RCA REVIEW

a technical journal

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

VOLUME IX MARCH 1948 NO. 1

RCA REVIEW

GEORGE M. K. BAKER
Manager

CHAS. C. FOSTER, JR.
Business Manager

SUBSCRIPTIONS:

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

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

rporation of America, RCA Laboratories Division*

ember, and December by Radio Corporation of
0 Rockefeller Plaza, New York 20, N. Y.

IEW, Radio Corporation of America,
inceton, New Jersey.

1946, at the Post
of March 3, 1879

A

LE

LE, *Treasurer*

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 | MARCH 1948 | NUMBER 1

CONTENTS

RCA REVIEW

FOREWORD

THE seventh volume in the **RCA Technical Book Series** was published by RCA REVIEW in January 1948. This book—FREQUENCY MODULATION, Volume I—includes technical papers on FM broadcasting and other FM applications written by RCA scientists and engineers and originally published during the years 1936-1947. Twenty-four full papers and twenty-one summaries are included together with an FM Bibliography and FM Station Placement and Field Survey Technique Summaries.

* * *

Other available volumes in the **RCA Technical Book Series** include:

TELEVISION, Volume III	(1938-1941)	— 1947
TELEVISION, Volume IV	(1942-1946)	— 1947
RADIO FACSIMILE, Volume I		— 1938

* * *

One volume in the **RCA Engineering Book Series** has been published:

PATENT NOTES FOR ENGINEERS — 1947

* * *

RCA TECHNICAL PAPERS—INDEX, Volume II(b) (1947) was published in February 1948. Volume I (1919-1945) and Volume II(a) (1946) were published early in 1947.

* * *

Information on any of the above material may be obtained by writing to:

RCA REVIEW
Radio Corporation of America
RCA Laboratories Division
Princeton, New Jersey

The Manager, RCA REVIEW

ELECTRO-OPTICAL CHARACTERISTICS OF TELEVISION SYSTEMS*†

By

Otto H. Schade

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

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 this 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, resolving power, response characteristics, and steady and fluctuating brightness distortions are discussed and related to the characteristics of external imaging systems and the television process.

Introduction

THE function of a television system is to generate optical images which create in the mind of the observer the illusion of seeing real objects and action scenes. The degree of technical perfection in the optical image required to create this illusion depends to a considerable extent on the televised subject matter and the skill exhibited in capturing and directing the interest of the observer.

---

* Decimal Classification: R138.3 × R583.11.

† 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 appear in this issue. The remaining parts are scheduled for publication in succeeding issues of Volume IX of *RCA REVIEW* during 1948.

The process of "tele-vision," illustrated by Figure 1, requires a complex electro-optical system to extend the optical imaging process in the eye to the lens of the television camera which sees the real object. The system components ((1) to (4), Figure 1) should not limit the capabilities of the eye; but on the other hand, they should not be required to pass information the eye cannot see. It is neither essential nor desirable for easy vision to reproduce stationary images with minute detail requiring inspection at closer than normal viewing distances, because magnification and "close-ups" of interesting detail are functions performed with greater reality and better perspective by the television camera at the real object.

Contrast, gradation, color, sharpness, and brightness distortions in the reproduced image are judged by the eye. The capabilities and optical characteristics of the eye determine, therefore, the optical standards for the reproduced image. It would be rather hasty to conclude that the same optical standards of quality apply as well to the optical image formed by the camera lens, even in a purely photographic process. This assumption would require that the quality of the entire conversion process between the optical and reproduced images in the system exceed the quality of both images by a substantial factor leaving but a very small margin for degradation in each individual process. Losses and degradations, however, occur in every stage of a practical process, and it is by no means obvious what degree of perfection should be attained in optical or electrical components where an unbalance of performance is economically or technically sound, and where a correction may best be introduced to compensate for a deliberate unbalance or a limitation. The answers to these questions can hardly be given in a usable form unless a carefully planned examination of all processes in the system, including the process of vision is made.

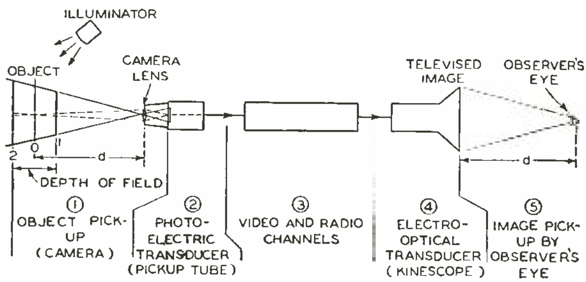


Fig. 1—Optical and electrical components of television process.

It might appear logical to analyze the system in an order beginning with the camera and progressing through the various stages of conversion to the final optical image. This course, however, would lead to a very specific or an unnecessarily broad treatment of processes which can be avoided by first establishing the guiding standards for optical images which are based on the characteristics of the visual process. After the transfer characteristics of the kinescope are determined, these optical standards can be translated into electrical specifications for the television process and the "video" signal. The generation of signals can then be treated in a normal sequence. Beginning with a discussion of the photoelectric conversion process, the properties of lenses, camera tubes, and video signals can be reviewed on a unified basis permitting coordination in a final evaluation of the television process.

All components of a television system are fundamentally converters or "transducers" of energy. In the simplest case a transducing process can be specified by a single number, which will be termed the "transfer factor." The transfer factor (g) is the ratio of the output "signal" to the input "signal" in appropriate units and often implies the efficiency of the process. In most cases, however, the transfer factor is a variable dependent on one or many parameters such as the average signal level, the input signal intensity, the output "load", the frequency and form of the signal, and others. The transfer properties of a transducer are, therefore, more accurately described by one or several characteristic curves; i.e., the "transfer characteristics". Examples of well-established transducer characteristics are the static characteristics of electron tubes and of photographic film, and the transfer characteristics of electrical coupling networks and filters.

The methods of determining or measuring transfer characteristics and their use in the graphic determination of signal distortion and dynamic operating conditions are fundamentally similar in all of these cases. Moreover, the same methods can be applied to all transducing processes in a television system. When the transducing system of the eye and vision is discussed later on, the similarity of the treatment to that of electrical transducers will be apparent and useful, although the treatment may seem at first glance to be a rather unorthodox method of covering a much discussed and somewhat controversial subject. In this and other instances, the derived characteristics are based on data and observations reported in the literature by specialists in the particular field, although the form of presentation may, at times, be different.

Aperture Considerations and Detail Visibility

One property of particular interest in the analysis of image-transducing systems, discussed in detail in Parts II and IV but referred to throughout this paper, is the loss of resolution and quality which occurs in each process of optical or electrical conversion. In a broad sense each conversion process has a "frequency response characteristic" which shows a decrease of "signal output" in response to "input signals" of increasing spatial or electrical frequency when analyzed by scanning.

In electrical terms, transducers act as low-pass filters. When the signal is of electrical origin (as in the case of fluctuation "noise" signals) the entire transducing system is advantageously treated as an electrical filter system.

In most cases, however, it is of advantage to retain the physical concepts of scanning when treating resolution and the response of transducers to detail signals. According to these concepts the "resolving element" of optical and electrical devices can be regarded as a small "aperture". This "aperture" may be a real scanning aperture as in a Nipkow scanning disc or in an image-dissector tube; it may be a multiplicity of reduced optical images of one larger aperture simultaneously forming countless figures of confusion as in the case of a defocussed lens; or again it may just be a fictitious or equivalent aperture, moving or stationary, which determines the cross section of elemental beams of light or electrons.

If the size and flux distribution of this "aperture" is known, the response characteristics or "aperture effects" can be computed by analysis with the scanning process and vice versa. Mathematical evaluations of the aperture response based on a Fourier analysis of a step function or pulse wave can be found in the literature[1-4], but the results do not apply directly to practical aperture shapes or test patterns.

A more practical approach requires a method of expressing the aperture response in a form which can be checked by measurements of the response to signals generated with normal test objects producing repetitive square-wave flux patterns (optically such as the standard bar or line pattern). The detail area in these test objects is defined with respect to the picture area or a unit area. When the resolving aperture is symmetric, the resolved area can be specified by its length

---

[1] H. A. Wheeler and A. V. Loughren, "The Fine Structure of Television Images", *Proc. I.R.E.*, Vol. 26, pp. 540-576, May, 1938.

[2] Pierre Mertz, "Television—The Scanning Process", *Proc. I.R.E.*, Vol. 29, pp. 529-537, October, 1941.

[3] M. Cawein, "Television Resolution as a Function of Line Structures", *Proc. I.R.E.*, Vol. 33, pp. 855-864, December, 1945.

[4] R. E. Graham and F. W. Reynolds, "A Simple Optical Method for the Synthesis and Evaluation of Television Images", *Proc. I.R.E.*, Vol. 34, pp. 18W-30W, January, 1946.

in one direction (horizontal or vertical), but most commonly the length is expressed by the reciprocal value: the line number per unit length (N/mm) or the line number per picture frame height (N).

When the resolving aperture is asymmetric, a square having an equivalent detail area can be specified by a balanced line number $\overline{N} = \sqrt{N_H \times N_V}$. It is well known that the balanced line number $\overline{N}$ indicates for a television system the minimum electrical frequency channel Δf required for its reproduction. For a given frame time T_f, a balanced line number $\overline{N}_{co}$ at channel cut-off, and normal blanking percentages, the frequency channel is given by the relation

$$\Delta f = 0.85\, \overline{N}_{co}^{\,2} / T_f \tag{1}$$

This equation establishes a connection between optical and electrical picture information.

Response characteristics have been computed for various aperture types as a function of the line number in square-wave flux patterns in order to compare them with the measured characteristics of optical and electrical processes. A series of photographs illustrating the response of optical apertures give quantitative proof of the theory.

Any point on a response curve is expressed by a response factor, which is the ratio of the aperture signal output at a line number N to the normal signal output at $N = 0$. Absolute signal values, such as the optical detail contrast, are easily derived, but should not be confused with the response factor which only indicates such values on a percentage basis. It is important to remember, especially in the treatment of the eye, that a small response factor may represent signal intensities well above threshold values when the "signal" to the aperture is sufficiently large provided it does not cause overload or saturation of the "indicating device."

With these principles in mind it appears logical to review the characteristics of the eye and vision as a transducing system and to develop, if possible, some of its transfer and aperture response characteristics as a guide in determining desirable optical specifications for the final image of a reproducing system. Comparison of these standards of quality with those of the graphic arts will be made on various occasions for reference to the performance level of accepted practices.

The subject of brightness distortion, especially the visibility of undesired optical detail signals resulting in graininess of the image, is treated in greater detail. The visibility of these "random brightness fluctuations" is observed by experiment and evaluated by taking into account the "filtering effect" due to the response characteristics of eye and kinescope. The results of this analysis are expressed by grain-visibility constants which are optical "signal-to-noise" ratios.

With this background a number of characteristics and parameters of the television process can be examined more specifically in relation to optical performance standards. Translation of these standards into specifications for the electrical or video signal is based on the electro-optical transducing process: the transfer and aperture characteristics of the "kinescope".

Electrical signal-to-noise ratios for high-quality images and various optical effects arising from scanning-line structures and response characteristics of limited electrical channels can then be evaluated.

The relative sharpness of images, reproduced by a television process with certain characteristics and by photographic processes with limited resolving power can be compared by subjective methods as described by Baldwin[5].

The results are of considerable importance because they indicate certain television system specifications necessary to achieve equality with accepted motion picture performance. Because of differences arising from the use of small scanning apertures, sharply defined electrical channels, and the influence of grain or "noise" in practical images, a re-evaluation of the relative sharpness by a subjective method is of interest.

Video Signal Generation and Factors Determining the Over-all Performance of the Television System

As will be discussed in Parts III and IV, present methods of generating video signals for transmission over a single electrical channel are based on the scanning process. Fundamentally, a projected area of the field of view is inspected through a small "aperture" moving with uniform speed along adjacent parallel paths. The spatial distribution of light along these paths, placed end to end, is thus converted into an order of time distribution and the light fluctuations "observed" through the scanning aperture can readily be transduced into electrical signals.

An early application of this principle is the light-spot scanning system which has again become of interest for certain uses. In modern applications, the mechanical aperture of the old Nipkow scanning disc is replaced by an equivalent aperture defining the size of the light-exciting electron beam in a high-voltage kinescope. The intense electrically-deflected light spot is focused onto opaque or transparent

[5] M. W. Baldwin, Jr., "The Subjective Sharpness of Simulated Television Images", *Bell Sys. Tech. Jour.*, Vol. 19, pp. 563-587, October, 1940.

objects. The light reflected or transmitted from the elemental object area under the scanning light spot is picked up by a phototube which transduces the received light-flux variations into electrical signals. To overcome various limitations imposed by this special method of object illumination, later efforts have been directed toward solutions permitting the use of a normal continuous illumination.

The principle of electrical charge storage as first applied by V. K. Zworykin in his iconoscope[6,7] overcame the serious lack of efficiency which occurs when continuously illuminated objects are "observed" with an instantaneous photoelectric transducer through a scanning aperture. Because only the small fraction of light energy under the scanning aperture is converted into signals at any one instant, the conversion efficiency of instantaneous transducers (with respect to the total light flux) decreases in direct proportion to the size of the scanning aperture.

Because of the energy storage between repetitive periods of signal development, the signal current and efficiency of storage-type transducers under continuous exposure conditions are basically independent of the size of the scanning aperture. The iconoscope thus made possible the first direct pickup of scenes with good resolution under natural lighting conditions.

The need for still higher sensitivity remained, and it is largely due to the work of Rose, Iams, Weimer, and Law[8-11] that the efficiency of all three stages of the transducing process has been increased to the present high level obtained in the image orthicon.

Although further increases of sensitivity are possible and desirable, the factors controlling the quality of television signals and images have become more important. Of particular interest are the relations of scene illumination, sharpness, and signal-to-noise ratio which depend on the combined characteristics of optical and photoelectric processes in the television camera.

---

[6] V. K. Zworykin, "The Iconoscope—A Modern Version of the Electric Eye", *Proc. I.R.E.*, Vol. 22, pp. 16-32, January, 1934.

[7] V. K. Zworykin, G. A. Morton, and L. E. Flory, "The Theory and Performance of the Iconoscope", *Proc. I.R.E.*, Vol. 25, pp. 1071-1092, August, 1937.

[8] H. A. Iams, G. A. Morton, and V. K. Zworykin, "The Image Iconoscope", *Proc. I.R.E.*, Vol. 27, pp. 541-547, September, 1939.

[9] A. Rose and H. A. Iams, "The Orthicon", *RCA REVIEW*, Vol. 4, No. 2, pp. 186-199, October, 1939.

[10] A. Rose, "The Relative Sensitivities of Television Pickup Tubes, Photographic Film, and the Human Eye", *Proc. I.R.E.*, Vol. 29, pp. 293-300, June, 1942.

[11] A. Rose, P. K. Weimer and H. B. Law, "The Image Orthicon—A Sensitive Television Pickup Tube", *Proc. I.R.E.*, Vol. 34, pp. 424-432, July, 1946.

The efficiency of the optical imaging process performed by the camera lens is not only a function of the field depth which is to be imaged sharply, but it is also a function of detail size. The properties of camera lenses as transducers of light can be expressed in terms of optical transfer factors, transfer characteristics, and aperture response characteristics. Because the conversion of light energy into physical information on film is not exactly comparable to the purely photoelectric process in television camera tubes, the principles and limitations of optical imaging deserve reviewing. Variation of optical parameters causes in some respects dissimilar effects which must be considered in the use and selection of lenses and pickup tubes for the television camera.

The optical transfer factor of the television camera, for example, is independent of image size and is determined by the lens diameter (not the f: number) which has a fixed value for a specified viewing angle and sharpness in depth. The minimum "plate" size for practical cameras, however, is limited by optical and electrical difficulties in obtaining adequate resolving power.

Observations on television signals and difficulties experienced in making accurate test patterns for television purposes have indicated that the aperture response of camera lenses may depart considerably from the theoretical curve expected from their limiting resolution. As data on the aperture response of lenses are practically non-existent, a series of optical tests were made with a variety of lenses to indicate the order of the deviations and the type of equipment best suited for direct measurements.

A "television micro-photometer" was developed which, in principle, is an optical lens-bench arrangement except that observation and measurements are made through a television system. The lens under test is set up to form a greatly reduced image of an intensely illuminated line test pattern taxing its resolving power. The optical image from the lens is inspected through a high-powered microscope over a television camera chain and is seen as a highly magnified image on the television screen. One cross section through this image is made visible on an oscilloscope by means of a "line selector". The waveform represents the transduced light-flux variations in the lens image and is a photometric trace of its aperture response.

The response characteristics obtained in this manner furnish exact numerical values for expressing the sharpness of the optical camera image and permit correlation with the performance of other system components.

Given the aperture characteristics of the components, the over-all aperture response of the process can be computed and approximated by the response of one equivalent aperture having an effect similar to the combined effect of all aperture processes in the system including that of the eye. The conditions for equivalence are debatable in some processes, requiring verification by other methods, but the aperture effect of single components can be judged by its influence on the over-all response of the system. It will be apparent that much can be gained by improving the response of nearly all system components.

For a better understanding of both its strong points and limitations, the transducing process in television pickup devices will be examined in somewhat greater detail. In storage types, the storage capacity and efficiency of the signal-developing process determine latitude and obtainable signal-to-noise ratios which, in turn, determine gradation and range of light values in the final image. The shape of the over-all transfer characteristics showing light output as a function of light input is not fixed by individual characteristics because it can be controlled in the electrical channel. Modifications are limited, however, by the character and magnitude of the fluctuation noise signals.

The question of whether or not the television process should have a transfer curve like a photographic process cannot be answered with a definite yes or no, because both processes have strong and weak points and both will give optimum performance only when properly used within their limitations.

* * *

PART I—CHARACTERISTICS OF VISION AND VISUAL SYSTEMS

(A) Some Characteristics of Vision and Visual Systems

It is impossible to simulate accurately the process of normal vision by an equivalent device or to formulate the operational characteristics of the process of vision except for relatively simple and properly defined viewing conditions and test objects.

The viewing conditions for television images have no exact precedent in the visual arts, although they are quite similar to the conditions prevailing when small motion pictures are viewed in the home. A formulation of the viewing conditions is attempted but they are based on the characteristics of the eye and not necessarily on present standards of related visual arts.

1. *Viewing Angle and Viewing Ratio*

The normal field of vision covers an angle (2α) of 30 to 40 degrees (see Figure 2). A considerably larger field (approximately 90 degrees for color vision) is imaged on the retina but it is rather poorly resolved by the eye. It is well known that only the central area (fovea centralis) extending over hardly more than 2 degrees is capable of high resolution. The resolving power decreases rapidly, away from the optical axis for the remaining area which acts primarily as a view finder.

For sharp vision the eyeballs are, therefore, moved continuously to enclose and follow sections of interest in the viewing field within this small angle. Comfortable vision over extended periods should not require movement of the head and the field under observation should, therefore, not exceed a certain angle. The length of one line in a book of normal size indicates this angle to be of the order of 25 degrees. Pictorial objects are usually viewed under a smaller angle such as 15 to 20 degrees in order to obtain a more simultaneous impression of the entire object. Standard television images have the dimensional ratios V:H:D = 3:4:5 (Figure 2). It is convenient to express the diagonal

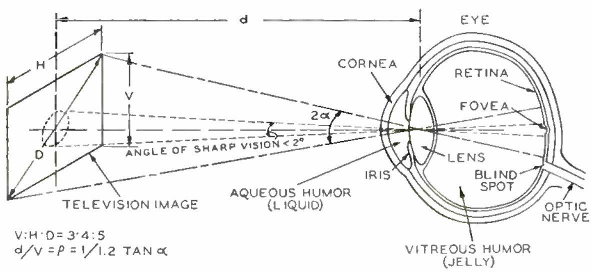


Fig. 2—Pickup of television image by the eye.

viewing angle 2α by the ratio ρ of viewing distance d to picture height V; hence

$$\rho = d/V = 1/1.2 \tan \alpha \qquad (2)$$

Comfortable vision requires a viewing ratio $\rho \geqq 4$. The values generally found in motion picture theaters range from $\rho = 1$ for the front seats to $\rho \simeq 7$ for the rear seats. The ratio $\rho = 4$ will thus be considered as the *minimum* standard viewing ratio at the receiver.

2. *"Static" Transfer Characteristics of the Signal Conversion Process in the Eye.*

Light entering through the lens system of the eye is transduced into stimulating signals perhaps by the photoreceptors of the retina (cones and rods). The signals are transmitted over nerve channels to the visual center of the brain for analysis and interpretation. The sensation of brightness (S) may be regarded as one of the "output signals" of this process and is a function of the retinal illumination (E_r). The curve of S as a function of E_r is a "static transfer characteristic" of this transducing process. Its shape, range, and slope, even though approximate, are of considerable interest in evaluating requirements for external imaging processes. The characteristic curves $S = f(E_r)$ cannot be measured directly but they can be generated by the integration of incremental slope values which are known for specified viewing conditions. The reciprocal of these slope values is the "minimum perceptible brightness difference" $\triangle B_{min}/B$ which has been measured[12] as a function of field brightness B. One such characteristic for adjacent areas is shown in Figure 3a. Characteristics with larger values for the unit S will be obtained for separated areas; further variations occur when the brightness of the background surrounding the viewing field is changed.

Figure 3a is a general type of transfer characteristic, but the value of the sensation unit ($S = 1$) should not be considered as absolute. The unit of equivalent retinal illumination E_r corresponding to an external field brightness B is the "troland" (Td.)[13] The ratio of E_r to the viewing field brightness B is determined by the optical constants of the eye lens system and is proportional to the effective area (A) of

---

[12] John W. T. Walsh, PHOTOMETRY, Constable & Co., Ltd., London, Eng., 1926. (Page 53, Figure 26—The "Troland" supersedes the "photon"; both have the same definition.)

[13] L. T. Troland, "Absence of the Purkinje Phenomenon in the Fovea", *Jour. Frank. Inst.*, Vol. 182, pp. 111-112, July, 1916.

There is a good deal of laxity in the presentation of characteristics of the eye and use of the "troland" (formerly photon) unit. It is often impossible to correlate various data because of omission of information. Some authors use the troland unit, assuming a fixed pupil area, which is called "normal" and may be anything between 1 to 10 millimeters. Data on flicker, color sensitivity acuity, and, in general, all information on eye characteristics are of little use unless viewing angle, average field brightness, and/or the iris opening are specified because these values determine light flux density and internal eye illumination. Without definite specification, the data are as useless as film measurements made with great care with a camera having an f:2 lens but which neglects the fact that the camera has an automatic iris varying the aperture from f:2 to f:10 according to an unspecified law which may be a function of intensity, total light flux, flux distribution, color, and exposure time, or all combined. It is only to be expected that the results obtained for the film lack correlation and that curves of film characteristics show peculiar variations from normal steady functions.

Conversely, it must be expected that curves of film (or eye) performance versus external field brightness will exhibit the variations caused by the action of the automatic iris.

In this paper the conversion into troland units is based on the iris area as a function of the average field brightness regardless of color as given in Figure 4 which may be a reasonable assumption for a viewing ratio (ρ) of 4. In view of this, the author does not claim a high degree of accuracy for the general characteristics of the eye which may represent but a rough approach to accurate facts.

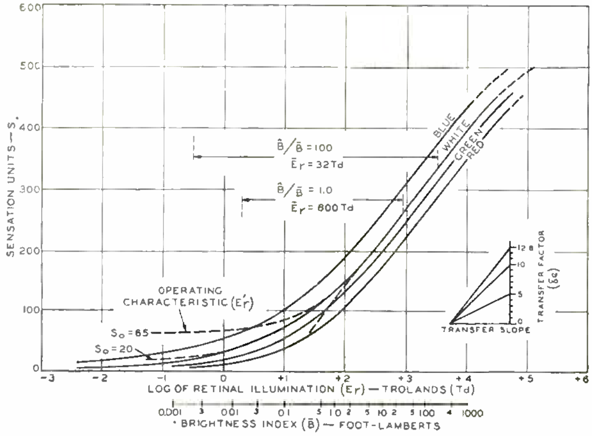


Fig. 3a—Static characteristics of the eye.

the lens of the eye. By definition

$$E_r/B = 3.43A \tag{3}$$

where A is in square millimeters, B in foot-lamberts, and E_r in trolands. Lens diameter and area A are controlled automatically by the iris of the eye in response to field brightness and follow the relation shown by Figure 4.[12] The triangles in Figure 4 indicate check points obtained

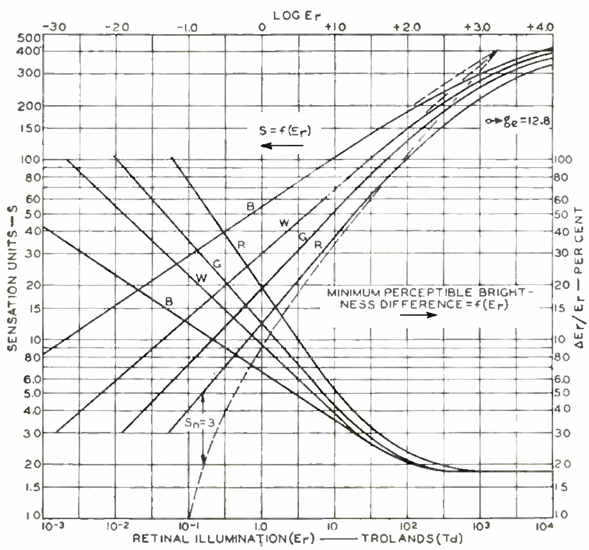


Fig. 3b—Static characteristics of the eye and threshold response to light increments.

with a viewing ratio (ρ) of 4 on an 8×10-inch kinescope raster of uniform brightness in a dark room. The action of the iris causes the brightness ratio (Equation (3)) to shift at most by a factor of two (see Figure 4) for a 10:1 change in the brightness of the object field. Because the shift is small, it is justifiable to make the simplifying assumption that it is the average value $\overline{B}$ of the field brightness which determines aperture and brightness ratio E_r/B.

The transfer curves (Figure 3a) obtained by integration of incremental slope values permit many graphic solutions and their use is in many respects similar to that of electron tube characteristics or sensitometric curves for photographic film. The transfer characteristics of the eye cover the enormous "input signal" range of 10 million to 1 (approx.). This range cannot be seen in one image because of various

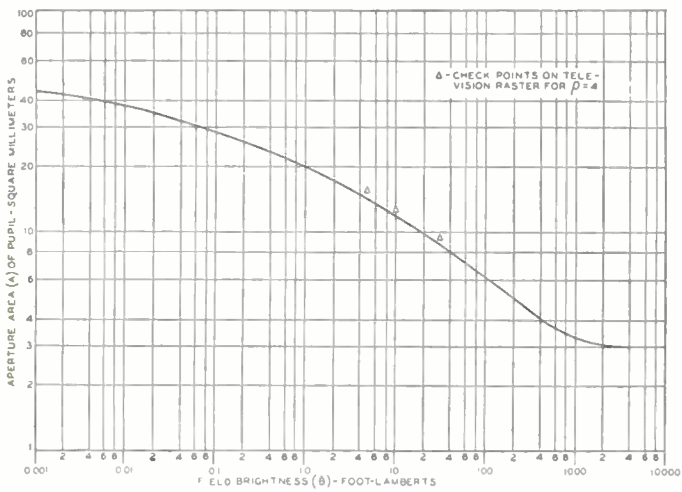


Fig. 4—Variations of aperture area of pupil with field brightness.

effects in the optical system and "development" process such as the optical scattering of light and the interaction between strongly and weakly exposed sections of the image surface which is similar to the electron redistribution in iconoscopes and image orthicons or to film development in limited solutions.

When the incremental slope values are integrated for the particular characteristic, in Figure 3a, the integration constants, i.e., the values of the curves for $S = 0$ are uncertain, although the unit $S = 1$ is defined as the minimum perceptible brightness difference. By replotting $S = f(E_r)$ on a logarithmic scale (Figure 3b) it was found that the toe sections follow a power law because the curves become straight lines upon addition (or subtraction) of certain constant values ($\Delta' S$)

to S. This addition represents merely a vertical shift of the curves of Figure 3a without deformation.

The process of determining the integration constant is analogous to the process of determining the contact potential (E_o) of electron tubes. Their characteristics are known to follow the power law $I = K(E)^{1.5}$ which is a straight line with a slope of 1.5 on log-log coordinates. In this case, the measured curve $I = K(E + E_o)^{1.5}$ differs by an additive constant E_o, the contact potential.

One may speculate that the true additive constant for the straight line section of the eye characteristic in Figure 3b should express the fluctuation "noise" level which determines the threshold excitation of visual sensation. The constant is determined as the value at which the minimum perceptible brightness difference $\Delta E_r/E_r$ becomes unity and, according to Figure 3b, has the value of $S_n \simeq 3$. Subtraction of $S_n = 3$ furnishes then the actual transfer curve (indicated for red light by the broken line in Figure 3b) with $S = 0$ when $\Delta E_r/E_r = 1$.

