal{F}$ or the signal input energy Q to a process introducing additional fluctuation energy must be increased (to maintain a given ratio $|R|$) to the values

$$Q' = Q(1 + Q_f/Q) \text{ and } \mathcal{F}' = \mathcal{F}(1 + \mathcal{F}_f/\mathcal{F}) \qquad (48)$$

or the additional fluctuation energy will cause a decrease of $|R|$ to the lower value given by Equation (46).

Addition of a *constant* fluctuation energy alters the manner in which $|R'|$ varies as a function of signal energy.

The effect of inserting various *constant* amounts of fluctuation

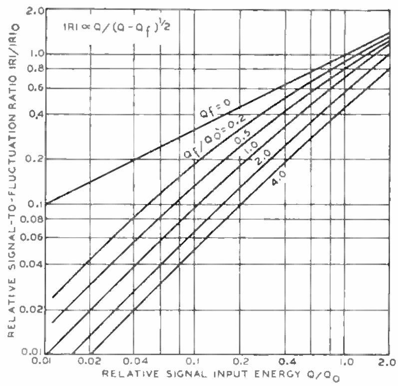


Fig. 43—Effect of introducing constant fluctuation energies on the normal signal-to-fluctuation ratios.

energy into a system is shown in Figure 43 where the ratio $|R|/|R|_o$ is plotted from Equation (46) as a function of the relative signal energy Q/Q_o with Q_f/Q_o as parameter. For $Q_f = 0$, the normal ratio $|R|$ decreases with the ½ power of the signal. The addition of a constant fluctuation energy may occur by introducing the shot "noise" of a beam or amplifier current in a television camera or the "fog" caused by a developing process, or stray illumination (haze, flare) in a photographic process. The addition causes a more rapid decrease of $|R|$ which decreases in proportion to the signal for the signal range where $Q_f \geqq Q$.

5. *Non-Linear Transfer of Signals and Fluctuations*

The relations between the signal-to-fluctuation ratio $|R|$ and the signal forms Q or $\mathcal{F}$ in Figure 43 are not disturbed by amplification in transducers having a linear transfer characteristic, i.e., a constant transfer factor (g) independent of signal amplitude and "operating point". The transfer of energy over a non-linear characteristic, however, alters $|R|$ because large and small signals are no longer affected in proportion.

The effects of signal distortion by non-linear transfer characteristics are well known in electron tube engineering. The transfer factor G for large unidirectional signals $\Delta\mathcal{F}$ measured from a zero point differs from the differential small-signal transfer factor g as illustrated by Figure 44. The large-signal transfer factor is the ratio of the increments $\Delta\mathcal{F}$ measured from the operating point 0, $G = \Delta\mathcal{F}_{\text{out}}/\Delta\mathcal{F}_{\text{in}}$ (49a)

while the differential transfer factor is determined by the slope of the characteristic

$$g = d\mathcal{F}_{\text{out}}/d\mathcal{F}_{\text{in}}. \quad (49b)$$

It is apparent that G as well as g are functions of $\mathcal{F}$. The choice of the zero-signal point or operating point 0 has a considerable effect on the relative values of signals with superimposed fluctuations.

When the operation point 0 coincides with the value $\mathcal{F} = 0$, the "transfer ratio" G/g has values smaller than unity when g is increasing steadily (Figure 44(a)) and values larger than unity when g is decreasing (Figure 44(b)). The signal-to-fluctuation ratio $|R|$ is then increased or decreased by the factor

$$|R|_{\text{out}} = G/g|R|_{\text{in}}. \qquad (50)$$

For power-law characteristics, the transfer ratio has a constant value. A square-law characteristic has the value $G/g = 0.5$, and for a cube-law characteristic (kinescope characteristic, see Figure 41, Part II, page 279), $G/g = 1/3$. The relative fluctuation increase in these cases is accompanied by a gradual expansion of signals ($\Delta\mathcal{F}$) and a correspondingly larger expansion of differential signals at all levels. Characteristics with increasing slope (g) increase, therefore, the contrast in the highlights as well as the detail contrast* although they decrease

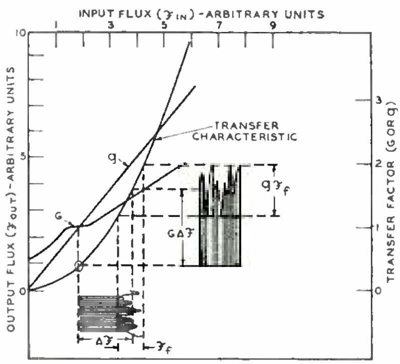


Fig. 44(a)—Decrease in signal-to-fluctuation ratio due to transfer characteristic with increasing slope.

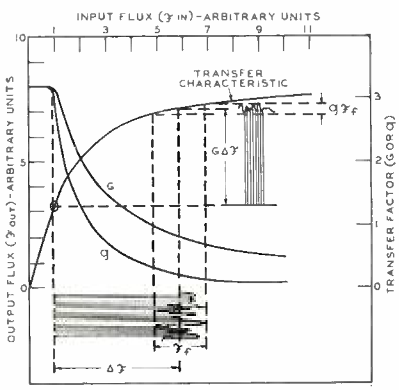


Fig. 44(b)—Increase in signal-to-fluctuation ratio due to transfer characteristic with decreasing slope.

the signal-to-fluctuation ratio $|R|$; an exchange which should be expected from Equation (42b). The opposite effect occurs with transfer characteristics of decreasing slope which cause a compression of signals. A ½-power characteristic, for example, has a constant transfer ratio $G/g = 2$. Fluctuations are reduced, i.e., $|R|$ increases by this factor, but the high-light contrast as well as the detail contrast at all levels is decreased by the same factor.

Practical transfer characteristics depart more or less from a true power law. Many processes also cause a reversal of signal polarity, i.e., "negative" image signals, where "black" represents a high flux value and "white" a low flux value. The reversal of polarity may then cause

* This detail contrast increase is not an "aperture" effect because the aperture response is a ratio measured at one given signal level.

the transfer ratio G/g to become inverted because G is a different function of $\Delta\mathcal{F}$.

When the operating point O is placed at a value $\mathcal{F} > 0$, which is the case in most amplifying processes as well as in the use of photographic negative film, the transfer ratio G/g is, in general, a function of the signal flux. Its initial value is always unity. With larger signals G/g increases or decreases depending on the curvature of the non-linear characteristic. Characteristics following a constant power law cause the transfer ratio to change from unity at small signals to the values obtained for $O = 0$ when the signal amplitude has become sufficiently large. The range of signal amplitudes in which this change occurs is small when the operating "bias" is small, i.e., when O is only slightly above zero level. A more gradual change of G/g from unity to its final value is obtained with larger "bias" values. It follows, in general, that the signal-to-fluctuation ratio $|R|$ may vary as a function of signal in many different ways depending on the curvature of the transfer characteristics and the position of the operating point O on the characteristics. Expansion of a signal or tone range decreases $|R|$ and increases the detail contrast within the expanded section, and vice versa. Restoration of linearity in the over-all "amplifying" process results again in the original ratio $|R|$.

6. *The Transfer Characteristic, Storage Capacity, and Signal-to-Fluctuation Ratio in Primary Signal-Conversion Processes*

Of particular interest are the effects of a non-linear transfer characteristic on the properties of the primary signal-conversion process in the camera system, i.e., the process which determines the lowest energy level and sample number in the signal generating system. A signal-conversion process may be linear or have a transfer characteristic following a law of diminishing returns. In the latter case, sample number and energy (see Equations (40) to (42)) may increase according to the law $Q = Q_{max}\,(1 - e^{-kQ_{in}})$
which is usually modified by secondary effects: retarding forces which increase with the number of transduced or stored samples as shown by the characteristics in Figure 45. Typical examples of transducers with diminishing returns are the eye, photographic film, and television pickup tubes such as the iconoscope, and the image orthicon. It is evident that characteristics of this type can cover a large range of input energy with a relatively small change in the total number of transduced or stored samples. The small range of Q can be expanded again in later processes by transfer characteristics with increasing slope.

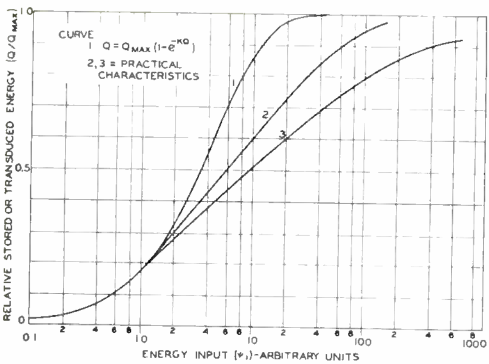


Fig. 45—Transfer characteristics with diminishing returns.

A primary signal-conversion process with a non-linear transfer characteristic of the above type offers basically no advantage over a linear process with respect to sensitivity or the total input (light) energy. It does require, however, only a fraction of the energy-handling or *storage capacity* of the linear process. To cover an input range of 200 to 1, the "sample" capacity of the linear process (1) in Figure 46 must be 12.5 times larger than that of the "logarithmic" process (2); for a given limited "capacity," therefore, the linear process is restricted in range. The storage capacity of photographic film, for example, is determined by the number of developable grains per unit area. It increases with grain density and area. The energy storage in television transducers (camera tubes) occurs in form of electrical charge storage

$$Q = CE_c. \tag{51}$$

The total electrical capacitance C of the storage area and the potential E_c to which it may be charged increase, therefore, with signal quality and range. The required capacitance can be related to resolution ($\bar{N}_{co}$ or Δf) and $|R|$ by substituting Equations (42a) and (39) into (51) yielding

$$C = 2.72\,|R|^2 \Delta f\, T_f \times 10^{-19}/E_c.$$

Because the developing process may introduce additional fluctuations

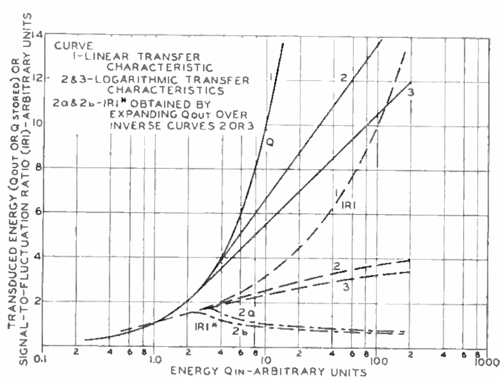


Fig. 46—Primary signal-to-fluctuation ratios $|R|$ and modified signal-to-fluctuation ratios $|R|^*$ for several transfer processes.

by an auxiliary current I_f, we obtain according to Equation (48) the more general equation

$$C = 2.72\,|R|^2 (1 + I_f/I)\,\Delta f\; T_f \times 10^{-19}/E_c \quad (52)$$

where C = capacitance in farad, I_f = auxiliary current (beam current) required for obtaining the particular signal current I at the ratio $|R|$, and E_c = maximum charge potential on C in volts. (The performance of practical camera tube types is discussed in Section B.2).

The signal-to-fluctuation ratio of both linear and logarithmic characteristics is plotted against input energy in Figure 46. It increases in proportion to the square root of the output energy (Q), but in the logarithmic process (Curve 2), $|R|$ is no longer proportional to the square root of the input energy (Q_{in}) and increases only by a factor of two from $Q_{in} = 4$ to $Q_{in} = 200$. The slow increase results from the "saving" in output energy but is not necessarily a disadvantage. When expanded in later transducing stages to give a *linear* over-all response, the ratio $|R|$ is decreased as shown by Curve (2a). The expansion can be made in several stages, as long as the product produces the inverse transfer characteristic of the primary process.

Transfer characteristics with lower slope (Curve (3) in Figure 46) will cause a larger decrease of $|R|$ as a function of signal when the over-all response is linearized (Curve 3a). A certain slope can, therefore, be found for which $|R|$ will remain fairly constant.* This condition permits a maximum in performance with a minimum output energy and storage capacity because the constant incremental sensitivity of the eye in the range above 1 foot-lambert indicates that a constant $|R|$ in this range is probably satisfactory. (Discussed further in Part IV.)

The logarithmic response of the visual process (Figure 3a, Part I, page 16) suggests that the sensation scale (S) may be related by a similar mechanism to the sample energy Q transduced by the light conversion process and that $|R|$ is constant for $B > 1$ foot-lambert. Threshold conditions imposed by the fluctuation ratio $|R|$ in the toe region of the transfer characteristic have been discussed recently by A. Rose in several papers[26,27].

The transfer characteristics, transfer efficiency, and aperture re-

---

* If a certain degree of compression is retained (as in motion pictures), the optimum slope of the primary characteristic has a lower value than required for a linear reproduction.

[26] A. Rose, "A Unified Approach to the Performance of Photographic Film, Television Pickup Tubes and the Human Eye", *Jour. Soc. Mot. Pic. Eng.*, Vol. 47, No. 4, p. 273, October, 1946.

[27] A. Rose, "The Sensitivity Performance of the Human Eye on an Absolute Scale", *Jour. Opt. Soc. Amer.*, Vol. 38, No. 2, pp. 196-208, February, 1948.

sponse characteristics of optical and electrical components in practical camera systems will be evaluated in subsequent sections.

B. Transfer Characteristics and Aperture Response of Camera Components

1. *Optical Characteristics*

The camera lens performs two principal functions: the collection of light energy emitted or reflected from an object or scene; and, the transfer of this energy to a photosensitive surface on which it must form an accurate image of the energy distribution in the object. The energy level in the image is determined, therefore, by parameters controlling the collection and transfer efficiency of the lens, while the accuracy in reproducing the energy distribution is specified by the transfer and aperture-response characteristics of the lens.

a. Collection Factor and Transfer Factor ($g_{(o)}$)

The camera lens covers a field of view with the diagonal angle 2θ. The field is in sharp focus at a certain lens-to-object distance d (see Figure 47). For a standard television aspect ratio ($H/V = 4/3$), the

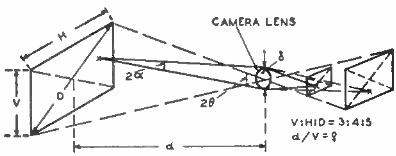


Fig. 47—Light flux pickup of camera lens.

projected dimensions of the object area (A_o) is $A_o = (4/3)\,V^2$. (53)

Expressed by the viewing ratio $\rho = d/V = 1/(1.2 \tan \theta)$ (54)

the object area is given by $A_o = (4/3)\,(d/\rho)^2$. (55)

The light flux reflected from A_o is determined by measuring the average flux density $\overline{B}_o$ (foot-lamberts) with an exposure or foot-candle meter. The object light flux is given by $\overline{\psi}_o = \overline{B}_o\,A_o/144$ lumen or with Equation (55)

$$\overline{\psi}_o = 0.92\,\overline{B}_o\,(d/\rho)^2\,10^{-2}\ \text{lumen}\ (d \text{ in inches}). \quad (56)$$

The camera lens admits to the camera a cone of light with the solid angle 2α from the hemispherical radiation of each elemental object point. Assuming a Lambert distribution of the object-point flux, the fraction K admitted by the lens diameter δ to the camera is

given by $\psi_{2\alpha}/\psi_{2\pi} = K = \sin^2\alpha;\ K = 1/(\,(2d/\delta)^2 + 1)$ (57)

The collection factor K decreases for points near the edges of the field especially for thick lenses and large angles of view (θ)*. With few exceptions the object distance is much larger than the lens aperture

$(d >> \delta)$ permitting the simplification $$K = (\delta/2d)^2 \tag{58}$$

In passing through the lens, the light flux is reduced by reflections on glass-air surfaces and by absorption in cement layers. These losses are specified by the transmission factor. For normal angles, a good three-piece lens with six glass-air surfaces and two cement layers has a transmission factor $\tau \simeq 0.92$ (coated) and $\tau \simeq 0.75$ (uncoated) The product τK specifies the efficiency of the lens in collecting and transducing light flux to the image surface.

$$\overline{\psi}_i/\overline{\psi}_o = \tau\ (\delta/2d)^2 \tag{59}$$

Combining Equations (59) and (56) yields an expression for the light flux in practical units

$$\psi_i = 2.3\ \tau\ (\delta/\rho)^2\, B_o\, 10^{-3} \text{ lumen (Lens diameter } \delta \text{ in inches)}. \tag{60}$$

The ratio $$g_o = \Delta\psi_i/\Delta B_o = 2.3\ \tau\ (\delta/\rho)^2\, 10^{-3} \tag{61}$$

is *the optical transfer factor of the television camera*. It is proportional to the lens-stop diameter squared but independent of f: number and focal length (F) of the lens.

The optical transfer factor for photographic film is more adequately expressed by an illumination-to-brightness ratio

$$g_{of} = \Delta E_i/\Delta B_o \tag{62}$$

because the film exposure depends on light flux density, the total flux ψ_i being, therefore, a function of image area. By dividing Equation (59) by the ratio of the areas which equals the linear magnification squared: $$A_i/A_o = M^2$$

and with the substitutions $\delta = F/f\colon$ *and* $d = (\dfrac{1}{M} + 1)\, F$

The optical transfer factor for film is obtained in the familiar form:

$$g_{of} = \Delta E_i/\Delta B_o = \tau/(2f\colon(M+1))^2 \tag{63}$$

A specific photographic "signal" requires a given energy Q per

* (See Section B.1(c).)

unit of film area to effectively serve as an attenuator of light after development. The total energy increases, therefore, with film area, but the signal quality increases because of a corresponding increase in resolution.

The television signals from an f:2 lens with 2-inch focal length to an image "plate" with a 1.5-inch diagonal, or from an f:6 lens of 6-inch focal length to an image plate with a 4.5-inch image diagonal, however, are identical because the diameters (δ) of these lenses are equal and, also, the value of light flux ψ_i. (See Equation (60)). The difference is caused by the fixed resolution in a given frequency channel which specifies a constant energy flux (See Equation (45)) for a given signal quality.

The selection of a large or small television camera image depends, therefore, on factors such as the relative aperture response of lenses and pickup tubes of different size in the range below the cutoff resolution $\overline{N}_{co}$ of the system.

b. Optical Transfer Characteristics and Contrast Range in Camera Images

The light transfer characteristic of optical imaging systems departs from the ideal linear relationship $E_i/B_o =$ constant because of reflection and scattering of light in the camera.

Reflections and diffusion of light in the camera lens can be treated as an "aperture" effect caused by a large diffusion disc of very low intensity (See Part II, Figure 19, page 252) which reduces the flux response at low line numbers. This effect, termed "lens flare", is of a different order than the detail response factor at higher line numbers (See next Section). Lens flare is reduced considerably by "coating" the glass-air surfaces of the lens, but depends also on the general correction of the lens which determines its detail response. The large diameter of the diffusion disc (in the order of centimeters) causes the optical transfer characteristic to become a function of image content. This function is best illustrated by the transition curve from a unit function boundary which can be measured statically by projection in an enlarger (Figure 48). For a rough but useful approximation in evaluating the effects of lens flare the back of the lens surface may be considered as a source of diffused image light flux $\bar{\psi}_b$ which is a certain fraction

$$b = \bar{\psi}_b/\bar{\psi} \tag{64}$$

of the total light flux ($\bar{\psi}$) transmitted through the lens. The value of b is in the order of $b \simeq 0.05$ for uncoated lenses and $b \simeq 0.01$ for coated lenses.