This correction has not been made on the curves because no absolute validity is claimed for the particular characteristic, although both the value $S_n = 3$ as well as the corresponding threshold level E_{r_o} for color vision are of the correct order of magnitude. The relative position of the characteristics for light of different color in Figure 3a is thus obtained by a replot of the corrected auxiliary curves of Figure 3b.

The center sections of the curves of Figure 3a have a substantially constant slope $\Delta S/\Delta E_r = 1/0.018E_r$. If the sensation unit itself is considered a log unit, the slope or "transfer factor" (g_c) of the logarithmic characteristic may be defined as

$$g_c = \Delta S/10\ (\log E_r - \log\ (E_r + \Delta E_r)) \tag{4}$$

The value of g_c decreases for Figure 3a slowly from 12.8 to zero in the long "toe" of the curves.

3. *Operating Characteristics*

The brightness range which can be seen in one object field is a function of the ratio of the peak to average brightness $\hat{B}/\overline{B}$. This ratio approaches unity for small objects in a large white field. As a test object, a logarithmic step tablet (10×2 centimeters approximately) may be placed over a white 8×10-inch field (kinescope raster) for viewing at a distance of 32 inches ($\rho = 4$). When the field surrounding the tablet is covered with black paper, the other extreme $\hat{B}/\overline{B} \simeq 100$ is obtained.

If an internal eye-reflection factor of 1 percent is assumed and other effects neglected, the zero-sensation level S_o will occur at $E_o = 0.01\bar{E}_r$ and the operating curve (E'_r) may be constructed from the static curve by subtracting E_o from E_r $(E'_r = E_r - 0.01\bar{E}_r)$. See Figure 3a. A 10:1 decrease in transfer factor, i.e., when $g_c \simeq 1.3$, at the dark end of the light range may be considered as a *practical* contrast visibility limit corresponding to a black level raised above S_o by approximately 6 sensation units.

For a peak brightness of $\hat{B}$ equal to 30 foot-lamberts, the limits $\hat{B}/\bar{B}$ equal to either 1 or 100 give $\bar{B}_1$ equal to 30 foot-lamberts ($\bar{E}_r = 880$ Td) and $\bar{B}_2$ equal to 0.3 foot lamberts ($\bar{E}_r = 32$ Td), respectively. The corresponding operating curves are shown in Figure 3a and cover the ranges 350:1 for the test with white background and 10,000:1 for the test with dark background.

These ranges agree substantially with those observed in a dark room with the step-tablet test. The compression of tone values in the lower third of the ranges below $g_r \simeq 1.3$ is quite evident and observation indicates that this section may be combined into one level. The essential contrast range varies from 100 to 1000 depending upon the size and distribution of the light and dark areas.

It is difficult to satisfy this remarkable capability of the eye. Fortunately, there are few really black objects (deep cavities) in normal scenes. White snow with nearly 100 percent reflection factor and black velvet with 1 percent reflection represent, in general, the extremes in range, excluding specular reflections. It must, however, be considered that even larger differences in object brightness can occur when the illumination differs greatly in parts of the scene (light and shadows), although each part may have a range considerably less than 100 to 1. An image brightness range of *100* to 1 can probably be considered for most cases as an adequate standard.

4. *The Over-all Transfer Characteristic of Television Systems*

Because the operating characteristic of the eye is identical when viewing object or image, any imaging process capable of reproducing an object with natural brightness must have a linear transfer characteristic for *truthful* reproduction of tone values. The transfer characteristic will be linear also for an image of reduced brightness so long as both object and image cause the eye to operate with similar characteristics. It is evident from Figure 3a that an object-contrast range of 100 to 1 with $B/\bar{B} = 5$ and a peak illumination $\hat{E}_r$ at 1000 Td up to 10,000 Td will meet these conditions. Objects can thus be reproduced by a linear system for all values of object or image peak brightness in the corresponding ranges of $\hat{B} = 20$ to 1000 foot-lamberts. It is to be understood that this statement applies only to a *true* reproduction of light ratios over a 100 to 1 range. Specular reflections or highlights exceeding this range have to be compressed for optical or electrical reasons. *It may, however, be very desirable to expand or contract the transfer characteristic in sections of the operating range to create artistic effects* in the image which may give a more pleasing illusion than a true reproduction of object tone values would give.

The peak brightness of modern kinescopes permits operating with $\hat{B} \geqq 20$ foot-lamberts. Studio monitors should, therefore, be operated at similar brightness levels or correction of their transfer curve is necessary because uncorrected low-level operation in a dark monitor room with $\hat{B} \simeq 5$ foot-lamberts is not exactly comparable due to the increased black compression by the eye at the lower brightness value. This effect, however, may be considered small when compared to changes in the transfer characteristic of the over-all system which may be introduced by a non-linear characteristic of one or several system components such as the expansion of brightness values in normal kinescopes or a dark-range compression in black-level setters or amplifiers. These deviations should be compensated for by an inverse transfer characteristic in the (electrical) system *if* a substantially linear over-all relation between object and image brightness is desired. The degree and ratio of the precompression of signal amplitudes at the transmitter are largely dependent on the over-all expansion in the transfer characteristic of the electro-optical receiver and a factor governed by receiver "noise" conditions. They are not entirely dependent on gradation requirements. However, it should again be stated that additional compression or expansion (not necessarily logarithmic) of the over-all system response may be desired for intentional changes of the brightness scale.[14,15]

5. *Color Response*

Reproduction of colored objects as black-and-white images with a natural brightness scale requires that the spectral response of the photo-sensitive surface in pickup cameras be similar to that of the eye (Figure 5). Larger deviations from the eye curve (often in the blue, red, and infrared regions) must be corrected by filters or by adjustment of color temperature in illumination. For natural appearance of tone values in artificially illuminated scenes, the product of the spectral characteristics of light source, photosensitive surface, and matching

---

[14] D. G. Fink, "Brightness Distortion in Television", *Proc. I.R.E.*, Vol. 29, pp. 310-321, June, 1941.

[15] W. Mortensen, ON THE NEGATIVE, Simon & Schuster, Inc., New York, N. Y., 1940.

filter (if required) should equal the product of the luminous sensation curve of the eye (Figure 5) and light of normal color temperatures, i.e., the temperature of light which would be used on the object for direct viewing by the eye. (A point-by-point match of the frequency characteristic is required to determine normal color temperature.)

The reproduction of objects by an image in natural colors requires information on the frequency of light within the range of visible energy covered by the luminosity curve. According to the trichromatic theory, the eye analyzes the colors of the spectrum by a triple receiving mechanism, each covering widely overlapping sections of the frequency range with a different frequency response. The frequencies of maximum color stimulation are red, green, and blue light. The brain weighs these response curves; equal energy stimulation of each of the primary sensations is analyzed as white light. The corresponding stimulation or mixture curves are shown in Figure 5. The eye does not discriminate between a monochromatic color and one resulting from a com-

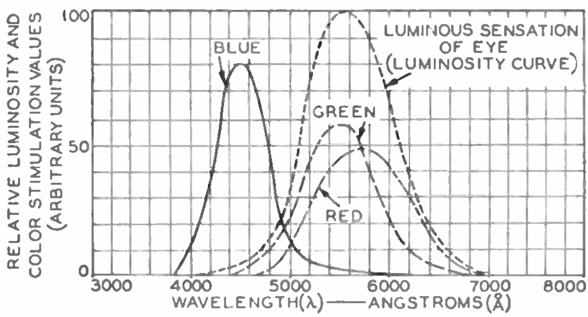


Fig. 5—Mixture curves.

bination of frequencies or frequency bands causing the same relative stimulation of the three primary color sensations. For this reason all colors can be synthesized, theoretically, by mixing suitable amounts (including negative amounts) of light from the three principal regions of the spectrum (monochrome or bands).

Color-reproduction processes based on only three primary colors are not perfect because they require the existence of hypothetical negative values. The values required decrease rapidly toward zero as the number of saturated primary color bands is increased. Much theoretical and experimental work has been done in color printing and photographic processes to determine the best practical trichromatic set of color filters for analysis and synthesis. One of these sets is the Wratten tricolor filter series A#25 red, B#58 green, and C5#47 blue for use with daylight and panchromatic (Type B) film. Deviations in the spectral characteristics of light sources or photosensitive surfaces require amplitude and/or frequency-response correction to obtain a similar over-all color response. At the camera, it is generally desirable that white light should cause signals of equal amplitude from each primary color band; at the reproducer, equal signals should again produce color intensities combining to white light. The over-all brightness response for each component should again be linear with modifications as stated for black-and-white image reproduction.

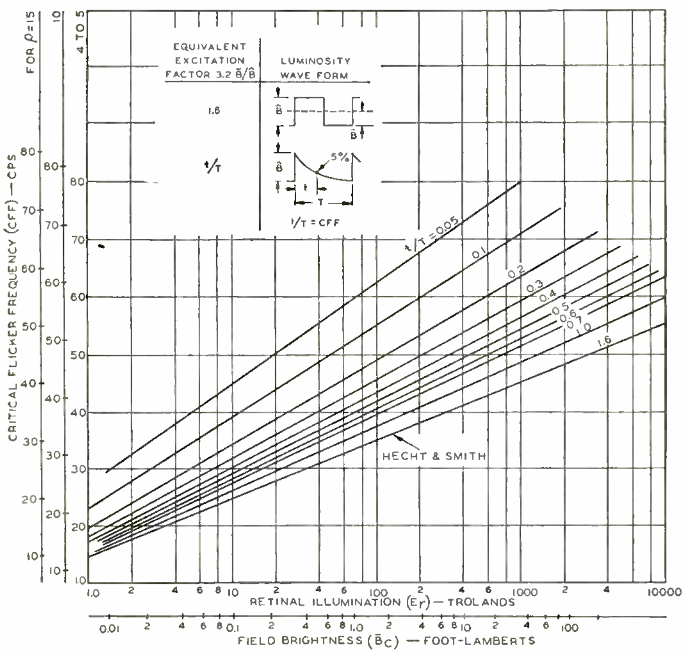


Fig. 6—Threshold flicker values for intermittent illumination.

6. *Persistence of Vision and Flicker*

The response of the eye to light, though rapid, is not instantaneous. Cyclic fluctuations in light intensity cause the sensations of flicker. This sensation disappears at low light intensities (low g_c) and when the fluctuation frequency is increased beyond certain values.

The critical frequency (C.F.F.) at which flicker is just noticeable in large areas depends on viewing ratio (ρ), luminosity, amplitude, and wave shape of the brightness fluctuation. The curve family given in Figure 6 has been constructed for exponentially decaying impulse waves of white light from the known curve for a square-wave brightness fluctuation. Assuming that the critical peak brightness $\hat{B}_c$ at any particular frequency is substantially constant for waveforms with $\hat{B}_c/\overline{B}_c > 3$, it follows that the critical average brightness $\overline{B}_c$ in an area may be expected to decrease in proportion to the excitation time of the eye (decay time of kinescope phosphor) during one cycle.

This proportionality is modified by the decrease in g_c toward low light values which require intersection of the curves near $S = 0$. Convergence and spacing of the curves are somewhat uncertain according to different investigators. The general distribution was taken from Engstrom.[16] The slope of the square-wave fluctuation curve, however, seems to be well established. For a 10-to-1 change of B_c, the change in C.F.F. is 10 cycles.

The impulse excitation caused by a rapidly moving light spot of high intensity scanning a television screen is a rather special case of integration. It may be expected to follow the same general trend and will be discussed in Part II.

The flicker sensitivity of the eye to intermittent white and colored light from a steady source is constant for a given luminosity[12] at normal brightness levels. This statement does not apply to light sources in which the impulse wave shape and duration are functions of color or wave length. Evaluation of B_c by direct measurement is thus indicated for kinescopes in general and especially when composite screen materials are used because simulated conditions with optical projectors are not equivalent.

7. *The Resolving Power of the Eye*

The interference of light waves sets a limit to the resolving power of an optical system, because *two* points cannot be resolved when forming an angle (α_o) with the eye smaller than given by

$$\sin \alpha_o = 1.22 \lambda / \delta \tag{5}$$

For a pupil diameter (δ) of 0.3 centimeters and the wave length (λ) of 0.000055 centimeters, this angle is 0.77 minutes.

At a focal distance (F) of 15 millimeters, the two points are imaged 0.0033 millimeters apart on the retina which corresponds approximately to the diameter of one photoreceptor at the fovea centrals (cone diameter $\approx$ 0.003 millimeter). For viewing ratio, ρ, of 4, this diameter limits eye resolution N_e to 2200 television lines for the optical system of the eye alone and to 1500 lines approximately if the cone structure is considered also. Effects of aberration, diffusion, and

[16] E. W. Engstrom, "A Study of Television Image Characteristics", *Proc. I.R.E.*, Vol. 23, pp. 295-310, April, 1935.

fluctuation phenomena (noise) in the visual process, however, are neglected.

The over-all "frequency response characteristic" of the visual process for small optical signals, i.e., its ability to translate optical detail into sensory response as a function of detail area or line number (N), remains to be determined by subjective measurements. At very low light levels this function is expected to be controlled by fluctuation phenomena in the process of transducing signals to the brain; at normal light levels and for small angles it should follow the law of optical "aperture effects."

8. *Detail Response Factor* (r_e) *and Response Characteristics of the Eye*

The "aperture" response* for the optical system of the eye cannot be determined separately from its transfer effects because response measurements include the transducing process which culminates at the brain in a light sensation. Various occurrences in this process alter the over-all detail response characteristic as in sensitive television pickup transducer systems in which saturation effects, leakage, and interactions between mosaic areas may outweigh the optical "aperture" response.

A representative response characteristic $r_e = f(N)$ of the "eye" for television conditions can be obtained by measuring the brightness difference and contrast required for threshold visibility of detail. This method is analogous to the variable-input/constant-output method employed for determining the frequency characteristics of electrical networks. The operating point (light bias) on the transfer characteristic (Figure 3a) should remain substantially constant. This requirement calls for a constant** average brightness $\overline{B}$ of representative value ($\overline{B} \simeq 7$ foot-lamberts), a fixed viewing ratio and field size, and a test object with calibrated detail size (line wedge) and adjustable contrast. A simple optical test setup includes a white screen with a 4:3 aspect ratio illuminated by a fixed source of light to an average brightness $\overline{B}$. A vertical line wedge covering about 15 per cent of the picture area on a transparent slide is "faded" in optically by projection. The projector brightness at which the first outlines appear has a value $\triangle B_o$; finer detail, i.e., higher line numbers (N) become successively distinguishable at increased brightness values $\triangle B_N$. The measurement should be made with white lines on a dark slide and dark lines on a white slide. The latter test requires correction of $\overline{B}$ at high values of projector

* See INTRODUCTION and also its specific treatment in Part II.
** Not very critical as long as g_e remains constant.

brightness. (Care must be taken to obtain uniform field illumination from either light source). The increments ΔB_N are found to be small signals with values $\Delta B_N < 0.1\,\bar{B}$ up to $N \simeq 400$. The absolute value is determined at a higher line number at which $\Delta B = \bar{B}_2 - \bar{B}_1$ is easily measured.

The measurement has also been made under actual television conditions by viewing a white kinescope screen ($\bar{B} = 7$ foot-lamberts) and electrically fading in the test-wedge image by video signal control. The system used 525-line interlaced scanning, a 20-megacycle video channel with signal correction adjusted to give a substantially constant wedge sharpness (horizontally) on a 12-inch high-definition kinescope

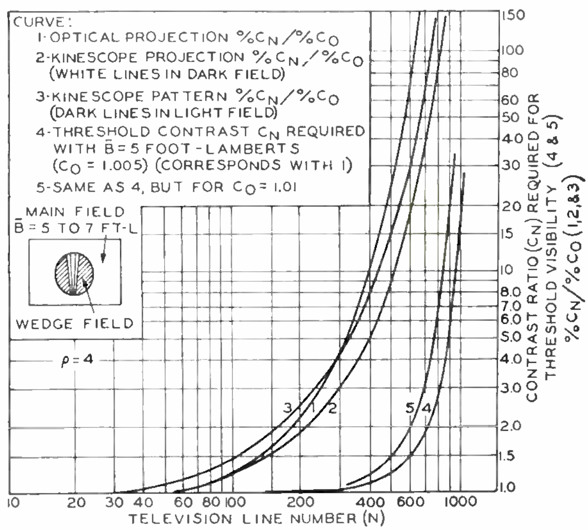


Fig. 7—Object contrast and ratio required for threshold vision.

up to approximately 800 lines. In this case the small signal ratio $\Delta B_N / \Delta B_o$ is measured by the video signal ratio. Positive and negative wedge images are easily obtained by reversing the signal polarity.

The results of a number of measurements are plotted in Figure 7 in terms of the per cent contrast ratio $\%C_N/\%C_o$ which shows the required optical input signal ratio for a constant minimum sensation output as a function of the television line number N. The object contrast ratio C_N required for threshold visibility of N is shown also.

The pickup response factor of the eye (r_e) as a function of N is expressed by

$$r_e = \Delta S_N / \Delta S_o \tag{6}$$

Because of the logarithmetic transfer characteristic (Equation (4))

$$r_e = g_{e(o)} \left[\log (\bar{B} + \triangle B_o) - \log \bar{B}\right] / g_{e(N)} \left[\log (\bar{B} + \triangle B_N) - \log \bar{B}\right] \quad (7)$$

which is equivalent to $r_e = g_{e(o)} \log C_o / g_{e(N)} \log C_N$ (7a)

For moderate signals and constant average brightness, the transfer factor is constant. With $g_{e(o)} = g_{e(N)}$ and for particular values $\bar{B}$

$$\left.\begin{aligned} r_e &= \log C_o / \log C_N \\ r_e &= (\log (\bar{B} + \triangle B_o) - \log \bar{B}) / (\log (\bar{B} + \triangle B_N) - \log \bar{B}) \end{aligned}\right\} \quad (8)$$

For small signals $\triangle B < 0.1\,\bar{B}$ the small section of the transfer characteristic approaches linearity. It can be shown that the following ap-

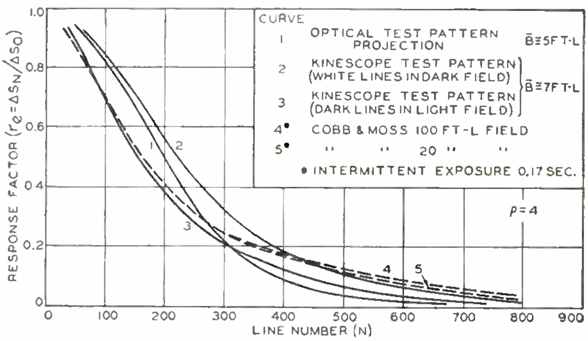


Fig. 8—Detail response factor of the eye.

proximations hold with a maximum error of less than 10 per cent.

$$\text{For } \triangle B < 0.1\,\bar{B} \quad \left\{\begin{aligned} & C \simeq 1 + (\triangle B / \bar{B}) \text{ and } \%C \simeq 100\,\triangle B / \bar{B} \\ & \log C_o / \log C_N \simeq \triangle B_o / \triangle B_N \\ & \text{hence:} \quad r_e \simeq \triangle B_o / \triangle B_N \end{aligned}\right. \quad (9)$$

The simple relation Equation (9) is, therefore, useful for $N < 400$ in determining the pickup response factor r_e; higher line numbers require computations of r_e with Equation (8).

The sharper cutoff obtained with the optical method (See Figure 8) is probably caused by limited detail contrast of the 35-millimeter film and projector used in the test.

The curves 1 and 2 in Figure 8 indicate an "effective" resolving aperture in the over-all pickup system of the eye equivalent to $N \simeq 200$

lines at $\rho = 4$ with a limiting resolution (1 per cent contrast) near 900 television lines. ($\alpha_o \approx 1$ minute for $\rho = 4$ and $\overline{B} \approx 7$ foot-lamberts).

Curves 4 and 5 in Figure 8 have been plotted from data given in a paper by Cobb and Moss.[17] These curves show the same general position even though they were measured with a short subject exposure of 0.17 second and a different test object. At very high values of brightness, the response characteristic r_e is displaced toward the cutoff limit of the optical system ($N_{co} \approx 1500$). As brightness decreases, r_e shifts towards lower line numbers and at low values of brightness it becomes very poor because of fluctuation noise. It is further affected by superposition of reflected light from other parts of the viewing field. All these effects have their counterparts in photography and in television pickup tubes in which a small light bias may improve detail signals, but a strong light bias (glare) may cause saturation effects in the signal development process and the reduction of detail signals.

The significance of the response characteristic r_e may be illustrated by a representative operating condition on the transfer characteristic of the eye, Figure 3a. If an average brightness $\overline{B}$ of 4 foot-lamberts and a peak brightness $\hat{B}$ of 20 foot-lamberts is assumed, the operating section in Figure 3a extends from $\hat{E}_r = 1000$ Td downwards. Subtracting 1 per cent (2 Td) of the average level for reflected light inside the eye, we obtain a total of approximately 200 sensation units in a 100-to-1 brightness range from 10 to 1000 Td. The size of the sensation unit decreases as a function of line number as indicated by the response factor and is

$$S_N = S_o\, r_e \tag{10}$$

For 500-line detail, the sensation scale comprises only $200 \times r_e = 18$ steps and shrinks to 6 units at $N = 700$, 4 units at $N = 800$, and 2 units at $N = 900$ lines. If the original unit size for $N = 0$ is correct, a difference of one single unit $S_o\, r_e$ should be perceptible upon careful observation at the average level (200 Td). The number of simultaneously observable gradation steps over the entire range, however, is probably smaller and the essential and easily visible number of steps is even considerably smaller.

Brightness ranges of 100 to 1 are rather exceptional in optical reproductions. Prints and theatre motion pictures seldom exceed a large area contrast ratio of 30 to 1. The contrast of *fine detail* areas is further substantially reduced by "aperture" and diffusion effects.

It may be concluded that an imaging system having a substantially flat over-all response up to $N = 500$ will furnish an image of excellent sharpness for a viewing ratio $\rho = 4$ because it cuts off only a small

[17] P. W. Cobb and F. K. Moss, "The Four Variables of the Visual Threshold", *Jour. Frank. Inst.*, Vol. 205, pp. 831-847, June, 1928.

percentage of object detail which is visible to the eye only at stationary contours of high contrast.*

Equation (1) indicates an electrical channel (Δf) of 6.35 megacycles for a frame time (T_f) of 1/30 of a second and a balanced line number ($\overline{N}_{co}$) of 500.

Tests were conducted with many observers who viewed both live-pickup and still pictures over a high-quality variable-channel television system with a maximum band width of 20 megacycles. These tests show that up to $\overline{N} = 800$ lines slight increases in sharpness can be detected by the eye at a ρ of 4 on fine detailed stationary subjects of high contrast such as small type, *but little preference was indicated for increases beyond* $\overline{N} = 500$ *when viewing motion scenes, persons, and other normal television subjects.**

It is of interest to note that satisfactory half-tone prints in magazines employ rasters producing 85 to 133 points per inch. Gradations from black to white are obtained by variation of the point size from small white points in a black field (dark) to alternate black-and-white squares for 50 percent white, and small black dots for near white. If both black-and-white points in a 50 percent tone are counted, the prints have 170 to 266 television lines per inch and for a ρ of 4 at the normal viewing distance of 12 inches, they correspond to a 3×4-inch television raster with 510 to 800 lines. The points in the 800-line print cannot be resolved with the naked eye while a 600-line print shows a barely visible structure.

Color prints employ, side by side or superimposed, the same number of points for each of the printing colors (3 and higher) which results in 3 or more times as many picture elements as used for black and white.

The television line process gives continuous tone variations in the horizontal direction without structure, while in the vertical direction a fine spot size** with a diameter of ½ the line width is desirable for high resolving power and will give a 1000-line structure (counting again spaces and lines) with 500 scanning lines. The present television raster with 525 scanning lines is, therefore, in comparison with good-quality printing standards quite satisfactory.

(B) Brightness Distortions (Steady and Fluctuating)

When optical images are viewed, brightness values are compared

* The subject of image sharpness with respect to line number and resolving power in different processes will be treated on various occasions as it can be approached in many ways. A more precise evaluation will be given in the last part of this paper when the over-all "aperture response" of cascaded transducing systems is treated.

** See Part II.

with mental pictures of real objects. Eye and mind detect rather quickly an error in brightness if it is of sufficient magnitude and is inconsistent with the illusion caused by the image. Transfer factor (g_e) and field brightness fluctuations (flicker) have been discussed. This section deals with steady and fluctuating brightness distortions in limited areas.

1. *Steady Deviations*

The steady deviation ($\triangle S$) permissible from normal sensation values, as readily seen from Figures 3a and 8, is a function of area, the brightness distribution or "shape" of the variation ($\triangle B$), its color, and the average brightness level $\bar{B}$. That the eye will tolerate a sinusoidal or other variation with a gradual gradient change ($\triangle S$) of 10 to 15 sensation units ($\triangle B = 20$ to 30 per cent at $g_e = 12.8$) when extending over the entire viewing field, especially when some detail is present is exemplified by the variations of field brightness found in commercial projectors. For smaller areas in the order of $N = 10$ to 50, the deviation should not exceed 1 to 2.5 sensation units ($\triangle B \simeq 2$ to 5 per cent at $g_e = 12.8$) with gradual distribution.

Areas with fine detail such as spots and scratches are of lower visibility because of attenuation due to the low-pass filter characteristics (Figure 8) of the eye. The deviation caused by such areas should remain below $1/r_e$, where r_e is the detail response factor. For the eye transfer factor (g_e) of 12.8, this brightness deviation is

$$\triangle B \simeq 0.018\, B / r_e \qquad (11)$$

The above values apply to black-and-white images or white light. Images in natural colors have substantially the same tolerances for non-uniformity of color brightness, but are affected, in addition, by non-uniform color response. The eye is equally sensitive to deviations in color saturation or mixture, but they are seldom detectable in black-and-white images.

2. *Random Fluctuations and Visibility Factors (Grain and "Noise")*

A statistical amplitude-distribution sample of random occurrences or impulses is shown in Figure 9. Electrical fluctuations of this type[18] are caused by current fluctuations termed "noise" in elements of the electrical channel, especially in the pickup tube and in first amplifier stages. It is well known that the frequency components and power of the complex fluctuation wave are uniformly distributed in pass bands with constant amplitude response. Fluctuation peaks occur when all components are in phase and the peak duration is determination by the highest-frequency component in the pass band. These random impulses cause, therefore, brightness fluctuations on the viewing screen appearing as a moving grain structure of picture element size.

Picture element and apparent grain area are inversely proportional to the balanced line number $(\overline{N}_{co})^2$ in a television channel. The "grain", however, is seen through low-pass filters, i.e., the eye and kinescope, which attenuate the high-frequency components in the visible fluctuation wave. It is generally accepted[18] that fluctuations within any one small section $\triangle f'$ of a very wide frequency band occur at a rate proportional to the frequency and within the same amplitude limits. Oscillograms of fluctuations in any one section of the frequency spectrum are identical when taken with a writing speed proportional to the section frequency.

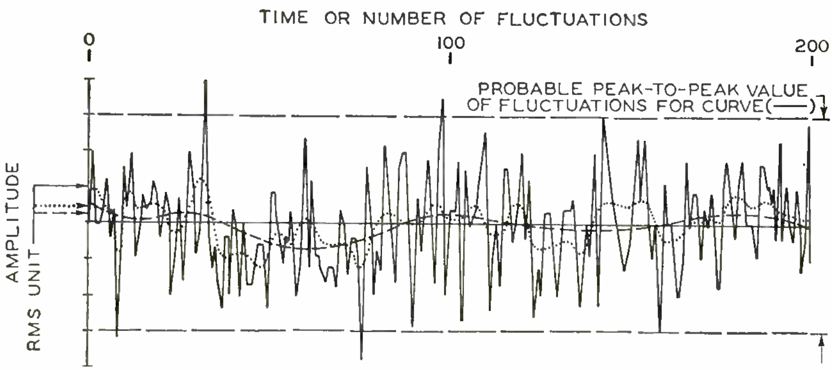


Fig. 9—Sample plot of random fluctuations.