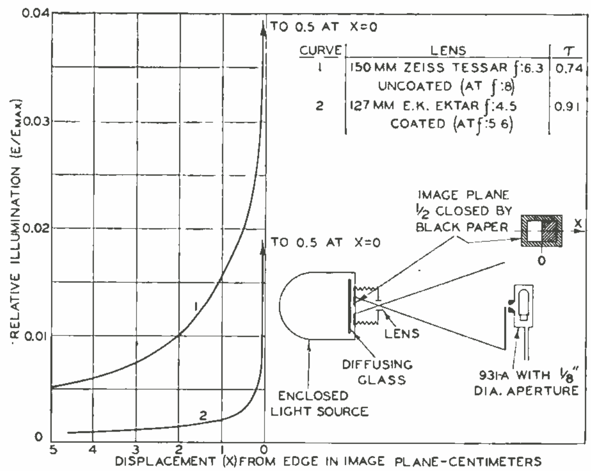


Fig. 48—Transition curves for coated and uncoated lens.

The light flux $\bar{\psi}b$ is superimposed as a more or less uniform "light bias" on the image flux, displacing the transfer characteristic by an additive constant $b\,\bar{E}$ as shown in Figure 49a. The value $b\,\bar{E}$ is, therefore, the minimum illumination in the transferred image which limits the maximum contrast in the image to $C_{\max} = (\hat{E} + b\,\bar{E})/b\,\bar{E}$. (65)

The average illumination $\bar{E}$ of the image plane can be considerably higher than the average image illumination $\bar{E}_i$ *inside* the useful image frame when light sources are imaged by the lens outside the frame area. To illustrate this case, assume a normal ratio of peak-to-average illumination $\hat{E}_i/\bar{E}_i = 6$ in the image. A bright sky imaged outside the frame may, however, raise the total average to $\bar{E} = 10\,\bar{E}_i$. For an uncoated lens with $b = 0.05$, Equation (65) furnishes the maximum image contrast $C_{\max} = 13$. A coated lens ($b = 0.01$) would give a maximum contrast $C_{\max} = 61$.

The contrast range is increased considerably by restricting the light pickup of the lens to the useful image area by means of a rectangular shade of proper dimensions. The average values $\bar{E}$ and $\bar{E}_i$ are then identical which, for the above example, increases the image contrast to $C_{\max} = 121$ and 601, respectively.

In television transducers, such as the image orthicon, a lens shade is particularly effective because the deterioration of the tone scale in the "blacks" is augmented by "electron reflection" from saturated high-light areas in the transmitted image, or imaged on unused por-

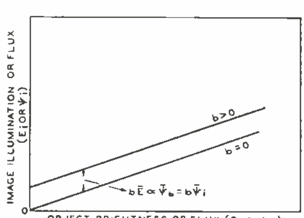


Fig. 49(a) — Transfer characteristics and "flare" light-bias of a lens.

tions of the photocathode. To prevent further loss of contrast, the camera walls should be treated to have a very low reflection coefficient. Surfaces should be rough, broken up into cavities by grooving, or coated with fluffy dead black material (velvet, etc.). Light baffles or masks must cover unused portions of the image surface, which itself should be made as non-reflecting as possible. When the image surface is transparent, as in modern transducers, all tube parts exposed to transmitted light should be treated similarly. Most important, however, is the use of lens shades to prevent excess light from passing through the lens and entering the camera.

Measured values of maximum image contrast ranges are shown in Figure 49(b) as a function of the peak-to-average ratio $\hat{B}/\bar{B}$ (illumination ratio $\hat{E}/\bar{E}$ in Figure 49(b)) for simple test objects.

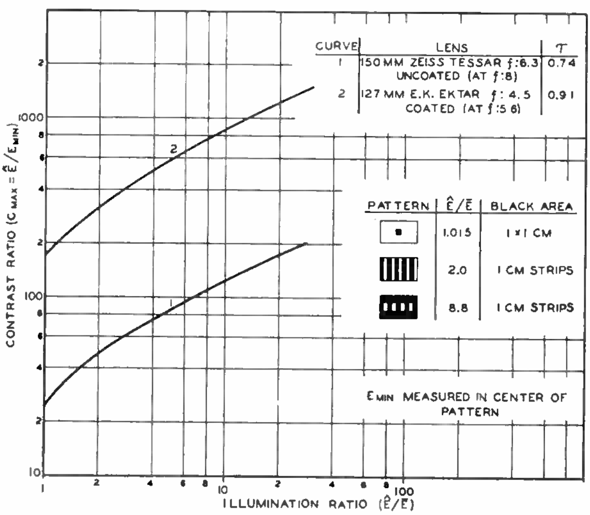


Fig. 49(b)—Large-area contrast in optical camera images.

Correction of the transfer curve

It is electrically quite simple to subtract a light bias by adjusting the electrical black level (clipping level or shading), provided b is a constant or varies smoothly from top to bottom or side to side of the image. If the transducer has a linear transfer characteristic (in the blacks especially), the original linear tone scale can be restored in this manner.

In non-linear processes the correction is more difficult requiring a curved transfer characteristic. In both cases the "bias" light flux contributes nothing to the useful signal and is, therefore, a fluctuation energy (Q_f or $\mathcal{T}_f$) decreasing the ratio $|R|$ as shown in Section A.4.

c. Resolving Power and "Aperture" Response of Camera Lenses

The resolving power of a perfect lens is limited by the finite wavelength (λ) of light. The optical elements filling the lens opening may be perfect for directing an undistorted wavefront from an object point to a point focus, but the finite boundaries of the lens give rise to phase differences and spreading of light at the edges of the limited wavefront section. This "diffraction" of light causes interference effects. The point image is spread out over a disc area and surrounded by dark and

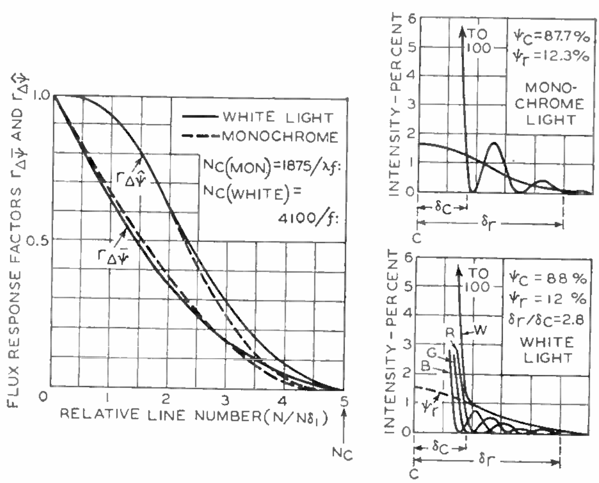


Fig. 50—Aperture flux distribution and response factors of theoretically perfect lens.

light rings of low intensity. (See Figure 50). The effect of diffraction decreases with distance and increasing ratios of lens area to boundary length. The diffraction disc diameter δ_c is, therefore, proportional to the focal length F and inversely proportional to the lens or stop diameter δ[28],[29]

$$\text{For } F/\delta > 1.5 \qquad \left.\begin{array}{l} \delta_c = 1.22\ (2\lambda/\delta) F \\ \delta_c = 2.44\ \lambda\ f: \end{array}\right\} \qquad (66)$$

[28] M. Born, OPTIC, EIN LEHRBUCH DER ELECTROMAGNETISCHEN LICHTTHEORIE, Springer, Berlin, 1933 (p. 207).
[29] C. P. Shillaber, PHOTOMICROGRAPHY IN THEORY AND PRACTICE, John Wiley & Sons, Inc., New York, N. Y., 1944.

The perfect lens has a resolving power limited by the diffraction disc size, which is its circle of confusion and effective resolving "aperture". The light flux[28] within the central disc area can be approximated by a cosine-square distribution (Figure 50) with the equivalent diameter $\delta_c \simeq 2.2\,\lambda f\!:$. See (Part II). Because the limiting resolution is $N_c = 4.35\,N\delta$ the diffraction cutoff line number for monochromatic light in television lines per millimeter is

$$\text{For } f\!: > 1.5 \qquad N_c = 4.35 \times 10^3/\delta_c = 1875/\lambda f\!:$$

$$\text{and for } f\!: < 1.5 \qquad N_c = (4100 \times \text{numerical aperture})/\lambda \tag{67}$$

with λ expressed in microns (1 micron = 0.001 millimeter).

The aperture response characteristic of a perfect lens is a complex curve which can be approximated by the sum of two response characteristics. The flux rings are approximated by a coaxial diffusion disc $\delta_r = 2.8\delta_c$ with a flux amplitude $\hat{\psi}_r/\hat{\psi} = 0.016$ and a flux distribution

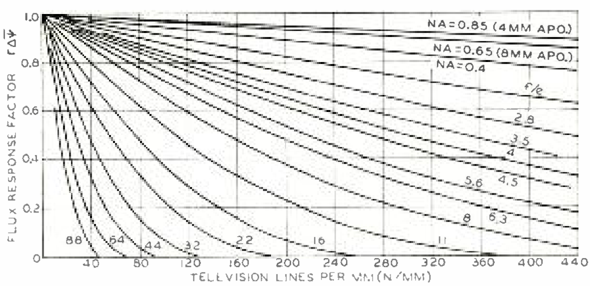


Fig. 51—Flux response characteristics of theoretically perfect lens as function of $f\!:$ number.

equal to that of the main "aperture" δ_c as shown in Figure 50. (See complex apertures, Part II). The theoretical aperture characteristics* of the perfect lens for monochromatic and white light obtained by this synthesis are shown in Figures 50 and 51, and Table V. The aperture response for white light has been computed by superposition of three coaxial monochromatic "apertures" with equal flux at $\lambda_1 = 0.45\,\mu$, $\lambda_2 = 0.55\,\mu$, and $\lambda_3 = 0.65\,\mu$. Figure 50 furnishes the relations $N_c/N\delta = 5$. For $f\!: > 1.5$, $N_{c(w)} = 4100/f\!:$ (Television lines per millimeter). (68a)

For microscope objectives with $f\!: < 1.5$

$$N_{c(w)} = 8400 \times \text{numerical aperture} \tag{68b}$$

* The resolving "aperture" should not be confused with the lens stop "opening" which is normally designated as the lens aperture.

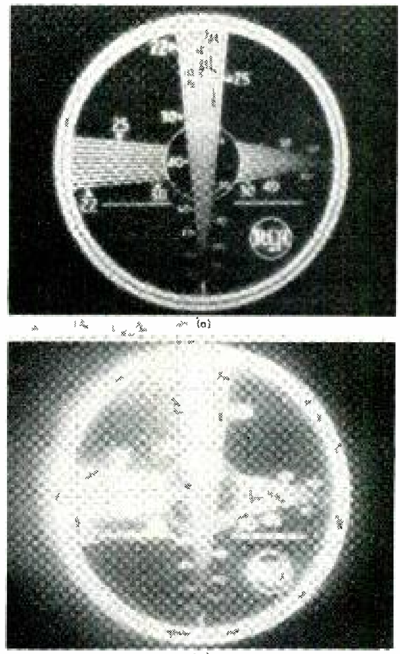
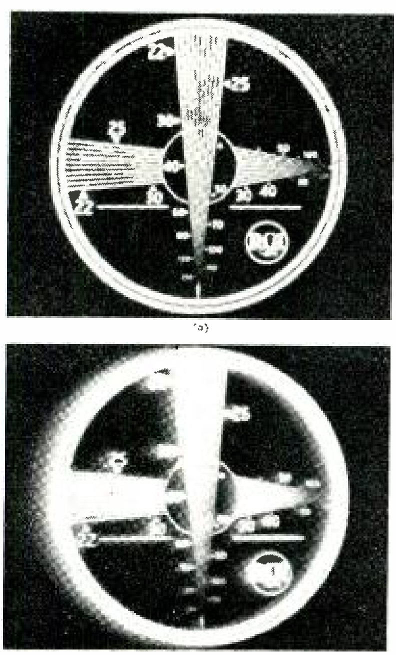


Fig. 52—Photomicrographs of test-pattern image (½ millimeter) taken at f:1.9 with white light.

a. On lens axis ($N_c \simeq 1000$)

b. 10 degrees off lens axis

Theoretical cutoff values N_c are obtained in the center of the field with good small-diameter lenses and with high-quality lenses of large diameter at two to three stops below maximum opening. The theoretical aperture response shown in Figure 51 is not attained with practical lenses and especially not with open large-diameter lenses becauses of distortions and phase delays which occur in portions of the light wavefront.



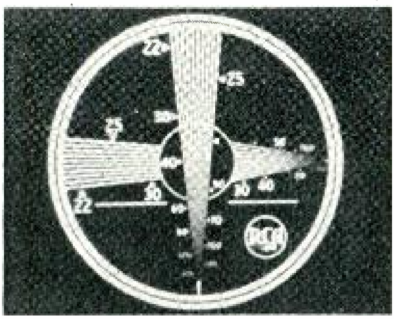
As a result, the effective resolving aperture diameters δ_c and particularly δ_r increase and the "aperture" flux response $r\Delta\bar{\psi}$ remains low, (exhibiting a pronounced "knee") far below the limit of resolving power which is often maintained by the density peak in the "aperture" flux. This point may be illustrated first by the photographic performance of a well-corrected coated lens.

The photomicrographs, Figures 52, 53, and 54 show the

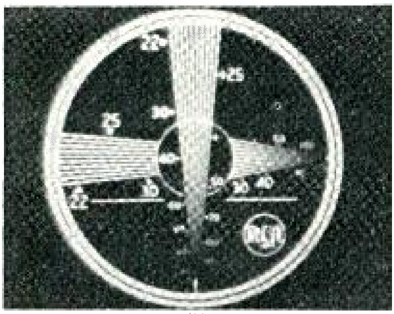
Fig. 53—Photomicrographs of test-pattern image (½ millimeter) taken at f:4 with white light.

a. On lens axis ($N_c \simeq 900$)

b. 10 degrees off lens axis



(a)



(b)

Fig. 54—Photomicrographs of test-pattern image (½ millimeter) taken at f:8 with white light.

a. On lens axis ($N_e \simeq 480$)

b. 10 degrees off lens axis

relative excellence of an f:1.9 50-millimeter lens, as diffraction cutoff is obtained up to f:4 for axial points.# Larger stops show less resolving power and contrast due to spherical aberration and flare. The degradation for angular rays is evident for $f: < f:8$.

These photographs were taken with a microscope focussed normal to incident rays on the image point. They do not show the effects of field curvature. The loss of resolution caused by angular rays and field curvature is illustrated by Figure 55, showing a test object photographed on an Eastman Kodak high-resolution plate.** The small test patterns in the center and the top right corner have again a diameter of 1/2 millimeter on the plate, an angular separation of approximately 19 degrees, and are shown highly magnified in Figures 56 and 57, respectively.

Because the eye and the photographic process have non-linear transfer characteristics, it is difficult to estimate the aperture response of the lens from optical images. A photometric trace of the light flux variations in a square-wave test-pattern image can be obtained by observing the image with high magnification through a television system. This "television microphotometer" transduces optical flux variations into electrical current variations permitting accurate measurement of the aperture response of lenses. A brief description of the apparatus is given in a subsequent section.

The pattern image is 0.5 millimeter in diameter; the numbers on the pattern multiplied by ten give television lines per millimeter. The diffraction cutoffs appear in the original photographs but have been lost in the illustrations.

** Type 649 GH Spectroscopic plate.

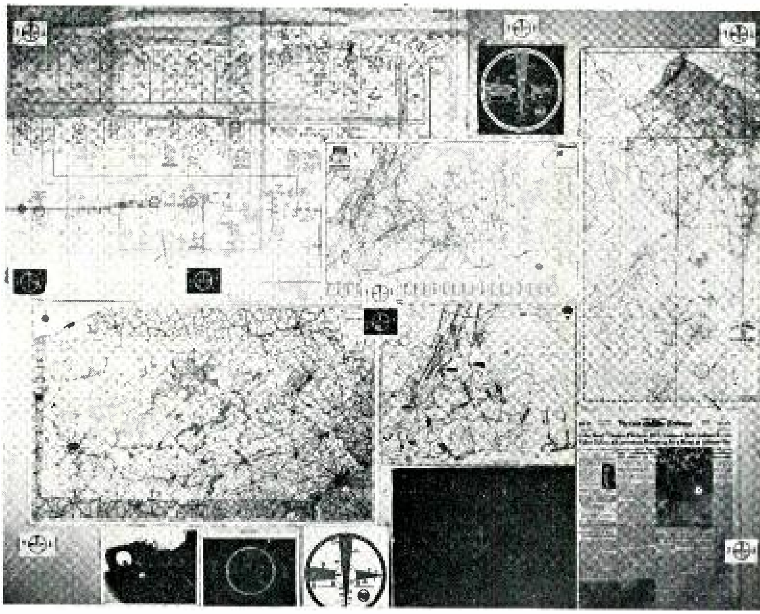


Fig. 55—Test object photographed on Eastman Kodak high-resolution plate with f:1.9, 50-millimeter lens at f:4.

Because of the variety of parameters and factors which control the performance of a lens, a number of curve families are required to describe its aperture response characteristics and to form an opinion on the lens quality. The effect of *chromatic aberration* on the aperture response is illustrated in Figure 58(a) and (b) on two similar high-quality camera lenses of different design at one (optimum) lens stop. The lenses are focused and measured with green light (Wratten filter #58). The measurement is repeated with different light colors (white, #47 blue, #26 red) without disturbing the focus adjustment. The lens in Figure 58(a) is not in focus for red light (zero point at $N = 140$ lines per millimeter) because of chromatic aberration. The chromatic error increases with lens stop diameter and causes a reduction of the white-light aperture response as shown. Both characteristics indicate a "diffusion" disc in the light flux distribution ("knee" in the curves). The relative performance of the same lenses *at different angles* from the optical axis is shown in Figure 59(a) and (b). The situation is now reversed. The lens with better chromatic correction is not as well in focus at the image plane at larger angles partly because of a larger "field curvature", but even when "focused in" there is little difference at 20 degrees between the two lenses. These errors increase normally with larger stops. The curves show also the decrease of the collection

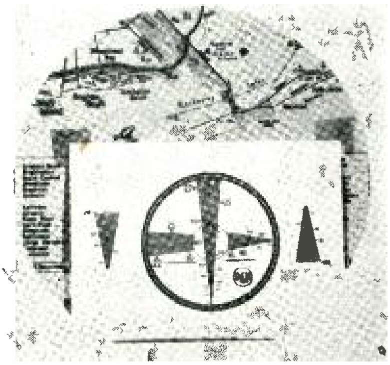


Fig. 56—Enlargement of ½-millimeter test-pattern circle in center of plate (Figure 55) ($N_c \simeq 760$).

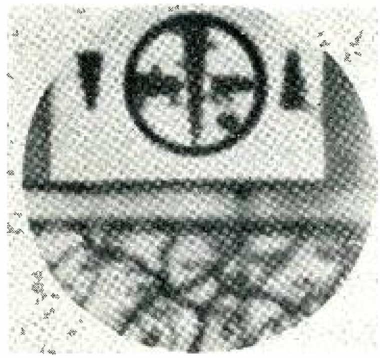


Fig. 57—Enlargement of ½-millimeter test-pattern circle at corner of plate (Figure 55) at $\theta = 19$ degrees.