Power and voltage developed by a fluctuating current in a constant impedance have, therefore, constant values for a given increment $\triangle f'$, independent of the value f. The incremental power $\triangle P$ can be written

$$\triangle P = K \triangle f' Y^2 = \bar{a}^2 P_o \tag{12}$$

[18] a. B. J. Thompson, D. C. North and W. A. Harris, "Fluctuations in Space-Charge-Limited Currents at Moderately High Frequencies", *RCA REVIEW*, Vol. IV, No. 3, pp. 269-285, January, 1940.

b. D. O. North, "Diodes and Negative Grid Triodes", *RCA REVIEW*, Vol. IV, No. 4, pp. 441-472, April, 1940.
Vol. V, No. 1, pp. 106-124, July, 1940.

c. D. O. North, "Multi-Collectors", *RCA REVIEW*, Vol. V, No. 2, pp. 244-260, October, 1940.

d. B. J. Thompson and D. O. North, "Fluctuation Caused by Collision Ionization", *RCA REVIEW*, Vol. V, No. 3, pp. 371-388, January, 1941.

e. W. A. Harris, "Fluctuations in Vacuum Tube Amplifiers", *RCA REVIEW*, Vol. V, No. 4, pp. 505-524, April, 1941.
Vol. VI, No. 9, pp. 114-124, July, 1941.

where K is a constant, Y a significant amplitude or a root-mean-square value of the fluctuation wave in the frequency limits $\Delta f'$, and P_o the normal power value. The factor $\bar{a}$ in the second form is an amplitude or gain coefficient specifying the mean deviation from a normal value Y when the frequency is varied. The factor $\bar{a}$ applies only to the frequency characteristic of system components located between the point of fluctuation insertion and the point of observation.

The total fluctuation power (P_X) in a band width extending from f_a to f_c is then:

$$P_X = \int_{f_a}^{f_c} P_o \bar{a}^2 df \qquad (13)$$

which may be evaluated as the sum of incremental powers:

$$P_X = \Sigma P_1\bar{a}_1^2 + P_2\bar{a}_2^2 + \cdots + P_n\bar{a}_n^2 \qquad (13a)$$

(a) System components with flat frequency response have a constant amplitude or attenuation factor a. It follows from (13) that

$$P_X = K(f_c - f_a) = K\,\Delta f \qquad (14)$$

The fluctuation wave amplitude is, therefore, $Y_X = K'\,\Delta f^{1/2}$ (15)

(b) System components with a frequency response proportional to frequency have amplitude factors $\bar{a}_1$, $\bar{a}_2$, $\bar{a}_n$ which increase in proportion to frequency. Equation (13) furnishes the relations

$$P'_X = K\,\Delta f^3 \qquad (16)$$

and

$$Y'_X = K'\,\Delta f^{3/2} \qquad (17)$$

(c) System components with non-uniform frequency response will change the power distribution from the "normal" values P_X or P'_X given in Equations (14 and 16). The modified power distribution is obtained by subdividing the frequency band into equal increments $\Delta f'$ and multiplying the normal power increments ($P_n = \Delta P_X$) by the corresponding mean deviation factors squared ($\bar{a}_n^2$) (See Figure 10). Summation of the products $P_n\bar{a}_n^2$ (Equation 13a) furnishes the total power, the square root of which is the fluctuation voltage for comparison with the normal voltage.

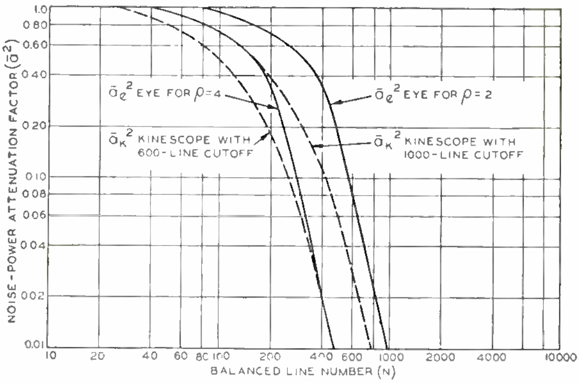


Fig. 10—Noise attenuation due to filter effect of eye or kinescope.

"Fluctuation filter factors" (m) may be determined by which normal small fluctuation values (brightness, voltage, or current) are multiplied to specify the effect of electrical or optical low-pass filters such as kinescope and eye. The filter factor (m) is defined as the square root of the ratio of the modified noise power to the normal noise power.

Filter factors have been computed as a function of the balanced resolution number for various combinations of kinescope and eye with the attenuation factors $\bar{a}^2$ as shown in Figure 11. Equivalent electrical channel widths are given by Equation (1). A reasonable unbalance[5] of horizontal and vertical resolution does not materially affect the result.

The relative filter effects of eye and kinescope on fluctuations of small amplitude from "flat" and "peaked" channels are shown by Figure 12. Curve 1 in the lower group shows the normal proportion of the fluctuation or "noise" amplitude to the balanced-resolution line

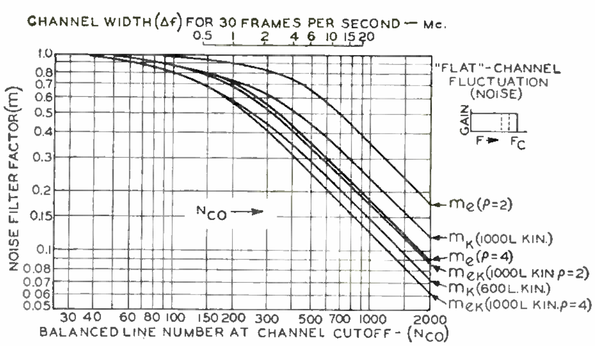


Fig. 11a—Effect of channel width on fluctuation filter factors for a flat channel.

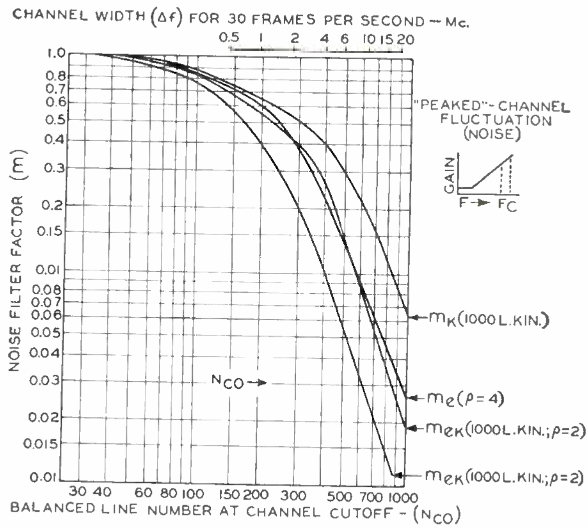


Fig. 11b—Effect of channel width on fluctuation filter factors for a peaked channel.

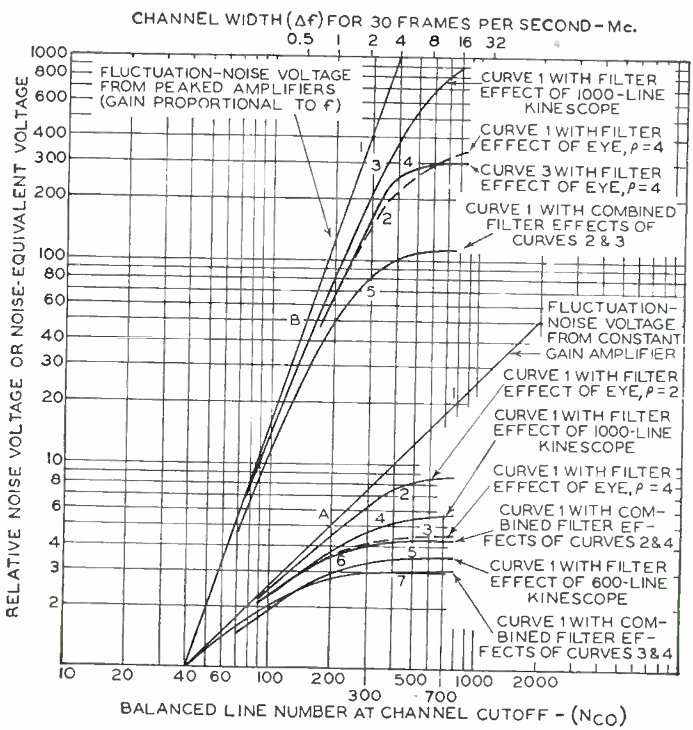


Fig. 12—Integrated filter effects of eye and kinescope on fluctuations of small amplitude.

number at cutoff for constant gain channels; curves 2 to 7 show the equivalent noise amplitude after attenuation of high-frequency components by the low-pass filter action of the eye, kinescope, or kinescope and eye.

The upper group has been computed similarly for fluctuation noise which has previously passed through electrical channels with an amplitude response proportional to frequency, such as used in camera amplifiers with capacitive input impedance. This "peaked" noise increases in proportion to the 3rd power of the balanced resolution ($\overline{N}_{co}$) in the channel (curve 1) and is attenuated by eye, or by eye and kinescope, as shown by curves 2 and 5, respectively.

Figure 12 shows that the filtered fluctuation amplitude, i.e., the visibility of small fluctuations for a viewing ratio of 4, approaches a constant value for channels passing more than 400 lines. This condition is confirmed by observations of noise visibility in television channels over 400 lines, indicating again that higher resolution contributes little to image detail. It is further seen (Figure 11) that the relative attenuation and *visibility of noise in peaked and flat channels differ by a factor of three for the present channel width of 4.25 megacycles.*

The ratio of peak values to root-mean-square values is expected to remain constant for any section of a wide frequency band. This ratio has been checked for flat-channel fluctuations by measurement of root-mean-square voltage and the (representative) peak-to-peak deflection voltage on an oscillograph (20-megacycle flat response) with bandwidths up to 20 megacycles. The peak-to-peak voltage measured normally is fairly well defined because theory and observation indicate a rapid decrease in the occurrence of peaks exceeding a certain level.

The ratio of peak-to-peak values to root-mean-square values is 6 to 1 for practical measurements as observed for channels with a band width of 4 to 20 megacycles. This value is indicated on the statistical amplitude distribution sample (Figure 9).

3. *Grain-Visibility Constant and Signal-to-Noise Ratios*

Observations indicate that the threshold sensitivity of the eye to low-frequency components in random brightness fluctuations is equivalent to approximately one sensation unit or a 2 per cent brightness change (peak value) at normal field brightness values ($g_c = 12.8$ in Figure 3a).

The visibility of complex brightness fluctuations such as caused by motion-picture-film grain or electrical fluctuations seen through a linear transducer (ideal kinescope) should be obtainable by multiplication with appropriate filter factors (m) because effective area and amplitude of the fluctuation peaks depend on the combined effect of *all* components. The optical signal-to-noise ratio *at the source of the image light flux* for threshold visibility of grain or noise fluctuation at normal operating points on the eye characteristic ($g_e = 12.8$, $\overline{B} > 2$ foot-lamberts) is therefore:

a) with respect to the peak-to-peak fluctuation value

$$\hat{R}\phi = \overline{B} / \triangle B_{N(p\,p)} = m/0.04 = 25m \tag{18}$$

b) with respect to the root-mean-square fluctuation value

$$|R|\phi = 6\hat{R}\phi = 150m \tag{19}$$

The significance of Equations 18 and 19 will be illustrated by practical cases.

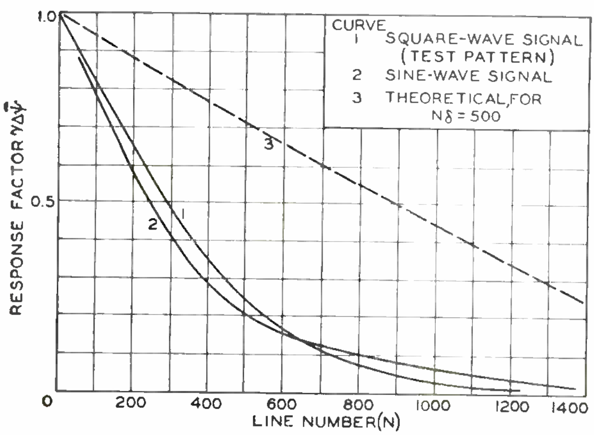


Fig. 13—Equivalent square-wave aperture response of "1000-line" kinescope.

The over-all resolution of an optical motion picture projection* is represented sufficiently well by a kinescope characteristic with 1000-line cutoff (Figure 13) for which filter factors m_K have been computed. Threshold ratios for reproduction over an ideal linear transducer (kinescope) and various flat channels are obtained from Equations 18 and 19 and Figure 11 and are listed in Table I. The threshold ratios of the 1000-line channel indicate values which may be considered representative for a purely optical transmission of the motion picture image. All conditions apply also to television systems with ideal linear transducers. The third column applies to film noise measurements and

---

* This includes deterioration in camera, printing, and projecting processes on normal 35-millimeter film; not on special test reels.

appears to be in substantial agreement with values observed on high-quality 35-mm film.

The ratio of signal to peak-to-peak noise $\hat{R}_{\phi} = \frac{1}{2}$ may well be considered as indicating the grain size of the film itself. If the ratio of signal to peak-to-peak noise is equal to 4.3 (See Table I) in a 1000-line channel, it will decrease to ½ when the channel is increased to $N_c = 8600$ lines. At this bandwidth the vertical frame dimension of 15.7 millimeters can accommodate $N_c = 8600$ television lines or 548 television lines per millimeter. This line number corresponds to 274 elemental light pulses spaced by 274 grains in one millimeter with an

Table I

Channel Resolution N_{co}	Kinescope (representing 35-millimeter film) Filter factor of system m_k (1000)	$\hat{R}_{\phi}$	$\lvert R \rvert_{\phi}$	Eye and Kinescope Filter factor of system m_{ek}	$\hat{R}_{\phi}$	$\lvert R \rvert_{\phi}$	Eye Filter factor m_e	$\hat{R}_{\phi}$	$\lvert R \rvert_{\phi}$
410	0.5	12.5	75	0.29	7.3	44	0.41	10.3	62
500	0.45	11.2	68	0.24	6	36	0.35	8.8	53
800	0.29	7.3	44	0.15	3.8	23	0.25	6.3	38
1000	0.23	5.8	35	0.12	3	18	0.17	4.3	26
	For $\rho < 1$, $m_e = 1$			For $\rho = 4$			For $\rho = 4$ Optical ratio on screen or electrical ratio at grid of linear ideal transducer.		
	Optical ratio on film in motion								

average intensity B (See Equation (18)) equal to ½ the peak-to-peak noise fluctuations $(\triangle B_{N(p\text{-}p)})$. The resolving power of the random grain structure of film is approximately 6 grain diameters or 90 lines per millimeter. (See Part III)

It is significant that grain size and resolving power of 35-millimeter motion picture film are in this order. It is stated further by film manufacturers that a resolving power of 50 lines per millimeter permits enlargement to 10 or more diameters "without objectionable graininess". This resolving power corresponds to 760 or less television lines in 3 inches for which $\rho = 4$ at close viewing distance (12 inches). This figure again agrees with the limiting resolution of the eye.

It should be pointed out that the relation between limiting resolution and grain diameter stated above is not necessarily general.

When sharp signals are superimposed on a fluctuating grain pattern from a separate source as in many television images, recognition of elemental signal pulses is possible with $\hat{R}_\phi \simeq 0.5$ $(|R| = 3)$; line patterns with elemental line width $1/N_c$ are visible with even much lower signal-to-noise ratios as the eye integrates along the line. It is also observed that motion of the ground-glass view plate in the focal plane of a camera allows observation of much finer detail (with a magnifying glass) than when stationary.

The values given by Equations 18 and 19 refer to signal-to-noise ratios at particular brightness levels B_o. They refer to the entire gradation range only (for $g_e = 12.8$) when signal and noise remain proportional.

The noise decreases with signal in film and phototubes, although not in proportion. The random noise currents, however, generated in video amplifiers or by the scanning beam of present storage-type pickup tubes have constant values for a given operating condition. The threshold ratio R_ϕ at a brightness level B_o requires, thus, for a constant noise source and linear transducers the relationship

$$R_{\phi\max} = R_\phi \hat{B}/B_o \tag{20}$$

Threshold signal-to-noise ratios in the video signal channel require specification of the function $B = f(E)$ and the electro-optical transfer factor g_k (usually a variable) between observation points, such as kinescope grid signal and screen brightness.

Therefore: $$\hat{R}_{\max} = \hat{R}_\phi g_{k(o)}\, \hat{E}/B_o \tag{21}$$

and with (18) $$\hat{R}_{\max} = 25 m\, g_{k(o)}\, \hat{E}/B_o \tag{22}$$

where $\hat{E} \doteq$ peak signal voltage required for $\hat{B}$ (i.e. $\hat{B} = f(E)$), and $g_{k(o)} =$ transfer factor (foot-lamberts/volt) at the brightness B_o for which threshold visibility of noise is desired.

Numerical values for the signal-to-noise ratio of high-quality television signals will be given in Part II in which general specifications for television systems and components are discussed.

FREQUENCY STABILIZATION WITH MICROWAVE SPECTRAL LINES*†

BY

W. D. HERSHBERGER AND L. E. NORTON

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

Summary—*Absorption lines of gases at reduced pressure exhibit Q's of 100,000 in the 24,000 megacycle range, and the center frequency is unaffected by pressure and temperature. Stabilization of a K-band klystron has been effected, using the 23,870.1 megacycle line of ammonia contained in a short section of matched waveguide, both at the center frequency of the line itself and at frequencies removed from the line frequency by a controlled intermediate frequency. Indications are that the frequency stability attained compares favorably with that of quartz crystals but with the added advantages that arise from the inherent stability of spectral lines. Applications to a wide range of frequencies in the microwave range are indicated as well as to a clock of high precision.*

THE ABILITY to make increasing use of microwaves hinges in part on the ability to stabilize the frequency of oscillators operating anywhere in the microwave range. The successful operation of relay links or the satisfactory application of beacons in navigational systems which serve large numbers of homing aircraft will depend in large degree on progress in methods of and techniques for frequency control. Methods are needed in the microwave range which are comparable in flexibility and effectiveness to quartz crystal techniques used at lower frequencies. Several methods of stabilization have been described in the literature[1,2] in which a metal cavity resonator is used as a reference standard. Such a method will be effective only if the temperature of the cavity is held constant or if the cavity is made of a material having a low coefficient of expansion. Again a cavity may be sealed to prevent atmospheric detuning, or an absorbing gas may be used to modify the cavity resonance and the combination used as a discriminator in conjunction with a d-c amplifier in a stabilizing

* Decimal Classification: R141.1 × R210.

† Presented at the 1948 I.R.E. National Convention in New York, N. Y., on March 25, 1948.

[1] R. V. Pound, "Electronic Frequency Stabilization of Microwave Oscillators", *Rev. Sci. Inst.*, Vol. 17, No. 11, pp. 490-505, November 1946.

[2] V. C. Rideout, "Automatic Frequency Control of Microwave Oscillators", *Proc. I.R.E.*, Vol. 35, No. 8, pp. 767-771, August 1947.

system.[3] Stability to one part in 50,000 has been reported for one of the cavity methods and to one part in 25,000 for the gas discriminator.

The use of atomic and molecular transitions in an "atomic" clock was suggested by I. I. Rabi in the Richtmyer lecture, given in January, 1945 before a joint meeting of the American Physical Society and the American Association of Physics Teachers, and arose in particular in connection with the method of atomic beams. In the present work, stabilization has been effected both at the mid-frequency of one of the numerous spectral lines that lie in the range of a K-band klystron and at other frequencies offset from line frequency by a controlled and

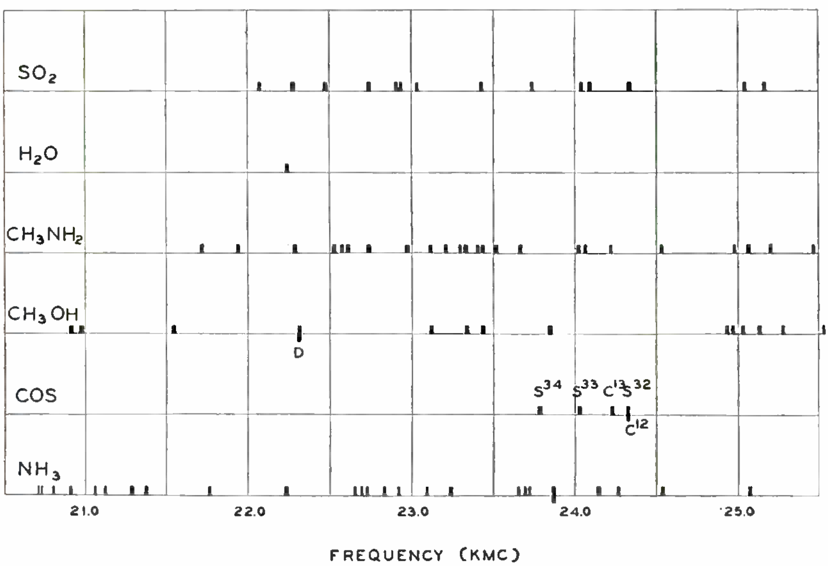


Fig. 1—Absorption frequencies for six gases.

known intermediate frequency. By the use of known techniques of frequency multiplication and addition the method lends itself to stabilization of a wide range of microwave frequencies. Stabilization with the present scheme has been effected to one part in a million but this figure is set, not by any observed instability, but rather by the limitations of present frequency monitoring equipment. While quartz crystal standards with frequencies in the region of 100 kilocycles are available with stability of 1 part in 10 million or better, there is no evidence that this degree of accuracy is retained in the extensive

[3] W. V. Smith, José de Quevedo, R. L. Carter, and W. S. Bennett, "Frequency Stabilization of Microwave Oscillators with Spectrum Lines", *Jour. Appl. Phys.*, Vol. 18, No. 12, pp. 1112-1115, December 1947.

multiplier chain required for instantaneous frequency comparison at K-band frequencies.

Selective Absorption in Gases

The list of gases which exhibit selective absorption in the microwave range is increasing rapidly as the results of new studies are reported. Absorption itself may arise from a variety of causes such as simple molecular rotation of a linear molecule, as carbonyl sulphide, from the tunnel effect as in ammonia, or, for a complex molecule as methylamine, from some combination of pure rotation, the tunnel effect, and hindered rotation about a carbon-to-carbon bond. Figure 1 presents in graphical fashion some of the absorption frequencies of a few substances which display lines that fall in the range of a K-band klystron. The absorption of carbonyl sulphide, methyl alcohol, methylamine, and sulphur dioxide were first reported in earlier studies undertaken at these laboratories.[4, 5] More recent studies[6, 7] have resulted in interesting data showing that the highest frequency indicated for COS is found when the molecule consists of one nucleus of sulphur 32, one of oxygen 16, and one of carbon 12. The absorption frequency becomes lower first when carbon 13, situated quite near the center of gravity of the molecule, is substituted for carbon 12, and then sulphur 33 and 34 in turn substituted for sulphur 32. The weak water absorption line[8] is given as a matter of general interest. The most powerful absorber known—ammonia—also was the first in point of time to be studied.[9] The rotational fine structure of ammonia was first reported by Bleaney

---

[4] J. E. Walter and W. D. Hershberger, "The Absorption of Microwaves by Gases II", *Jour. Appl. Phys.*, Vol. 17, No. 10, pp. 814-822, October 1946. W. D. Hershberger, E. T. Bush, and G. W. Leck, "Thermal and Acoustic Effects Attending the Absorption of Microwaves by Gases", *RCA REVIEW*, Vol. VII, No. 3, pp. 422-431, September 1946.

[5] W. D. Hershberger and J. Turkevich, "Absorption of Methyl Alcohol and Methylamine for 1.25 cm Waves", *Phys. Rev.*, Vol. 71, No. 8, p. 554, April 15, 1947.

[6] T. W. Dakin, W. E. Good, and D. K. Coles, "Bond Distances in OCS from Microwave Absorption Lines", *Phys. Rev.*, Vol. 71, No. 9, pp. 640-641, May 1, 1947.

[7] R. E. Hillger, M. W. P. Strandberg, T. Wentink, and R. L. Kyhl, "The Microwave Absorption Spectrum of Carbonyl Sulphide", *Phys. Rev.*, Vol. 72, No. 2, p. 157, July 15, 1947.

[8] C. H. Townes and F. R. Merritt, "Water Spectrum Near One-Centimeter Wavelength", *Phys. Rev.*, Vol. 70, Nos. 7 and 8, pp. 558-559, October 1 and 15, 1946.

[9] C. E. Cleeton and N. H. Williams, "Electromagnetic Waves of 1.1 cm Wavelength and the Absorption Spectrum of Ammonia", *Phys. Rev.*, Vol. 45, No. 4, pp. 234-237, February 15, 1934.

and Penrose[10] and two accurate determinations of the line frequencies themselves have been published.[11,12] The frequencies of these lines are now known to about one part in a million. The heavy marks indicating spectral lines in Figure 1 are more than one hundred times as wide as the lines themselves on the scale used and there is room for more than 20,000 resolved lines in the frequency range shown. The (3.3) line of ammonia was used in the present work, but any one of a number of other lines could have been used equally well. Its frequency has been measured as 23,870.13 ± 0.02 megacycles.

Molecular spectral lines possess two properties which make them particularly suitable for use in frequency control work and also as frequency standards. The mid-frequency of such an absorption line has the absolute or invariable character possessed by other spectroscopic standards. Thus the wavelength of one of the spectral lines of cadmium is employed as a standard of length and its wavelength is given to eight significant figures. The frequency of the spectral lines used in the present work, in common with that of lines in other regions, is unaffected by either temperature or pressure, over the range of interest. However, frequency may be shifted by application of appropriate electric or magnetic fields. A gas sample in the interior of a waveguide or a resonator may readily be shielded against unwanted electric fields, while the effect of a magnetic field as weak as that of the earth may be neglected. Thus, one may take the frequency of a spectral line just as it is found, or if so inclined one may shift the frequency of a line by the use of a field. In the former case, the frequency of a beacon stabilized at the center of a line, or at a frequency simply related to the line center by frequency multiplication and addition, is the same whether the beacon is placed at the edge of an airfield in New York or placed on a rocket directed to the moon. Moreover, the frequency will be the same when the experiment is repeated in twenty-five years.

The second characteristic of spectral lines which makes them suitable for frequency control of microwaves is the high inherent Q of a line. At a pressure of the order of 10^{-5} atmospheres and at room temperature the Q of an ammonia line at K-band frequencies is approximately 100,000. That is, a line with center frequency of 24,000 megacycles is approximately 250 kilocycles wide at the half-power points on its resonance curve. Line width increases linearly with pres-

---

[10] B. Bleaney and R. P. Penrose, "Ammonia Spectrum in the 1 cm Wavelength Region", *Nature*, Vol. 157, p. 339, March 16, 1946.

[11] M. W. P. Strandberg, R. Kyhl, T. Wentink, and R. E. Hillger, "Inversion Spectrum of Ammonia", *Phys. Rev.*, Vol. 71, No. 5, p. 326, March 1, 1947.

[12] W. E. Good and D. K. Coles, "Microwave Absorption Frequencies of $N^{14}H_3$ and $N^{15}H_3$", *Phys. Rev.*, Vol. 71, No. 6, pp. 383-384, March 15, 1947.

sure over a wide pressure range. The various parameters which determine the absorption coefficient of one of the ammonia lines are shown in the expression for this coefficient. α is the power absorption coefficient for plane waves at the center of the line and is given by

$$\alpha = \frac{2\pi^2 \mu_o^2 f_o^2 K^2 g(J, K) N}{ckT J(J+1) F \Delta f} \tag{1}$$

μ_o is the dipole moment of the gas,

f_o is the resonant frequency,

J and K are quantum numbers identifying the rotational state of the molecule ($J = K = 3$ for the particular line here considered),

$g(J, K) = (2J + 1) \times 4$ when K is divisible by 3,

k is Boltzmann's constant,

T is absolute temperature,

F is the classical rotational partition function,

c is speed of light,

N is the number of molecules per cubic centimeter, and

Δf is line width in cycles.

Since N and Δf both increase linearly with pressure over an extended range, α itself over this range is independent of pressure but Q varies inversely with pressure. Q also varies inversely with the square root of absolute temperature. The important factors which place an upper limit on Q are intermolecular collisions, collisions between molecules and the wall of the gas container, and Doppler effect arising from thermal motions of the molecules. The effect from the first factor may advantageously be reduced until the mean free path in the gas becomes comparable to cavity or guide dimensions. Further reduction in pressure gives rise, not to an increase in Q, but to a reduction in the absorption coefficient. The contribution to line width arising from Doppler effect depends on temperature only and not on pressure. However, the fact that line width does depend both on temperature and pressure, though the mid-frequency of the line is independent of these factors, makes it essential that stabilization be effected at the line center rather than off center at a point of maximum slope on the gas resonance curve. An oscillographic trace of the (3.3) line, which also shows the quadrupole fine structure is shown in Figure 2. The satellite lines are displaced from line center $\pm$ 1.7 megacycles and $\pm$ 2.3 megacycles.

After pressure has been reduced to a value such that optimum Q has been realized without an excessive reduction in the absorption of the gas, the value of the absorption coefficient at room temperature is approximately 3×10^{-4} nepers per centimeter. When such an absorption coefficient obtains, a wave at 24,000 megacycles transmitted through a conventional matched waveguide with inside dimensions 0.430 by 0.180 inch will lose 13 per cent of its initial power on traversing 12 feet of guide, or 7 per cent on traversing 6 feet. Either length is quite adequate for the purpose of the present stabilization method; or a resonant cavity may be used as a convenient gas container, provided due care is taken to avoid intensities great enough to give rise to saturation effects.[13] In the earliest work a 12-foot section of guide

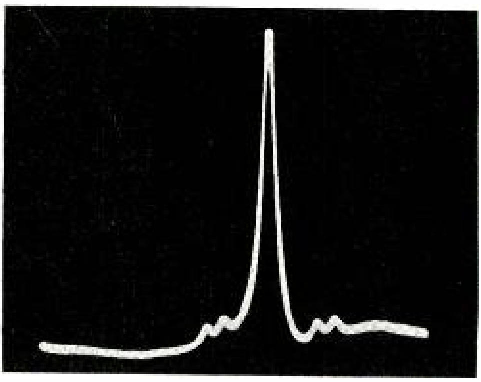


Fig. 2—The (3.3) absorption line of ammonia.

was used. Mica windows 0.002 inch thick are employed to make the system vacuum tight and the pressure used is approximately 0.01 millimeter of mercury.

Method and Apparatus

At the outset one is faced with a choice between several systems of stabilization. One method employs an interaction at the microwave frequency between the high Q absorbing gas and the generator somewhat akin to the interaction between a quartz crystal and its associated amplifier or oscillator. This method is useful in particular when the load on the stabilized oscillator is both fixed and small, and accordingly

[13] C. H. Townes, "The Ammonia Spectrum and Line Shapes Near 1.25-Cm. Wave Length", *Phys. Rev.*, Vol. 70, Nos. 9 and 10, pp. 665-671, November 1 and 15, 1946.

presupposes the existence of buffer and power amplifiers. In view of the present lack of such components, a method has been devised in which the frequency of an oscillator working into normal and variable loads is stabilized.

In this method, the frequency of the low Q generator with poor stability is compared cyclically with the frequency of a high Q resonator—the gas in its container—with high inherent stability. As a result of this frequency comparison an "error" voltage is developed and fed back to the oscillator under control to maintain equality between the gas absorption frequency and a second frequency which is derived from that of the oscillator being stabilized by addition and multiplication. Figure 3 shows in a schematic fashion the opera-

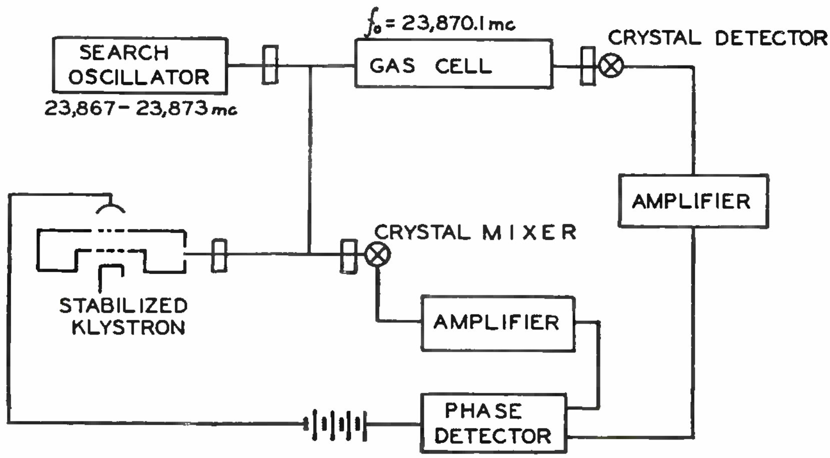


Fig. 3—Block diagram of stabilization system.

tion of the system as a whole. The frequency of a search oscillator is swept through both the gas absorption frequency and through that of the oscillator being stabilized. Power from the search oscillator is fed into two loosely-coupled microwave transmission paths, one containing the gas cell, and the second a crystal mixer also energized by the oscillator under control. The output of the detector which receives power from the search oscillator through the gas section is a series of pulses which follow each other at intervals determined by the sweep repetition rate. One such pulse is shown in Figure 2. After amplification these pulses are led to a phase detector or, more precisely, a coincidence detector. The output of the mixer which receives power both from the search oscillator and the oscillator being stabilized is a variable beat frequency which is amplified. The pass characteristic of the amplifier serves as a frequency discriminator to convert the frequency modulation present in the output of the search oscillator into an amplitude modulation. Thus, the output of the detector which follows the amplifier is a set of pulses similar to those of the first set but with a constant time difference between the two sets. This time difference varies linearily with the difference between the frequency of the stabilized oscillator and the frequency of the gas absorption line. The amplifier for this second set of pulses has a bandwidth slightly less than that of the gas line and it may be either an audio-frequency amplifier or an intermediate frequency amplifier depending on whether stabilization is to be effected at the line frequency itself or at some other frequency offset from line frequency by a known and controlled difference. The output pulses from the second amplifier are also fed to the phase detector. In the phase detector, both sets of pulses are differentiated and limited to ensure operation at the midpoint of the generated pulse to make system performance quite independent of gas pressure and temperature in the one channel and of the amplifier bandwidth in the second channel.

The output of the phase detector is impressed on the reflector of the klystron under control completing the feed back loop, and it now serves to keep the two sets of pulses in time coincidence, which also means the desired equality between the gas frequency and a frequency derived from the klystron frequency. The system in reality is a servo-mechanism, using this designation in a broad sense; in fact, is of that particular variety to which the term *regulator* is properly applied as well as the adjective *sampling*.

Methods of monitoring frequency with high precision in the microwave range are not as yet particularly well developed. In the present work, energy from the stabilized oscillator was coupled very loosely into a second and independent system employing its own sweeping oscillator and ammonia cell. The intermediate frequency, chosen quite arbitrarily, actually used is 30 megacycles, thus stabilizing the klystron frequency at 23,900.1 megacycles. An independent and stable 24-megacycle signal is injected in the mixer in the monitor together with energy from the stabilized klystron to give rise to side bands at $(23{,}900.1 \pm 24)$ megacycles. As a result, in the monitoring system, the narrow pip which results on sweeping through the side band at 23,876.1 megacycles is viewed on an oscilloscope screen in proximity to the trace from the (3.3) line of ammonia arising in the second system. Such a trace is shown in Figure 4. The width of the pip from the stabilized oscillator is approximately 200 kilocycles and is determined by the bandwidth of the amplifier used in the frequency monitor. The steadiness of the trace from the stabilized oscillator with respect to the trace of the ammonia line in the second system is observed and one concludes that the position of this pip is constant with respect to the center of the ammonia line to better than 25 kilocycles. This represents a frequency stability of better than one part in a million. Further refinements in monitoring technique are essential before one can with certainty place an upper limit to the precision with which frequency has been stabilized. The statement that frequency has been stabilized to better than one part in a million means in the present instance that one requires of the servo system, and in particular of its critical element—the phase detector, the ability to keep in time coincidence two wave forms to better than one tenth of the width of the wider of the two wave forms. In this case, this is the wave form which is generated by sweeping the frequency of the search oscillator through a gas absorption line with an inherent Q of 100,000. This is certainly not a stringent requirement.

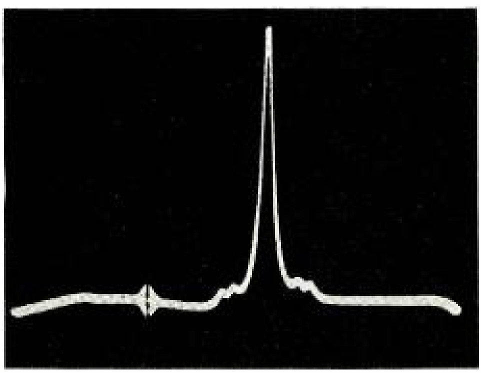


Fig. 4—Oscilloscope trace showing (3.3) line and stable frequency in a monitor.

Discussion

One factor characteristic of the present method of frequency stabilization is that the oscillator under control with its low inherent Q—of the order of 500 for a klystron—may be heavily loaded and moreover, the load need not be fixed. The high inherent Q of the gas spectral line is used for stabilization in conjunction with the low Q of the klystron for power generation and for certain species of modulation. Lest the impression arise that a solution has been found to the problem of "simultaneously eating yet having a piece of cake," it becomes needful to inquire into two closely related topics: (1) the factors which determine and place an upper limit on the search or interrogation rate, and (2) the frequency spectrum of disturbances against which the system is protected—and unprotected.

The upper limit of the search rate is determined by two factors: (1) the frequency range over which search is conducted, and (2) the need for searching slowly enough to permit a high Q microwave component to respond with fidelity during the search cycle. It is of course a technical advantage in many instances to sweep the searching oscillator over a relatively wide frequency range to insure a degree of control over the widest possible variety of operating conditions. Use of a wide sweep minimizes the warm up period of an oscillator, and permits its use with poorly regulated power sources, over a wide ambient temperature range and with a minimum amount of attention. However, use of a wide frequency sweep reduces both the rate at which error information is collected and the quality of the information itself, and eventually, if carried too far, places too severe operating conditions on the phase detector. When it is possible to hold ambient temperature within narrow limits and to use stable power sources, a narrow sweep range is permitted, which makes possible improvements in both the quality and rate of accumulation of error information. Proposed use will largely dictate the number of line widths over which search is made. It is necessary to sweep across the gas resonance curve slowly enough to elicit a response which embodies precise frequency information. The effect of sweep speed and of sweep range has been studied in some detail in a paper which treats of the transmission of frequency modulated signals through selective networks.[14] In particular, effects which arise because one sweeps through a high Q circuit too rapidly are shown in a number of oscillograms. In the present instance, the most selective element in the transmission channel involving the the gas is of course the absorption line itself. If p time constants are spent in sweeping across the gas resonance curve, and search is made over a frequency range of n line widths, while a sawtooth sweep with negligible return time is employed, the "search" frequency is given by

$$f_s = \frac{\pi \, \Delta f}{np} \tag{2}$$

and the "search" range by

$$D = n \, \Delta f \tag{3}$$

where Δf is the width of the line at the frequencies for which the power absorption coefficient attains one-half of its maximum value.

The function of the searching oscillator is to collect accurate error

[14] W. J. Frantz, "The Transmission of a Frequency-Modulated Wave Through a Network," *Proc. I.R.E.*, Vol. 34, No. 3, pp. 114-125, March 1946.

information, rapidly if possible, and then pass it on in suitable form to the reflector of the klystron for correction of frequency. The delay or lag suffered by the information depends upon the speed with which the phase detector assimilates the information and in particular on the filter used in its output circuit. In a numerical case, line width may be 250 kilocycles, while n and p are chosen such that the search frequency is 10 kilocycles. If the output filter of the phase detector has a time constant of 500 microseconds, the system is protected against frequency modulation at frequencies lying in the range from approximately 2 kilocycles down to and including zero frequency. However it is not protected against frequency modulation at frequencies lying between 2 kilocycles and an upper limit set by the Q of the klystron cavity. Performance of this character may be interpreted by those who like to make use of analogies, in terms of the behavior of a low pass filter.[15] From the characteristics of this filter, one may infer immediately the transient response to a step-function, which in this case is a disturbance which tends suddenly to shift frequency by a fixed amount, or its steady state response to frequency modulation at frequencies which may lie either within or without the pass band of the filter.

In some applications, it may be desirable to use a low value of search frequency, such as 10 searches per second, perhaps in conjunction with a wide search range, when the emphasis is on long term stability of the center frequency of an oscillator together with the ability to modulate it, subject of course to the restriction that the sidebands generated shall fall outside the range of the search oscillator. Under these conditions one sacrifices stability for other desired characteristics. In other applications, it is desirable to use the highest permissible value of search frequency, a narrow search range, stable sources of power, together with maximum protection from external disturbances which tend to shift frequency. This procedure is followed when one wishes to realize the maximum possible degree of stability, and will find application in particular to problems involving frequency standards based on the stability of spectral lines.

One attractive feature of the present method of stabilization arises from the ability to stabilize at any chosen frequency in a wide range, rather than only at those specific frequencies for which a strong absorption line happens to exist in nature. The use of a controlled intermediate-frequency amplifier, controlled by a quartz filter if desired, in one of the transmission paths in the system permits some flexibility in choice

[15] L. A. MacCool, FUNDAMENTAL THEORY OF SERVOMECHANISMS, D. Van Nostrand Company, New York, N. Y., 1945 (pp. 14 and 35).

of frequency. Stability requirements on the frequency of this amplifier are not at all stringent, since any uncertainty in its frequency has but a small effect on the sum or difference obtained on combining the absorption frequency of the gas, which is of the order of tens of thousands of megacycles, and the mid-frequency of the amplifier, which may be of the order of megacycles. In addition, the use of frequency multiplication materially extends the utility of the method. For example, if it is desired to stabilize a klystron at 5970.00 megacycles with the (3.3) line of ammonia the following procedure may be used. A crystal detector is used to multiply klystron frequency by four yielding a frequency of 23,880.00 megacycles. This quadrupled frequency is employed in the servo-system and the value of the intermediate frequency to be used is 9.87 megacycles. The output of the phase detector then is fed back to the reflector of the 5970.00 megacycle klystron. In essence, the fourth harmonic of the low frequency klystron is now held constant but with its frequency offset by a fixed amount from the frequency of the ammonia line.

The problem of determining the long term stability of an oscillator whose frequency is controlled by use of a microwave spectral line leads presently to the development of a clock with unique characteristics. Electronic counting circuits at present work effectively from an input frequency of 10 megacycles down to low frequencies where comparison with standards, astronomical in origin, may be made. One approach to the problem of bridging the gap from 24,000 megacycles down to 10 megacycles is to make cascaded use of the scheme of frequency division outlined in the preceding paragraph. That is, a cyclic comparison is made in a second divider between the stabilized 5970.00 megacycle oscillator and, say, the fifth harmonic of a 1194.00 megacycle oscillator whose frequency is subject to control by application of a voltage. A second servo system is used to keep the fifth harmonic of the 1194.00 megacycle in coincidence with the 5970.00 megacycle oscillator whose fourth harmonic was initially stabilized by use of the ammonia line. This process may be repeated until 10 megacycles is reached, whereupon methods already developed become available. Spectral lines have already found application as standards of length, and there is every reason to believe that they will now find application as standards of time.

PRINCIPLES OF FREQUENCY-MODULATED RADAR*†

By

IRVING WOLFF AND D. G. C. LUCK

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 useable radio-frequency band. FM radar avoids need of high peak power by using a narrow noise band, and facilitates use of 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.*

I. INTRODUCTION

EVEN before Breit and Tuve[1] applied radio-wave pulses to echo sounding of the ionized layers of the atmosphere, Appleton and Barnett[2] had used a frequency-modulated signal for the same purpose. Both methods meet the requirement, basic for radar detection and ranging of reflecting objects, that the returned echo shall be distinguishable despite the transmitted signal. In pulse radar this requirement is met by separation in time of reception of transmitted and

* 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.

[1] G. Breit and M. A. Tuve, "A Test of the Existence of the Conducting Layer", *Phys. Rev.*, Vol. 28, pp. 554-575, September, 1926.

[2] E. V. Appleton and M. A. F. Barnett, "On Some Direct Evidence for Downward Atmospheric Reflection of Electric Rays", *Proc. Royal Soc.*, (sec. A), Vol. 109, pp. 621-641, December 1, 1925.

reflected signals, but in frequency-modulated radar both signals are present at once. The presence of the reflected signal is distinguished in the FM case by the difference of its frequency or phase from that of the transmitted signal.

Only a very small fraction of the total effort applied to radar development has been in the field of frequency-modulated radar. Some of the rather special principles on which FM radar operates, and the ways in which they determine its performance, are consequently not widely known. This paper sets forth briefly some of these principles and their consequences. The ideas here presented arose in the course of work done by many members of the RCA organization. It is not practicable to apportion credit among all those who took part, but mention should be made of the many important contributions of R. C. Sanders, Jr.

Broadly, the practical advantages of frequency-modulated radar equipment are the following:

1) Ease of utilizing wide radio-frequency bands to obtain high resolution.
2) Freedom from high peak-power requirements.
3) Direct availability of information on both range and speed of target.
4) Simplicity of means capable of providing automatic control based on such information.

The price paid for these advantages is technical difficulty in separating the effects of multiple targets and in duplexing transmission and reception on a single antenna.

Frequency-modulated radar has been used mainly for aircraft altimetry, though it has also proved especially well suited for certain military fire-control uses. Use in searching an area for targets has often been suggested and has been tried experimently. Techniques and equipment for such uses, embodying the principles here developed, are described in later papers. ‡

II. Basis of Range and Speed Determination

(A) Effects of Frequency Change.

A radio signal transmitted to and reflected from a target at distance or range R will return to a receiver adjacent to the transmitter after a time interval τ#, where

$$\tau = 2\,R/c \qquad (1)$$

‡ It is planned to publish these papers, "FM Radar Techniques" and "Some Applications of FM Radar" in the June, 1948 and September, 1948 issues of *RCA REVIEW*.

\# The notations used in this paper are listed on pages 74 and 75.

and c is the radio-wave propagation velocity (c is 983.2 feet per microsecond in normal sea-level air). If the radio frequency F of the transmitted signal is varying uniformly with time at a rate $\dot{F}$, a signal that was transmitted at an instant t_1 with frequency F_1 will return from reflection by a stationary target at a time $t_1 + \tau$, when the frequency being transmitted has reached an altered value F'_1.

The frequency F'_1 being transmitted at time $t_1 + \tau$ is of course just $F_1 + \dot{F}\tau$. If mixed with the signal of frequency F_1 transmitted at t_1 and received after target reflection, it will produce a beat note at the difference frequency $\dot{F}\tau$. Using Equation (1) for τ, this *radar range-beat frequency* f_R is

$$f_R = 2\dot{F}R/c. \tag{2}$$

Figure 1(a) shows the variation with time of transmitted frequency (solid line) and frequency received after reflection from a stationary

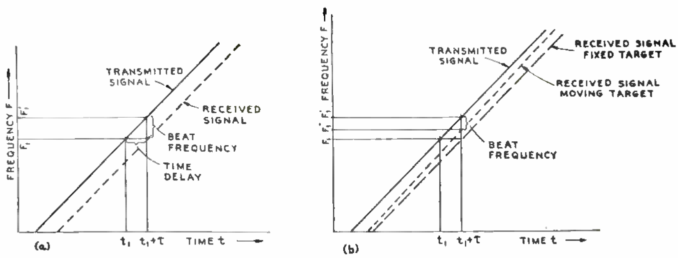


Fig. 1—Variation of transmitted and received frequencies with time.

target (broken line). It is evident from the figure that so long as the variation of transmitted frequency with time is linear, that is at a constant rate $\dot{F}$, the range-beat frequency is constant.

If the radar is moving relative to the target, the current induced in the target will differ in frequency from that in the transmitting antenna, and the current induced in the receiving antenna by target-reflected signal will in turn differ in frequency from that in the target. This is the well known Doppler Effect, and results in reception at frequency

$$F'' = F\ (c + S)/(c - S) \tag{3}$$

of a signal transmitted at frequency F, if the radar and target are approaching one another at speed S. Since all practical speeds of radar or target motion are extremely small compared to c, the modified frequency F'' is very close to $F + 2FS/c$. If an unmodulated signal transmitted at frequency F is mixed with the same signal as received at frequency F'' after reflection from a target moving at speed S, there will result a steady beat note at the *radar speed-beat frequency*

$$f_S = 2\,FS/c. \tag{4}$$

If a radar transmitting at uniformly increasing frequency is at range R from a target and is moving at relative speed S toward that target, the frequencies of simultaneously transmitted and received signals will vary as shown respectively by the solid and dotted lines of Figure 1(b). The net radar beat frequency will be $f_R - f_S$, with f_R and f_S still given by Equations (2) and (4).

(B) Alternative Viewpoints.

Particular conditions can occur under which other explanations of FM radar phenomena are more convenient than the time-lag and frequency-shift explanation just given. So long as the change in transmitted frequency during the signal-travel time τ is slight, a "standing-wave" pattern will be set up around the target and will move slowly as transmitted frequency varies. The receiving antenna immersed in and moving relative to this pattern will observe a resultant signal which varies in amplitude according to the standing-wave envelope.

Variation of receiver output may also be regarded as resulting from combination with varying phase of a signal voltage arriving directly from the transmitter and one received after target reflection. This is true even though it is often desirable in practice to minimize radiation directly from transmitting to receiving antenna, and to introduce the reference-phase signal into the receiver over a controlled artificial path instead. On this point of view, the variations occurring can be illustrated by conventional rotating-vector diagrams.

The number N of standing waves between radar and target separated by distance R is

$$N = R/(\tfrac{1}{2}\lambda) = 2FR/c, \tag{5}$$

with transmission at radio wave length λ or frequency F. Phase difference ψ between direct- and reflected-signal vectors at the receiver is

$$\psi = 2\pi N + \psi_1 \tag{6}$$

radians, where initial phase ψ_1 depends upon phase change at target reflection and upon any fixed phase shift incurred by artificially introducing reference signal into the receiver.

Figure 2 illustrates both of these alternative viewpoints. It shows how new waves effectively slip past the radar into the standing pattern as transmitted frequency increases, cyclically varying the resultant-signal amplitude at the radar as they do so. It shows also the corresponding rotation of the reflected-signal vector e_2 relative to the direct-signal vector e_1, together with the change in amplitude of resultant e_3 so produced.

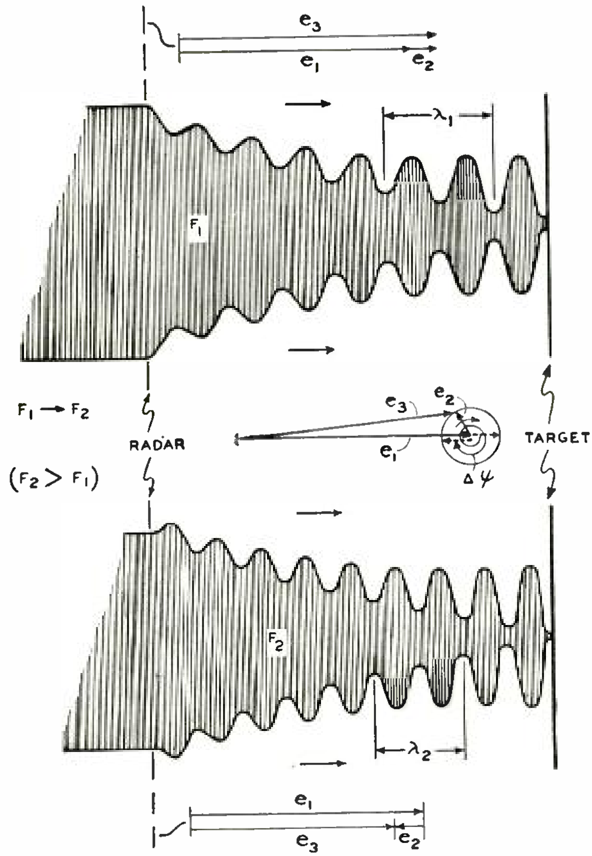


Fig. 2—Effect of frequency change on standing-wave pattern and signal-vector phase.

For a given frequency, the standing-wave pattern is fixed in space relative to the target. Relative motion of pattern and radar can occur in two ways: the pattern can grow or shrink past the radar as frequency is varied, or the radar can actually move through the pattern relative to the target. Either motion will produce a cyclic variation of receiver output as the number N of waves standing between radar and target changes. The frequency f of output variation is simply, by differentiation of Equation (6),

$$f = dN/dt = 2\,\dot{F}R/c + 2\,F\dot{R}/c = f_R - f_S. \qquad (7)$$

Since rate of change of range $\dot{R}$ is the negative of approach speed S, this is exactly the typical result found on another basis in Section II(A).

These simple pictures are valid so long as the frequency modulation used is gradual enough to permit the radio signal to be regarded as a single substantially pure sinusoid of slowly varying frequency. Fortunately, this has always been the case in actual FM radar systems. Fourier analysis of the FM signal into an unmodulated sinusoidal carrier and a set of fixed side frequencies has not been found helpful.

(C) Periodic Modulation.

To restrict the transmitted frequency to a limited band while maintaining a flow of information from the radar, periodic frequency modulation, rather than the simple one-way sweep of the above discussion, is customary. The detailed properties of the radar output depend on the characteristics of the particular modulation used, and extremely simple properties result if the frequency varies in symmetrical-sawtooth fashion. Variation of rate of change of frequency $\dot{F}$ and of transmitted frequency F in this type of modulation are illustrated by graph (a) and the solid-line graph at (b) of Figure 3. Frequency of the signal received from a stationary target is shown by the dotted line of Figure 3(b) and time variation of radar beat-note frequency, represented by the vertical separation of solid and dotted lines of graph (b), is shown in more detail at (c).

A typical wave form for the actual radar beat-note output voltage with target stationary is shown at (d) of Figure 3. This peculiar output wave, consisting of successive sinusoidal sections with sudden large phase shifts between them, is a striking characteristic of radar operation with triangular frequency modulation. The sinusoidal output variation during each single frequency sweep, or half cycle of modulation, is evidently the mirror image of that during adjacent half cycles. That this must be so is easily seen by considering the cyclic motion of the standing-wave pattern past the radar during modulation.

Figure 3(e) shows the time variation of transmitted (full line) and received (dotted line) frequencies for the case of a target moving toward the radar, at a moderate speed or large range such that range-beat frequency f_R exceeds speed-beat frequency f_S. Time variation of beat-note output frequency is shown for this case by graph (f). The output beat signal is itself frequency modulated, having a fixed frequency $f_R - f_S$ during the upward modulation sweep of the transmitter frequency and a fixed frequency $f_R + f_S$ during the downward sweep. Full information as to both range and speed of the target is evidently contained in this single output signal. Figure 3(g) shows the actual output wave form for a slowly moving target, with alternate sinu-

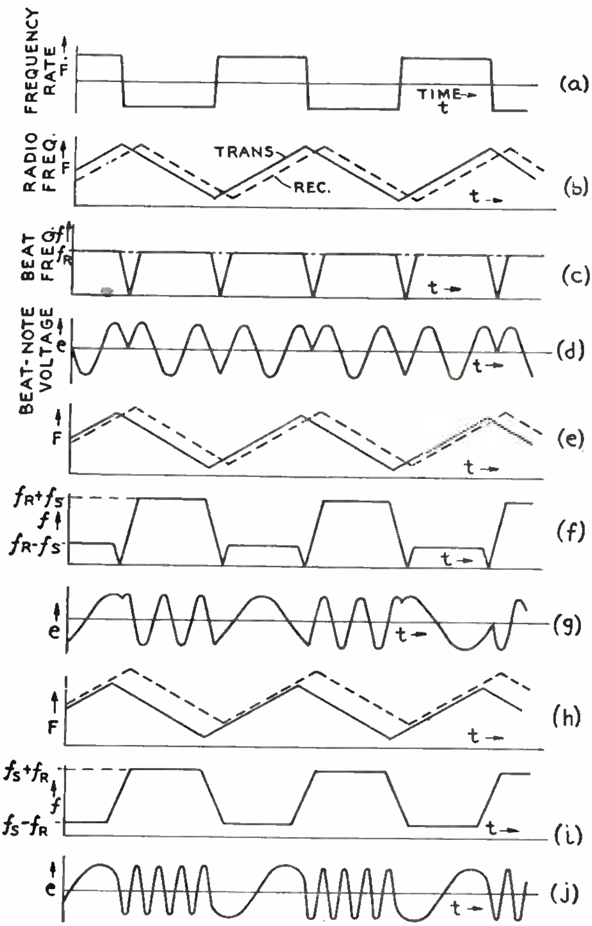


Fig. 3—Frequency and voltage wave forms for symmetrical-sawtooth or triangular modulation.

soidal sections of two different frequencies and sudden large phase shifts between adjacent sinusoidal sections.

If the target approaches the radar at high speed or if $\dot{F}$ is small, so that f_S exceeds f_R, the reflected-signal frequency varies according to the dotted line of Figure 3(h). This results in the beat-note frequency variation shown by graph (i) of the figure, with the output-voltage wave form of graph (j). This output wave differs from those with zero or moderate speed in that no phase jumps appear at the ends of the modulation sweep. Phase jumps occur only if relative speed of radar and target is low enough or modulation rapid enough so that the total relative motion of radar and standing-wave pattern reverses during the cycle of frequency modulation.

Further details of the triangular frequency modulation are shown in Figure 4. Since the total width W of the modulation band is uniformly swept in time $t_m/2$, the frequency rate $\dot{F}$ is

$$\dot{F} = 2W/t_m = 2f_m W \qquad (8)$$

for a modulation frequency f_m. Range-beat frequency is then [see Equation (2)]

$$f_R = 4\,W f_m R/c. \qquad (9)$$

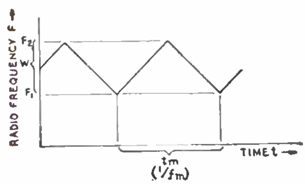


Fig. 4—Characteristics of triangular modulation.