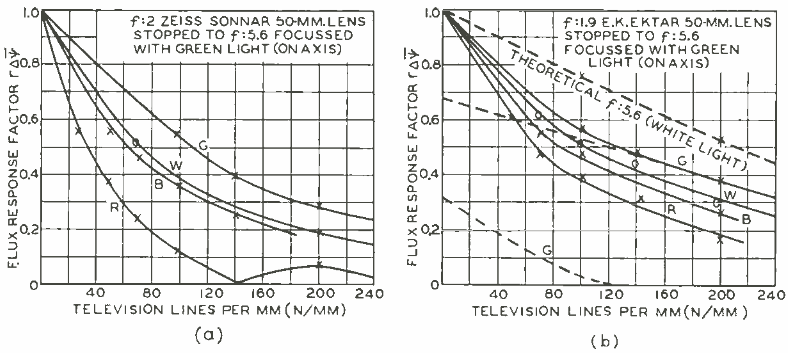


Fig. 58—Flux response characteristic of miniature-camera lenses illustrating chromatic aberration.

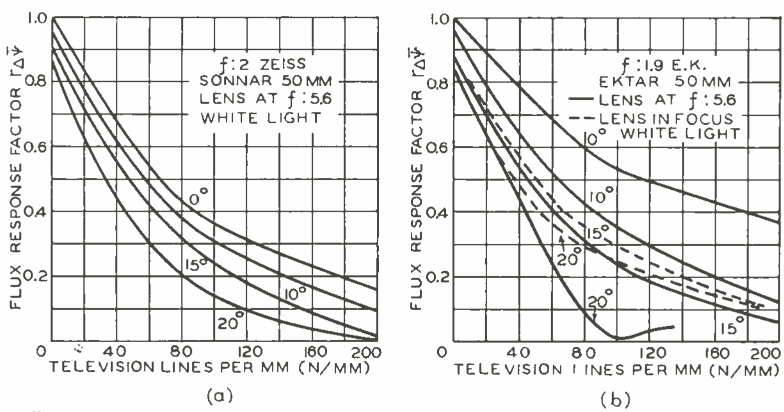


Fig. 59—Flux response characteristic of miniature-camera lenses for angles from 0 to 20 degrees off axis.

Table V—Theoretical Aperture Response Factors of Lenses (Figure 50)

Monochrome Light			White Light*		
$N/N\delta_1$	$r\Delta\hat{\psi}$	$r\Delta\bar{\psi}$	$N/N\delta_2$	$r\Delta\hat{\psi}$	$r\Delta\bar{\psi}$
0.5	0.99	0.83	0.5	0.99	0.815
0.75	0.97	0.74	0.75	0.965	0.725
1.	0.93	0.66	1.0	0.93	0.64
1.5	0.80	0.51	1.5	0.80	0.49
2.	0.615	0.39	2.	0.625	0.36
2.5	0.42	0.27	2.5	0.45	0.26
3.	0.25	0.16	3.	0.30	0.175
3.5	0.114	0.08	3.5	0.176	0.11
4.	0.035	0.022	4.	0.088	0.056
4.35	0.	0.	4.5	0.03	0.02
			5.	0.	0.

λ	450 $m\mu$**	550 $m\mu$	650 $m\mu$
δ	0.82	1.	1.18
$\hat{\psi}$	0.47	0.31	0.22
ψ	1/3	1/3	1/3

* Composite response of three coaxial spots.
** Millimicrons.

factor K with angle, because the response does not increase to unity at $N = 0$. The curve families in Figure 60 illustrate the change of aperture response as a *function of the f number*. The response increases at first when the lens is stopped down (increase of f number) because spherical and chromatic aberrations are decreased; for smaller stops it decreases as may be expected from theoretical considerations. (See Equation (68a) and Figure 51). A photograph of the aperture-response changes as seen on the oscilloscope of the television microphotometer is shown in Figure 61. This figure illustrates the progressive diffraction cutoff from $f/8$ to $f/11$. It should be noted that the response at lower line numbers ($N = 100$/millimeter) increases from $f/2$ to $f/8$.

Although most lenses have a much higher limiting resolution (N_c) than can be used with normal photographic film or a television channel, their aperture response departs considerably from theoretical values at low line numbers. For an image height $V = 25$ millimeters, which is the approximate size in miniature cameras and the present image orthicon (2P23 and 5655), the line number $N = 500$ represents 20 lines per millimeter. The response factors at this line number for a number

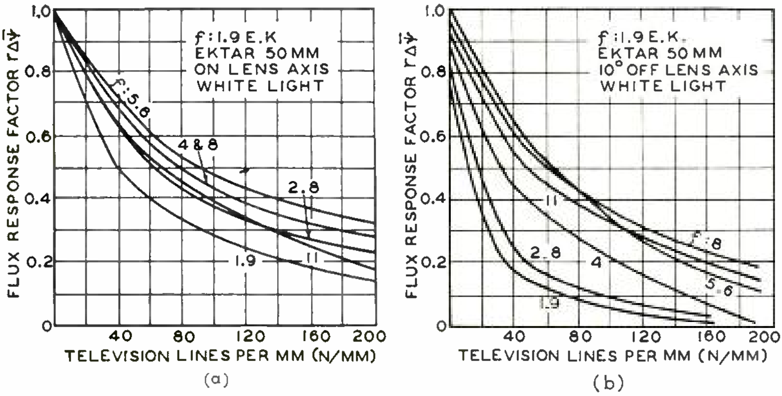


Fig. 60—Flux response characteristics of miniature-camera lens as function of f: number. (Compare Figures 52, 53, and 54).

of high-quality camera lenses are shown in Figures 62(a) and 62(b) as a function of lens diameter (δ/F) and f number. If a 10-degree angle is selected as an average value, it can be seen that the response $r\Delta\bar{\psi}$ at 20 lines per millimeter ($N = 500$ for $V = 25$ millimeters) averages $r\Delta\bar{\psi} = 0.7$ at small lens stops ($f: > 4$) and decreases below this value for "faster" lenses. For the purpose of computing the over-all aperture response of the system, the line number $N_{0.5}$ at which $r\Delta\bar{\psi} = 0.5$ has been plotted as a function of f number in Figure 63(a) and (b).

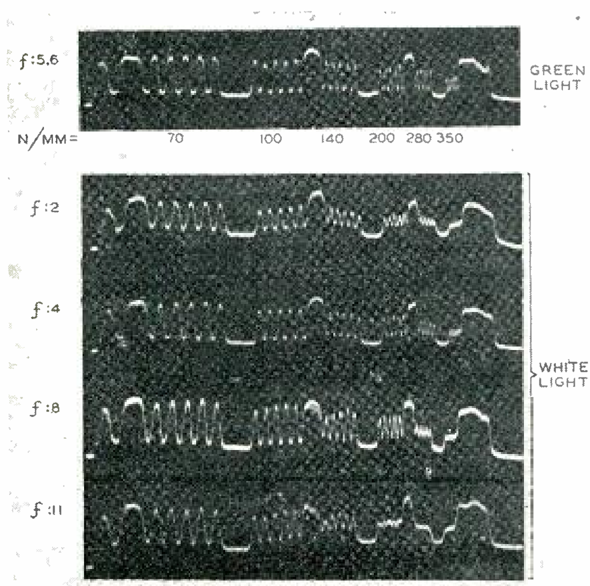


Fig. 61—Oscillograms of aperture response of f:1.9 miniature-camera lens ($0 = 0$ degrees) taken with television microphotometer (Figures 71 and 72).

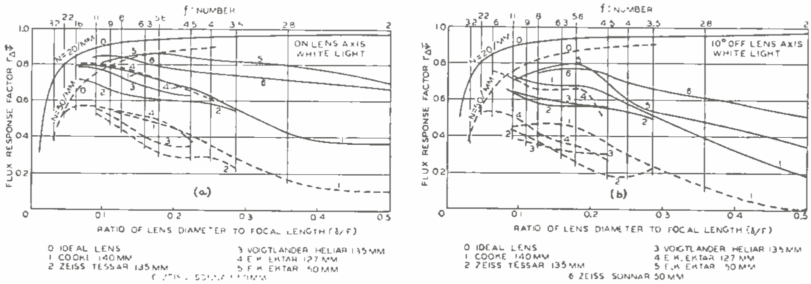


Fig. 62—Flux response characteristics of various high-quality camera lenses at specified detail size.

An examination of various lens characteristics indicates that aberrations decrease, in general, with the focal length F of the lens although not in proportion. A larger image size results, therefore, in a generally higher response factor for a given f number and line number N in the image dimension. Because the transfer efficiency of the television camera lens is constant for a given lens diameter δ, an increase of image dimensions corresponds to an increase in the f number of the lens ($f{:} = F/\delta$) and a correspondingly rapid increase in the relative response factor of the lens. As the largest lens diameters and collection factors (See Equation (58)) are obtainable in lenses with a long focal length the sensitivity of the television camera increases with image size. The selection of practical image sizes is, therefore, governed by factors such as weight, availability of lenses, the required mechanical precision in tube and camera construction as well as certain electrical operating requirements which are a function of size.

Lens Diameter and Aperture Response for Deep Fields of View

When imaging three-dimensional fields, the camera image can be

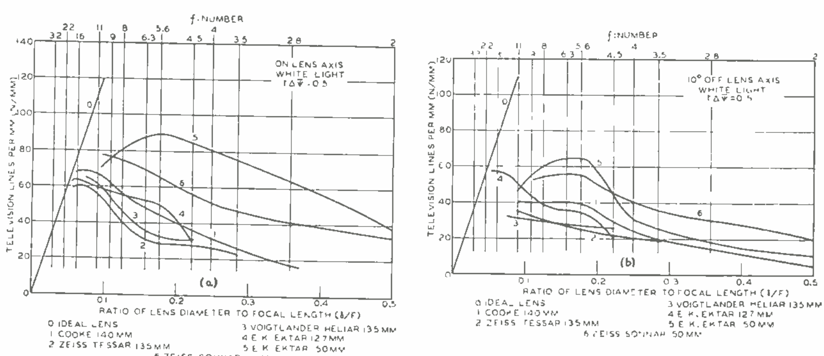


Fig. 63—Line number reproduced with 50 per cent response by various high-quality camera lenses.

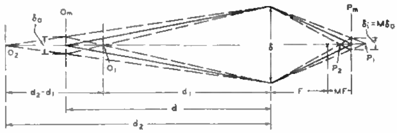


Fig. 64 — Relations between lens diameter, circle of confusion, and depth of field for finite distances d_1 and d_2.

in focus at one distance only. The effective resolving "aperture", therefore, is enlarged for planes at the front and rear limits of deep fields to a "disc of confusion". The relations between lens diameter, "depth of field", and disc of confusion are easily established by geometrical optics from the ray diagram Figure 64.

The object plane O_m is assumed in sharp focus for an image plate placed at P_m. Points in all other planes between O_1 and O_2 are out of focus and are imaged as discs of varying diameter but not exceeding the diameter $\delta_i = M\,\delta_o$. This diameter is obviously determined by the out-of-focus distance and the lens stop diameter, and has a minimum value in a given "depth of field" when the lens is focused to the mean distance

$$d = 2d_1\,d_2/(d_1 + d_2). \tag{69}$$

The depth of field $(d_2 - d_1)$ is related to δ and δ_o by

$$d_2 - d_1 = (d_1 + d_2)\;\delta_o/\delta. \tag{70}$$

The theoretical disc of confusion (δ_o) is a round resolving "aperture" with cosine-square flux distribution (See Part II). To maintain an aperture response $r\Delta\overline{\psi} = 0.71$ within the field limits, its largest diameter should not exceed the value

$$\delta_o = V_o/\overline{N}_{co}. \tag{71}$$

$(V_o =$ Vertical object field dimension)

By expressing the object depth by its ratio to the field height:

$$p = (d_2 - d_1)/V_o \tag{72}$$

and combining Equations 70, 71 and 72, the lens stop diameter required for $N\delta_o = \overline{N}_{co}$ is specified by $\delta_{\max} = (d_1 + d_2)/(p\overline{N}_{co})$ for the mean distance $d = \rho(d_2 - d_1)/p$ (73)

A representative depth of field may be obtained from the consideration that three-dimensional scenes or objects are frequently extending over a depth equal to their vertical dimension (i.e., $p = 1$). Table VI has been computed with Equation (73) for equal depth and height $(p = 1)$ of the "sharply" imaged field for a fixed distance $d = 18'$ and the sharpness limit $\overline{N}_{co} = 400$. (Camera equipped with lens turret and fixed position.) The lens stop diameter $(\delta_{\max})$ is practically constant

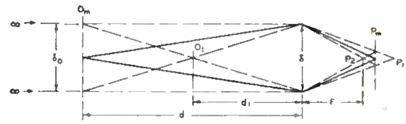


Fig. 65 — Relations between lens diameter, circle of confusion, and depth of field for hyperfocal conditions.

for all conditions. The lower section of the table lists focal length (F) and f: number for this stop diameter for four different image plate diagonals D; all give the same sharpness and, for a given viewing ratio ρ, equal television signals. *The optical transfer factor g_o (Equation (61)) is, therefore, a constant specified by the required sharpness in depth and the viewing angle.* This fact remains substantially unchanged for variations of camera distance. When the rear distance limit (d_2) is moved to infinity, a depth of field from d_1 to infinity requires focusing

to the "hyperfocal" distance (See Figure 65) $\quad d = 2d_1. \quad$ (74a)

Obviously $\delta_o = \delta$ and with (71) and (54) $\quad \delta_{max} = d/(\rho \overline{N}_{co}). \quad$ (74b)

The above evaluation has assumed the hypothetical condition that the lens can produce a mathematical point image. Actually, the image is of finite size because of diffraction and aberrations. The disc of

Table VI — Camera and Lens Constants

(for $d = 18'$, $p = 1$, $N\delta_o = 400$)

Viewing Ratio	ρ	2.5	4.	8.	
Mean distance	d	18	18	18	feet
Depth and height of field	$d_2 - d_1 = V_0$	7.2	4.5	2.25	feet
	$d_1 + d_2$	37.4	36.7	36.3	feet
Lens stop diameter (Equation 73)	δ_{max}	1.12	1.1	1.09	inches
Optical transfer factor (Equation 61)	$g_0 = \Delta\psi_i/\Delta B_0$ ($\tau = 0.9$)	1/2400	1/6350	1/25800	Lumens/foot-lambert

Viewing Ratio		2.5			4			8		
D	V	M	F	f:	M	F	f:	M	F	f:
Inches			Inches			Inches			Inches	
1.5	0.9	1/96	2.22	2.	1/60	3.54	3.2	1/30	6.95	6.4
3.	1.8	1/48	4.4	3.9	1/30	6.95	6.3	1/15	13.5	12.4
6.	3.6	1/24	8.7	7.8	1/15	13.5	12.3	1/7.5	25.4	23.3

confusion is, therefore, somewhat larger in size and is decreased in first approximation by the cascaded value of the in-focus resolving aperture of the lens. A viewing ratio $\rho = 4$ for a 1.5-inch plate diagonal (Table VI, 2nd column) is obtained, for example, with an f:3.2, 3.54-inch lens, which has an (in focus) aperture response of 50 per cent at $N \simeq 20$ lines per millimeter (See Figure 63(b)) corresponding to $N = 500$ for this image size. The disc of confusion (δ_o) selected according to Equation (73) gives 71 per cent response at $N\delta_o = \overline{N}_{co} =$ 400 and has a 50 per cent response at $N = 1.75\ N\delta = 700$ (See Part II, Appendix Table II, page 285). The line number $N_{0.5}$ for which $r\Delta\bar{\psi} = 0.5$ at the field limits is obtained with Equation (30) Part II

$$N_{0.5} = 1/\sqrt{1/700^2 + 1/500^2} = 406.$$

It may be concluded, therefore, that the aperture response of the optical system cannot be neglected in its effects on the television system quality and especially not when the image size is small.

2. *Photoelectric Transducers (Camera Tubes)*

The construction and operating principles of television camera tubes have been adequately described in the literature.* The following treatment will be limited to an evaluation of their characteristics as transducers of light into electrical signals.

a. *Storage Capacitance and Transfer Characteristics*

The quality of signals developed by photoelectric transducers such as the iconoscope, orthicon, and image orthicon, depends, in principle, on the total energy Q which can be stored and transduced into electrical signals. The storage capacity of transducers is determined by the electrical capacitance C and the maximum charge potential E_c of the storage surface as expressed by Equations (51) and (52). The mechanism effecting the transfer of charges into television signal currents, however, imposes limitations on the maximum energy value and is, therefore, a controlling factor in the design of practical camera tubes. To retain image definition it is essential that each incremental area Δa of the storage surface (target) in the transducer be associated with an individual storage capacitance ΔC. In iconoscopes and orthicons, this association is accomplished by breakup of the photosensitive image surface into a mosaic of minute silver islands. The "space" capacitance of these elements is increased by addition of a parallel grounded plate, termed "signal plate". Dependent on its position, the signal plate is a solid conductor (iconoscope), a conductor transparent to light (orthicon), or a conductor transparent to electrons (collector screen in image orthicons).

* See References (7), (8), (9), (11) in Part I.

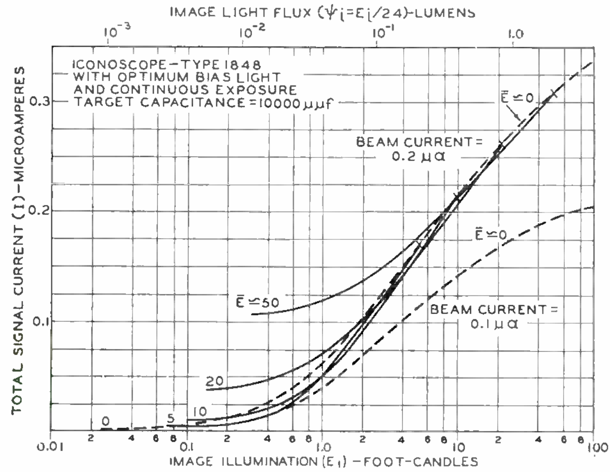


Fig. 66—Dynamic transfer characteristics of an iconoscope.

Orthicons operate with saturated photoemission. Element potential and charge increase in proportion to the image light flux (ψ_i), but are limited by insufficient beam current or secondary emission to certain maximum values. The transfer characteristic $I = f(\psi_i)$ is linear but breaks off when E_c is approximately 6 volts. Higher potentials cannot be neutralized by low-velocity beams because of excessive generation of secondary electrons.

In iconoscopes, the photoemission from elements takes place largely without strong collecting fields and E_c is limited by emission velocity distribution and partial discharge by secondary and photoelectrons to a maximum of approximately 2 volts. The function $E_c = f(\psi_i)$ and the transfer characteristics follow a curve of diminishing returns as shown by the characteristics in Figure 66. The gradual signal compression is a desirable characteristic for well-lighted subjects of good contrast and requires little electrical correction in the amplifier. High average levels of light flux (See $\bar{E}$ in Figure 66) cause an excess of photoelectrons which discharge small signal charges thereby reducing the signal range.