Dependence of the beat frequency per unit range only on the product of modulation frequency f_m and width W of band swept in modulation is an important characteristic of FM radar.

As transmitted radio frequency varies from F_1 to F_2 in modulation, the total number of standing waves between the radar and a target at fixed range varies from N_1 to N_2 in accordance with Equation (5). So long as frequency change proceeds steadily from F_1 to F_2 without reversals, the total number of cycles of radar-output variation produced in consequence is simply $N_2 - N_1$, or $\triangle N$, independently of the exact way in which F varies with time. That is, the number of range-beat cycles per modulation sweep depends only on the frequency limits of the sweep and not on the modulation wave form, so long as the latter has no re-entrant kinks. Equation (9) therefore holds for the time average of the range-beat frequency during periodic modulation, independently (within limits) of the modulation wave form. Where only range of a single target is to be determined, without regard to its speed, any convenient form of modulating wave, such as the easily produced sinusoidal form, may be used equally well. Many other types of periodic modulation are of course possible, and may indeed prove desirable for particular applications.

In many FM radar applications, the wave-propagation delay τ is an extremely small fraction of the modulation period. The portion of the time during which the radar beat frequency with triangular modulation departs from its proper values, greatly exaggerated in graphs (b), (c), (e), (f), (h), and (i) of Figure 3 is, in such cases, negligibly small. The phase change of the output-voltage wave at the turning points of the transmitter-frequency variation is then sensibly instantaneous, as indicated in graphs (d) and (g) of Figure 3.

Because of their small magnitude, secondary propagation-delay effects are usually negligible. However, recognition of their existence is important in avoiding conceptual difficulties. Exact analysis[3] shows, in brief, that Equation (2) represents the range-beat frequency exactly for any chosen instant if the value of frequency rate $\dot{F}$ used is the average over the propagation-time interval τ just preceding the instant in question. Similarly, Equation (5) gives the number of standing waves exactly at every instant if the average value of radio frequency F over the interval τ just preceding the instant in question is used.

III. Single-Target Operation

(A) General Characteristics.

Applications of frequency-modulated radar have normally involved operation with only a single reflecting target. In the case of aircraft altimeters, the single target has been the surface of the earth; in the case of automatic radar bombing, the single target has been an isolated surface vessel; in supervisory control of aircraft movements, the single target has been an isolated aircraft. Such applications have in common the exploitation of certain characteristics of FM radar data, and employ very similar apparatus for utilization of the data to produce the desired final result.

It is convenient to characterize the radar itself, for purposes of measurement and control from observation of a single target, by two operating parameters. The range-beat frequency per unit range may be called the *radar range sensitivity* k_R, and from Equation (2) or (9) is

$$k_R \equiv f_R/R = 2\,\dot{F}/c = 4\,Wf_m/c. \tag{10}$$

[3] P. Giroud and L. Couillard, "Sondeur Radioelectrique pour la Mesure des Hauteurs des Aeronefs au-dessus du Sol", *Ann. de Rad. Elect.*, Vol. 2, No. 8, pp. 150-172, April, 1947.

If modulation of transmitted frequency is non-linear, Equation (10) gives the average range sensitivity for the complete modulation cycle. Similarly, the speed-beat frequency per unit speed may be called the *radar speed sensitivity* k_S, and from Equation (4) is

$$k_S \equiv f_S/S = 2\,F/c. \tag{11}$$

Speed sensitivity changes as radio frequency F varies over the modulation cycle, but in most practical cases this variation is slight; in any case, Equation (11) gives the correct average speed sensitivity if the value of F used is the time average F_o of the radio frequency over the modulation cycle.

If range information only is desired, it has already been shown that the details of modulation wave form are unimportant so long as re-entrant kinks are avoided. If speed information only is desired, the details of frequency modulation are again of no direct importance, so long as frequency rate $\dot{F}$ always remains sufficiently small. If information on both range and speed is desired, however, the use of symmetrical-sawtooth modulation such as is shown in Figure 3 becomes highly advantageous.

In terms of the radar sensitivities k_R and k_S, the beat frequency f_u during frequency upsweep in modulation was found above to be (Figure 3(f))

$$f_u = k_R R - k_S S, \tag{12}$$

and the beat frequency f_d during downsweep to be $f_d = k_R R + k_S S$. (13)

Equations (12) and (13) describe the typical frequency-modulated beat-note data output of the radar itself when operating against a single target. If the relative approach speed S of radar and target is so high that Figure 3(i) applies, the effect is simply to interchange R and S throughout Equations (12) and (13).

Beat-note frequency data must be converted to some other form, such as voltage, for practical application. Conversion may be on the basis of an average over the complete modulation cycle, in which case the data-converting apparatus may be characterized by a *range conversion sensitivity* h_R in volts per unit range-beat frequency. This results in a range-voltage output $k_R h_R R$. Alternatively, conversion may be on the basis of difference between upsweep and downsweep beat frequencies during linear modulation, in which case the data-converting apparatus may be characterized by a *speed conversion sensitivity* h_S in volts per unit speed-beat frequency. This results in a speed-voltage output $k_S h_S S$. If operating conditions are those of Figure 3(h) and

(i) rather than of Figure 3(e) and (f), effects of speed and range converters are simply interchanged.

Range and speed voltage outputs may be used separately or combined; they may be combined, also, with any predetermined bias voltage. The most general single-target output of a frequency-modulated radar and associated data converters is the voltage

$$e = k_R h_R R - k_S h_S S + e_o, \tag{14}$$

where e_o is an added bias voltage. The algebraic sign of each term of Equation (14) may be made positive or negative as desired. This resultant output voltage may be indicated directly, or it may be made to exert selective control action according to whether e is greater than, equal to, or less than some predetermined value e_1.

Output voltage e may be held automatically at a value e_1 if all or part of the radar and data converter is included in a servo feed-back loop. This may in principle be done by servo control of any one or more of the variables R, S, k_R, k_S, h_R, h_S, e_o, and e_1. Such control of R, k_R, and e_o singly has been used quite successfully, and control of S, h_R, or h_S should also be practicable. Control of k_S would involve change of radio-frequency channel, usually a difficult matter, and control of e_1 has been unsatisfactory for technical reasons.

(B) Signal Strength.

Signal strength is controlled by the same factors in FM radar as in the better known case of pulse radar.[4,5,6] Some important properties of received-signal strength will therefore merely be stated here without derivation. Their effects on single-target operation of FM radar may then be pointed out. Type of target and conditions of operation control the result obtained.

In altimetry, the target is the earth and the effective target area is determined by transmitter directivity. The power P_r available at the receiving-antenna terminals after ground reflection is

$$P_r = KP_t A'^2_a \; R'^2, \tag{15}$$

where P_t is power fed to the transmitting-antenna terminals and A_a

[4] K. A. Norton and A. C. Omberg, "The Maximum Range of a Radar Set", *Proc. I.R.E.*, Vol. 35, No. 1, pp. 4-24, January, 1947.

[5] E. M. Purcell, "The Radar Equation", RADAR SYSTEM ENGINEERING, Vol. 1, chapter 2, Radiation Laboratory Series, McGraw-Hill Book Co., New York, N. Y., 1947.

[6] A. J. F. Siegert, L. N. Ridenour, and M. H. Johnson, "Properties of Radar Targets", RADAR SYSTEM ENGINEERING, Vol. 1, chapter 3, Radiation Laboratory Series, McGraw-Hill Book Co., New York, N. Y., 1947.

is effective antenna area, usually the same for transmission and reception. Primes indicate that range is measured in units of wave length of the radio signal and antenna area is in square wave lengths. K is a numerical coefficient, dependent on the character of the reflecting surface and on the transmitter directivity, which has the value ¼ for the limiting case of a perfect plane mirror.

For an isolated target in free space, with a definite echoing area of A'_e square wave lengths,

$$P_r = P_t A'^2_a A'_e / (4\pi R'^4). \tag{16}$$

Effective areas of antennas and target normally depend strongly upon the direction of the radar-to-target line relative to their respective orientations. The best way to take into account the effect on Equation (16) of reflection from the ground depends upon conditions of operation in FM just as it does in pulse radar. Both Equations (15) and (16) exclude losses in lines between antennas and equipment, as well as losses due to atmospheric absorption; the methods of allowing for such losses are well known.

Range beat-note frequency is directly proportional to range and received-signal power is, in the case of an altimeter, inversely proportional to range squared. It is therefore advantageous in FM altimetry to use a beat-note amplifier having a gain proportional to frequency (i.e., a gain rising 6 decibels per octave of frequency increase); this produces a final beat-note signal strength which is independent of altitude. To produce an output level independent of range in the case of a free-space target of definite area requires a beat-note amplifier gain proportional to the square of the beat frequency (i.e., rising at 12 decibels per octave), since received-signal power is in that case proportional to the inverse fourth power of range.

Reflected signal from the desired target must compete with undesired signals reflected from the surroundings of the target and the surroundings of the radar, as well as with undesired modulation of the transmitted signal and with random noise generated in the receiver. Only if random noise is the limiting factor will increased transmitter power increase the maximum range for useful operation. It is technically quite difficult to reduce other disturbing factors to such a degree that random noise is the major factor limiting useful range.

The ratio of desired-signal strength to strength of undesired disturbing signals that is necessary for satisfactory operation depends upon type of operation and accuracy desired. With the data-conversion method usually used, automatic control operation accurate to a few per cent requires a desired-signal amplitude at least 10 times the root-mean-square amplitude of all disturbing signals combined. This high signal noise ratio represents the price paid for foregoing the judgment of a human operator and seeking instead fast and quantitatively accurate automatic operation.

(C) Fixed Error.

Averaging cycle-rate counters, such as are now found in commercial audio-frequency meters, have been widely used for conversion of FM radar beat-frequency data. These devices transfer a fixed increment of charge to a current-measuring element each time the voltage of the signal whose frequency is to be measured passes upward through its average value. Such average-value passage is an integral event which may occur or may not occur: partial occurrence is, except for technical imperfections, impossible. The counter can therefore register only an integral number of beat-frequency cycles per radar modulation cycle. If target range is truly constant and radar modulation truly periodic, the radar range-beat frequency will necessarily be truly periodic also. The counter must then register the same number of counts on each modulation cycle, and the average beat-note frequency indicated can only be some integral harmonic of the modulation frequency, no matter what the actual target range.[3] Fourier analysis of such a beat signal would show only harmonics of the modulation frequency.

Since the indicated range-beat frequency can only change in jumps equal to the modulation frequency f_m, Equation (9) shows that the

indicated range must change in jumps of $\delta R = \frac{1}{4}c/W = \frac{1}{4}\lambda_w$. (17)

λ_w is the wave length of a radio signal having a frequency equal to W, the width of the frequency band swept in modulation, and may be called the *sweep wave length* of the radar transmission. The indicated range may evidently be in error by as much as $\frac{1}{4}\lambda_w$; since this error does not vary with range, it has sometimes been called the *fixed error* of the radar. Details of the occurrence of fixed error are most easily explained by the study of some special cases.

The average number N_0 of standing waves present between radar and target during the modulation cycle is $R/(\frac{1}{2}\lambda_0)$, where wave length λ_0 corresponds to mean radio frequency F_0. The change ΔN in number of standing waves during each modulation sweep is correspondingly $R/(\frac{1}{2}\lambda_w)$. The number of range-beat cycles per modulation sweep is ΔN, which depends only on range and sweep wave length and not on mean radio frequency F_0 or modulation frequency f_m. The change $\Delta\psi$ in relative phase angle between direct and target-reflected signal vectors during modulation is $2\pi\Delta N$ radians, while mean phase ψ_o is $2\pi N_o + \psi_1$ and does not depend on sweep width W or modulation frequency f_m.

Figure 5 illustrates conditions at selected ranges, for the particular case in which modulation sweep W is $F_o/20$, or λ_w is $20\lambda_o$. Graph (a) of the figure illustrates variation with range of standing-wave amplitude at mean frequency F_o, in the vicinity of a perfectly conducting target; dependence of signal strength on range is neglected to simplify the graph. Seven specific ranges selected for study are indicated by numbered vertical lines, and the portion of the standing wave

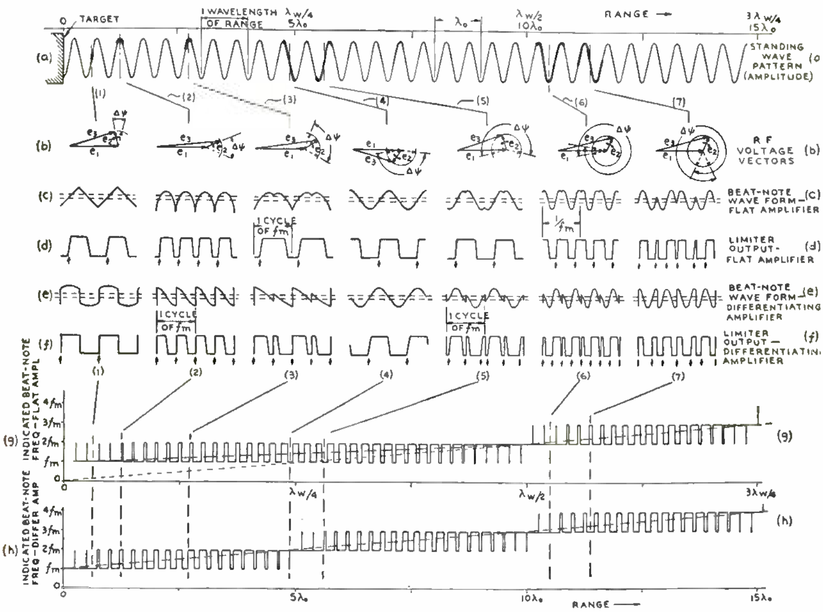


Fig. 5—Origin of fixed error, for the case $W = F_o/20$.

which sweeps past each selected range during modulation is indicated by a heavy section of the graph. Figure 5(b) gives for each range selected a vector diagram showing the phase relation between direct reference signal e_1, target-reflected signal e_2, and resultant signal e_3. The variation $\Delta\psi$ of relative phase over the modulation cycle is indicated in each diagram and the corresponding variation of resultant-signal amplitude is evident. The diagrams are drawn for an initial phase ψ_1 of π radians, or 180 degrees.

Figure 5(c) shows a small graph of time variation of radar beat-note output voltage for each selected range; each single-sweep section of such a graph is that portion of the standing-wave envelope swept over in modulation and shown by a heavy line in Figure 5(a). Figure

5(e) shows corresponding graphs of the output of a beat-note amplifier with gain rising at 6 decibels per octave, such as would be used in an altimeter. This amplifier characteristic has the effect of differentiating the signal voltage with respect to time, so that the ordinate of any graph (e) at any particular instant represents the slope of the corresponding graph (c) at the corresponding instant.

Figure 5(d) shows the effect of passing the signals of Figure 5(c) through a voltage limiter, to prepare them to actuate a cycle-rate counter; each graph (d) is, in effect, merely a magnified view of the portion of the corresponding graph (c) enclosed by the dotted lines. Each small arrow in (d) indicates the transfer of a measured charge increment to register a single count. Figure 5(f) similarly shows the effect of a limiter on the differentiated signal of Figure 5(e).

The indicated beat frequency is of course simply the product of the number of counts registered per modulation cycle by the modulation frequency. The crinkly solid lines in Figure 5(g) and (h) are graphs of the variation of indicated frequency with range, respectively for direct radar beat-note output and for output differentiated by an amplifier with rising gain-frequency characteristic. These graphs are to the same range scale as Figure 5(a), and the seven ranges selected for study are again indicated by numbered vertical lines. For ranges in the immediate neighborhood of any particular value, it is evident from Figure 5 that the indicated range-beat frequency may have either of two values different by f_m, and that in general neither of these values is f_R as given by Equation (2) or (9). It is further evident that a trifling change in range, perhaps 1/30 r-f wave length at a range of five wave lengths for the conditions illustrated, can change the indicated frequency by a factor of two.

Fixed error occurs only for strictly constant range and strictly periodic modulation. In actual operation, random or steady target motion will almost always result in indication of the time average of a rapidly varying range-beat frequency. Careful examination of the behavior of the direct radar-output signal at ranges adjacent to point (4) of Figure 5 will show that the indicated frequency will be f_m for just half of such range values and will be $2\,f_m$ for the other half. The average indication for a range fluctuating about point (4) will therefore be $\frac{3}{2}f_m$. The dashed-line graphs of Figure 5(g) and (h) indicate the variation with range of such averaged indications. It is evident from the upper part of the dashed line of Figure 5(g), and its dotted extension, that the averaged frequency indicated for ranges in excess of $\frac{1}{2}\lambda_w$ will be f_R for the direct radar output; at shorter ranges, $\frac{1}{2}f_R + f_m$ will be indicated. Similarly, the dashed line of Figure 5(h) shows that a beat frequency of $f_R + f_m$ will be indicated for any fluctuating range if the radar output is differentiated before being fed to the counter.

If a linear scale of indicated range must be obtained down to extremely short ranges, the use of a differentiating beat-note amplifier is evidently suggested; use of differentiated signal produces on the average one extra count per modulation cycle, increasing all range indications by $\frac{1}{4}\lambda_w$. If speed-beat frequency exceeds range-beat frequency, as in Figure 3(h), (i) and (j), no beat-note phase jumps occur at the turning points of the modulation cycle, the cusped wave forms of Figure 5(c) and (e) never occur to give extra counts, and the average of beat frequency over the modulation cycle gives target speed with no counterpart of fixed error. Fixed error from moving targets for which range-beat frequency exceeds speed-beat frequency is possible, but is extremely improbable in actual operation.

Fixed error is a particularly striking characteristic of FM radar using cycle-rate counters under laboratory conditions. Beside being of little importance in practical operation, however, fixed error appears not to be an intrinsic property of FM radar. Data-converting methods can readily be imagined which would single out for chronographic measurement a few beat-note cycles near the middle of each modulation sweep, where conditions at the turning points of the modulation have negligible effect. Such devices might be impracticably complex, but should in principle be entirely free of fixed error.

IV. Multiple-Target Operation

(A) General Characteristics.

Selective means of utilizing FM radar beat-frequency data are required when several targets are present at once. One target may then be selected and followed for operation of single-target type, or else overall information may be provided as to the individual ranges and speeds of the several targets. The first of these possibilities has not been fully investigated, but the second provides an FM radar searching device in fairly direct competition with its pulsed counterparts, so that comparisons are in order. As far as directional search of the space around the radar is concerned, pulse and FM systems do not seem significantly different, so attention is here devoted to range and speed search in one particular direction only, often called *searchlighting.*

A frequency-selective element fed with the beat-note output of an FM radar can provide information as to the presence or absence of a target within a predetermined narrow region of range. This is analogous to the action of a timed "gate" in pulse radar. Multiple gates may be used with either FM or pulse radar to investigate a number of different range channels at once. Either multiple indicators or a single indicator scanning the outputs of the several gates may be used in either case. An example of a multiply selective frequency indicator presumably applicable to FM search radar is the well-known Frahm reed tachometer.

Alternatively, a single movable gate may scan a number of range regions or channels successively. In FM search technique, investigation of the radar beat output with a wave analyzer would typify this method. Still another possibility in the FM case, which seems to have no close analogue in pulse technique, is to scan numerous range channels past a single fixed gate. This may be accomplished by progressively altering the rate of change of transmitted frequency over a suitable number of successive modulation sweeps, which requires separate control either of width of band swept or of duration of sweep for each individual modulation sweep.

Important operating parameters to be specified for searchlighting use of search radar are:

1) maximum useful range;
2) range resolution;
3) minimum range;
4) time required for complete range search; and
5) directional resolution.

Some relations among these characteristics will be developed. Important apparatus characteristics of FM search radar, which must be related to operating characteristics as a basis for design, are:

1) power;
2) antenna directivity;
3) noise bandwidth;
4) receiver noise factor;
5) number of range channels used;
6) frequency of modulation; and
7) width of modulation sweep.

(B) Range Resolution.

The range resolution, or minimum range interval at which two targets are distinctly indicated as separate, is determined in the pulse case by the duration of the receiver-output pulse, which is in turn determined jointly by transmitted-pulse duration and pass band of the receiver. Resolution is determined somewhat differently in the FM case. Because of the phase jump in radar beat-note output at the end of each single frequency sweep of the modulation, the total range information available for each target comprises only ΔN beat-note cycles at a frequency $2\Delta N f_m$, where from Equation (5) and Figure 4

$$\Delta N = 2WR/c = R/(\tfrac{1}{2}\lambda_w). \tag{18}$$

If two targets present at once are to be resolved, their values of ΔN must be sufficiently different so that the total radar-beat output wave will give, during only ΔN cycles, clear evidence of being complex.

It is directly evident that if the difference $\delta(\Delta N)$ for two targets is only a small fraction of a beat cycle altogether, its existence will not be clearly shown. It is equally evident that if the difference $\delta(\Delta N)$ is many cycles, its existence, and thereby the presence of more than one target, will be unmistakably shown by an amplitude modulation of the resultant radar beat-note output. The exact value of $\delta(\Delta N)$ required for resolution will depend upon apparatus used, but a value near one cycle may be expected, with a resulting range resolution δR of the order of one-half sweep wavelength.

The exact behavior of a selective circuit, such as may be used to resolve multiple targets in actual operation, depends on details of circuit design. To indicate satisfactorily whether or not a signal at its maximum-response frequency is present, the selective-circuit response must build up to substantially its final steady-state value during the time that signal input is available. The build-up time for the response of a circuit with noise bandwidth Δf is proportional to and of the same order of magnitude as $1/\Delta f$, and such a circuit will discriminate clearly between two signals having frequencies different by an amount of the same order as Δf. With a build-up time of only one-half modulation cycle at frequency f_m available, selector bandwidth must be at least $2f_m$, and minimum signal-frequency difference for definite separation by such a selective gate will be assumed for purposes of discussion to be $3f_m$. By Equation (9), the range resolution so obtained is $\tfrac{3}{4}\lambda_w$.

(C) Working Range.

As in the single-target case, desired-target signal must compete with undesired signals reflected from surroundings of target and radar, with effects of undesired transmitter modulation, and with random noise generated by the receiving system. Maximum range of pulse-radar operation is usually limited by random noise only, and FM search radar should be considered on a comparable basis, even though technically not yet developed to such a degree of refinement.

If an average power P_t is applied to the transmitting antenna in discrete pulses which are repeated at a frequency f_r, the total energy

U_t transmitted per pulse is P_t/f_r. Received energy U_r per pulse for free-space radar transmission is given by Equation (16), with U_t in place of P_t. Empirical study of operator response for searchlighting operation[7] of pulse radar has shown the minimum received-pulse energy U_{min} perceptible in the presence of random receiver noise to be

$$U_{min} = \overline{NF}kT(f_r/1670)^{-1/3}, \tag{19}$$

if the optimum receiver band width is used. $\overline{NF}$ is receiver noise figure, k is Boltzmann's constant of 1.37×10^{-23} watt-seconds per degree centigrade, and T is normally the room temperature of 300 degrees Kelvin.

Equating U_r to U_{min}, the maximum range for pulse operation at radio frequency F_o (or wave length λ_o) is found to be

$$R_{max} = \lambda_o \left(\frac{P_t A'_o{}^2 A'_r}{2\pi F_o \overline{NF} kT} \right)^{1/4} (F_o/3340)^{1/4} (f_r/1670)^{-1/6}. \tag{20}$$

This inverse-sixth-root dependence of maximum range on pulse-repetition frequency for constant average power differs in form from the better known relation[1] for constant pulse energy. Pulse duration t_p is determined by the range resolution required, and if the receiver band width Δf is so chosen that $t_p \Delta f = 1$, an optimum signal/noise ratio will prevail. So long as this optimum selectivity is maintained, maximum range for pulse operation shows no dependence upon range resolution δR.

Conditions in frequency-modulated radar operation may be studied by considering that the result of each single linear modulation sweep is a relatively long pulse of beat-note signal, and then following the same procedure as above. The pulse-repetition frequency f_r is just twice the modulation frequency f_m, but the pulse duration is diminished by a fraction ρ because of the time lost at the beginning of each sweep, while signal is going out to and returning from the target. From Equation (1), the lost-time fraction τf_r is

$$\rho = 2f_r R/c; \tag{21}$$

only a fraction $1 - \rho$ of the total transmitted energy per sweep, P_t/f_r, appears as useful pulse energy U_t. The optimum noise bandwidth of selective beat-note filters for such pulses is

$$\Delta f = f_r/(1 - \rho). \tag{22}$$

---

[7] A. V. Haeff, "Minimum Detectable Radar Signal and Its Dependence Upon Parameters of Radar Systems", *Proc. I.R.E.*, Vol. 34, No. 11, pp. 857-861, November, 1946.

Range search can be accomplished by examining one distinct range element on each of N successive single sweeps of transmitter frequency. Repetition frequency f_s of the complete search, to be used instead of f_r in Equation (19) for minimum discernible energy, will then be f_r/N. Equating U_r to U_{min} now gives a relation among R_{max}, f_s, and ρ_m, the lost-time fraction at maximum range. Using Equation (21) to eliminate either f_s or R_{max} from this relation gives respectively

$$R_{max} = \lambda_o \, (3340N/F_o)^{-1/10} \left(\frac{P_t A'^2_a A'_c}{2\pi F_o \overline{NF} kT} \right)^{3/10} \rho_m^{1/10} \left(\frac{1}{\rho_m} - 1 \right)^{3/10} \qquad (23)$$

and

$$f_s = 1670 \, (3340N/F_o)^{-9/10} \left(\frac{P_t A'^2_a A'_c}{2\pi F_o \overline{NF} kT} \right)^{-3/10} \rho_m^{9/10} \left(\frac{1}{\rho_m} - 1 \right)^{-3/10} \qquad (24)$$

These two equations together give a relation analogous to Equation (20) between search-repetition frequency and maximum noise-limited range for FM radar, and determine as well the maximum lost-time fraction ρ_m. Maximum range is again independent of range resolution δR if optimum selectivity is maintained.

If the complete range search is accomplished in a single modulation sweep by examining the radar output simultaneously with M filters of equal bandwidth $f_r/(1-\rho_m)$, a range interval

$$R_{max} - R_{min} = (M-1)\ \delta R \qquad (25)$$

will be covered with uniform range resolution δR (see last part of Section IV(B)) of $\frac{3}{4}\ \lambda_w/(1-\rho_m)$, which in turn determines the required sweep width W. Maximum range is still given by Equations (23) and (24), with $N = 1$; if f_s is low enough so that $\rho_m << 1$, Equations (23) and (24) combine in this FM case to form the same simpler Equation (20) already found for pulse radar. With operating parameters R_{max}, R_{min}, δR, A'_c, f_s, and λ_o specified, the factors $P_t A'^2_a \overline{NF}$, M, W, and f_r required for apparatus design are fully determined.

If range search is accomplished by examining the radar output with a wave analyzer with fixed pass band, which is tuned to a different frequency on each of N successive sweeps, maximum range is given by Equations (23) and (24) for the value of N used. Range interval searched is given by Equation (25) with N in place of M, and δR is again as above.

Searching in range may also be done by sweeping transmitter frequency linearly across a band of different width during each of N successive and equal time intervals of duration $1/f_r$, and applying the beat-frequency radar output to a single fixed filter. Maximum range is again given by Equations (23) and (24), but the range resolution δR is in this case not fixed but proportional to range. The fixed fractional range resolution $\delta R/R$, or δ, is proportional to the ratio of bandwidth to center frequency of the fixed filter, and the range region searched is now

$$R_{\max}/R_{\min} = (1+\delta)^{N-1}. \tag{26}$$

The sweep width required follows a stepped-exponential law with respect to time elapsed after the start of each search.

Search may also be developed by varying the duration t_r of each of N successive sweeps over a frequency band of constant width W, with radar output applied to a single filter tuned to a fixed frequency f_o. The filter pass band must be changed from sweep to sweep if best signal/noise ratio on the long beat-frequency pulses at long range and adequate response to the short pulses at short range are both to be attained. Eliminating R from Equation (21) by use of Equation (9) for range-beat frequency f_o, ρ is found to be f_o/W, which is constant throughout the search. δR is also constant, and the range region searched is again given by Equation (25). Sweep duration must follow a stepped-parabolic relation to elapsed time from the start of each search cycle, and search-repetition frequency

$$f_s = \frac{c\rho}{N(R_{\max}+R_{\min})}. \tag{27}$$

Because Equation (27) is more complicated than Equation (21), the signal-strength relations for variable-duration modulation are more complex and less general than Equations (23) and (24); $R_{\min}$ or δR as well as N must be specified to determine the dependence of $R_{\max}$ on search frequency f_s.

(D) Comparison of Systems.