The image orthicon combines characteristics of the iconoscope and orthicon. Its saturated photoemission forms a photocurrent image focused on a thin glass target having the properties of a storage surface with incremental capacitances which are charged by secondary electrons from bombardment of high-velocity photoelectrons, but which can be discharged from the opposite side by a low-velocity scanning beam as in the orthicon. An electrically transparent signal plate (fine mesh screen) collects the secondaries from the image current. It per-

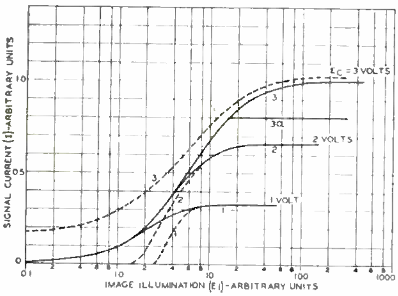


Fig. 67(a) — Dynamic transfer characteristics of image orthicon as function of target potential.

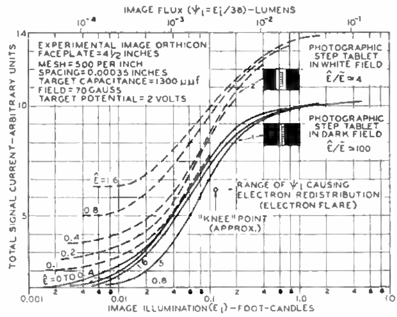


Fig. 67(b) — Dynamic transfer characteristics of an image orthicon at optimum target potential.

mits adjustment of the capacitance (by spacing) and limits the charge potential maximum of the storage surface (target) according to the applied potential. With the charge potential under control, the transfer characteristic does not end abruptly and can be changed from a linear function to one similar to that of an iconoscope.

Although a particular characteristic curve of signal-current output as a function of light input shows a definite saturation value (See published curve of Type 2P23), it is not necessarily describing the transfer characteristic of the tube. This signal curve represents the transfer characteristic only when the range of light values does not exceed the "knee" of the curve. At higher light-flux values, the actual dynamic transfer characteristic is determined by electron redistribution and becomes a function of image content and the peak-to-average ratio of the light flux values. The effects of electron redistribution on the useful contrast range are in some respects similar to "lens flare" as shown by the transfer characteristics in Figure 67(b) which have been measured on an experimental high-capacitance tube to be discussed later.

It is logical to conclude that increases in the target capacitance C of storage tubes will result in a larger range of the transfer characteristics and in higher signal-to-fluctuation ratios $|R|$, provided the stored charges can be transduced into signal currents. A camera tube with low element capacitance is more suitable for low light operation because it develops higher potentials with small photocurrents and low values of light flux which are easier to transduce than low-potential charges in a tube with large capacitance. The low-capacitance tube, however, cannot store large charges because of the voltage limit (E_{cmax}) and, thus, is limited to a lower image quality. High-capacitance elements cannot be discharged as completely as low-capacitance elements containing an equal charge. This condition is particularly true for orthicons and image orthicons because of the velocity distribution in electron beams. An incomplete discharge forces the element potential to build up in successive exposure periods until the scanning beam is capable of removing the number of electrons emitted during one exposure period. The consequence of this action is a delayed slow appearance of under-exposed objects and their delayed disappearance or "smearing" (following "ghosts") in motion. *The maximum useful capacitance* of storage-type tubes is limited by difficulties in generating electron beams (especially of the low-velocity type) having a small velocity distribution and sufficient current density to affect a substantially complete discharge of low-potential elements. This restriction limits the performance of present pickup tubes. Small storage surfaces require a higher beam-current density for a given light flux than larger surfaces. Larger tubes are thus capable of discharging higher capacitances and attaining higher resolution and signal-to-fluctuation ratios.

The transfer characteristics of the iconoscope (Figure 66) and the image orthicon (Figure 67) are plotted on semi-logarithmic coordinates to permit a direct comparison with the *primary transfer characteristic* of photographic film. The total signal current including "flare" or level currents from these camera tubes is a measure of the stored and transduced energy Q_e. The corresponding stored energy in the photographic process is the quantity of silver "specks" caused by photoelectric reactions and "amplified" chemically in the development process into a quantity of silver grains Q_s. This silver quantity is measured by the "density" of the film to which it is directly proportional[30]. Because the density D is plotted to a linear scale in sensitometric curves of film, the characteristics Figures 66 and 67 permit a direct comparison with photographic film.* (Discussed in Part IV of this paper).

The dynamic transfer characteristics of high-capacitance image orthicons depend on a number of parameters[11]. Two of these parameters, beam current, and collector-mesh potential, E_c, are controllable. Others,

[30] C. E. Kenneth Mees, THE THEORY OF THE PHOTOGRAPHIC PROCESS, The Macmillan Company, New York, N. Y., 1942.

* The film, however, is used as an attenuator in a second light transducing process (copying) in which the density effects light modulation by absorption. As an arithmetic increase of density causes a *geometric* increase of light absorption, the linear scale D is also a logarithmic scale $D = 1/\log$ transmission when both transducing processes are considered. The two processes are invariably connected together and the log/log slope of the film curves is defined as the *gamma* (γ) of the film. It is apparent that the slope of the electrical transfer characteristics, i.e., their *transfer factor g* is not a "gamma" and cannot be manipulated mathematically like the gamma of film, because the relation between Q_e and light output is dependent on the entirely different transfer characteristics of separate transducers (amplifiers and kinescopes).

such as the secondary-emission ratios of target (glass) and collector-mesh materials cannot be adjusted in a finished tube. The dynamic characteristics of the image orthicon depend on these parameters as illustrated qualitatively in Figure 67(a). The solid curves 1, 2, and 3, show the effect of varying the mesh potential E_c which controls the maximum charge potential of the storage capacitance (glass target). The beam current is increased in proportion to E_c. When the current is insufficient, the high-light region is compressed as indicated by curve 3a. These characteristics are obtained with "low key" lighting as represented by a small photographic step tablet in a dark field in which redistribution effects are small. When the field surrounding the step tablet is light, the transfer characteristics remain unchanged as long as the peak illumination $\hat{E}_i$ remains below the "shoulder" of the curves. When $\hat{E}_i$ is increased and approaches the flat portion of the shoulder, electron redistribution increases and modifies the low-light range as indicated by the dotted "high key" characteristics for the following reasons. Assume that the mean velocity of secondary electrons emitted by interception of high-velocity photoelectrons on the collector-mesh wires is two volts. A "white" background in the image causes a large number of such secondaries which can "land" on the glass storage surface when its potential E_c with respect to the mesh is equal to or smaller than two volts. The secondaries from the mesh discharge especially, therefore, the small charges in the low-light region for $E_c < 2$ volts and shorten the transferred light range by "black compression". At higher mesh potentials, $E_c > 2$ volts, this discharge action decreases and, as shown by the dotted curve 3, may be over-compensated by the effect of a "flare" light bias due to the optical system. It is apparent that the secondary emission velocity determines, in general, the optimum mesh bias E_c for a particular tube. (There are other parameters which have been neglected).

The transfer characteristics, Figure 67(b) of an image orthicon were obtained by measuring the signal from a small logarithmic step tablet in both a dark field (curves 1) and a light field (curves 2). The light field covers only a portion of the image area as indicated in Figure 67(b). These two conditions are representative of extremes in actual images. In this particuluar case both "black level" and "white level" are raised considerably with increasing exposure in "high key" images, (curves 2), the added exposure contributing nothing to the actual signal (ΔI) above $\hat{E}_i = 0.4$. Over-exposures cause the center of black areas to "wash out" due to "electron flare", their edges remaining dark which, of course, is undesirable for good quality. For similar reasons, higher exposures than $\hat{E}_i = 0.8$ for the case of dark (low key) images (See curves 1) must be considered as undesirable over-exposures. Over-exposure (curves 2) requires an increase of beam current to accommodate the black-level current. Because of the low modulation factor in present tubes, (See following section), the signal-to-fluctuation ratio $|R|$ at different signal levels in one image varies substantially in proportion to the signal level (ΔI). It can be shown that $|R|$ decreases, therefore, at all levels by the square root of the black-level current.

Image orthicons of the 5655 type have storage capacitances, and transfer characteristics similar to the experimental tube on which the transfer characteristics, Figure 67(b) were taken. The light range transduced by these camera tubes is in the order of 100 to 1 as in normal photographic processes. (Discussed in Part IV).

b. Transfer Efficiency and Transfer Factors

The generation of video signals in storage-type camera tubes occurs in two stages: the photoelectric conversion of image light into electrical charges and the development of signal current from the charges. The corresponding transfer equations have these forms:

$$Q_e = g_1 \psi_i T_{ex} \text{ (photoelectric process)} \tag{75}$$

and

$$I = g_2 Q_e / T_f \text{ (signal development process)} \tag{76}$$

where the symbols have the following meanings:

Q_e = stored electrical charge energy (coulombs)

ψ_i = image light flux (lumens)

T_{ex} = exposure time (seconds)

$g_1 = \tau s S$ = photoelectric transfer factor

τ = optical transmission factor of elements absorbing light

S = photosensitivity in microamperes per lumen

s = storage factor of camera tube

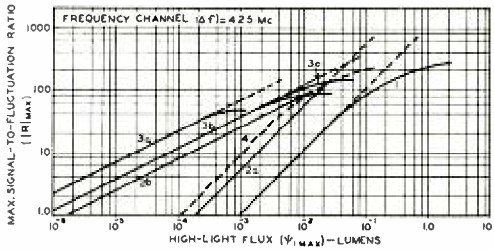
I = primary signal current (microamperes) before amplification

T_f = frame period (seconds)

g_2 = transfer efficiency of the signal developing process (excluding amplification).

For a continuous exposure (no shutter), the exposure time equals the frame time; $T_{ex} = T_f$ and Equations (75) and (76) combine to the simple expression

$$\psi_i = I / g_1 g_2 \text{ (lumen).} \tag{77}$$



CURVE	CAMERA TUBE	TYPE
1	ICONOSCOPE (I_b = 0.2 μA)	AMPLIFIER
2a	ORTHICON (4½″)	AMPLIFIER
2b	ORTHICON (4½″)	MULTIPLIER
3a	IMAGE ORTHICON 2P23	MULTIPLIER
3b	IMAGE ORTHICON 5655	MULTIPLIER
3c	EXPERIMENTAL IMAGE ORTHICON (4½″)	MULTIPLIER
4	IMAGE ICONOSCOPE	AMPLIFIER

Fig. 68—Maximum signal-to-fluctuation ratios of television camera tubes.

By substituting Equations (45) and (48) the image light flux is expressed in terms of the signal quality, ($|R|$ and Δf determining $\overline{N}_{co}$).

$$\psi_{i(R)} = 2.72\,|R|^2\,\Delta f\,(1 + I_f/I) \times 10^{-13}/(g_1\,g_2) \text{ (lumen)}. \qquad (78)$$

The factor $(1 + I_f/I)$ is required, because signal development by a beam current $I_b = I_f$ introduces additional current fluctuations. In camera tubes with electron multiplication (orthicon and image orthicon types), the reciprocal ratio $I/I_f = I/I_b$ is the modulation factor of the beam by the signal current I which seldom exceeds 25 per cent in practical tubes. Because the beam current must remain constant and have a value sufficient to discharge the maximum signal charge, the factor $(1 + I_f/I)$ equals 5 for the high-light flux $\hat{\psi}_i$ only. The flux values $\psi_{i(R)}$ computed by Equation (78) for a standard channel $\Delta f =$ 4.25 megacycles and various values $|R|_{\max}$ have been plotted in Figure 68. These curves do not represent the change of $|R|$ versus light flux changes in a given image because I_f is then a constant and $|R|$ varies as shown by the curve $Q_f/Q_o = 4$ in Figure 43.

Approximate values for the transfer factors $g_1\,g_2$ and contributing factors (See Equation (76)) are listed in Table VII for several camera tube types. Upper limits for $|R|_{\max}$ computed from the capacitance of the storage surfaces by Equation (52) are indicated by the length of the solid line curves in Figure 68.

A certain percentage of charges may be lost for signal generation because of unsaturated photoemission or a partial discharge by secondary electrons. Both effects are small in the orthicon types. ($g_2 = 1$, $s = 1$) In iconoscopes, however, they decrease the effective storage (s) to 60 per cent (approx.) at moderate light levels and to less than 20 per cent at high light levels. Because of incomplete collection of the

Table VII—Transfer Factors ($g_1 g_2$) of Camera Tubes

Camera Tube Type	Approx. C-$\mu\mu f$	S μa/lumen	τ	sg_2	Avg. Value $g_1 g_2$	Spectral Response	Signal Amplification
1 Iconoscope	10,000	7-10	1.	0.12*	1.	Normal	Amplifier
2 Orthicon	750	6-8	0.5-0.9	1.	5.	Normal	Amplifier or Multiplier
3a Image Orthicon 2P23	250	20-30	1.	1.5**	35.	High red and infrared	Multiplier
b 5655	750	5-10	1.	1.5**	10.	Normal	Multiplier
4 Image Iconoscope	5000	7-10	1.	0.2×5**	8.	Normal	Amplifier

* Decreases above a certain light flux.
** Includes gain in image multiplication and decreases above a certain light flux.

signal current, the transfer efficiency of iconoscopes is further reduced as indicated in Table VII.

Incomplete storage and partial cancellation of charges by redistributed secondary electrons* occur also in image orthicons, especially at high-light conditions. (See Figure 67) The effective storage of this tube varies as a function of light and collector mesh potential from near 100 per cent at low light levels and normal collector potentials to 20 per cent and less under high-light conditions and low collector potentials as evident from the transfer characteristics shown in Figure 67. Representative photosensitivities (S) and utilization factors of practical transducer types suitable for direct pickup in television cameras are given in Table VII.

Signal amplification by amplifiers—In some camera tube types the video signal is developed across a resistance in the signal plate lead without the use of an electron multiplier. Because of the large number of electrons required as a "carrier" current for small signals in amplifier tubes, their modulation factor is extremely low and the signal-to-fluctuation ratio is dominated by the constant amplifier fluctuation current I_f and requires, therefore, large signal energies for good quality. The current I_f is a function of tube and circuit conductances and may be computed from the equivalent noise resistance of the amplifier tube.[18e] For a minimum fluctuation current the input circuit capacitance C_1 of the amplifier (camera tube plus amplifier grid circuit capacitance) should be made the controlling factor of the circuit impedance.

* There are also other causes.

The amplifier is compensated to give constant voltage output with constant signal-current input and has, therefore, a gain proportional to frequency over most of its range Δf (See Part I). Assuming $C_1 = 25$ micromicrofarads as the dominating input impedance, the amplifier noise current in the type 6AC7 amplifier tube is expressed by

$$I_f = 3.7 (\Delta f)^{1.5} \times 10^{-19} \text{ amperes.} \tag{79}$$

The character of the resultant "peaked" "noise" and its effect on visibility have been discussed in Part I and must be considered when comparing signal-to-fluctuation ratios of multipliers and amplifiers. For visual equivalence of the "peaked" noise from this amplifier with flat channel noise from multipliers, $|R|$ is to be multiplied by the ratio K of the fluctuation filter factors for the two types of noise

$$K = m_{ek} \text{ (peaked)} / m_{ek} \text{ (flat)} \tag{80}$$

shown in Part I, Figure 11, as a function of Δf. The signal current in terms of equivalent flat-channel noise exceeds the fluctuation current by the ratio $|R|$ and is, therefore, given by:

$$I' = 3.7 \, |R| \, K (\Delta f)^{1.5} \times 10^{-19} \text{ amp.} \tag{81}$$

The required light flux is obtained with Equation (77) being reduced slightly by the camera blanking ratios to

$$\psi'(R) = 3.14 \, |R| \, K (\Delta f)^{1.5} \times 10^{-13} / (g_1 g_2). \tag{82}$$

A comparison of camera tube performance and a table of the scene brightness or luminance (B_o) required for their operation will be given in Part IV.

c. The Aperture Response of Television Camera Tubes

Television camera tubes contain a number of elements in cascade which may restrict the resolving power. In storage types, the scanning beam and physical apertures in conductors (collector mesh in image orthicons) are, in general, the principal limiting apertures. Secondary aperture effects such as those caused by unstable focus of the electron image or by graininess of photoelectric and storage surfaces can be made negligible by proper design of tubes and equipment.

The over-all aperture response of the transducer is measured by observing amplitude and wave shape of the electrical signal. The optical signal source is a large (8 x 10-inch) square-wave test pattern picked up by a high-quality lens, with known aperture response. The electrical signal is observed on an oscilloscope with cross-section selector (horizontal or vertical).

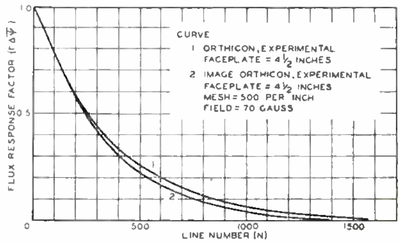


Fig. 69—Flux response characteristics of two camera-tube types.

The measurement in the horizontal direction gives the combined aperture response of transducer and electrical channel. To obtain the aperture response of the transducer directly, the electrical channel must be capable of reproducing the signal waveforms generated by the transducer. This requirement specifies a frequency band exceeding by at least a factor of five the frequency ($f\delta$) or line number ($N\delta$) at which the transducer response becomes sinusoidal, i.e., the point where its amplitude response begins to depart from unity.

The aperture response $r\Delta\overline{\psi}$ of two developmental storage transducers operating into a flat twenty-megacycle channel is shown in Figure 69. Although representing peak performance of present constructions their response decreases to approximately 22 per cent at 500 lines. The aperture response of transducers can be corrected by optical or electrical processes having a "negative" aperture characteristic as discussed in Part IV of this paper.

3. A Television Microphotometer for Measuring the Aperture Response of Lenses and Grain Structures.

Aperture response measurements require, in principle, a scanning process. An arrangement for scanning an optical image and transducing light-flux variations into electrical current variations for observation and measurement is a television system used as a photometer. Because of the relatively high resolving power of lenses, it is expedient to magnify the optical image considerably by a microscope before it is scanned in a television camera tube. With adequate optical magnification of the image the resolving power of the television system is removed as a limitation, and the response limit of the microphotometer depends only on the resolving power of the microscope objective.

A photograph and block diagram of the principal parts of the television microphotometer are shown in Figures 70 and 71.

A test pattern (2) (Figure 70) illuminated by a projection lamp and condenser system (1) is imaged by the lens under test (3) with a reduction equal to or greater than 20 to 1 to remain within the normal correction range of the lens. With this reduction the optical input signal from a good test pattern is sufficiently accurate at the lens cutoff (N_c).

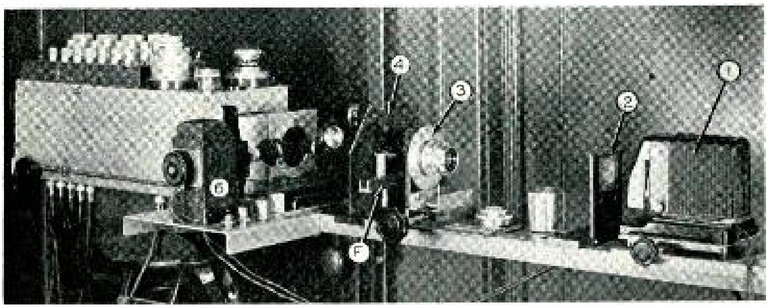


Fig. 70—Photograph of television microphotometer.

A small part of the image is observed with a television camera through a microscope (4) giving high optical magnification ($M = 100$ to 1000 as needed) and having a resolving power much higher than the lens under test (A 4-millimeter Apochromat, $NA = 0.95$, for use without cover glass resolves $N_c = 8000$ television lines per millimeter). The magnified image is scanned in the television pickup transducer (5) and translated into electrical aperture-response signals.

Amplitude response and wave form of these signals are observed and measured on an oscillograph with line selector, the microscope image being visible on the monitor kinescope for inspection and focussing. (Figure 72). To avoid errors by non-linear transfer characteristics of the camera tube or amplifying system, the amplitude of the light flux wave is measured optically at the input to the television camera. A fine light bar produced by a slit mask in a calibrating projector, (6) in Figure 70, is superimposed on a dark "line" of the test pattern image (See Figure 72), appearing as a pulse signal in

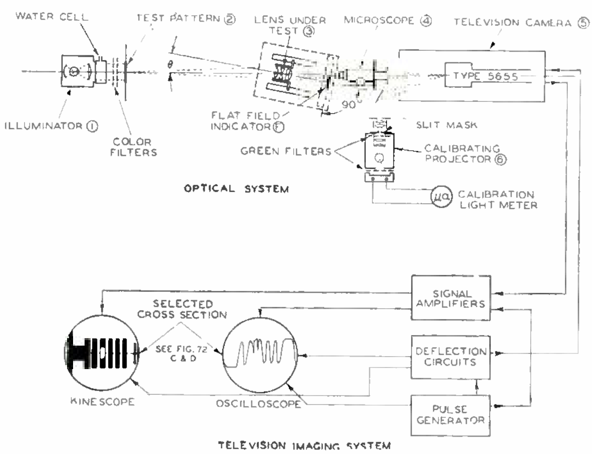


Fig. 71—Block diagrams of television microphotometer system.

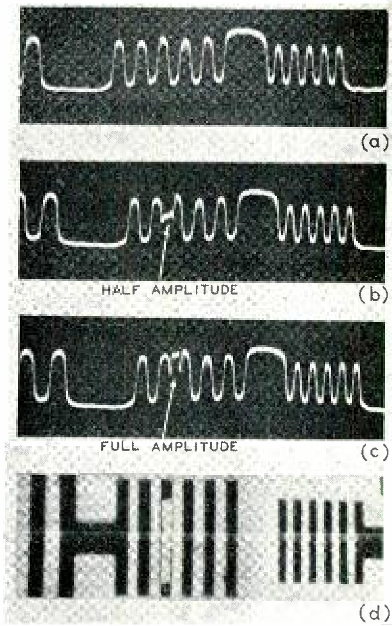


Fig. 72 — Oscillograms and kinescope image showing measurement of flux response with television microphotometer.

the oscilloscope trace. The calibrating light intensity is increased until the negative wave peak is raised to the level of the positive wave peaks. The difference ΔB introduced by the calibrating source is the peak-to-peak brightness difference in the flux wave which is thus measured by reading the relative light output of the calibrating source with photocell and meter (See Figure 71). By the use of sharp green filters in front of the cell and the slit mask, the calibrating light can be controlled by varying the lamp current.

All measurements are made on the same line of sight along the optical axis of the system to avoid errors due to variations in light distribution of the projector or the sensitivity of the camera tube. The line number, hence, is changed by moving the test pattern slide to bring different line sizes under the calibrating light slit which remains in a fixed position.

The base plate carrying the lens and focussing adjustment is equipped with a "flat field" indicator ((F) in Figure 70), a knife-edge slider which can be moved in a guide set perpendicular to the lens axis. The lens image is brought into focus coinciding with the knife edge for axial rays. This lens setting with respect to the slider guide is maintained throughout one measurement series. For angular measurements the entire mount base is rotated and the slider, moved to indicate the image plane at the selected angle, is brought into focus by moving the mount base with respect to the microscope objective. This operation requires several trials to line up calibration slit, slider, and the image section under observation. It solves automatically for the nodal point of the lens. The object-to-image distance is of course increased by $1/\cos\theta$ for the particular angle.

The aperture response of photographic film (test patterns) is measured in a similar manner by placing the film directly in front of the microscope objective. The aperture effect of translucent grain structures has been measured by projecting a sharp optical image on the grain structure which is placed in the focal plane of the microscope.

The number of bars in test-pattern line groups should be sufficient to allow the flux wave to build up to a steady value (See Part II, Figure 14, page 248). It is, therefore, recommended to use *at least* four black bars at line numbers between $N\delta$ and N_c; two bars being sufficient below $N = N\delta$. The line groups should be separated by alternate black and white spaces of sufficient length which serve to give "level" signals. (See Figures 61 and 72).

By definition, the aperture response factors are independent of the absolute image contrast. The contrast reduction by microscope objectives is, therefore, of no consequence. Measurements with test patterns of low contrast ($C = 1.5$) have shown in some cases a slight increase of $r\Delta\overline{\psi}$ in the high-resolution section of the response curve in comparison with high-contrast patterns ($C \simeq 1000$). Positive and negative patterns gave identical results. The reference calibration $r\Delta\overline{\psi}$ at $N = o$ and zero angle is obtained by measuring the white-to-black signals between the shadow cast by the knife edge slider in the image plane and the white background of the test pattern image. Lens flare has normally little effect on the measurement and should be determined separately. It should be mentioned that the color temperature of the light source should be adjusted by filters to result in a reasonably uniform response from the camera tube for red, green, and blue light of normal spectral energy to avoid misinterpretation of chromatic aberration effects in "white"-light-response measurements.

SOME APPLICATIONS OF FREQUENCY-MODULATED RADAR*†

By

IRVING WOLFF AND D. G. C. LUCK

Research Department. RCA Laboratories Division,
Princeton, N. J.

NOTE: The previous papers in this series, entitled PRINCIPLES OF FREQUENCY-MODULATED RADAR, and FREQUENCY-MODULATED RADAR TECHNIQUES, were published in *RCA REVIEW* in March and June 1948. The summaries of these papers are reprinted herewith for reference purposes.

PRINCIPLES OF FREQUENCY-MODULATED RADAR

(Reprinted from RCA REVIEW March, 1948)

Summary—*The principles of operation of FM radar systems are developed for the determination of range and relative speed of reflecting objects. Quantitative expressions are derived relating the radar output frequency to the range and speed being measured and the transmitted-signal characteristics. It is shown that range and speed may be independently determined. In single-target systems, measurement of output frequency by means of cycle counters has been frequently used. This leads to a type of error which may appear in laboratory calibration of the equipment but which is usually averaged out in field operation. Search operation against more than one target requires either a multiplicity of selective range gates or scanning in range with a single gate. The former system leads to rather complicated apparatus and the latter to unduly slow operation. For the same average transmitted power, pulse and FM radar are capable of similar maximum ranges; however, the simple FM system which scans in range with a single gate uses time inefficiently and either requires longer than the pulse system to obtain data or operates at reduced maximum range. In general, operating with the same average power both FM and pulse systems have their maximum ranges determined by the time allowed to obtain the complete range data and their range resolution by the usable radio-frequency band. FM radar avoids need of high peak power by using a narrow noise band, and facilitates use of a wide radio-frequency band for high resolution, but requires great care in protecting reception against its own transmission. Techniques for automatic control on the basis of range and speed, described in later papers, are very simple, but equipment yet devised for search is either complex or slow in action.*

---

* Decimal Classification: R537 × R148.2.

† This paper covers work initiated in 1938 by RCA Laboratories Division and carried on since 1941 with the support of the U. S. Navy, under Contracts NOs-87822, NXsa-25337, and NXsa-35042.

FREQUENCY-MODULATED RADAR TECHNIQUES

(*Reprinted from RCA REVIEW, June, 1948*)

Summary—The FM radar development program has given rise to a number of interesting techniques. Two superheterodyne systems have been devised which remove the transmitted modulation from the received intermediate-frequency signal. An electro-mechanical device using a vibrating capacitor and an electronic device using a beam of electrons have proved useful for frequency modulating radar oscillators. Several adaptations of the well-known cycle-rate counter for utilization of beat frequencies as data have by their simplicity made FM radar very useful, where automatic indication of or control by range and speed of a single target is needed.

* * *

There follows the concluding paper in this series—SOME APPLICATICNS OF FREQUENCY-MODULATED RADAR.

The Manager, RCA Review

SOME APPLICATIONS OF FREQUENCY-MODULATED RADAR

Summary—Application of principles and techniques already described has yielded light, compact FM radar equipment for airborne use. Characteristics of motion of aircraft and bombs render FM radar data particularly suitable for providing automatic bomb release from low altitutes. Applications of FM radar have been made to aircraft altimetry, automatic flight control, automatic missile release, control of aircraft landing approach, and radar search.

I. Introduction

THE principles on which the frequency-modulated radar art is based, and some useful techniques that have been developed in the process of applying these principles, have been described in the earlier papers of this series.[1,2] It is the purpose of the present paper to describe the way in which these principles and techniques have been integrated to produce complete radar systems that meet specific operating requirements. Since each system is devised for the purpose of fulfilling specific requirements, it can be described only in relation to those requirements. System design is a matter of detailed synthesis rather than of following broadly valid general rules.

II. Fully Developed Systems

A. *Altimeters*. The major operational use of FM radar systems has been as aircraft altimeters. A series of designs ending with that

[1] I. Wolff and D. G. C. Luck, "Principles of Frequency-Modulated Radar", *RCA Review*, Vol. IX, No. 1, pp. 50-75, March, 1948.

[2] I. Wolff and D. G. C. Luck, "Frequency-Modulated Radar Techniques", *RCA Review*, Vol. IX, No. 2, pp. 352-362, June, 1948.

designated *AN/APN-1 found very wide use by all Allied forces during the war, while a somewhat different type was developed and used by the Germans. Use of radar altimeters by civil transport aircraft may become mandatory in the United States. Originally developed as an aid to instrument landings on water and for control of low-flying pilotless aircraft, the *AN/APN-1 has also proved very useful in maintaining safe clearance for release of parachute troops.

A single main unit, shown in front and top views in Figure 1, contains at its left end an acorn-triode push-pull transmitter with vibrating-capacitor modulator, with an acorn-diode balanced detector in a separate shielded compartment next to the transmitter compartment, a beat-note amplifier at the center rear, modulation source and data-utilizing circuits at center front, and a dynamotor power supply at the right end. The gain of the beat-frequency amplifier rises at 6 decibels per octave over the useful frequency range, to compensate variation of received-signal level with altitude, and falls off rapidly at frequencies above the operating range, to reduce noise. Except for the indicators and the antennas, which are simple dipoles or slots in the skin of the aircraft, the unit shown is the whole of the *AN/APN-1 altimeter system.

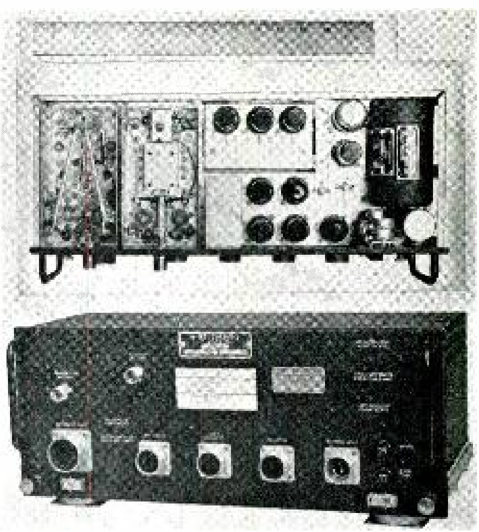


Fig. 1—Main unit of radar altimeter *AN/APN-1.

The radar portion of the altimeter is organized according to the simple block diagram of Figure 1 of Reference 2, but feeds two separate data-utilizing systems. One of these is of the type shown in Figure 7 of Reference 2, with a milliammeter, calibrated directly in feet of altitude, connected in the follower cathode circuit. The counter-diode return tap on the cathode resistor in this circuit is set to provide a controlled degree of counter non-linearity. This opens out the lower part of the indicator scale and permits altitude to be read very closely at the instant of landing. The second or "altitude limit" data-utilizing system employs a null counter of the type shown in Figure 8 of

Reference 2, with the balancing voltage divider located in a separate small unit and normally of a step type preset by hand to a chosen altitude. Relays actuated by null-counter unbalance then permit visual warning of departures from the preset altitude or, by means of a small auxiliary unit feeding data to a standard automatic pilot, permit an aircraft to be held automatically in level flight at the chosen radar altitude. Another small auxiliary unit, which provides servo balancing of the null-counter output, may alternatively be connected to the altitude-limit circuits to feed altitude data automatically to fire-control devices.

The separate indicator and manual altitude-limit-switch units, as well as warning lights to show departure from preset altitude, appear in Figure 2; the desired non-linearity of indicator scale is clearly evident. The main power switch and the switch selecting either a 0-400 or a 400-4000 foot range of operation are the only necessary operating controls and are mounted directly on the indicator unit, which in turn is normally mounted on the instrument panel of the

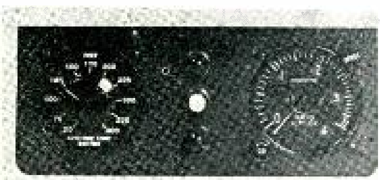


Fig. 2—Auxiliary units of altimeter.

aircraft; the manual altitude-limit switch for level flight, if used, is a third and last operating control. Rather heavy indicator damping is used to give an acceptable compromise between sluggishness and jumpiness of indication.

Transmitted power is approximately 100 milliwatts, at an average radio frequency of 440 megacycles, and the primary power consumption is 3 amperes of direct current at 27 volts. Modulation frequency is 120 cycles and modulation sweep 39 megacycles, giving a radar range sensitivity k_R of 19 cycles per foot and a radar beat frequency of 8000 cycles at maximum altitude. Shifting to the high range involves primarily the reduction of modulation sweep by a factor of ten. The altimeter measures the length of its own antenna cables in addition to the true altitude of the antennas. This avoids excessively low beat frequencies, even at landing, but requires the application of suitable bias to the indicator-counter load and to the limit-counter balancing potentiometer. These biases must be suitably set in each particular type of aircraft to give correct zero readings.

Absolute accuracy at landing is very high, but errors of a few per cent may be present at higher altitudes. Over very poorly conducting land, indicated altitude will occasionally be low on the high range because of inadequate signal reflection from the ground. Over water, however, a full-scale indication of 4000 feet is sustained up to altitudes exceeding 8000 feet. Automatic control of level flight at a preset altitude is smooth and accurate, the aircraft flying to follow terrain contour at constant clearance. The FM altimeter appears to be one radar system that was developed during the war to a point which will permit it to remain widely useful in peace.

B. *Low-Altitude Bombing.* Neglecting air resistance, a bomb released from an aircraft traveling in a straight line at constant speed will remain directly beneath that aircraft throughout its fall; the aircraft will then be directly over the point of impact of the bomb when the latter strikes. Release of the bomb must occur prior to the instant of passage of the aircraft over the desired target by just the time interval taken by the bomb to fall, if the bomb is to hit the target.

Time T_f taken by a bomb to fall from altitude A under the normal acceleration of gravity g is $\sqrt{2A/g}$, neglecting air resistance. If there is a radar target at flight altitude and directly over a bombing target on the ground, the time T required for an aircraft at range R to reach the radar target which it is approaching at constant speed S is R/S. A bomb released at the instant when $\sqrt{\tfrac{1}{2}g/A}\,R - S = O$ (1) will therefore strike the bombing target.

FM radar with switched counters gives, as described in References 1 and 2, an output voltage $e = k_R h_R R - k_S h_S S + e_o$. (2)

This is evidently the type of data needed for automatic bomb release in the above simple case. It is only necessary, for this use, to adjust the radar and counter sensitivities so that $k_S h_S / k_R h_R = \sqrt{2A/g}$, (3)

and to provide a release relay that will operate when total counter-output voltage e reaches bias voltage e_o, as output decreases with decreasing range during approach.

Actually, there is no radar target conveniently set up over each bombing target. An aircraft at point P of Figure 3 measures by

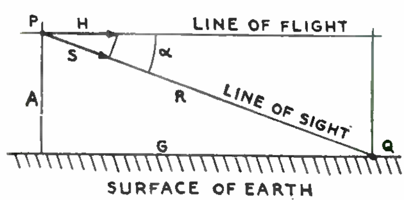


Fig. 3—Geometry of level-fight bombing.

radar its slant range R and slant speed S relative directly to the bomb target on the ground at Q, rather than the ground range G and horizontal speed H. The time required to reach a point above Q from P is G/H and not R/S. The exact slant range R at which a bomb should be released to strike Q does not vary linearly with slant speed S. For sufficiently low altitudes and a practically useful range of speeds, however, the exact relation between R and S for release turns out to be extremely well approximated by the modified linear relation

$$R/T' - S + S_o = O. \tag{4}$$

Both the correction factor T'/T_f, which is close to unity, and the added term S_o, which is fairly small, depend only on altitude.

Comparison of Equations (2) and (4) shows that FM radar is capable of giving highly accurate automatic control of release in actual low-altitude bombing as well as in the idealized case first described. Release must be arranged to occur at a total counter-output voltage e_1, preset to exceed bias e_o by a small amount dependent on altitude and corresponding to the required correction intercept S_o. A correction must also be made to e_1 or e_o for length of radio-frequency cables, which is measured by the radar in addition to target range; this correction also depends only on altitude, for any given cable length. If it is desired that the first bomb of a train release shall fall short of the target, correction for such "range lead" is included with that for cable length. In addition to the geometric factor T'/T_f, a further correction is required to time of fall, to allow for the time lag due to counter-output smoothing, relay action, and bomb-release mechanism; this lag is fixed for a given system, and alters slightly the relation to be set up between altitude and range/speed sensitivity ratio. In low-altitude bombing at moderate speeds, neglect of the effects of air resistance on motion of the bomb is quite permissible.

Figure 4 (page 538) shows a complete FM radar system for automatic low-altitude bombing. This system, designated AN/APG-4 and commonly called the "Sniffer", requires only that the bombing craft be flown level at a chosen altitude and headed so as to pass directly over the target, but use of such auxiliaries as the *AN/APN-1 altimeter and a search radar is necessary, when the target cannot be seen, to provide data for maintaining such flight. Given a correct flight path, the Sniffer itself releases the bomb at the correct instant to insure a hit, with no further attention than the throwing of an arming switch at the beginning of the bombing run.

Like the altimeter, the AN/APG-4 system is self-contained, except for antennas and a small control unit, in a single main unit, which looks very much like the main altimeter unit of Figure 1 both internally and externally. Antennas are, as shown, Yagi end-fire arrays, mounted pointing forward in well-separated positions on the aircraft wings. The Sniffer system is organized according to the block diagram of Figure 5 (page 539), and a complete circuit diagram showing functional groupings of components is given in Figure 6 (pp. 538-539). The circuits of the radar portions of the Sniffer, which occupy the left half of Figure 6, differ from those of corresponding portions of the altimeter only in details of the modulation source and audio amplifier.