The results obtained above permit an instructive comparison of performance capabilities of the various search systems. Since a number of assumptions have had to be made, comparisons based on the relations developed have only approximate significance. The following representative conditions will be used as a basis of comparison:

Mean Radio Frequency F_o:	4000 megacycles per second
Wavelength λ_o:	7.5 centimeters
Average Transmitted Power P_t:	1.0 watt
Effective Antenna Area A'_a:	20 square wavelengths
Target Echoing Area A'_c (medium freighter):	1,420,000 square wavelengths (8000 square meters)
Receiver Noise Figure $\overline{NF}$:	40 (32 decibels), with no image rejection
Selectivity:	Set for best signal/noise ratio at maximum range
Range Difference Resolved δR:	1000 feet (at mean range if not constant)
Frequency of Complete Search f_s:	10 per second
Number of Successive Range Elements N:	1 and 100

Under these conditions, the dimensionless signal/noise coefficient $\frac{P_t A'^2_a A'_c}{2\pi F_o \overline{NF} kT}$ has the value of 1.38×10^{17}.

The results of the maximum-range calculations for all five systems considered are displayed in Figure 6, where R_{max} is plotted against f_s, both on logarithmic scales. The inverse-sixth-root relation of Equation (20) gives the straight line marked "R_{max}, pulse." For multiple-selector FM search, Equations (23) and (24) together give the full-line curve marked "R_{max}, freq. mod. $(N = 1)$", as well as the lost-time fraction shown by the light-line curve marked "$\rho_m (N = 1)$." For either wave-analyzer or variable-sweep-width operation with 100 range elements searched successively per range scan, Equations (23) and (24) together give the dashed curves marked "R_{max}, freq. mod. $(N = 100)$" and "$\rho_m (N = 100)$." The dashed and full-line curves have the same shape but different coefficients for the coordinates. In fact, it is evident from the form of Equations (23), (24), and (20) that if the coordinate scales of Figure 6 are read as x and y, with the significance given in the figure caption, the solid curves and lines are generally applicable for all values of P_t, A'_a, A'_c, $\overline{NF}$, F_o, and N. For search by variable sweep duration the dotted curves marked "R_{max}, var. dur. mod. $(N = 100, \delta R = 1000)$" and "$\rho_m$, var. dur. mod. $(N = 100, \delta R = 1000)$"; have been calculated, but these apply only for the special values of N and δR indicated. R_{min} as well as R_{max} is determined in this case, and the dotted curves end abruptly when R_{min} falls to zero. Excessive lost time produces ambiguous indications with any system, setting the limits marked "R_{max} $(\rho = 1)$" for all systems using fixed repetition frequency.

Pulse and multiple-channel FM systems evidently have comparable noise-limited performance. The former is favored by its high peak power and the latter by its low noise bandwidth, while time lost during propagation penalizes the FM system somewhat at high repetition frequencies. Theoretical maximum ranges for these two systems under the conditions tabulated are respectively 25 and 21 nautical miles at 2000 pulses per second, but at 10 pulses per second both systems operate to 60 miles. Relative to the fast but complex multiple-selector FM

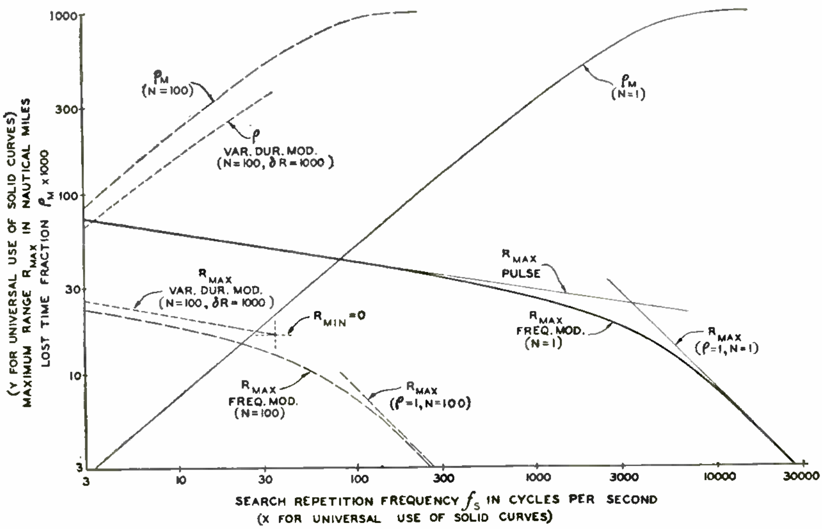


Fig. 6—Dependance of maximum radar range on search repetition frequency. For universal use of solid curves,

$$x = 2.13\,(3340\,N/F_o)^{9/10}\left(\frac{P_t A'^{\,2}_a A_c'}{2\pi F_o \overline{NF} kT}\right)^{3/10} f_s$$

and

$$y = 22.7\,(3340\,N/F_o)^{1/10}\left(\frac{P_t A'^{\,2}_a A_c'}{2\pi F_o \overline{NF} kT}\right)^{-3/10} R_{max}/\lambda_0$$

system, all the relatively simple but slow single-selector FM search systems are very heavily penalized by inefficient use of time, any single desired signal falling outside the noise-pass band during most of the search. The variable-sweep-duration system shows a slight range advantage, perhaps because more of the search time is applied to the longer ranges, but requires a filter with variable selectivity. Ranges at 10 search repetitions per second are respectively 18, 18, and 21 nautical miles for the three single-selector systems considered. Slowness of the single-selector FM systems is a serious drawback if directional as well as range scanning is required.

Radio-frequency channel width for 1000-foot resolution is about one megacycle per second for all systems, with lost time penalyzing the FM systems quite heavily if f_s and consequently ρ are made high. Constant fractional range resolution, as found for the variable-sweep-width system, is uneconomical for operation to extreme short ranges but may otherwise be well suited to meet particular operational requirements for search resolution.

Linear variation of transmitted frequency is essential if the full resolution capability of FM radar is to be attained, as even a slight variation of $\dot{F}$ during sweep may smear the beat frequencies produced by a single target over several adjacent range channels. Linear circuit characteristics are also essential, as intermodulation among multiple beat notes will cause indication of ghost targets.

(E) Effects of Speed.

The foregoing discussion of FM search radar has assumed stationary targets and radar. Target motion will shift the radar beat frequencies in just the same way as it does in single-target operation. If each range element is examined throughout a full modulation cycle, an approaching target will be found to decrease the beat frequency during modulation upsweep and to increase frequency during downsweep. Any single moving target will thus appear split in range by an amount proportional to its speed, with the center of the pair of indicated targets at the true range. This will require the speed of search to be cut in half, so that each target will appear on both an upsweep and a downsweep, but can be quite valuable in distinguishing motion.

Rangewise motion of the radar itself will split all targets equally, which is usually uninformative and should be avoided. Correction for speed of radar can be made, in the case of triangular modulation, by making durations of upsweep and downsweep unequal. This correction must be made proportional to speed, but fortunately turns out to be independent of range and sweep width; the correction factor is independent of modulation frequency also.

V. Conclusions

Frequency-modulated radar can easily provide separate determination of both range and speed of a single isolated target. The data provided is well suited for automatic measurement and control by simple means.

In range search for multiple targets, frequency-modulated radar with multiple selective channels should give performance substantially equal to that of pulse radar for equal average transmitted power. FM search radar using repeatedly a single selective channel should be markedly inferior to pulse radar in noise-limited maximum range. At the present state of the art, multiple-selector FM equipment would be unduly complex.

Because of its efficient use of a relatively narrow noise band, frequency-modulated radar does not require the high peak powers used in pulse radar. Because transmission and reception occur simultaneously, the problem of protecting reception against disturbance by the transmitted signal is more difficult for FM than for pulse radar, particularly if a single antenna is to be used for both transmission and reception.

High range resolution or operation at very short ranges require a very wide radio-frequency band. The technique of using such bands is at present more satisfactory for frequency-modulated than for pulsed systems.

SYMBOLS

A'_a	Effective area of antenna in square wave lengths.
A'_e	Effective echoing area of target in square wave lengths.
c	Speed of radio-wave propagation, 983.2 feet per microsecond for normal sea-level air.
e	Data-voltage output of FM radar.
e_0	Bias component of data voltage.
e_1	Threshold voltage for control operation; also, direct-signal radio-frequency voltage vector.
e_2	Target-reflected radio-frequency signal-voltage vector.
e_3	Resultant radio-frequency signal-voltage vector.
f	Frequency of radar-beat signal.
f_o	Center frequency of filter pass band.
f_d	Net radar-beat frequency during downward sweep of frequency modulation.
f_m	Frequency at which transmitted-signal frequency is modulated.
f_r	Repetition frequency of pulse or single FM sweep.
f_R	Radar-beat frequency due to range.
f_s	Repetition frequency of complete range search.
f_S	Radar-beat frequency due to speed.
f_u	Net radar-beat frequency during upward sweep of frequency modulation.
$\triangle f$	Noise bandwidth of receiver or filter.
F	Frequency of radio signal.
F_o	Average radio frequency of FM signal.
F_1, F_2	Particular values of radio frequency, especially those at limits of modulation sweep.
F', F''	Radio frequencies modified by target range or speed.
$\dot{F}$	Time rate of change of modulated radio frequency.
h_R	Range-data conversion sensitivity, in volts per cycle per second.
h_S	Speed-data conversion sensitivity, in volts per cycle per second.

k	Boltzmann's constant, 1.37×10^{-23} watt-seconds per degree centigrade.
k_R	Range sensitivity of FM radar, in cycles per second per unit range.
k_S	Speed sensitivity of radar, in cycles per second per unit speed.
K	Numerical factor describing effect of reflecting surface on signal strength in radar altimetry.
M	Number of range elements examined simultaneously in search operation.
N	Number of standing waves between radar and target; also, number of range elements examined successively in one complete range search.
N_0	Average number of standing waves during modulation cycle.
N_1, N_2	Standing-wave members at limits of frequency modulation.
$\overline{NF}$	Noise figure of receiver.
$\triangle N$	Change in number of standing waves during single frequency sweep of modulation.
$\delta(\triangle N)$	Difference of $\triangle N$ for two targets present at once.
P_r	Average power in received signal.
P_t	Average transmitted power.
R	Range or distance between radar and target.
R_{max}, R_{min}	Limits of range search.
R'	Range in wave lengths.
$\dot{R}$	Time rate of change of range.
δR	Difference in range of two targets just resolved; also, fixed range error due to cycle counting.
S	Relative speed of approach of radar and target.
t_1	Particular instant of time.
t_m	Period of frequency modulation.
t_p	Duration of radar pulse.
t_r	Duration of single frequency sweep.
T	Noise temperature.
U_r	Total useful energy in received pulse.
U_t	Total useful energy in transmitted pulse.
U_{min}	Minimum received-pulse energy discernible over noise.
W	Width of frequency band swept in modulation.
δ	Fractional range difference between two targets just resolved.
λ	Wave length of radio signal, c/F.
λ_0	Wave length corresponding to mean frequency F_0.
λ_w	Sweep wave length, corresponding to frequency W.
ρ	Fraction of duration of frequency sweep lost in radio-signal propagation.
ρ_m	Lost-time fraction at maximum range.
τ	Time required for signal to propagate from radar to target and back.
ψ	Phase difference between direct- and reflected-signal radio-frequency voltage vectors.
ψ_0	Average phase difference over modulation cycle.
ψ_1	Initial phase difference.
$\triangle\psi$	Change in phase difference between limits of frequency modulation.

SIMULTANEOUS FIELD STRENGTH RECORDINGS ON 47.1, 106.5 AND 700 MEGACYCLES*

BY

W. L. CARLSON

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

Summary—*During the summer of 1946 radio transmissions from New York City on 47.1, 106.5 and 700 megacycles were simultaneously recorded at Princeton, N. J. The recordings were continued intermittently until May 21, 1947 on 47.1 and 106.5 megacycles. During the early afternoon hours the signals usually held the same steady values which are assumed to be the field strengths which occur when the dielectric gradient of the medium is normal. Abnormal dielectric gradients were more favorable to reception on the higher frequencies. The reception on 700 megacycles was characterized by occasional high field strengths particularly during rain storms at the receiving location.*

INTRODUCTION

AT THE REQUEST of the Federal Communications Commission, equipment was set up at Princeton, N. J. for automatically recording field strengths from the following stations located in New York City.

WBAM	47.1 megacycles	Bamberger Broadcasting Service
WBAM	106.5 megacycles	Bamberger Broadcasting Service
W2XCT	700 megacycles	Columbia Broadcasting System

The investigation provided an excellent opportunity to observe the relative affect of atmospheric conditions on wave propagation at different frequencies over a distance of 45 miles. The information obtained is useful in predicting television and FM broadcast reception near the optical horizon.

TRANSMITTERS AND RECEIVERS

The antennas for the 47.1 and 106.5 megacycle transmissions were each located 570 feet above street level at 444 Madison Avenue. These transmissions were frequency modulated. The antenna for the 700 megacycle transmission was located on the Chrysler Building, 909 feet above street level. This station used pulse modulation with a repetition

* Decimal Classification: R270.

rate of 300 cycles per second and a duty cycle of 1 to 10. The polarization was vertical for 700 megacycles and horizontal for the lower frequencies. The transmitting schedules were usually from Monday through Friday of each week.

The receiving antennas were located 50 feet above ground on the north-east side of the main RCA Laboratories building, free from local obstructions. A single dipole antenna was used for receiving both 47.1 and 106.5 megacycle signals. The 700 megacycle receiving antenna consisted of 4 dipoles and a reflector. Model S-36 and S-27 Hallicrafter receivers were employed in recording the lower frequencies. The Navy Model AN/APR-1 receiver was employed in recording the 700 megacycle signals. The outputs of the receivers were connected to an Englehard recorder which made a record of each signal once every minute.

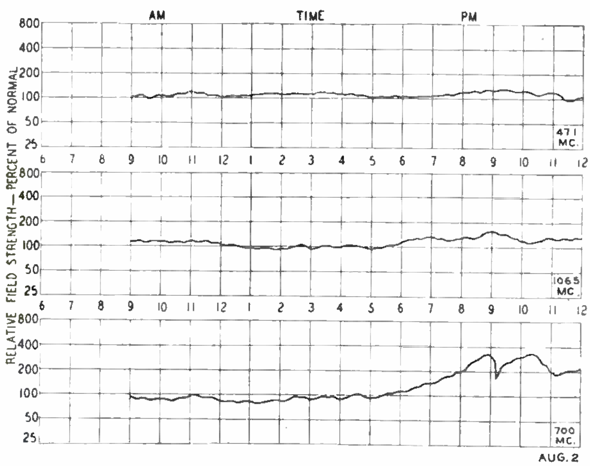


Fig. 1—Field strength patterns—August 2, 1946.

Recorded Data

During the early afternoon hours the recordings usually reached steady value which is assumed to be the normal field strength which occurs when the dielectric gradient is that of "standard atmosphere".

The field strength patterns for the three frequencies are shown in Figure 1 for August 2nd. This represents a typical comparison of relative field strengths often observed between the hours of 9 a.m. to midnight. Daily recordings similar to those in Figures 2 and 3 for July 17th and August 1st also occurred fairly often.

The most definite correlation between field strengths and atmo-

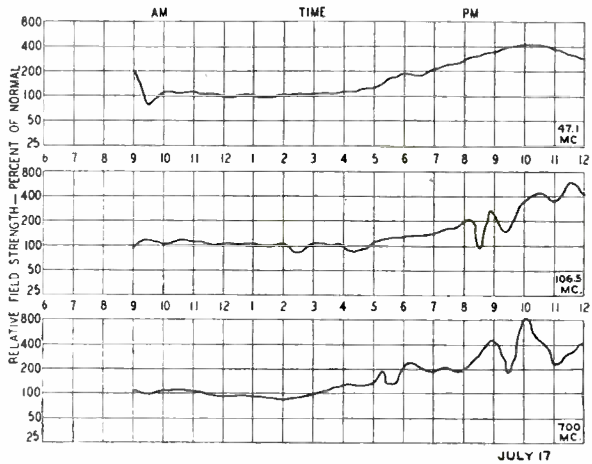


Fig. 2—Field strength recordings—July 17, 1946.

spheric conditions observed was that the 700 megacycle signal *increased* 2 to 8 times *during local rain storms* at Princeton. The 47.1 megacycle signal usually dropped off at intervals during these periods. The performance on 106.5 megacycles was less predictable.

During the afternoon of June 5th a thunder storm broke over Princeton at 2:10 p.m. This was followed by another local storm at 5:00 p.m. The relation between the storm and the signal strengths is shown in Figure 4.

On June 29th a thunder storm broke over Princeton at 3:30 p.m. accompanied by a rise in the 700 megacycle signal which reached values

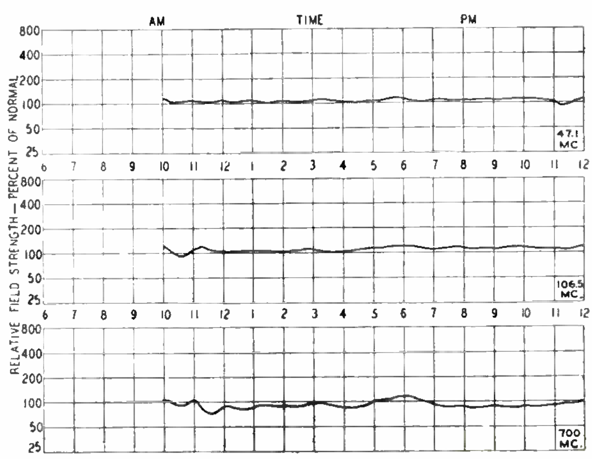


Fig. 3—Field strength recordings—August 1, 1946.

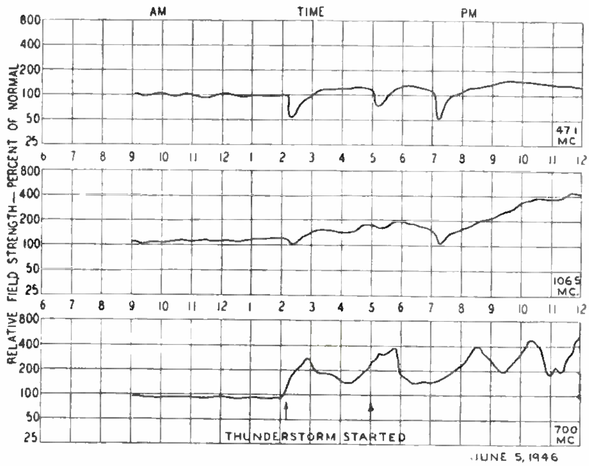


Fig. 4—Effect of thunderstorm on received signal strengths.

over 8 times normal intermittently during the remainder of the day. (See Figure 5.) Again on June 11th a similar rise was recorded for 700 megacycles during a thunder storm and heavy rains between 8:00 p.m. and midnight. The general trend was a reduction in signal strength for 47.1 megacycles during this period and a small rise for the 106.5 megacycle signal. (See Figure 6.)

A light rain at 4:45 p.m. on August 29th was followed by heavy rain at 5:30 p.m. and intermittent light and heavy rains until 8:30 p.m. The average field strength for 47.1 megacycles appears to have

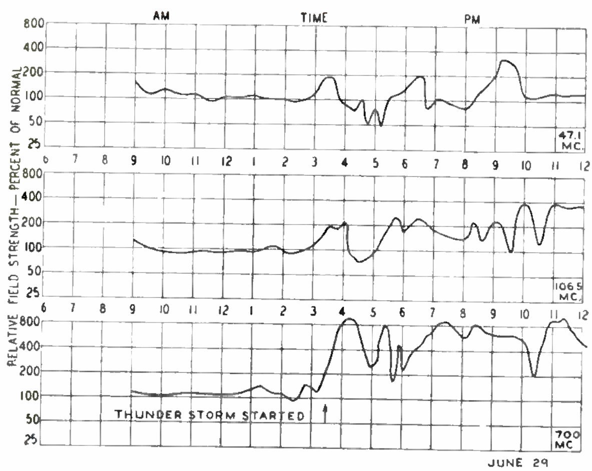


Fig. 5—Effect of thunderstorm on 700-megacycle signal—June 29, 1946.

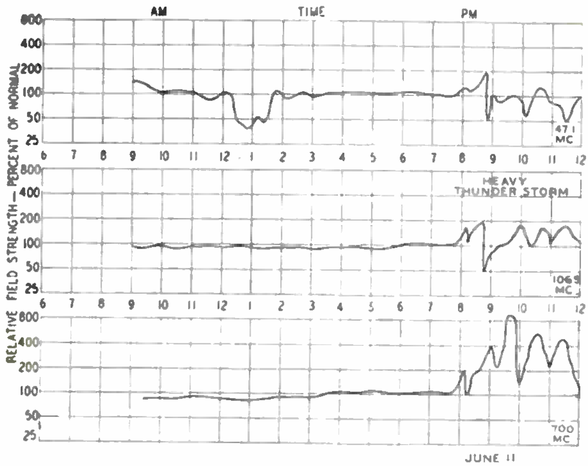


Fig. 6—Effect of thunderstorm and rains on signal strengths—June 11, 1946.

been depressed below normal by the storm (See Figure 7), while the 106.5 and 700 megacycle signals were increased above normal. Figure 8 shows a similar situation for July 22nd. On June 18th rain occurred about 3:00 p.m. This coincided with a rise in the 700 megacycle signal and a drop in the 47.1 and 106.5 megacycle signals. (See Figure 9.)

On a few occasions the signals fluctuated rapidly. For example, during the evening of July 9th, rapid variations occurred mainly on 47.1 megacycles. The weather was clear with moderate winds (New York at 1:30 a.m. July 10th). During the evening of July 10th the signal on 106.5 and 700 megacycles often varied 6 to 12 decibels within

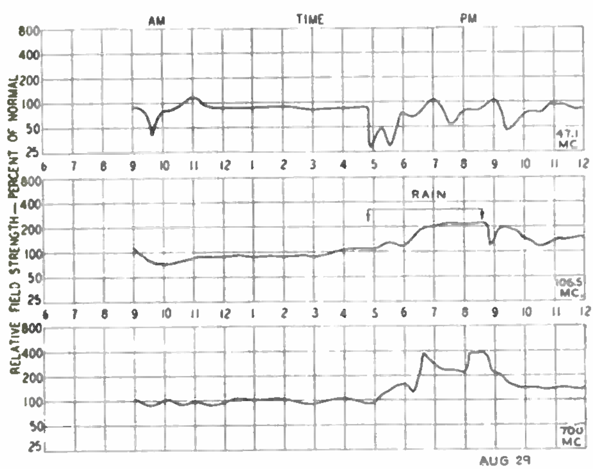


Fig. 7—Effect of rains on signal strengths—August 29, 1946.

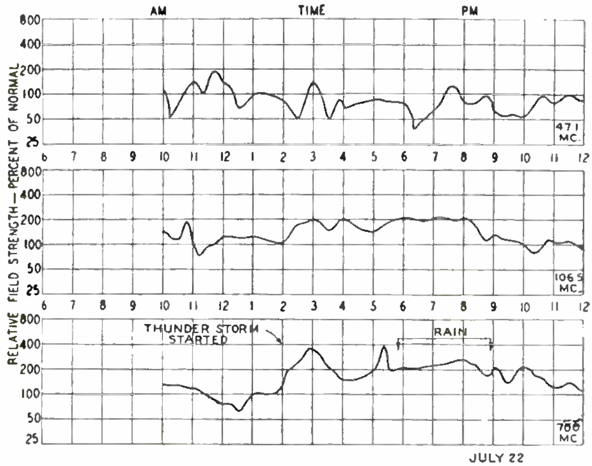


Fig. 8—Effect of rains on signal strengths—July 22, 1946.

a few minutes. The cloud ceiling was 400 feet with moderate winds (New York at 1:30 a.m. July 11th). The weather was generally warm and humid during July 9th and 10th.

During the morning of September 9th the signal strengths fluctuated about 6 decibels at 5 or 10 minute intervals on 106.5 and 700 megacycles. All frequencies varied in strength rapidly during the evening hours. The curves for this day in Figure 10 do not show the most rapid fluctuations in field strength that occurred. The weather was warm and humid. At 1:20 a.m. September 10th at New York the ceiling was 100 feet with no winds.

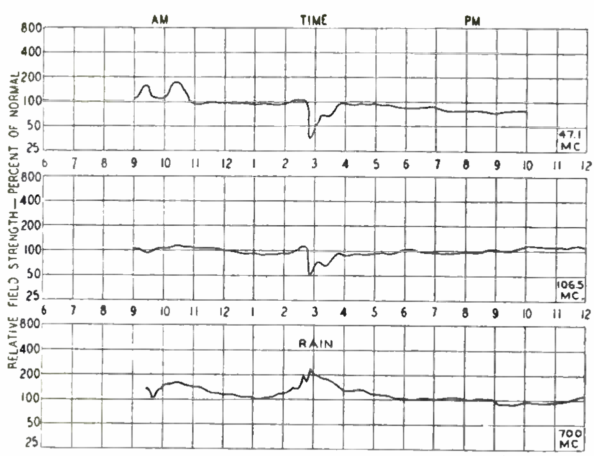


Fig. 9—Effect of rains on signal strengths—June 18, 1946.

Fig. 10—Signal strengths—September 9, 1946.

In analyzing the recorded data presented here, it is important to keep in mind that the 700 megacycle signal was favored by a transmitting antenna 340 feet higher than the antennas for the lower frequencies.

A summary of the recordings between June 3rd and September 12th is presented in Figures 11, 12 and 13. The data in Figure 11 is for 47.1 megacycles and 106.5 megacycles during the hours between 10 a.m. and 5 p.m. (daylight saving time). It is to be noted that both signals were 12 decibels above normal 0.1 percent of the total time. The 47.1 megacycle signal was 6 decibels below normal 3.5 percent of

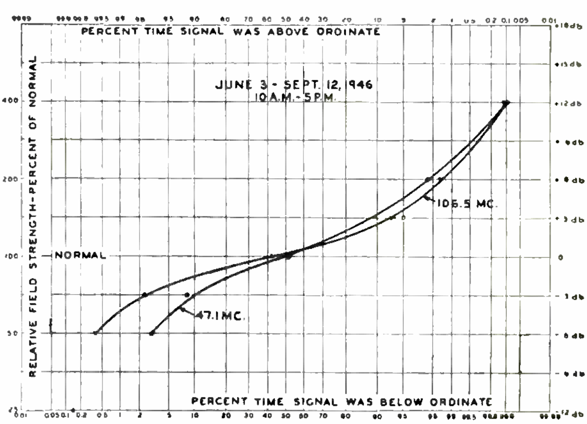


Fig. 11—Day recordings, 47.1 and 106.5 megacycles—June 3 to September 12, 1946.

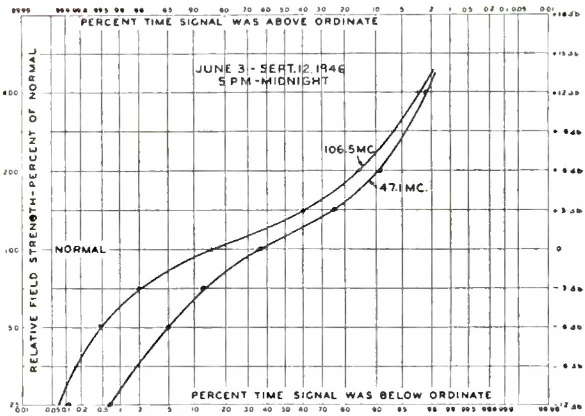


Fig. 12—Night recordings, 47.1 and 106.5 megacycles—June 3 to September 12, 1946.

the time while the 106.5 megacycle signal was 6 decibels below normal only .4 percent of the time. Figure 12 presents the performance for the hours between 5 p.m. and midnight. The signal dropped to − 6 decibels only one-tenth as often on 106.5 megacycles as compared to 47.1 megacycles.

Figure 13 shows the performance on 700 megacycles for the day time and evening. The signals were often abnormally high particularly during the evenings and rain storms. Transmissions on 700 megacycles were discontinued after September 1946.

Figures 14 and 15 are for the recordings between October 15, 1946 and May 21, 1947 on 41.7 and 106.5 megacycles. The field strengths

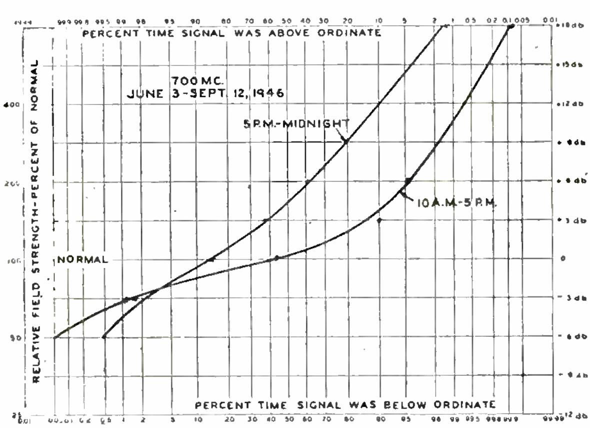


Fig. 13—Recordings on 700 megacycles—June 3 to September 12, 1946.

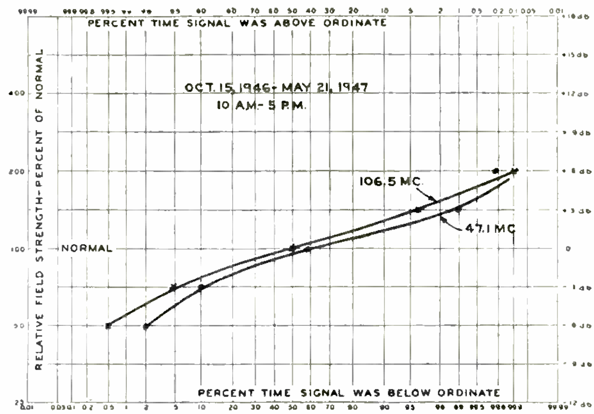


Fig. 14—Day recordings, 47.1 and 106.5 megacycles—October 15, 1946 to May 21, 1947.

were above normal less frequently during this period of recording and the trend in favor of 106.5 megacycles was less pronounced. Evidently the dielectric gradient of the medium departed less from "standard atmosphere" during these months than during the summer.

The recordings reveal a trend towards higher field strengths at higher frequencies when the "standard atmosphere" is disturbed.

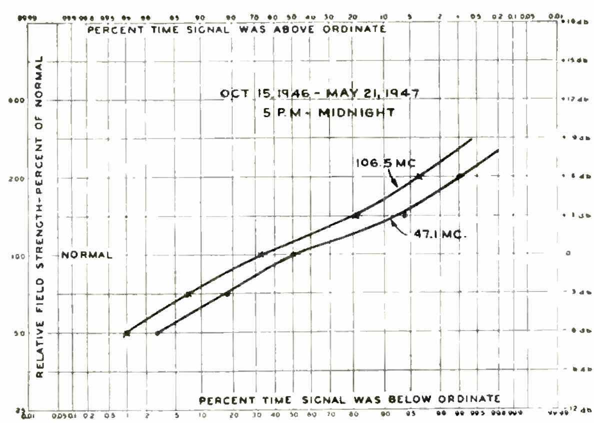


Fig. 15—Night recordings, 47.1 and 106.5 megacycles—October 15, 1946 to May 21, 1947.

TELEVISION DC COMPONENT*

By

K. R. Wendt

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

Summary—*Although important, and one of the oldest of the television techniques, the dc component is still one of the least understood. It is here explained in general, and with reference to transmitter and receiver applications. The various restorer circuits are described, with their advantages and disadvantages. Design considerations are given for equipment which handles the signal with the dc component present.*

Introduction

THE TRANSMISSION and reception of the picture dc component was one of the earlier television developments,[1-3] and the standards require that transmitted television signals include this component. It is paradoxial, then, that the dc component is one of the least understood of the television techniques and that received pictures often suffer in this respect.

During the development of television, transmitters and receivers have been operated quite satisfactorily for many long hours with test pattern modulation and no dc component. The dc component in this case remained fixed, and could easily be set manually at all points in the system. For program material, however, especially that including titles and fadeouts, the dc component becomes important and necessary. Poor blacks in the reproduced picture, caused mainly by poor dc insertion at the transmitter and poor restoration at the receiver, have often been responsible for low quality hazy pictures. This, like many another property judged subjectively, has been blamed on some innocent parameter such as bandwidth or number of lines.

Proper dc restoration is also important for receiver synchronization, since the failure of the dc restorer of the sync separator when noise is present on the signal, constitutes the real and limiting weakness in receiver synchronization.

---

* Decimal Classification: R583.1.

[1] F. Gray, U. S. Pat. 2,274,686, March 3, 1942 (Filed 9-16-31).

[2] R. D. Kell, U. S. Pat. 2,289,914, July 14, 1942 (Filed 9-26-31).

[3] R. S. Holmes, U. S. Pat. 2,251,677, August 5, 1941 (Filed 2-28-33).

Definition

The dc component has been defined as that portion of the television signal which contains information concerning the average light in the televised scene. For a fixed picture, such as a test pattern, this definition yields a definite number as the dc component. However, with a rapidly varying picture, the concept of averaging becomes complicated. The only definition adopted as a standard is that regarding dc transmission from the transmitter, which states, "It shall be standard that the black level be represented by a definite carrier level, independent of light and shade in the picture". This standard implies that the dc component is that portion of the signal which is necessary to maintain the black level constant. While this is a very useful concept, an examination of our meaning of the small letters dc can lead to a simple definition. The term dc, written as small letters, is not used as an abbreviation of the words "direct current", as illustrated by the often-used term "d-c voltage", which would be read as "direct-current voltage". Rather, the term dc component may be read as the "varying unidirectional" component. The definition of the television dc component then becomes the varying unidirectional portion of the complete video signal. A little reflection will show that the complete video signal must contain a dc component. That is, if the signal is to be representative of the light, then zero light, or black, must correspond to zero signal, or the "black level." All values of light must then be represented by various signal levels, extending in only one direction from zero, which signal must then possess a unidirectional, or dc component. Furthermore, as long as the light is maintained, the signal must also be maintained.

Figure 1a shows a signal which, for the first few fields, is all black and consists of only the sync pulses which extend beyond the black level. At time t_1 the signal goes to all white and remains there except as it is interrupted by the horizontal and vertical blanking pulses. If this signal were passed through a capacitor, and the dc component lost, it would appear as in Figure 1b. The white has slipped to dark grey, and the black is far beyond the true black level. A practical example would be one in which, for instance, a dancer in white might be picked up in front of a black curtain. Let us assume that the picture reproducing device, or kinescope, has been properly set so that the white costume is shown by a satisfactory highlight value, and the black areas of the curtain appear black on the kinescope. The rest of the troupe now comes into the scene. If the dc component is present on the grid of the kinescope, the scene will be reproduced properly. If it has been lost, the highlights of the dancers will fade to grey, and the dark grey area of the curtain will disappear into black.

Another example might serve to correct an erroneous impression. Suppose that for some reason not called for in the script, the studio light is reduced, giving an undesirable result on the kinescope. Since the average light, and therefore the dc component, are reduced, this is proof to some that the dc component is undesirable. Actually, however, the example involves not just the dc component but all ac components as well. The whole signal must be increased by means of the gain control to restore it to a level equal to that with the greater illumination. The dc component was no more at fault than the other portions of the signal.

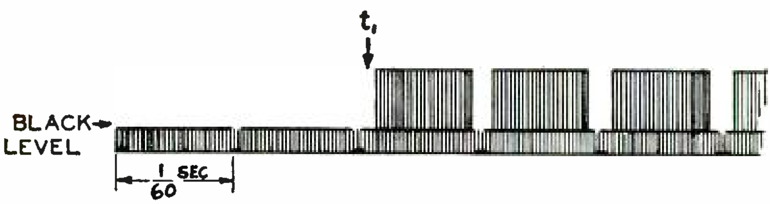


Fig. 1a—DC component present.

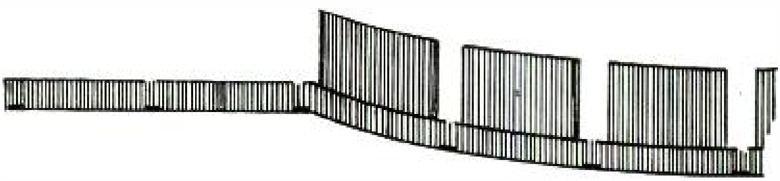


Fig. 1b—DC component lost.

The dc component is important and necessary. Actually, it may be dispensed with only in applications in which the average illumination remains constant, such, for instance, as the reproduction of indicating devices like meters, which change position but not brightness.

THE METHOD OF DC TRANSMISSION

Practically, it is impossible to use direct coupling through the whole system from the camera to the kinescope. The dc component, then, is essentially transmitted by a carrier system in which the ac portions of the video signal are amplitude modulated by the dc component. The dc component may be obtained by a peak detector from the ac video signal. Actually, a simple detector arrangement adds the dc component to the ac signal components, producing as a result the complete video signal. Such a circuit is called a dc restorer, since the dc components of the signal have been restored to it. It may be thought of simply as a device for holding the black level of the signal at a fixed dc value.

The sync peaks which operate the dc restorer occur at 15,750 cycles. Therefore, it should be possible to transmit, as parts of the dc component, all frequencies up to approximately 1/10 of the carrier, or about 1500 cycles. Such operation is quite possible, providing, as explained later, a keyed dc restorer is used, rather than the simple one. In spite of the complication, such operation is often used. It makes possible small coupling capacitors, and complete freedom from low frequency troubles such as hum and low frequency noise; the latter is often quite annoying on the end of a long cable running through metropolitan areas. Street cars and other power circuits with a ground return add a low frequency signal which can best be removed by a fast keyed dc restorer.

A simple dc restorer, however, is not able to restore frequencies above 60 cycles without considerable distortion.

The Transmission, or Initial Insertion

Before the dc component may be restored and used, it must be initially inserted into the signal in the studio control equipment. This should not be confused with the radiation of the dc component by the transmitter. Although the latter is definitely advantageous, it is not necessary. That is, satisfactory pictures could be obtained if the transmitter were modulated with only the ac portion of the signal. However, it is necessary, if any dc component is to be received, that it be originally inserted into the signal. This process consists of simply holding the black portions of the picture signal, as received from the camera, at the standardized black or blanking level of the video signal, to which level the signal returns at the end of each line and frame.

Figures 2a, b, and c show three types of signal in which the dc component has been properly inserted. Figures 2d, e, and f show similar signals in which the dc component has not been inserted. Here, as in all cases in which the dc component is not present, a picture which should be predominantly black or white appears as grey, and narrow peaks in the opposite direction extend beyond their proper levels.

It is possible for a skilled operator, with the aid of one or more rehearsals, to insert the dc component manually. This is almost necessary with iconoscope signals, although a complicated system using a photocell may be made to operate. With signals from orthicon type tubes, flying-spot pickups, or image dissectors, the dc insertion may be accomplished automatically. When properly operated, these pickup devices produce during the blanking interval, a signal which corresponds to that with no light. Some sort of dc restorer is then used to hold this "no light" or black level such that when the signal has been properly blanked and clipped, the picture black signals occur at the signal black or blanking level.

THE SIMPLE DC RESTORER CIRCUIT[4,5]

Figure 3 shows the simple or basic dc restoration circuit, and a signal as it would appear across R.

Figure 4 is a somewhat simplified diagram in which an ac generator feeding through the driving resistor R_L is substituted for the driving tube, and the grid is shown as a diode which may be considered as R_D.

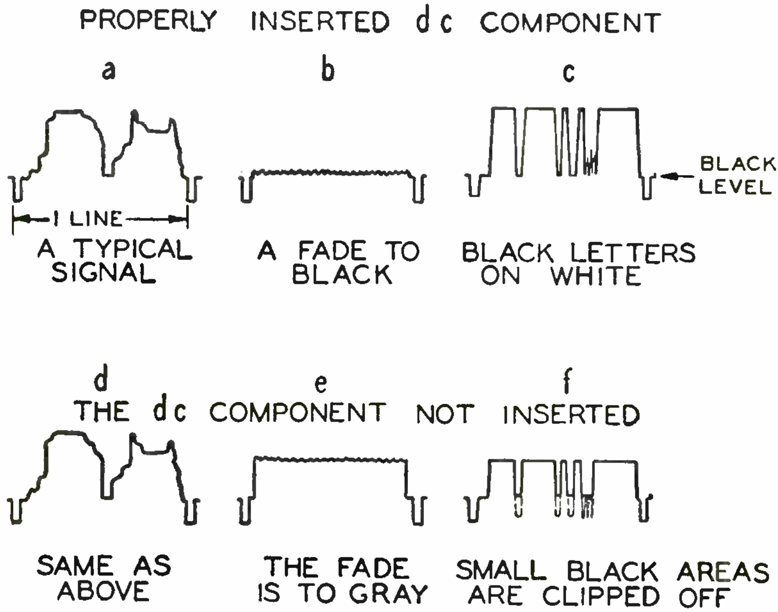


Fig. 2—Insertion of dc component.

R_D of course is non-linear, being the E_p-I_p diode characteristic. It varies from infinity to a few hundred ohms, depending upon the voltage across it.

The operation of this circuit may be described as follows:

When the plate is driven positive by a synchronizing peak, current flows through R_D, C, and R_L, and a charge is accumulated on C. During picture time R_D becomes open, and the capacitor partially discharges. The charge which flowed into C through R_D now flows out

[4] P. W. Willans, U. S. Pat. 2,252,746, August 19, 1941 (Filed Gr. Brit. 4-13-33).

[5] D. E. Foster and J. A. Rankin, "Video Output Systems", *RCA REVIEW*, Vol. V, No. 4, pp. 409-438, April 1941.

through R. This current establishes the dc component voltage across R, which, at equilibrium, is much larger than the incremental charges and discharges of C. The E_{dc} thus generated, appears across C, and may be considered as being added to the ac signal across R_L to produce the combined restored signal across R. The result is that the most positive portions of the signal, which are the sync peaks as shown in

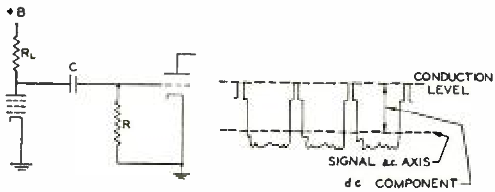


Fig. 3—The simple dc restoration circuit.

Figure 3, are held at the conduction level of the diode, and the rest of the signal extends negatively from this level. If some sync peaks reach too high, higher than average current flows, placing a greater charge on C, and conversely a low peak draws less current and the total charge on C will decrease. The circuit thus not only establishes a dc axis at the sync peaks and holds this axis at the conduction level, but also tends to line the peaks up evenly.

The necessary correcting signal is produced across C, which, when added to the input, gives the restored output. It does not matter what the source of the sync peak misalignment may be—hum, "clothesline", "bobble", too few or too many lows—the required correcting signal tends to be produced across C. Depending upon the circuit constants, however, there are limits beyond which the circuit is unable to function correctly.

This dc component which appears across C has superimposed upon it, a series of small sawteeth, occurring at sync-peak frequency, or 15,750 cycles. Since this sawtooth waveshape is added to the input to

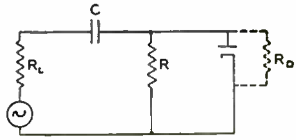


Fig. 4—Simplified equivalent of Figure 3.

produce the output, any knowledge of it would aid in explaining the operation or shortcomings of the restorer. However, it is quite difficult to calculate anything of value concerning it due to the non-linearity of the charging circuit and the complexity of the video signal. The discharge of the capacitor is through a linear circuit, however, and can be calculated. The peak to peak discharge is equal to

$$E_c = E_o \left(1 - e^{-\frac{t}{RC}}\right).$$

In most dc restorer cases, $\frac{-t}{RC}$ has a value smaller than .1, and the equation simplifies to

$$E_c \simeq E_o \left(\frac{t}{RC}\right)$$

in which E_c is the peak to peak value of the sawtooth, E_o the value of the charge at the start of the discharge, t the time of discharge in seconds, and R and C are in ohms and farads. The equation tells us that the sawtooth is proportional to the signal level and the time available for discharge and inversely proportional to the time constant RC. This is of some help, as will be explained in the following sections. However, the low frequency portions of the signal come mainly from variations in the charge of the capacitor, which follows no simple law.

Impedance Considerations

The impedances of a dc restorer circuit, meaning the value of R, and of the diode resistance R_D, are often considered as being relatively fixed. R is often made 1 megohm, and R_D assumed as a few thousand ohms. R, however, may in some cases be made much higher, with considerable improvement, and R_D should in all cases be made as small as possible. It can be shown that the ratio R_D/R should be made as low as possible, as follows: If R_D is other than zero, then a voltage E_D must exist across it if current is to flow. As the signal changes, and a different current is required, this voltage must also change, producing an objectionable irregularity in the alignment of the top of the sync peaks. This objectionable voltage E_D equals $I_D R_D$, which is proportional to $E_{dc} \frac{R_D}{R}$, since I_D is proportional to $\frac{E_{dc}}{R}$. To minimize E_D, R_D/R should therefore be as low as possible.

The ratio R_D/R can be changed, of course, by varying either of the two values. Unfortunately, however, when R is varied, R_D also changes in the same direction, and the ratio change is small. This is due to the non-linear E_g-I_g characteristic of triodes and pentodes, or E_p-I_p of diodes.

These characteristics, as shown in Figure 5, are quite curved near cutoff. Some of the difference in these curves is caused by a fixed

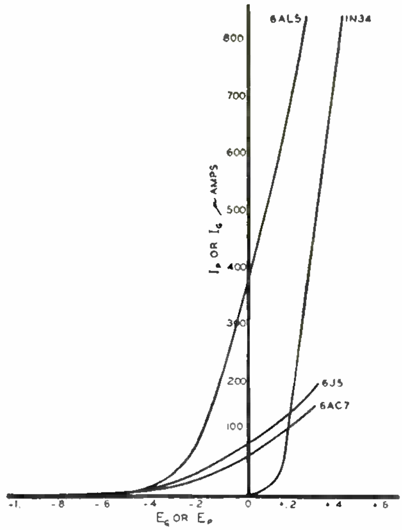


Fig. 5—Grid current curves.

series voltage composed of contact and thermal potentials. Figure 6 therefore shows these curves replotted with their cutoff points (approximately .1 microampere) coinciding. It is apparent that R_D varies considerably with the rectifier type, and that a IN34 germanium crystal has the sharpest bend in the curve.

However, the reverse or leakage current of the Germanium crystal appears in the circuit as a resistance across R, therefore limiting the

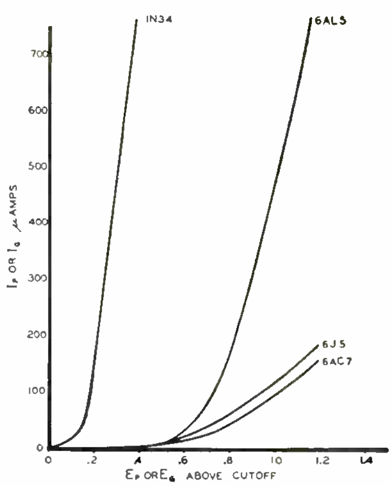


Fig. 6—Grid current curves.

value of R_D/R. As these crystals are improved in quality and uniformity, they may become quite important for dc restorer operation.

If an attempt is made to decrease R_D/R by increasing R, less total current will be required through R to produce E_{dc}. Therefore less pulse current I_D need flow, and the current will be drawn at a point on the curve which has a higher slope, or resistance, thus appreciably nullifying the intended change in the ratio of R_D/R. Stated in another way, if I_D were reduced from 100 to 50 microamperes, using a 6AC7, the voltage across the diode, which is E_D, would be reduced only from .8 to .6 volt.

This curved E_g-I_g characteristic has another important effect—R_D decreases with increasing signal amplitude. That is, as the signal level increases, and more diode current is drawn, the diode resistance R_D decreases, and the ratio R_D/R likewise decreases, giving improved performance. This leads to an important rule: *dc restorers using peak rectifiers should be operated at as high a signal level as possible.*

The value of R is important in another respect. Too low a value of R may appreciably load the video driving impedance. This effect is illustrated by the following numerical example.

Assume a circuit with the driving impedance R_L equal to 4700 ohms, the resistor R equal to 1 megohm, and a peak signal level (E_{dc}) of 10 volts. The current through $I_R = \dfrac{10}{1 \times 10^6} = 10^{-5}$ amperes $= 10$ microamperes. The sync peak is 8 per cent wide, and hence current flows into the capacitor C through the diode for 8 per cent of the time, and out through R for 92 per cent of the time. The current I_D is then 92/8 of 10 microamperes, or 115 microamperes. This I_D flowing through R_L then produces a voltage drop

$$\begin{aligned} E_{RL} &= 115 \times 10^{-6} \times 4700 = .54 \text{ volts} \\ &= 5.4 \text{ per cent of the signal.} \end{aligned}$$

An all white 10 volt peak signal will have approximately 3.5 volts of sync pulse. Hence .54/3.5 = 15 per cent of the sync signal will be lost in the video driving circuit due to the loading effect of the 1 megohm resistor. This loading effect varies as the dc component varies, thereby contributing to the overall distortion. In this case, the effect is essentially a shortening of the sync pulses, which appears as an irregular, or varying blanking level. The blanking level in fact varies as a result of the sum of the loading effect and the diode drop E_D, since both are maximum under the same conditions, and each produces a distortion of the blanking level in the black direction. That is, an increased diode drop causes the sync peaks to appear at a higher level, and increased loading effect shortens the sync pulses, thus doubly raising the blanking level.

One megohm is thus seen to be approximately the lower limit for the resistance *R*. The upper limit, with two exceptions, is dictated only by practical attainable values. With careful design and quality capacitors, *R* may be made as high as 100 megohms. The two cases in

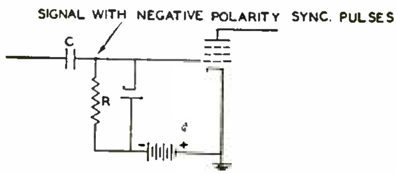


Fig. 7—Circuit in which gas current may cancel diode current.

which *R* must be kept low are, first, operation with adverse signal to noise conditions, which is treated separately below, and second, operation in a circuit in which gas current opposes the diode current. The latter condition is illustrated by the circuit of Figure 7, in which the dc restorer is a diode operating on a signal with negative sync pulses, and the restored output is applied directly to the grid of an amplifier tube. Gas current in the amplifier tube may be thought of as electrons leaving the grid. Current through *R* is thus divided, and the diode current is reduced by the gas current. Unless *R* is small, and I_D large, the gas current can make the diode and the dc restorer completely inoperative. For this reason, the circuit of Figure 7 should be avoided if possible.

A further example, taken from the circuit used in a prewar television set illustrates the magnitude of the diode drop and loading effect. The circuit was unsatisfactory due to the low signal level. Figure 8 shows a somewhat simplified diagram in which the second

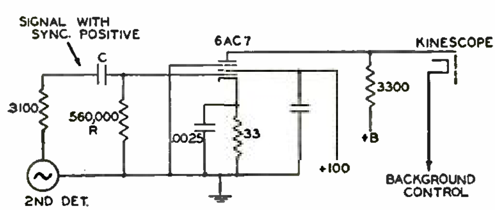


Fig. 8—A pre-war video output stage.

detector is shown as a voltage feeding through 3100 ohms, the approximate equivalent of the actual circuit. The restorer is slow, using a 1 microfarad capacitor. *R* is low, being 560,000 ohms. The 33-ohm cathode resistor protects the tube somewhat under no-signal conditions, and boosts the video high frequencies slightly. It also penalizes the dc restorer.

With an all-black signal, the current in the tube and through the 33 ohms is high, giving the cathode a plus voltage. The diode, or grid current, which is low because of the signal, is reduced even further by the cathode voltage, thus increasing the variation in I_D with picture content. This effect is opposed, however, by the grid current conduction with negative values of voltage. The grid current does not become zero until -1 volt is applied. This in effect applies 1 volt in series with R in the direction to cause conduction, considerably reducing the variation in I_D.

The following calculations show that the black level varies about .4 volt or nearly half the sync amplitude between an all-black and an all-white picture. It would cause objectionable variation in the picture blacks, except that the tube possesses some dc boost in its plate circuit. This dc restorer was replaced in most receivers by a diode working on the much larger signal at the kinescope grid.

	Black Picture	White Picture
	(Approximately)	
Maximum signal input peak to peak..	1 volt	3.5 volts
Sync	1 volt	1 volt
E_{dc}	.9 volt	3 volts
dc volts across 33 ohms	.4 volt	.1 volt
Cutoff voltage for I_g..............	1 volt	1 volt
Effective E_{dc} $(.9-.4+1)$, $(3-.1+1)$	1.5 volts	3.9 volts
I_R	2.7 micro-amperes	7.0 micro-amperes
Peak I_D $(I_R \times 11.5)$	31 micro-amperes	80 micro-amperes
E_D (taken from 6AC7 curve, Figure 6)	.73 volt	.96 volt
Difference between two E_D.........	.23 volt	
Peak drop across R_L $(I_D \times 3100)$....	.096 volt	.25 volt
Difference between two R_L drops....	.15 volt	
Total variation in black level........	.38 volt	

Slow and Fast Restorers

Although there is no sharp dividing line, dc restorer circuits are in general divided into slow and fast circuits.

Slow circuits are those in which RC is made large (usually .03 to .1 seconds). Sixty cycles are transmitted substantially as received, and only dc and very low frequencies can be generated across C. Slow circuits should be used in all applications in which the low frequencies are satisfactory on the input.

If the input contains an objectionable amount of incorrect low frequencies, such as poor signal lows, hum, or low frequency transients, a fast restorer, in which C is .01 microfarad or much less, is necessary.

The simple fast circuits, however, add characteristic errors to the signal, which for some purposes may be objectionable. The errors arise from two causes. First, from the change in wave form of the sync peaks during the vertical sync signal, and second from incomplete detection. The latter is made up mainly of the drop across R_L, and the diode drop, as described above.

The vertical sync error which occurs only in simple fast circuits appears as a countersinking of the vertical sync signal. It is due to the wide sync peaks which constitute the vertical sync signal. During the vertical sync signal the diode conduction to non-conduction time changes from the unfavorable ratio of 8 per cent to the very favorable ratio of 84 per cent. Due to this favorable ratio, the detection error is very low during the vertical, and in reality the vertical sync is not pushed in a hole, but the rest of the signal is pushed up by an amount equal to the detection error. The amount by which the vertical sync signal appears to be in a hole is therefore a measure of the detection errors. The error can therefore be reduced to zero only by reducing the detection errors to zero, requiring, for one thing, that R become infinity. Such is the case in keyed circuits, and a properly operating one has no vertical sync error, regardless of how fast it is.

As stated above, the vertical sync error does not exist in slow circuits, since C is large enough to prevent any appreciable change in its charge during the vertical sync interval.

A compromise, or medium time constant, or speed, is in general not as satisfactory as either a slow or fast circuit. The vertical error is not as large, but its effect lasts longer, preventing proper diode current from flowing for a number of lines after the vertical, thus causing poor restoration.

In simple fast circuits, the detection errors cause objectionable variations in the height of the sync peaks. The error is present regardless of the circuit speed, but in a fast circuit the error can change, due to picture content, within a vertical scan, resulting in horizontal streaks appearing in the picture. In slow circuits the error amounts to a partial loss of the dc component. The *variation* in this error due to picture content can be completely eliminated by the positive bias circuit described below under that heading. Unfortunately the positive bias circuit has no effect upon the vertical sync error.

NOISE CONSIDERATIONS

Performance on a signal containing noise is perhaps the most important design consideration for a dc restorer which must operate on such a signal. By its very nature, a dc restorer is sensitive to noise, and especially to impulse type noise. It operates on narrow 8 per cent peaks in the increased signal direction, in which direction the noise peaks also extend, and it responds quickly to a noise peak, but recovers slowly from it. In fact, for best performance the restorer must be aided by having the noise clipped from the signal as near as possible to the sync peaks.

There are, however, certain things which can be done to improve the performance of the simple dc restorer when the signal contains noise. If at all possible, the circuit should be slow, thereby integrating the effect of individual noise pulses over a long period, and thus reducing their amplitude effect. The value of R should be made as low as possible, limited only by the maximum allowable distortion under noiseless conditions. By this means, more energy is supplied by the signal to the circuit and the energy obtained from the noise becomes smaller by comparison. Such a low value of R further necessitates the use of a slow circuit, since the vertical sync error would become very large with low R values.

Greatly improved noise immunity is given by the keyed restorer, for at least two reasons. First, noise occurring between the keying pulses has no effect upon the circuit. Since the keying pulses are approximately 5 per cent, the improvement for this reason alone should be 20 times. Secondly, the circuit possesses as rapid a recovery as a response, which further greatly reduces the effect from any noise pulses which occur during the keying interval.

IMPROVED CIRCUITS

The Positive Bias Circuit[6]

This circuit, as shown in Figure 9, differs in the circuit arrangement from the conventional one only in having a positive voltage applied to the grid leak. It differs considerably, however, in operation. The dc voltage appearing across R is now made up of the sum of the signal dc and the applied bias. If the bias is several times the average signal dc, the total voltage, and hence the dc current through R, will vary only a few per cent from a white scene to the vertical sync signal, instead of several hundred per cent, as with zero bias. Since the cur-

[6] K. R. Wendt, U. S. Pat. 2,326,083, August 3, 1943 (Filed 3-1-41)

rent varies so little, all sync peaks will draw nearly the same peak current and reach to nearly the same height. By this method it is possible to reduce the *variation* in the diode drop and the loading effect, to practically zero.