Several special features of the circuit of Figure 6 may be noted. A mechanical switch actuated by a cam on the shaft of the plate-supply dynamotor develops a square wave from the regulated plate-voltage supply. This square wave is shaped by integrating and differentiating circuits and used to drive the vibrating-capacitor frequency modulator, to produce a symmetrical-sawtooth variation of transmitted frequency. Automatic gain control depending on beat-signal level is applied to an amplifier stage with selective feedback, cutting down gain at high beat frequencies to reduce noise and distant-target signals when the desired signal is strong, but not markedly affecting low-frequency response to the nearby desired target. The amplifier gain increases rapidly with increasing frequency in the working range, to reduce dependence of signal strength on target range, and falls off rapidly above the normal operating region. A two-stage limiter with interstage coupling emphasizing low frequencies is used to minimize disturbance by high-frequency random noise passed by the high-peaked amplifier. The switched counter is like that of Figure 9 of Reference 2, but switching signal from the square-wave modulation source is applied through resistance-capacitance couplings and a phase-inverting tube is required. False release of bombs, in the absence of desired signal, by such low-frequency noise as may result from detector unbalance or short-range reflections from the sea surface, is prevented by a circuit using automatic-gain-control voltage to disable the negative counter completely, and so to maintain counter output permanently above the release level, unless adequate signal is present.

The small control unit of Figure 7, with circuit shown at the lower right of Figure 6, is an altitude-compensation switch which sets up the Sniffer to produce release under the conditions required by Equation (4). Radar speed sensitivity k_S and counter range and speed sensitivities h_R and h_S are fixed, while a step-type voltage divider in

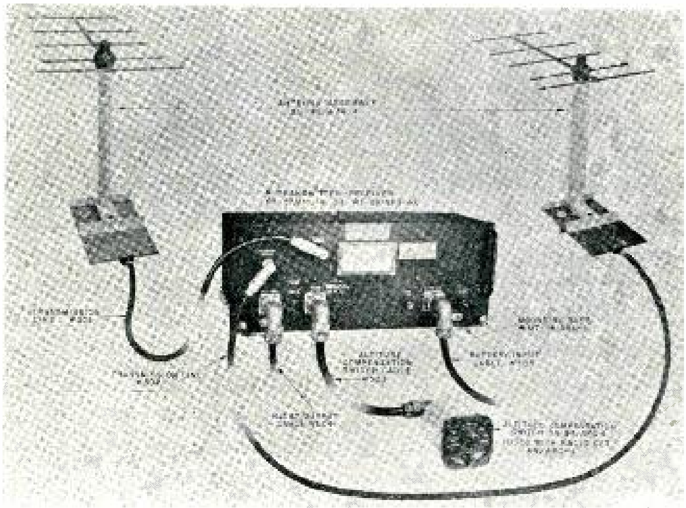
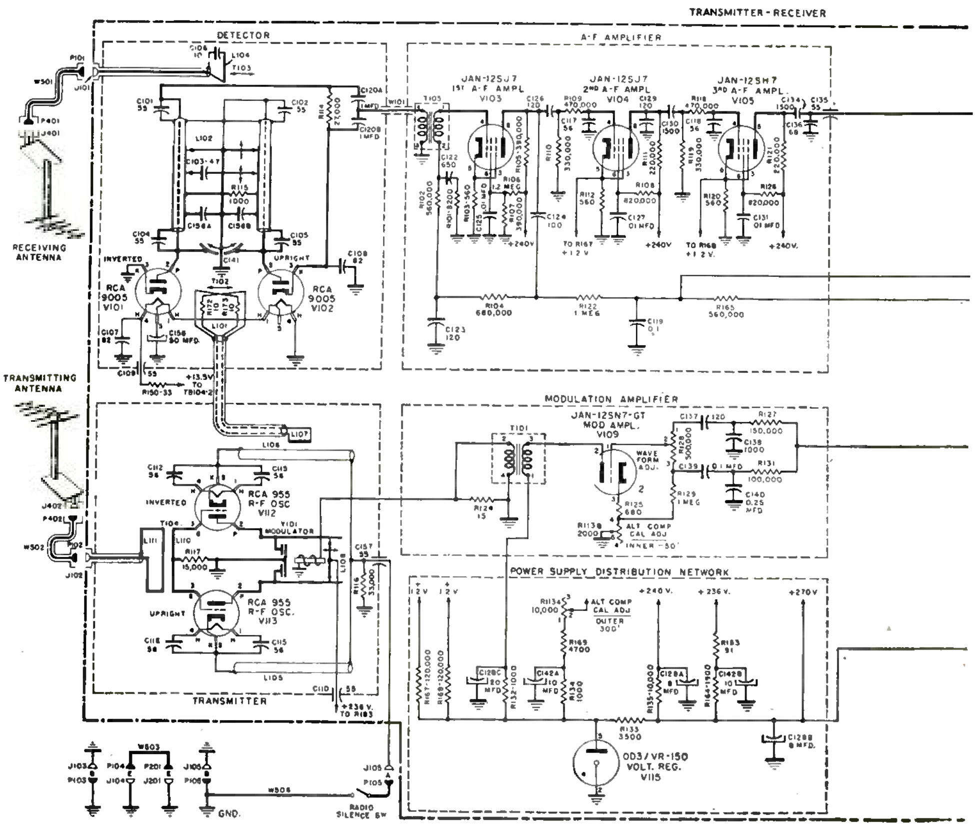


Fig. 4—Complete FM radar system AN/APG-4 for automatic bombing.

Fig. 6—Circuit diagram of automatic bombing radar.



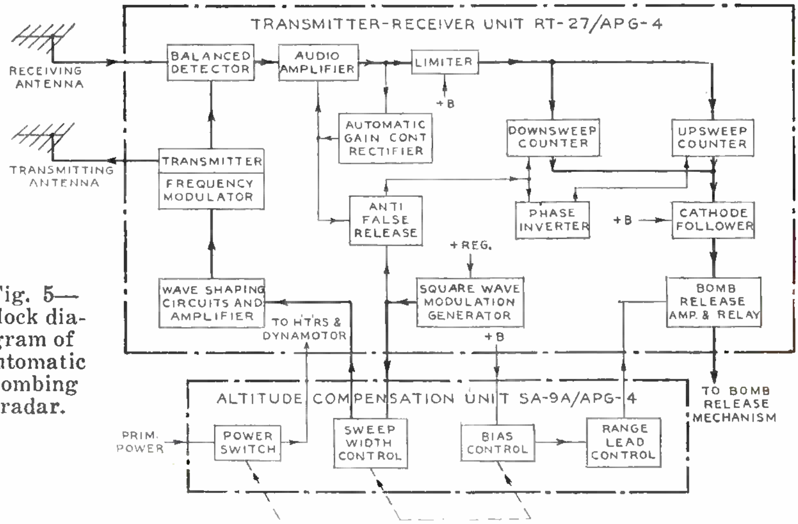


Fig. 5—Block diagram of automatic bombing radar.

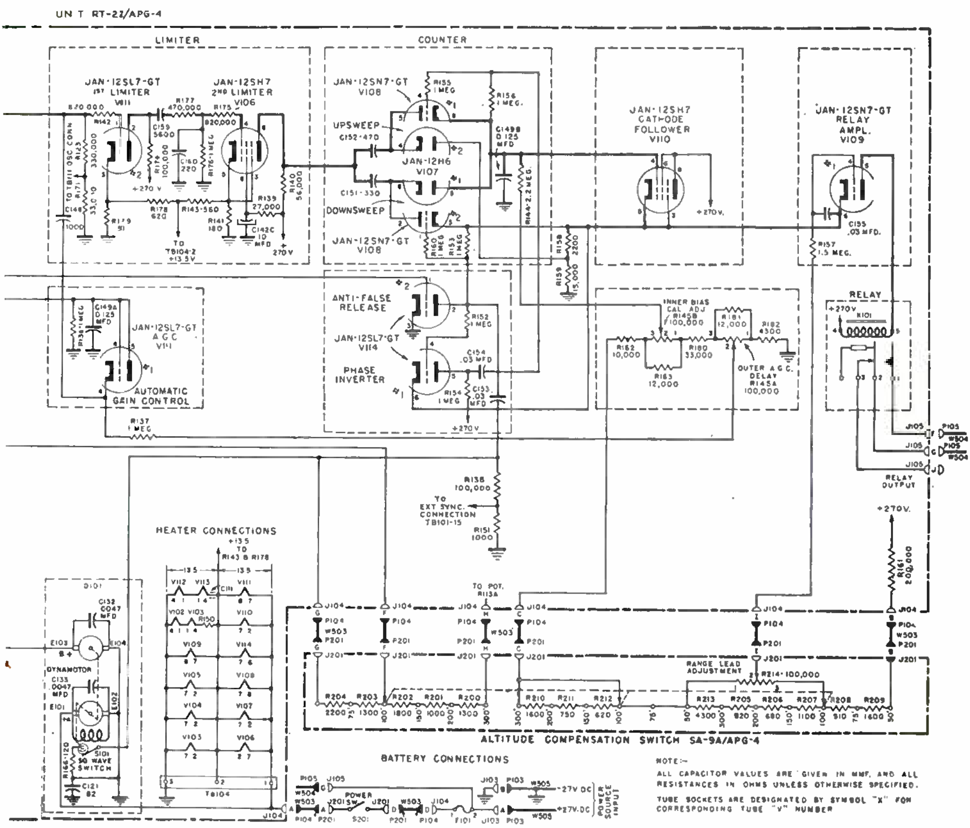


Fig. 7—Accessory to compensate bomb release for altitude.

the compensation unit sets amplitude of the square-wave modulation voltage, and thereby radar range sensitivity k_R, in accordance with the required value of T'. A step rheostat controls the difference $e_1 - e_o$ between release-relay threshold and counter-bias voltages, in accordance with the required value of S_o, while another step divider provides an additional altitude-compensated bias component to produce a "range lead" of the bomb impact relative to the target. These three step controls are ganged, together with the main power switch for the system, for manual setting by a single knob visible in Figure 7 and working against a scale of fixed altitudes from which bombing may be done. Magnitude of the range lead used may be set at will, between zero and 100 feet, by a separate manual potentiometer. This small compensation unit, together with a switch to arm the bomb-release circuits, is all that the pilot or other bombardier need see of the Sniffer system.

The requirement of level flight at one of a few predetermined altitudes limits the pilot's freedom. An alternative unit to that of Figure 7, of similar size and appearance, was therefore developed to provide automatic compensation for release from level flight at any altitude from 40 to 400 feet. This unit, connected to the limit circuits of the altimeter as well as to the Sniffer, includes a servo motor to set a shaft position in accordance with actual altitude. The servo shaft automatically adjusts smooth modulation-sweep and release-bias controls, corresponding to the step-type controls of the manual unit, and provision is again made for setting a range lead in manually. With this compensation unit, the pilot need only fly level at any preferred altitude within the operating limits to obtain automatic bomb release at the correct point for a hit.

AN/APG-4 equipment operates at a radio frequency of 410 megacycles, with a power of 200 milliwatts and 110-cycle frequency modulation, using total modulation swings up to 5 megacycles. The total installed weight, including antennas and shock mount but less cables, is 37 pounds; the power required is 2½ amperes of direct current at 27 volts. In extensive flight tests with practice bombs, automatic compensation of release for speed and altitude was found to occur as expected, and many strikingly accurate hits were observed; no adequate run of data under fixed conditions was taken to permit really definite statistical evaluation, however. In limited operational use of the Sniffer against submarines and shipping, a few successful results are believed to have been obtained.

Similar equipment operating at 1500 megacycles and bearing the designation AN/APG-17 was also developed and a production design made, but was never put into full production because of changing military requirements. This system comprises two main units, a trans-

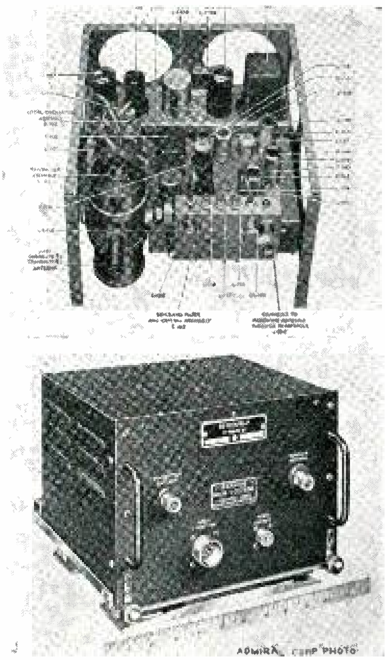


Fig. 8—Radio-frequency unit of 1500-Megacycle FM radar AN/APG-17.

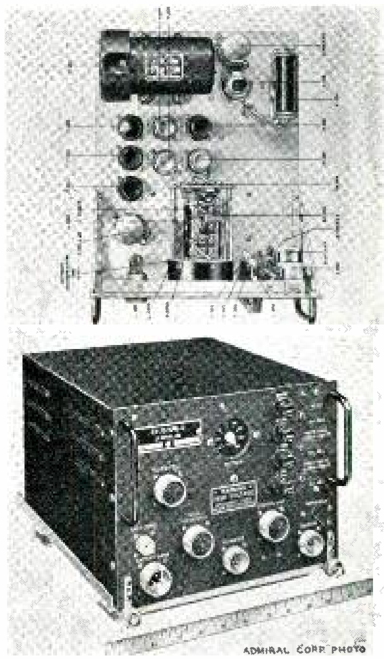


Fig. 9—Computer-power unit for bombing with 1500-megacycle radar AN/APG-17.

mitter-receiver unit and a computer-power unit, which are shown in Figures 8 and 9 respectively. Dipole antenna arrays in parabolic-cylinder reflectors, for mounting in the leading edge of the aircraft wing, are used. Power generated is 2 watts, but increased transmission-line losses and reduced antenna area at the higher frequency reduce the effective power to the same order as in the AN/APG-4.

A side-band superheterodyne receiver, organized according to Figure 2 of Reference 2, is used in the AN/APG-17 and leads to the circuit shown in Figure 10 for the transmitter-receiver unit. In the interior view of this unit, Figure 8, the cylindrical 1500-megacycle transmitter assembly appears at the left front, with the side-band filter and radio-frequency mixer unit at the right front and the intermediate-frequency amplifier assembly immediately behind the latter; wave-shaping and modulator-driving circuits appear at the rear of the unit. The heart of the receiver is shown in Figure 11. This single, small rod-and-cavity unit contains, from right to left of the interior view, a crystal modulator tuned to 1500 megacycles, a three-stage side-band

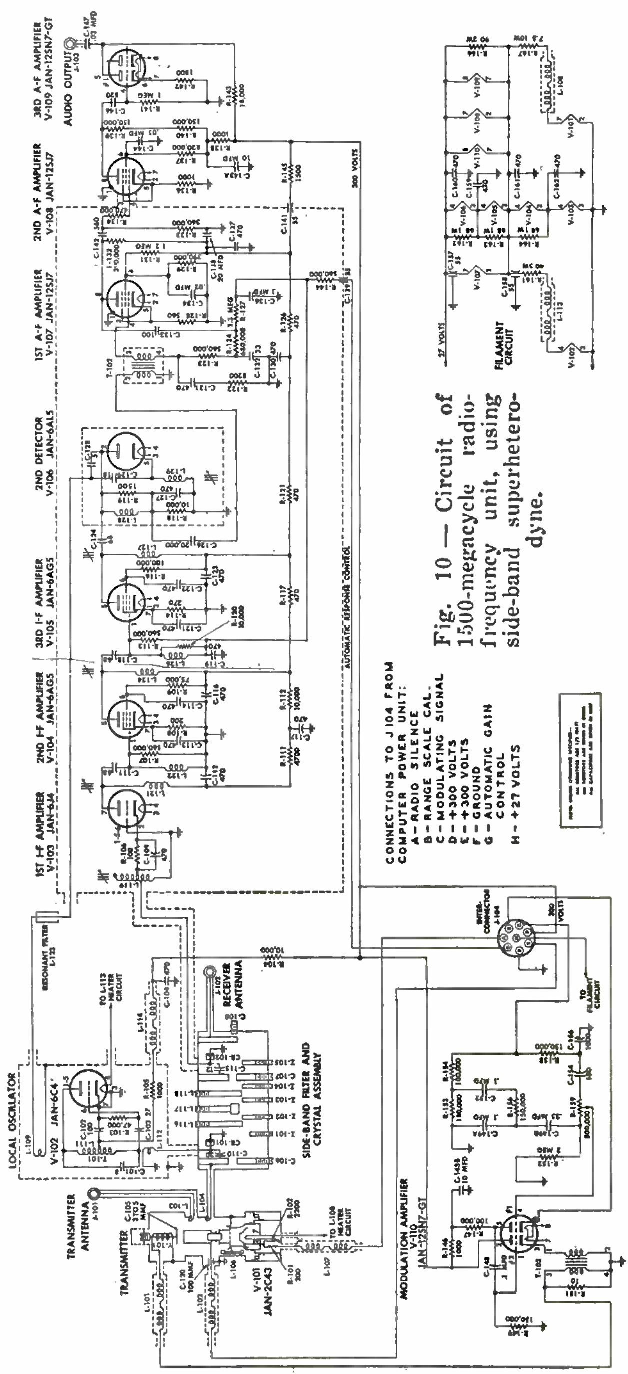


Fig. 10 — Circuit of 1500-megacycle radio-frequency unit, using side-band superheterodyne.

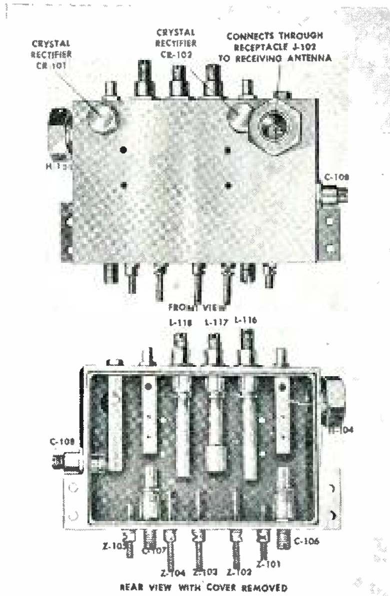


Fig. 11—Side-band filter unit of AN/APG-17.

filter passing 1380 ± 10 megacycles and rejecting 1500 megacycles, a tuned crystal first detector, and a 1500-megacycle receiver-input circuit. The 120-megacycle local oscillator amplitude modulates the frequency-modulated transmitter output in the mixer crystal, and the 120-megacycle received intermediate-frequency signal, with transmitter frequency modulation removed but radar frequency shift preserved, is taken from the first-detector crystal.

The circuits of the computer-power unit are substantially the same as those of the corresponding portions of the AN/APG-4, shown in the right half of Figure 6, except for inclusion of a warning accessory to be described later. A servo altitude-compensation element is built into the computer-power unit directly, and may be seen at the center front of the interior view in Figure 9. Performance of AN/APG-17 in flight tests was very similar to that of AN/APG-4, except for the advantage obtained by elimination of the need for critical detector balancing.

C. *Adaptations of Bombing Equipment.* Dependence of the Sniffer on data from some other source, for lateral guidance in the bombing run, is undesirable. Development of a single equipment capable of steering the bombing aircraft as well as of automatically releasing bombs was therefore carried through the stage of successful flight operation. This equipment, designated AN/APG-6 and often called the "Super Sniffer", was not put into production, however. It is built in a single main unit, like the Sniffer though somewhat larger, and operates at the same radio frequency as the Sniffer, but with a power of 2 watts. The radar portion proper differs from the AN/APG-4 in using a side-band superheterodyne receiver, while the bombing circuits are substantially similar except for addition of an anti-fading accessory.

Azimuth information is obtained in the AN/APG-6 by the well-known expedient of transmitting alternately through two differently oriented directional antennas, while receiving continuously through a single antenna. Selection of antennas is accomplished by a motor-driven mechanical switch synchronous with the modulation-generating switch but running at half speed, so that one full cycle of frequency modulation is transmitted over each antenna in turn. The effect is to produce a square-wave amplitude modulation of the radar-beat signals whenever the target is off the center line of the pair of transmitting antennas, the phase of this modulation indicating on which side of center the target lies.

Utilization of the azimuth data, appearing as amplitude modulation of the radar-beat signal which in this respect acts only as a low-frequency carrier, is obtained by first applying the beat signal to a grid-leak detector. The square-wave detector output is then fed to a transfer-capacitor arrangement, similar to the cycle-rate counters already described but using in place of the two diodes a mechanical switch driven in synchronism with the antenna switching. This provides, with very little equipment and good sensitivity, a steady voltage output proportional to the angular departure of the target from the center line of the antennas, with polarity determined by the direction of departure. Relays are easily operated in accordance with this data voltage to provide lateral guidance in flight.

Operation of the azimuth-determining circuits depends upon presence of the normal radar-beat signal to act as carrier for the antenna-switching modulation. Lateral guidance is needed at ranges considerbly greater than that for bomb release, however, and provision of a wide-band beat-frequency amplifier to accept the longer-range signals would result in poor signal/noise ratio. A circuit to provide automatic compression of the frequency-modulation sweep, and thereby of beat frequency for any given range, whenever the signal out of the beat-note amplifier is weak, is therefore a necessary part of the Super Sniffer. To insure accurate bomb release, this compressor, which uses the plate resistance of a tube as a shunt in the modulator-driving cricuits, must go out of action completely (by cut off) when signal strength reaches normal levels. Another necessary modification of the basic Sniffer circuits consists of momentary "blanking" of both bombing-circuit counters during the signal transients produced by antenna switching; since the counters are electronically switched anyway, such blanking requires only some modification of the switching wave.

Keeping the aircraft pointed directly at the target, which can be done with simple data derived as above, does not produce bomb hits in a cross wind or upon a crosswise-moving target. It is necessary instead to determine and steer an interception heading which will, if held without change, bring the aircraft directly over the target. To accomplish this, as well as to separate yaw of the aircraft from true crosswise relative motion of aircraft and target, requires availability of a steady reference direction such as can be provided by a gyroscope. The complete installation required for bombing is therefore that shown in Figure 12. Antennas are kept trained on the target by a servo motor, yaw being rapidly corrected under control of the gyro stabilizer and crosswise motion being slowly tracked under control of the AN/APG-6 azimuth circuits. Steering information is provided by a pilot-director indicator (PDI), also controlled by the AN/APG-6 and the stabilizer but in such a way that if the indicator is kept centered the aircraft will in time take up a heading that "leads" the crosswise target motion by the right amount for a hit. The pilot, human or automatic, has only to fly so that the indicated radar altitude is constant and the PDI remains centered.

Flight tests of the Super Sniffer have shown it to be able consistently to pick up a small vessel about three miles away and to bring the aircraft directly over it, even in the presence of cross wind and target

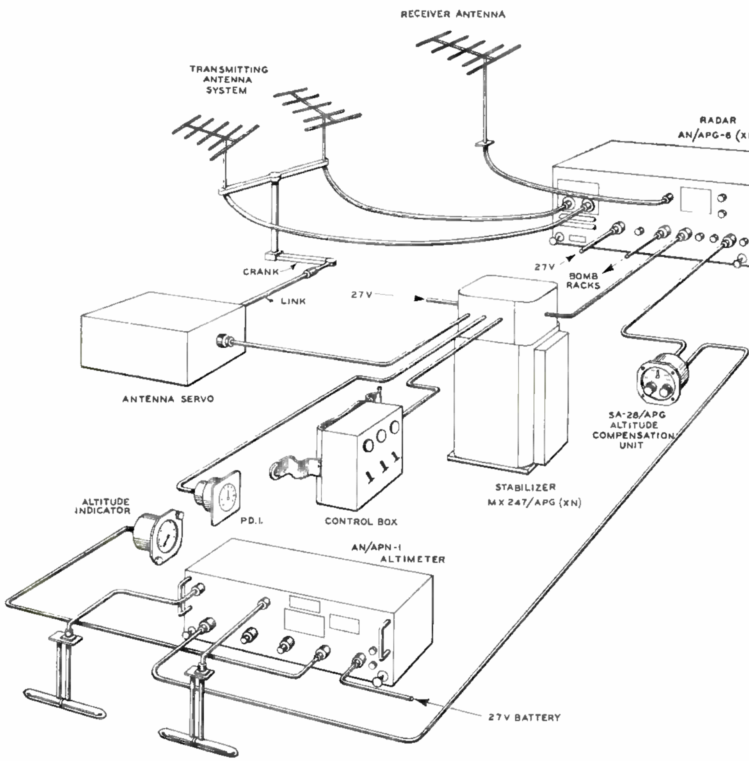


Fig. 12—Complete bombing system AN/APG-6 with azimuth determination.

motion. That this is considerably in excess of ranges obtained with the Sniffer, both at 410 and 1500 megacycles, is of considerable interest. Since the sea-returned, feed-through and detector-unbalance signals that normally appear to limit Sniffer operation increase with increasing power just as fast as does the desired target signal, it would seem that increased power should not increase useful range. Elimination of detector unbalance by the superheterodyne receiver and the especial care in reducing feed-through that is necessary to get clean working at higher power undoubtedly are the main reasons for the good performance of the AN/APG-6. It further appears, however, that the comfortable margin of working signal provided by the higher power makes operation satisfactory under a widened range of conditions.

Adaptation of the bombing equipment to a wider range of fire-control problems has also proved successful. Automatic firing of airborne rockets at invisible targets is desirable and the conditions making the target invisible are also likely to restrict the aircraft to a level firing flight. Study of rocket-ballistic data shows that correct firing range from level flight, while affected by many variables, varies with speed in a way that may again be well approximated, for any single combination of type of rocket and type of aircraft, by a straight-line relation. Variation with altitude of the slope of this approximating line fortunately proves to be strongly similar for a number of types of rockets and aircraft; the intercept S_0 of the line need not be varied with altitude. Propellant temperature has a considerable effect on behavior of rockets, but fortunately this can be adequately allowed for by adding to the normal firing range a ballistic correction that depends on temperature, type of rocket, and type of aircraft, but not on speed or altitude at firing. A very simple table for this correction may be carried in the firing aircraft.

The requirements for automatic level-flight rocket firing are evidently very similar to those for low-altitude bombing. Radar requirements change only in that longer ranges are used, so the versatile transmitter-receiver unit of the AN/APG-17 can be used without change other than revision of the frequency characteristic of its beat-note amplifier. Alterations required in the computing or data-utilizing unit are extensive but straightforward. To make best use of the pass band of the beat-note amplifier, adjustment of speed-counter sensitivity is necessary, so range and speed counters are separated. Graduated manual controls are provided for radar range-sensitivity scale factor, speed-counter sensitivity, reference level of the servo altitude compensator, and counter bias voltage. These four controls are not operating controls, but are preset on the ground to values tabulated for the particular type of missile and aircraft in use. Automatic altitude compensation of sweep width (radar range sensitivity), but not of counter bias, is again required.

Prototype equipment, modified from the AN/APG-17 in the above way and designated AN/APG-17A, was undergoing flight test when the war ended. The four set-up controls are provided in duplicate, so that adjustments for two distinct missiles, such as rockets of one type and bombs, can be made before flight and either missile can be selected for use by throwing a single switch while in flight. A large number of firings of one rocket type indicated that the dispersion of automatic level-flight firing controlled by the AN/APG-17A does not markedly exceed the ballistic dispersion of the rockets themselves.

While the automatic fire-control devices described make the task of flying a bombing approach quite simple, they also leave the pilot quite in the dark as to the progress of his attack. An accessory was therefore developed to warn the pilot that release will soon occur; its mode of operation is quite simple. Modulation sweep and counter bias are set up to operate the release relay prior to the proper moment for bomb release by the desired warning interval. But the release relay is not at first connected to the bombing circuits and the premature "release," when it occurs, is arranged only to light a warning signal and to readjust sweep and bias to the correct values for actual release. After the resulting transients have died out, the release relay is automatically connected to the actual bombing circuits and the actual release takes place normally at the proper instant. This adaptation, requiring only the addition of two relays, is built into the AN/APG-17. It has been found to work well so long as the target is adequate to give a good signal at the warning range.

In order to operate accurately, the sweep-width and bias scales of any of these fire-control systems must be calibrated under conditions simulating actual use. A distant target is most directly and simply simulated by connecting transmitter output to receiver input through an attenuating circuit which introduces an accurately known time delay; the range of the simulated target is determined by the time delay used. Relative motion of radar and target, with resulting Doppler shift of frequency, is simulated by a motor-driven continuous phase shifter introduced into the delay path. These methods have been quite successful, but excessive weight and attenuation of circuits giving relatively long delays have made it worth while to develop also less direct methods, which simulate range by producing an artificial beat-note signal with frequency determined by the actual condition of modulation of the radar transmitter being tested.

D. *Speed Measurement.* The transmitter-receiver unit of the AN/APG-17 has been used in a system for measuring the speed of approaching aircraft. This is only an FM radar system by courtesy, since it is derived from normal FM radar equipment by careful minimization of frequency modulation. Designated AN/SPN-2, this equipment is used in conjunction with precision pulse radar to aid in control of landing approaches to aircraft carriers under conditions of poor visibility. True air speed of the aircraft is required, but the Doppler radar gives only relative speed of aircraft and carrier, so carrier air speed or "wind over carrier" must be combined with the radar data to give the desired indication. Modulation of the reflected signal both in amplitude and frequency by the spinning aircraft propeller is a serious difficulty, and special measures are required to prevent degradation of accuracy by it.

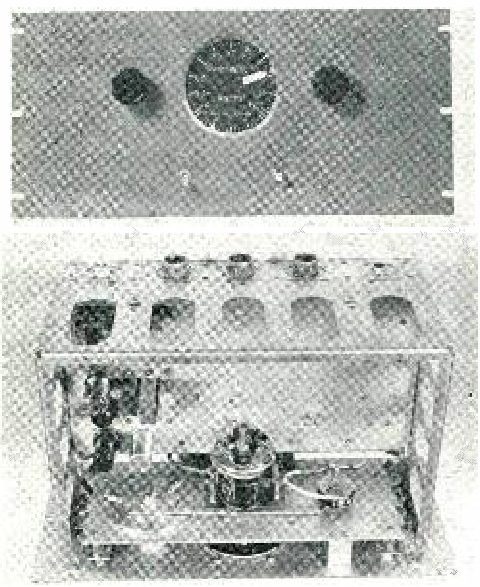


Fig. 13—Indicator unit of speed-measuring system AN/SPN-2.

Modification of the transmitter-receiver unit for speed-only use involves primarily replacement of the vibrating-capacitor modulator by a fixed capacitor, but the beat-note amplifier characteristic is also modified to meet the specified conditions of operation. Dipole antennas in large parabolic reflectors are used. The computer-power unit of the AN/APG-17 is also retained, unchanged in appearance from Figure 9, but its circuits are completely altered. Only a simple null-type unswitched counter, as in Figure 8 of Reference 2, remains; this is servo balanced, and the servo operates slowly to tune a feedback-amplifier filter to the average frequency counted. This filter is the means used to minimize damage by propeller modulation, and the filtered signal is the useful output of the computer-power unit. A relay actuated by signal level is also provided, to disable the system and prevent false indication unless adequate signal from a target is present.

Figure 13 shows the separate indicator and control unit of the

AN/SPN-2 equipment, which is mounted with the pulse radar portions of the complete system for carrier control of approach (CCA). This unit contains a servo-balanced null counter fed with the filtered beat signal from the computer unit. The servo output is combined mechanically with manually set data on wind over carrier, to give dial indication of air speed of the target aircraft. Cams on the output shaft actuate switches to give visual indication of whether air speed is less than, equal to, or greater than a chosen and manually preset value. Provision is made for remote indications, and for optional manual tuning of the filter in the computer unit to the chosen speed.

Field tests ashore and afloat have shown operation of this equipment to be very satisfactory, the indications being stable in spite of propeller modulation. Frequent disagreement with air speeds indicated in aircraft has been shown by timed flight over measured distances to be caused by errors in the usual differential-manometer airspeed indicators, the radar speed indications being always accurate.

III. Experimental Systems

A. *Higher Frequency Operation.* An experimental single-target system, working at 4000 megacycles with 20 watts output and relatively high antenna directivity, has been built and tested as a radar data source, but without specialized utilization of the data produced. As in the case of the AN/APG-6, increased power has been found to give improved operation at relatively long ranges, when the technical problems resulting from the higher power are solved.

This experimental system uses a magnetron as a continuous-wave transmitting oscillator, with magnetron plate current electronically regulated. Frequency modulation is accomplished by the simple but not highly linear process of varying the operating point of the current regulator. The receiver used is a signal-following superhetrodyne, arranged according to Figure 3 of Reference 2. The local oscillator is a reflex klystron, modulated to follow the transmitted signal by automatic control of its repeller voltage.

Because of the relatively high power used, especial care has been required to minimize the disturbing effect of feed-through signal reaching the receiving antenna from the transmitting antenna without travelling to a distant target. Beside reducing this feed-through signal as far as possible by careful decoupling of antennas, it has proved highly advantageous to reduce the damage it can do. This is accomplished by adjusting the internal mixing-signal path in the equipment to produce exactly the same time delay as does the external feed-through path. No disturbing beat signal then appears as a result of mixing these two frequency-modulated signals at the final detector. In the signal-following superheterodyne, the conditions for path equality require that the reference-channel and signal-channel intermediate-frequency amplifiers have matched phase characteristics.

B. *Bombing from Vertical Maneuvers.* Accurate bombing with the Sniffer equipments described in Section II. B above requires that vertical motion of the aircraft be carefully avoided when the bomb is released. There are distinct advantages, however, in being able to bomb from an aircraft maneuvering freely in the vertical plane.

Examination of the free motion of bombs shows that the linear approximation of Equation (4), between slant range and slant speed at release, remains very good even in the presence of a considerable vertical speed component at release. The main effect of vertical velocity is to alter the time of fall of the bomb. This means that Sniffer equipment can be used for dive and toss bombing, but the range/speed sensitivity ratio T' and the intercept bias S_o must then be adjusted as functions both of altitude and of vertical speed, rather than of altitude alone.

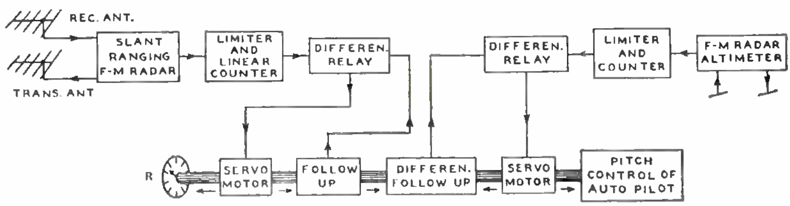


Fig. 14—Radar system to establish glide path.

Construction of a two-variable release compensator, using a three-dimensional cam surface to produce the required function T' of the two variables, altitude and vertical speed, proved practicable. Development of vertical-speed data from the radar altimeter also appears to be feasible. An experimental compensator was actually built, and an AN/APG-17 equipment modified for use with it, but the work had not reached the flight-test stage at the end of the war. Successful automatic bombing in maneuvers would probably have been accomplished if the work had reached the point of overcoming expected operating difficulties caused by loss of radar signal in maneuvers.

C. *Glide Path.* FM radar is well suited to solve the problem of bringing an aircraft down a predetermined inclined path, so long as the underlying earth surface is flat and an isolated target ahead is available. An arrangement for doing this is shown in block form in Figure 14. One radar automatically positions a shaft in accordance with slant range to the forward target. This shaft sets up the "altitude limit" system of the radar altimeter, so that the automatic pilot always flies the aircraft at an altitude varying in accordance with forward target range. If the follow-up devices have linear characteristics, the aircraft will follow a straight inclined path (constant ratio of altitude to range). This system was never flight tested, but experience with all its components leaves little doubt that it would work well.

D. *Time Tracking.* Tracking a target as it moves in range is a common step in the use of radar. Direct speed data such as is available from FM radar is helpful in this process, making it relatively easy to keep the motion of the tracking device in step with that of the target. This is a well known and straightforward application and requires no further comment. Tracking in range requires some sort of variable-speed drive, and such devices are seldom entirely satisfactory. FM radar data, however, lends itself well to a novel sort of target tracking, which does not require any variable-speed drive and may prove useful in a number of special problems.

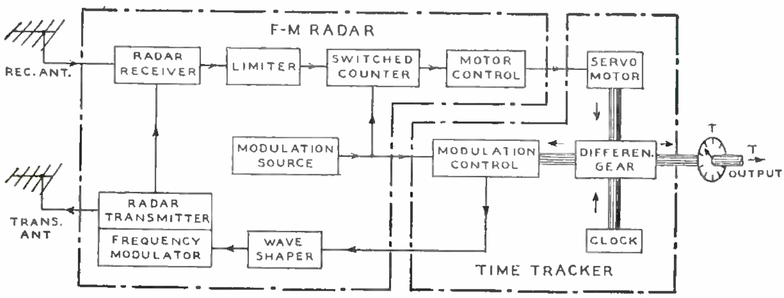


Fig. 15—System for tracking target on time basis.

Assume that at some instant it can be determined that a target will be reached in just 10 seconds. After 6 seconds then elapse, it is known that the target will be reached in just 4 seconds, no matter whether it is being approached at a high speed or a low speed. That is, an ordinary clock suffices to track the approach to a target in time, if only it can be set at some one instant to read correctly the time that must elapse until the target is reached. It was pointed out earlier that if a target at range R is directly approached at constant speed S, it will be reached after the lapse of a time T which is R/S. It was also pointed out that by setting the speed/range sensitivity ratio of an FM radar to equal this time T, the total output voltage e from the counters (see Equation (2)) will be made to equal the counter-bias voltage e_o. The FM radar data can therefore determine a correct clock setting.

Figure 15 shows the very simple basic organization of a time tracker. The clock, beside driving an output shaft, adjusts the modulation sweep width of the radar in inverse proportion to T, while departure of counter output from e_0 controls a servo motor that works through differential gearing to set the clock output to the correct value of time to reach target.

A simple tracker of the above type was built and flight tested. The time-to-target scale of the test unit is 12.5 seconds long. Within the limits of accuracy of observation of passage of aircraft over target, the time-to-target dial gave correct readings on a considerable number of runs. When started automatically upon reception of an adequate signal, the tracker exhibits an interesting sequence of operations. An initial large correction by the servo motor, based on radar data, is usually followed by a period of pure clock tracking with servo quiescent, then by a minor servo correction, another period of tracking without error by the clock, and so forth.

When tracking surface targets from an aircraft as in bombing, corrections for obliquity of radar line of sight are required, but the linear range-speed relation of Equation (4) remains a good approximation at low altitudes. The correction factor T'/T and intercept S_0 are no longer functions of altitude A only. The corrections do prove to be functions of the single variable A/T only, however, and this variable can easily be generated by feeding the follow-up element of the altitude-compensating servo with a voltage which is not constant but proportional to T. Slant corrections were never tried in the experimental tracker, but it does not appear that they would have presented any serious problem. Release of bombs is easily produced when $\frac{1}{2}gT$ becomes equal to A/T, the correct condition for a hit, and with a compensated time tracker release should be accurate even if the radar signal is fading violently.

E. *Search Radar.* Because it seems to involve either an unduly complex multiple-gate indicator or an unduly slow scanning-gate indicator, as discussed in Section IV. D of Reference 1, the application of FM radar to problems of area search has not been pursued very far. An experimental scanning-gate system, designated AN/APQ-19, has been built and given limited field tests on the ground. This system operates at 4000 megacycles and 20 watts continuous output, and uses a "B-type" indication, with target azimuth plotted horizontally and target range vertically on the face of the indicator oscilloscope.

The organization of the system is shown in block form in Figure 16. Range scanning is accomplished by varying the width of the frequency-modulation sweep from cycle to cycle of the modulation, thus bringing the beat notes corresponding to different ranges successively into the narrow fixed pass band of the receiver. The symmetrical-sawtooth wave with exponentially decreasing amplitude which is necessary to produce the frequency modulation for this type of scanning is obtained by integrating a square wave of exponentially decaying amplitude. This square wave is obtained by alternate stepwise discharge of two capacitors in a search-forming unit, under control of timing signals from the modulation generator. Modulation of the frequency of the magnetron oscillator is accomplished by control of approximately resonant helical electron beams within the anode cavities of the magnetron.[3,4] Dipole antennas in parabolic reflectors are used, and the receiver is a simple crystal mixer, feeding a highly selective beat-note amplifier which acts as a range gate. The B-type indicator follows entirely normal radar practice, but the transient buildup of response of the selective amplifier is blanked off from the indicator grid during the first half of each modulation sweep, thus avoiding false indications peculiar to FM operation.

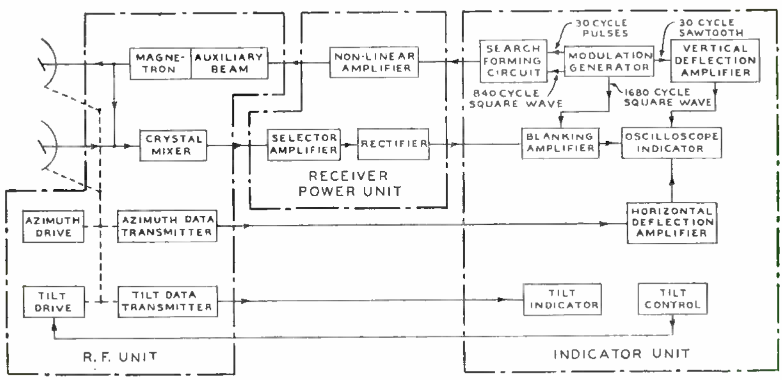


Fig. 16—Block diagram of experimental FM search radar AN/APQ-19.

Search with the AN/APQ-19 covers an azimuth sector of 150 degrees once per second, with 30 complete range searches per second at 5-degree azimuth intervals. During each range search, 56 single sweeps of transmitter frequency, all of the same duration but each of different width, permit separate examination of 56 range elements. Time spent on each range element is 600 microseconds and the time constant of the selective amplifier, tuned to 200 kilocycles, is 140 microseconds. Range search covers the region from 2 to 10 nautical miles,

[3] L. P. Smith and C. Shulman, "Frequency Modulation and Control by Electron Beams", *Proc. I.R.E.*, Vol. 35, No. 7, pp. 644-657, July, 1947.

[4] G. R. Kilgore, C. I. Shulman, and J. Kurshan, "A Frequency-Modulated Magnetron for Super-High Frequencies", *Proc. I.R.E.*, Vol. 35, No. 7, pp. 657-664, July, 1947.

with a constant fractional separation of range elements of 3 per cent as set by transmitter modulation and a fractional resolution of 5 per cent as set by amplifier pass band. Application of the considerations of Section IV. C of Reference 1 indicates that the noise-limited range of the AN/APQ-19 with azimuth scanning stopped should be about 31 miles against a target of effective echoing area 8000 square meters (medium freighter). The comparable range for pulse or multiple-gate FM radar, also operating at 30 complete range scans per second, is 106 miles.

Operation limited by essential noise was never attained with the simple equipment tested, microphonics being usually a major limitation. Small ships seen stern-on were lost at about 3 miles in tests, and a large ship broadside-on was seen clearly at 4 miles. Large fixed targets about 10 miles away were clearly and reliably indicated; this was about the range of the radar horizon for the testing site used. Spreading of indications over several range elements indicated inadequate linearity of frequency modulation in the system tested, though laboratory tests of the magnetron used show that good linearity can be attained by careful adjustment of operating conditions. Special tests on a very few targets showed strength of received signal to depend only slightly on whether transmitting and receiving dipole antennas were parallel or crossed, and limited tests of crossed-polarization duplexing with a single horn and lens radiator gave encouraging preliminary results.

IV. Conclusions

Data-utilizing techniques have been developed which make equipment for automatic control by FM radar data very simple and light. Superheterodyne receivers which use frequency-modulated local mixing signals to avoid excessive noise band have been successfully developed for FM radar use.

FM radar has been found in very extensive use to be well suited for aircraft altimetry, particularly at low altitudes and in landing.

Very accurate low-altitude bombing is possible if bombs are released at a slant range increasing linearly with slant closing speed. FM radar measures both range and speed, and is able to operate a relay when slant target range reaches a value depending linearly on slant closing speed. Fully automatic blind bomb release from low altitudes is therefore possible by using very light, simple FM equipment. Automatic steering control in the bombing run is also obtainable by suitable lobe-switching additions to the same equipment. By modifying its calibration, the bombing equipment can be used for automatic level-flight rocket firing.

Space search by FM radar has been demonstrated, but with a reasonably simple indicator is either unduly slow or insensitive.

The simplicity of automatic control by FM radar methods suggests such means for holding a craft on a predetermined glide path or for tracking a moving target. By combining speed and range data to track an approaching target on a time basis, the variable-speed drive required by other trackers is eliminated.

ACKNOWLEDGMENTS

Far too many individual contributions have been inextricably combined in the development of the systems here described to permit credit to be fairly apportioned to all who deserve it. No attempt will be made at such apportionment, but the important contributions of R. C. Sanders, Jr., throughout the earlier phases of the program cannot go unmentioned. All experimental prototypes were developed by these Laboratories, but the production models of *AN/APN-1 and AN/APG-4 series were designed by RCA Victor Division and of AN/APG-17 by the Admiral Corporation. Extensive flight tests were made possible only by whole-hearted cooperation of a number of naval activities.

RCA TECHNICAL PAPERS†

Second Quarter, 1948

Any requests for copies of papers listed herein should be addressed to the publication to which credited.

"An Accurate Photo Timer", W. C. Morrison, *American Photographer* (April) 1948

"Acoustic Problems in Studio Design", G. M. Nixon, *Electronics* (May) 1948

"Barium Titanate and Barium Strontium Titanate Resonators", H. L. Donley, *RCA Review* (June) 1948

"Comparative Propagation Measurements; Television Transmitters at 67.25, 288, 510 and 910 Megacycles", G. H. Brown, J. Epstein and D. W. Peterson, *RCA Review* (June) 1948

"Continuous Tropospheric Soundings by Radar", A. W. Friend, *Proc. I.R.E.* (Letter to the Editor) (April) 1948

"Electro-Optical Characteristics of Television Systems. Part II—Electro-Optical Specifications for Television Systems", O. H. Schade, *RCA Review* (June) 1948

"An Electronic Simultaneous Equation Solver", E. A. Goldberg and G. W. Brown, *Jour. Appl. Physics* (April) 1948

"Four-Channel Re-Recording System", H. Randall and F. C. Spielberger, *Jour. Soc. Mot. Pic. Eng.* (May) 1948

"Frequency-Modulated Radar Techniques", I. Wolff and D. G. C. Luck, *RCA Review* (June) 1948

"High Power U.H.F. Transmitter", H. C. Lawrence, *Radio News* (May) 1948

"How to Use the Television Stabilizing Amplifier", J. H. Roe, *Broadcast News* (May) 1948

"Improved Arrangement of Base-Pin Connections in New Miniature Tube Types", *RCA Application Note AN-133*, RCA Tube Department, Harrison, N. J. (May 17) 1948

"International Radiotelegraph Communications", S. Sparks, *Electrical Engineering* (June) 1948

"The Magic of Making Television Picture Tubes", Pamphlet, RCA Department of Information, New York, N. Y. (April) 1948

"Microwave Propagation Experiments", L. E. Thompson, *Proc. I.R.E.* (May) 1948

† Report all corrections or additions to RCA Review, Radio Corporation of America, RCA Laboratories Division, Princeton, N. J.

"Midget Field Intensity Meter for the AM Broadcast Band", R. E. Grimm, *Broadcast News* (May) 1948

"Minimum Detectable Absorption in Microwave Spectroscopy and an Analysis of the Stark Modulation Method", W. D. Hershberger, *Jour. Appl. Physics* (April) 1948

"Motion Picture Photography of Television Images", R. M. Fraser, *RCA Review* (June) 1948

"Note on Means of Measurement of Output Plate A.C. Voltage of a Television Deflection Circuit During Scanning Interval", J. M. Brumbaugh, *Broadcast News* (May) 1948

"Note on Ultraviolet Absorption of Zinc Oxide Phosphors", F. H. Nicoll, *Jour. Optical Soc. Amer.* (Letter to the Editor) (April) 1948

"Projection Screens for Home Television Receivers", R. R. Law and I. G. Maloff, *Jour. Opt. Soc. Amer.* (June) 1948

"Radiophoto Standards", S. H. Simpson, Jr. and R. E. Hammond, *Teleg. & Teleph. Age* (June) 1948

"Receiver Sensitivity and Gain Measurements at High Frequencies", *RCA Application Note AN-132* (May 17) 1948

"Sunspots and Radio Weather", A. Arzinger, H. E. Hallborg and J. H. Nelson, *RCA Review* (June) 1948

"Television Progress", David Sarnoff, *Television* (May) 1948

"Television Service. Part I. Using the Test Pattern", J. R. Meacher, *RCA Radio Service News* (May-June) 1948

"Theater Television—a General Analysis", A. N. Goldsmith, *Jour. Soc. Mot. Pic. Eng.* (February) 1948
Inter. Project. (April, May, June) 1948

"Theoretical Analysis of Various Systems of Multiplex Transmission: (Summary; Introduction; Signal Noise Ratios)", V. D. Landon, *RCA Review* (June) 1948

"The TP-16A 16mm Television Projector", R. V. Little, Jr., *Broadcast News* (May) 1948

"TV Field Equipment", J. M. Brumbaugh, *FM and Tele.* (April) 1948

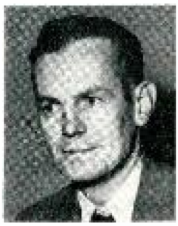
"Versatile Noise-Reduction Amplifier", K. Singer, *Jour. Soc. Mot. Pic. Eng.* (June) 1948

"A 10-Kilovolt Test Probe for Television", J. M. Brumbaugh, *Broadcast News* (May) 1948

NOTE—Omissions or errors in these listings will be corrected in the yearly index.

AUTHORS

EDWIN A. GOLDBERG received the degrees of B.S. in E.E. and M.S. in E.E. in 1938 and 1940 at the University of Texas. From 1938 to 1939 he was employed on a field seismograph crew by the Magnolia Petroleum Co. He joined the RCA Manufacturing Co. in 1940 as a student engineer, and was assigned to the Research Division in 1941. He was transferred to Princeton in 1942 with RCA Laboratories Division. Mr. Goldberg is a Member of Eta Kappa Nu, Tau Beta Pi, Sigma Xi, and the American Institute of Electrical Engineers.



WILLIAM A. HARRIS received his B.S. degree in Electrical Engineering from Rose Polytechnical Institute, Terre Haute, Indiana in 1927. He was employed by the General Electric Co. from 1927 to 1929, and was transferred to the RCA Victor Division in Camden in 1930 where he worked in the Receiver Design Laboratories. He has been with the Application Engineering Laboratory of the RCA Tube Department at Harrison, N. J. since 1931. Mr. Harris has been connected with much of the tube development and application work on frequency converters and on high-frequency circuit problems.

J. LEWIS HATHAWAY received his B.S. degree in Electrical Engineering from the University of Colorado in 1929. In the same year, he joined the National Broadcasting Company where, as a member of the Development Group, he has since been engaged in all fields of the Company's engineering activities. While on a leave of absence 1941 to 1944, he served as a Special Research Associate at Harvard University performing underwater sound development work. Mr. Hathaway was appointed a Staff Engineer of the National Broadcasting Company in 1945.

WILLIAM HOTINE attended Colgate University after which he was employed as a technician by RCA Marine Department in 1923. In 1925 he joined Brunswick-Balke-Collender Co. as a radio engineer, and was engaged in broadcast receiver work until 1930, when he entered the motion picture sound-on-film field with United Research Corporation. From 1935 to 1940 he was chief engineer of Films Inc., engaging in 16-mm sound-on-film projection problems. In 1940 he joined Federal Telephone and Radio Laboratories as a development engineer on U.H.F. radar. In 1942, he became associated with the Raytheon Manufacturing Company as a development engineer on microwave radar. He became a member of the Development Group of the NBC Engineering Department in 1946 where he has been engaged in radio broadcast equipment development.

GARDNER L. KRIEGER received a B.S. degree from Northwestern State College (Louisiana) in 1939 and did one year of graduate work at Louisiana State University. From 1940 to 1941 he worked as a student engineer with the RCA Manufacturing Company, and since then has been with the RCA Manufacturing Company and RCA Laboratories Division on the research staff of the Electronic Research Section. Mr. Krieger is a member of Sigma Xi and the American Physical Society.

VERNON D. LANDON attended Detroit Junior College. From 1922 to 1929, he was in charge of the Radio-Frequency Laboratory of the Westinghouse Electric and Manufacturing Company. In 1930 he was Assistant Chief Engineer of the Radio-Frequency Laboratories and in 1931 he became Assistant Chief Engineer of the Grigsby Grunow Company. In 1932, he joined RCA Manufacturing Company in Camden, N. J. Since 1942, he has been with RCA Laboratories Division at Princeton, N. J. engaged in research work on radio circuits. Mr. Landon is a Member of Sigma Xi, and a Senior Member of the Institute of Radio Engineers.

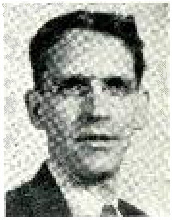
DAVID G. C. LUCK received his B.S. degree from Massachusetts Institute of Technology in 1927 and his Ph.D. in 1932. In 1927-28 he was a Swope Fellow in Physics, and a Malcolm Cotton Brown Fellow in 1928-29. He was an Assistant in the Department of Physics at Massachusetts Institute of Technology in 1929-32. He joined the Research Division of RCA Victor Company in 1932 and was with the Victor Division of RCA Manufacturing Company from 1935 to 1941, working on pulse communication, direction finding, and radio guidance of aircraft. He was transferred to RCA Laboratories Division upon its formation and has remained with the Princeton laboratories of the Division since that time. Dr. Luck is a Member of the American Physical Society, Sigma Xi, and the Institute of Radio Engineers.

GEORGE A. MORTON received his B.S. in Electrical Engineering in 1926; his M.S. in 1928 and his Ph.D. in Physics in 1932 from Massachusetts Institute of Technology. From 1927 to 1933 he was research associate and instructor at Massachusetts Institute of Technology. In 1933 he became a Member of the research division of RCA Manufacturing Co., and from 1941 to the present time has been with RCA Laboratories Division. In 1946-7 he was at Oak Ridge, Tennessee on a year's leave of absence working in the field of nuclear physics. He is active in the fields of television, special electron tubes, and the application of electronics to nuclear problems. During the war he served as Section Member of the National Defense Research Committee and is a member of the AAF Scientific Advisory Board. Dr. Morton is a member of Sigma Xi, the American Physical Society and the American Institute of Electrical Engineers and a Senior Member of the Institute of Radio Engineers.

JOHN E. RUEDY received an A.B. degree from Western Reserve University in 1924 and a Ph.D. degree in Physics from Cornell University in 1933. From 1925 to 1927 he was an engineer with Bell Telephone Laboratories, from 1927 to 1934 held teaching and research assistantships at Cornell University, from 1934 to 1936 was with the Evaporated Metal Films Company, and from 1936 to the present time has been with RCA Manufacturing Company and RCA Laboratories Division in the Electronic Research Section. Dr. Ruedy is a member of Sigma Xi and the American Physical Society.

OTTO H. SCHADE graduated from the Reform-Real-Gymnasium, Halle, Germany, in 1922. From 1922 to 1924 he was with the Telephonfabrik A. G. vorm. J. Berliner, Berlin and Dusseldorf; from 1924 to 1925, in charge of the laboratory in the radio manufacturing company "Ratag" in Berlin; and from 1926 to 1931, in the engineering department of the Atwater Kent Manufacturing Company. Since 1931 he has been with the Tube Department, RCA Victor Division at Harrison, N. J. He received a Modern Pioneer Award from the Radio Manufacturers Association in 1940. Mr. Schade is a Senior Member of the Institute of Radio Engineers.

A. C. SCHROEDER received the B.S. degree in electrical engineering from the Massachusetts Institute of Technology in 1937, and the M.S. degree from the same institution in the same year. He joined Radio Corporation of America in 1937, and is now engaged in television research at RCA Laboratories Division in Princeton, N. J. Mr. Schroeder is a member of the American Association for the Advancement of Science, and an Associate Member of the Institute of Radio Engineers.



KARL R. WENDT attended the Municipal University of Akron, Marquette University, and the University of Wisconsin. During 1928-1929, he was a research assistant in the chemistry department of the University of Wisconsin, and in 1929-1930 he was in the research laboratory of the Sun Oil Company. In 1930, he joined the RCA Manufacturing Co. and has been a member of the research laboratories of RCA since 1934. From 1942 to 1948 he was with RCA Laboratories Division in Princeton, N. J., engaged in television research and development. Since August, he has been with Colonial Radio Corp., Buffalo, N. Y. Mr. Wendt is a Member of Alpha Chi Sigma, Sigma Xi, and a Senior Member of the Institute of Radio Engineers.

IRVING WOLFF received the B.S. degree in physics from Dartmouth College in 1916 and the Ph.D. degree in physics from Cornell University in 1923. He was an instructor in physics at Iowa State College during 1919 and at Cornell University from 1920 to 1923, where he was a Hechscher Research Fellow in 1924. He joined the Technical and Test Department of the Radio Corporation of America in 1924. From 1930 to 1941, in the Research Division of the RCA Manufacturing Co., he worked on problems in microwaves, radar, and aviation. Since 1941 he has been connected with the RCA Laboratories Division, Princeton, N. J., where he is now Director of the Radio Tube Research Laboratory. Dr. Wolff is a Fellow of the Acoustical Society of America and the Institute of Radio Engineers, and a Member of the Physical Society, American Association for the Advancement of Science, and Sigma Xi.