The vertical sync error, however, is still present due to the variable discharge time, and can be reduced only by increasing C or R or the signal. When C is quite large, and the circuit is slow, the vertical sync error, of course, is not present.

The use of positive bias on the grid leak R requires that R be made quite large. This follows from the fact that the diode current should be kept to the same order as in the zero bias case. That is, the diode drop and the loading of R_L depend upon the diode current which, of course, all flows through R. When positive bias is to be applied to R, its value should be raised until I_R is approximately equal to its former value.

Suppose a signal with E_{dc} for a white picture of 10 volts is being restored by a circuit with $R = 2$ megohms, $I_R = 5$ microamperes. If

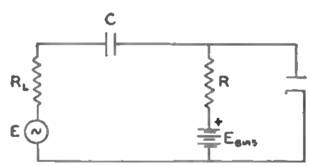


Fig. 9—Positive bias circuit.

the full $+B$ of 250 volts is applied to R, its value, with I_R maintained at 5 microamperes, would be

$$\frac{250 + 10}{5 \times 10^{-6}} = 52 \text{ megohms.}$$

When circuits with such high impedances are used, inadvertent circuit leakages may become important. However, in the circuit of Figure 9, leakage to ground causes almost no change in operation. Figure 10a shows the circuit of Figure 9 with resistor R_1 added, representing leakage to ground. Assuming the conditions above of $R = 52$ megohms, and $E_R = 250 + E_{dc} = 260$ volts, and $I_R = 5$ microamperes, suppose the leakage, R, to be equivalent to 10 megohms. The parallel combination of R and R_1 becomes 8.4 megohms, and the voltage at their junction, with no diode current, is 40.3 volts. The circuit is then equivalent to that of Figure 10b, in which $R_2 = 8.4$ megohms, and $E_2 = 40.3$ volts. The current $I_{R_2} = \dfrac{40.3 + 10}{8.4 \times 10^6} = 6$ microamperes, which is the same current that would have existed if the 10 megohms had paralleled the 2 megohms of the circuit with no positive bias. The bias voltage of course has been reduced, but is still large enough to obtain most of the benefit desired. Leakage to a voltage other than ground, such as through the coupling capacitor will of course change the current—but it can be shown to change the current in a positive bias circuit in practically the same amount as that in a zero bias circuit. Leakage is thus seen to be no more a problem in the high impedance positive bias circuit than in the lower impedance zero bias circuit.

The positive bias circuit has one weakness. It should not be used in an application in which the signal level may become quite low, such as under 1 volt, and the best operation is to be maintained. This comes

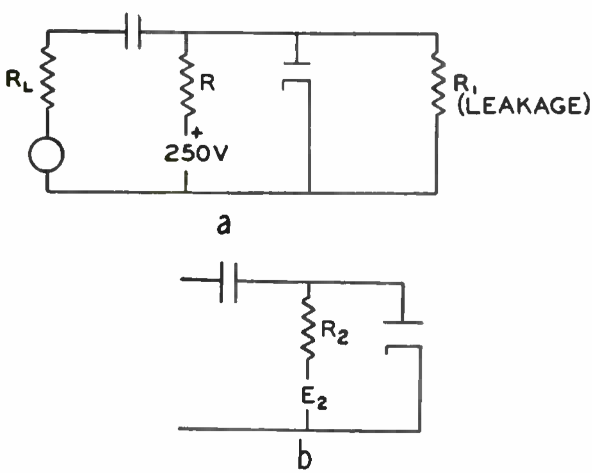


Fig. 10—The effect of leakage.

as a result of the distortion remaining constant with varying signal, and at low levels the distortion becomes too large a percentage of the signal.

The Low Boost Circuit

It has been found that the simple dc restorer works better if the signal supplied to it contains an excess of the low frequency (60 cycle) components rather than a deficiency. In explaining this action, it may be assumed that the dc restorer has failed, when one or several sync peaks fail to draw diode current. This happens when the downward slope of a line joining the sync peaks in the signal exceeds the rate of discharge of the capacitor. In a signal which has lost lows, a downward slope exists along the sync peaks during the vertical blanking interval, and an upward slope during the rest of the signal. The downward slope thus exists when the signal and E_{dc} are low, and the discharging current into the capacitor is small. In a circuit which has an excess of lows, an upward slope exists during vertical blanking, and a maximum downward slope during a white horizontal bar on a black background. However, the diode can more easily follow the downward slope on the signal with excess lows, because the slope occurs during a white signal, when a large voltage exists across R, allowing a faster discharge of C.

The exact amount of low boost is not critical, although a slow circuit can stand less than a fast one. Sufficient low boost to give an upward slope during the vertical blanking interval with an all-white picture will be found beneficial for all circuits except keyed ones, which are unaffected.

Figure 11 shows a circuit with low boost provided by an RC circuit, $R_F C_F$, in series with R_L in the usual manner in video circuits. Although

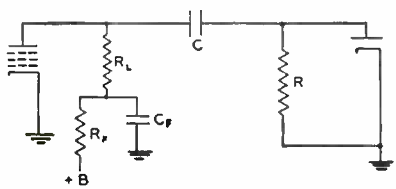


Fig. 11—Low boost circuit.

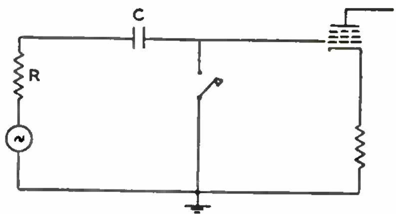


Fig. 12—The idealized clamp circuit.

the use of such a circuit might seem to belittle the advantage of the dc restorer, it is important to note that one low boost circuit and a dc restorer can properly restore lows which have been lost in a number of cascade circuits and according to different laws.

Keyed or Clamp Circuits

By using a keying signal, it is possible to obtain greatly improved dc restoration. Keyed circuits can be made very fast with low distortion and high noise immunity. They operate satisfactorily with signal levels much lower than for the simple restorer circuits.

Figure 12 shows an idealized equivalent of a keyed dc restorer. The key is operated, or closed, for a small interval during either the sync peak or the "back porch", which is the blanking interval following the sync peak. When the key is closed, the output voltage goes to ground. A charging *or* discharging current flows through C, limited only by R. C is small enough so that before the key is opened, it becomes completely charged, and current through it has dropped to practically zero. C now possesses a charge representing the difference between the signal voltage and ground. After the key is opened the charge can not change since no path exists for current to flow. The signal is transmitted through C as if it were infinite in size. When the keying interval again returns, the signal may be at an incorrect level, but the key when closed will force the output to the correct level, and the charge will be changed to agree with the new difference between the input voltage and the correct output voltage. If the level, however, needed no changing no current would flow into or out of C. The keyed circuit thus restores the dc component by holding the blanking level at a fixed voltage which may be considered the dc axis. The signal extends always in one direction from this axis, and has a dc component exactly as in all dc restorers.

Except for the difficulty in closing and opening the key at the proper time, the keyed circuit fulfills the requirements for a satisfactory dc restorer. In addition to its performance, it has operating advantages as follows: The dc level at which the sync or blanking is held may be easily adjusted. The same circuit will handle either positive or negative polarity signals, or, as stated above, it may be used to hold, or clamp, as it has been called, any flat recurrent portion of a signal, by propertly timing the keying to correspond with the timing of the desired portion of the signal. The blanking level may thus be held directly and thereby become independent of any variation in the height of the sync pulses.

The Double Keyed Diode Circuit[7,8]

This circuit most closely approaches the idealized circuit of Figure 12. As shown in Figure 13, it consists of two diodes driven into conduction by push-pull pulses which may come from a transformer, or usually, as shown, from the plate and cathode of a tube. For reasons given later, this tube is best driven by a negative pulse. The two diodes are driven through two capacitors, C_1 and C_2, and are connected by two resistors R_1 and R_2, the common point of which is grounded through some voltage E for supplying bias to the amplifier. The time constant R_1C_1 and R_2C_2 is long compared to the pulse time.

For an analysis, this circuit may again be redrawn, as in Figure 14, in bridge form. The circuit and letters are the same as in Figure 13, except that the two pulses are designated as P_1 and P_2, with the center point of the two pulses at ground. The two pulses are equal in this case, and of opposite polarity. R_1 and R_2 are equal, as are C_1 and C_2, which are much larger than C.

[7] K. R. Wendt, U. S. Pat. 2,299,945, October 27, 1942 (Filed 11-27-40).
[8] J. H. Roe, "New Television Field-Pickup Equipment Employing the Image Orthicon", *Proc. I.R.E.*, Vol. 35, No. 12, pp. 1532-1546, December 1947. (See appendix p. 1542).

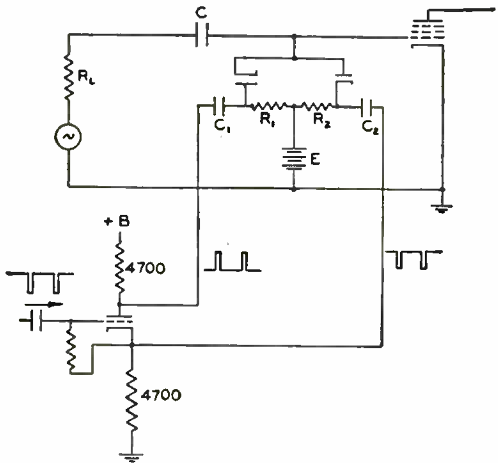


Fig. 13—The double diode clamp circuit.

During pulse time the diodes are driven into conduction by the positive pulse on the plate of D_1 and the negative pulse on the cathode of D_2. A current i_1 flows through the diodes and the capacitors C_1 and C_2. A current i_3 will flow into or out of C until point 1 is brought to an equilibrium voltage which is the clamping level. Between pulses, a current i_2 flows through C_1, C_2, R_1 and R_2, slightly reducing the charge on C_1 and C_2.

The clamping level or output voltage to which point 1 is brought during the pulse time is equal to E_1 under the conditions given above of equal pulses and equal R_1 and R_2. This voltage to which capacitor

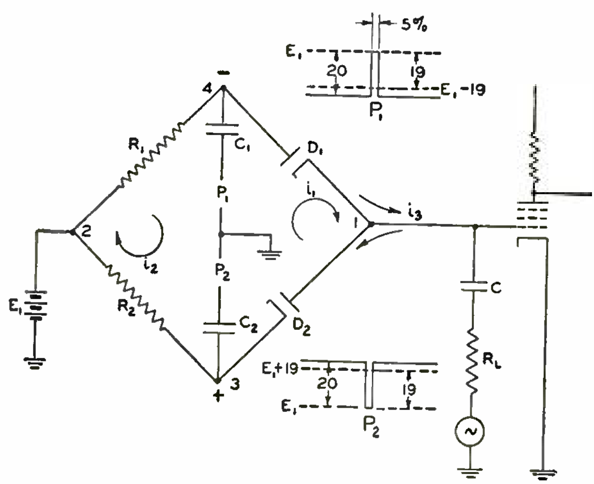


Fig. 14—Double diode clamp equivalent.

C is charged or to which point 1 is brought during the pulse time, depends upon the ac and dc voltages applied to the diodes at points 3 and 4. These voltages are shown in Figure 14. By way of illustration, the pulses are shown as 20 volts peak-to-peak, and of 5 per cent width, which would give them a peak value of 19 volts from the axis or dc values measured at points 3 and 4. During conduction time, points 3, 1, and 4 are all at approximately the same instantaneous voltage, which is the clamping level. The axis or dc voltage at 4 is then —19, and that at 3 is +19 volts from the *clamping level.* The clamping level is therefore midway between the dc voltages at 3 and 4. If $R_1 = R_2$, E_1 is also midway between these voltages, and equal to the clamping level. If the pulses are unequal, the clamping level will equal E_1 whenever $P_1/P_2 = R_1/R_2$.

$$\text{Clamp level } (C.L.) = E_1 + P_1 - \hat{\imath}_2 R_1 \qquad (\hat{\imath}_2 = \text{peak } i_2)$$

$$\text{since } \hat{\imath}_2 = \frac{P_1 + P_2}{R_1 + R_2}$$

$$\text{then } C.L. = E_1 + P_1 - \frac{(P_1 + P_2)\,R_1}{R_1 + R_2}$$

$$= E_1 + \frac{P_1R_1 + P_1R_2 - P_1R_1 - P_2R_1}{R_1 + R_2}.$$

Assume $P_1/P_2 = R_1/R_2$ or $P_1R_2 = P_2R_1$

$$\text{subst. } C.L. = E_1 + \frac{P_1R_1 + P_2R_1 - P_1R_1 - P_2R_1}{R_1 + R_2} = E_1.$$

If the resistors are unbalanced, C_1 and C_2 should be very large, or unbalanced such that

$$\frac{R_1}{R_2} = \frac{C_2}{C_1}$$

If the two pulse signals are constant in amplitude, E_1 may be obtained in effect from the diode circuit itself, and the voltage E_1 eliminated, point 2 being grounded. Such an arrangement is shown in Figure 15, where the center portion of the R_1R_2 combination has been replaced by a potentiometer. *There is a "point 2" on this potentiometer which has the same potential as the clamping level.* It is at the junction of the now hypothetical resistors R_1 and R_2 which have the same ratio as P_1 and P_2. If some other point on the potentiometer is grounded, a voltage will exist between the arm (ground) and point 2 since a dc voltage exists across $R_1 + R_2$. In the example shown, a negative voltage exists at point 2, since the positive portion of R_1, R_2 is ground. Any part of R_1, R_2 may be thus grounded, except that as the grounding point leaves point 2, an increasing portion of the pulse voltage also appears on the arm. A resistor R_A should therefore be inserted between the arm and ground in order to avoid loading the pulse.

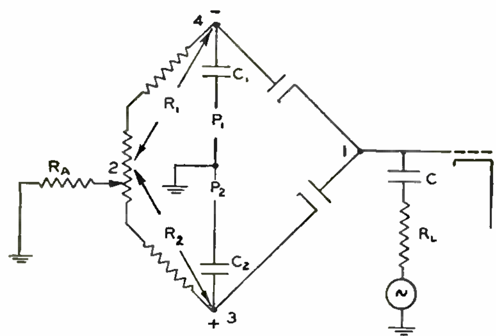


Fig. 15—Bias from double diode clamp.

Keying Pulse Requirements

A keyed circuit which has incorrectly timed pulses will give very poor results, the error usually showing up as strong dark or light horizontal streaking which varies erratically with the signal.

The pulses must of course be synchronous with the signal. The front edge of the pulse should occur approximately at the beginning of the portion of the signal which is to be clamped, such as the top of sync, or the "back porch". *The pulse should end well before the end of this portion of the signal.* A tolerance should be allowed, such that under no condition will the pulse last beyond the signal reference level, since a wrong charge will be put upon C, and small timing, or signal variations will produce large dc level variations.

The pulses must also be large enough such that under maximum signal swings, neither diode will conduct between keying pulses. It is apparent that even though the pulses place a large blocking voltage on the diodes, the signal also swings the other element of each diode, and will decrease at least on one of them the "bias" provided by the keying pulse.

As noted previously, it is preferable to connect the push-pull triode which supplies the pulses such that a negative pulse is required on its grid, as shown in Figure 13. The pulses thus occur when the triode is cut off, and the driving impedance consists of the plate and cathode resistors only. A circuit connected such that a positive pulse is required on the triode (the triode cathode feeds the diode plate, and the triode plate, the diode cathode) is more economical of triode plate current, but the triode is conducting during the pulses, and increased loading on one of the pulses caused by current into C has the effect of increasing the amplitude of the other due to the tube conduction. Incorrect charges are thus applied to C_1 and C_2, which return to normal quite slowly. The net result is that the circuit is very much slower in applying a new charge to C than would be expected.

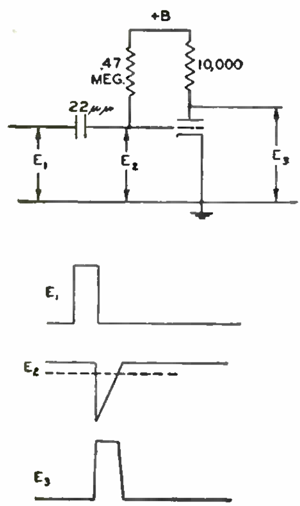


Fig. 16—Non-linear differentiator for delaying a pulse.

OBTAINING DELAYED KEYING PULSES[8,9]

It is often necessary in dc restorer circuits of the clamp variety to delay an existing pulse and thereby obtain a satisfactorily timed keying pulse. The usual example is that of delaying a sync pulse to obtain a keying pulse occurring during the back porch interval.

This may be easily accomplished by making a new pulse from the back edge of the sync signal. In Figure 16 is shown a diagram of a circuit which will produce such delayed pulses. The voltage at E_1 should be several times the cutoff voltage of the tube used. This is

---

[9] K. R. Wendt, U. S. Pat. 2,313,906, March 16, 1943 (Filed 5-25-40).

coupled into the grid with a very small capacitor. A grid resistor of approximately ½ megohm is connected to $+B$. Grid current, which is drawn most of the time, maintains the grid at approximately cathode potential. The front edge of the pulse at E_1 is therefore loaded down by grid current, and a charge is built up on the small coupling capacitor very rapidly. Only a very small portion of this front edge actually appears as a voltage on the grid. The back edge of the pulse, however, drives the grid negative and stops grid current and grid loading. The full swing of the back edge therefore appears on the grid. The grid resistor immediately starts charging the coupling capacitor toward $+B$ as shown at E_2. The charging of the capacitor is stopped by grid current. A sawtooth is thus produced on the grid from the back edge of the original pulse. Since this sawtooth is severely clipped, being much larger than the cutoff voltage, a clipped pulse as shown at E_3 appears on the plate. The width of

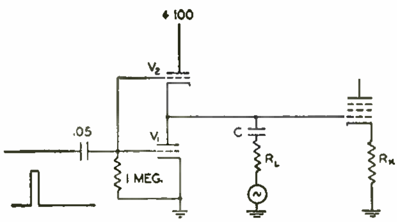


Fig. 17—The double triode clamp.

this pulse may be adjusted by varying either the grid coupling capacitor or the grid resistor. The width of this pulse also depends upon the amplitude of the input pulse. Therefore the input should be from a saturated tube in order to keep it constant. It is interesting to note that in this circuit the input is a positive pulse and the output is a delayed positive pulse.

The Keyed Double Triode

The keyed double triode was first described by the MIT Radiation Laboratory.[10] It is an improvement over the original clamp circuit of Browne and Blythen[11], who applied for their patent in September, 1934. The circuit is shown in Figure 17. The clamping is accomplished through the plate of V_1 and the cathode of V_2, and is thus "bidirectional", which is to say that the charge on C may be increased or decreased during the clamp interval. A positive pulse is applied

[10] M. I. T. Radar School Staff. PRINCIPLES OF RADAR, (2nd edition), pp. 2-35, McGraw-Hill Book Co., New York, N. Y., 1946.
[11] Brown and Blythen, U. S. Pat. 2,190,753, February 20, 1940 (Filed Great Britain 9-18-34).

to the two grids through the .05 microfarad coupling capacitor. Grid current in V_1 supplies a bias which holds the peak of the pulse at ground potential. The plate of V_1 must be above ground in order to draw current, which voltage also becomes the bias on V_2. The current through V_1 and V_2 must of course be equal. If the tubes are 6SN7's the equilibrium point during the pulse as found on the tube characteristic curves, is approximately +5 volts, for a plus B voltage of 100, as shown. The equilibrium point for V_1 is on the zero bias curve at a plate voltage (+5) equal to the bias on V_2 when both draw the same current (.2 milliamperes). If the grids are driven hard, they will be somewhat positive during the pulse, and the equilibrium will be lower with more current flowing. With 300 volts on the plate of V_2, the equilibrium will be about 15 volts, with about 1 milliampere peak flowing.

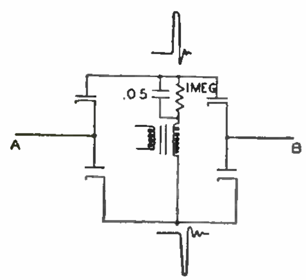


Fig. 18—The four diode clamp—used mainly for measurements.

This circuit has some advantages, and some disadvantages, over the balanced diodes. Its advantages are: It is simpler, in requiring only one polarity of pulse. The pulse need not be as large as the signal if negative peaks of the signal are clamped. Some power, in the form of current drawn by the following tube may be supplied without causing an appreciable change in the clamping level.

Its disadvantages: It is not balanced, and some keying pulse will appear on the signal. The clamping level can not be adjusted easily, always being a few volts positive from the cathode of V_1.

The Keyed Four Diode Clamp Circuit

The four diode clamp, as shown in Figure 18, is really a deluxe form of the two diode clamp. It is used mainly for *measuring* rather than *setting* a reference level, however. It was developed in 1938 for the lock-in circuit of the sync generator[12].

As shown in Figure 18, the circuit may be driven by a transformer

[12] K. R. Wendt, U. S. Pat. 2,250,284, July 22, 1941 (Filed 10-26-38).

which need not be carefully balanced, as no center tap is used. The center tap is supplied by the second pair of diodes, and unlike the two diode circuit of Figure 13, they supply both the ac and dc center. The circuit is thus inherently balanced. It acts essentially as a synchronous switch between the points *A* and *B*. A capacitor to ground at *B*, for instance, will be charged to a voltage equal to the absolute level existing on the signal at *A* during the keying pulse interval. This voltage may be used for automatic gain control of the signal at *A*, or as an indication of the phase between the signal and the pulse, and the phase thereby corrected, as in sync generator lock-in circuits, and automatic phase and frequency control of receiver deflection oscillators. The four diode circuit may be used as a normal dc restorer, although the simpler circuit of Figure 13 gives equivalent performance for most applications.

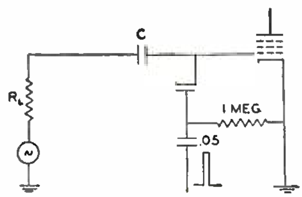


Fig. 19—The single keyed diode.

THE SINGLE KEYED OR UNBALANCED CLAMP CIRCUIT

The single keyed circuit[13] is the simplest of the clamp circuits. That is, if the word clamp is used with its original meaning, which is essentially that the circuit is held, or clamped, solidly from both directions by a synchronous signal, while between the clamping intervals no discharge path exists for the capacitor. During the war, the word clamp came to be used generally for all types of dc restorer circuits, since they were often used in applications in which there was no "original" to be "restored".

The circuit is as shown in Figure 19. The grid of the amplifier tube is used as a diode as in the simple dc restorers, but the grid leak is replaced by a keyed diode. Its operation is similar to the two diode circuit of Figure 13, in which the unbalance has been made complete.

The circuit possesses the speed and accuracy of clamp circuits. However, it has several disadvantages: Since it is unbalanced, some of the pulse will be applied to the signal. It can be used only for clamping on the positive peak of the signal, and the clamping level will always be at the cathode potential of the amplifier tube.

[13] K. R. Wendt, U. S. Pat. 2,299,944, October 27, 1942 (Filed 10-23-40).

USES OF THE DC COMPONENT

Although a number of uses of the dc component have already been mentioned, the more important ones are here summarized:

1. For the kinescope grid. To maintain picture black level with varying dc component.

2. For the sync separator. To maintain the sync at a fixed level to enable an amplitude separator to remove sync from the signal.

3. For the transmitter output.[14] To obtain maximum output with minimum distortion in the transmitter and receiver intermediate-frequency amplifier.

4. Blanking amplifier. To restore lows which may be advantageously lost in the camera, and to hold the black level constant to enable proper blanking, and original dc insertion.

5. Switching and fading. In order to avoid transients in the system due to sudden changes in the black level, all switching and fading operations should be done either by having the dc component present on the signals, and all black levels identical, or by following such operations with a very fast dc restorer.

6. Automatic gain control. The voltage for controlling the gain of a television receiver is obtained by a circuit which is similar to a dc restorer.

7. Sync stretcher or stabilizer. It is often advantageous to increase, and hold constant the sync amplitude into a transmitter. It is, of course, necessary to maintain the black level constant to stretch the sync only, and also to clip excess sync from the signal in order that a fixed amount of sync measured from the black level may be transmitted.

AMPLIFIER OPERATION WITH THE DC COMPONENT PRESENT

An amplifier which is fed a signal containing the dc component operates much differently than one with only an ac signal. When properly designed, a stage which contains the dc can produce more output with unobjectionable distortion than when it is designed to handle only the ac component. The important point here is the unobjectionable distortion. That is, appreciable distortion can be tolerated on a signal with the dc component, whereas, even small distortion may be objectionable on a signal with the dc component

[14] R. D. Kell, "Television Standards and Practice", National Television Systems Committee, p. 148, Edited by Donald G. Fink, McGraw-Hill Book Co., New York, N. Y., 1943.

missing. This is due to the fact that the television signal can stand appreciable distortion, providing the distortion *remains constant with changes in background.*

In designing a stage to handle a signal with its dc, three facts are important. First, the dissipation should not be excessive for continuous operation with an all-white or all-black signal—depending on the polarity—or for extended operation with no signal. Second, the dc regulation of the cathode, screen and plate voltages must either be zero or carefully considered in the design. Third, the peak-to-peak swing necessary is appreciably less for a signal with the dc component than for one which has lost the dc component. The latter is shown by Figure 20. The upper signals represent an all-white signal on the left, and an almost all-black on the right, both with a peak-to-peak amplitude of 4 units. When the dc component is not present, as in

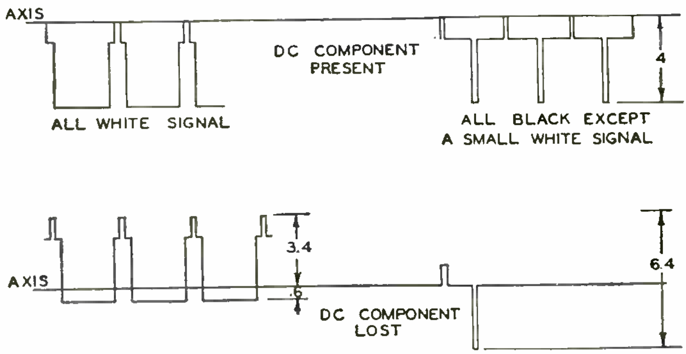


Fig. 20—Peak to peak signal with and without dc.

the two lower figures, the all-white signal swings up about 3.4 units and the white spike on the almost all-black signal swings down 3 units, thus requiring a peak-to-peak swing of 6.4 units, instead of 4 units, as with the dc component.

The amplifier may therefore be limited to lower currents, which for a screen grid tube means a lower screen grid potential. This works no hardship, since a smaller peak to peak swing is necessary for the same signal, and since more distortion can be tolerated. Using a 6AG7 for ac, a cathode bias of 2 volts and a screen potential of 125 volts are used. For the dc, the cathode is operated at zero, and the screen at 115 volts.*

The cathode of a dc stage should preferably be operated at zero, or, if not, it should be unbypassed, since the bypassing can be effective

* Operating conditions for a 6AG7 with both ac and dc signals are given in the RCA TUBE HANDBOOK.

of course only for ac. At any rate, any unbalance between the ac and dc gain has the same effect as the bypassed screen supply; both effectively lose some of the dc component. This loss can be regained in the plate filter, within limits. The dc gain is determined by the regular plate load resistor, plus the filter resistor. By an appropriate choice of filter resistor, loss at the screen or cathode may be compensated. The time constants of the circuits must then be made equal to avoid a transient mismatch. If a loss occurs at two places, then, in order to properly compensate, there should be two filter circuits in series in the plate circuit. If the dc component has been restored to improve the operation of the amplifier stage itself, it is unimportant to compensate for screen grid loss of the dc.

It is therefore helpful to restore the dc in all stages which might be nonlinear, also in all high level stages, since considerable current may be saved and smaller tubes be used as a result.

Summary of DC Restoration Circuits

Simple Circuit. Figures 3 and 4. Satisfactory for slow restoration from signal with good lows. *R* should be high, driving impedance low, and signal level high.

Positive Bias Circuit. Allows fast operation without introducing as much distortion as fast operation of simple circuit. Unsatisfactory for low signal levels.

Low Boost Circuit. Aids above circuits, especially when the signal has lost lows.

Keyed or Clamp Circuits. Excellent performance, even on very low level signals. May be very fast. Complicated in requiring accurately timed pulses.

Double Keyed Diodes. Balanced, adding very little pulse to signal. Requires push-pull pulses of greater amplitude than signal.

Keyed Double Triodes. Requires only one pulse signal. Can supply some power, or current, to output without changing dc level. Adds some pulse to signal.

Single Keyed Circuit. Simplest keyed circuit. Operates similar to keyed double triode, except can restore only on positive peaks of signal.

Keyed Four Diodes. Similar to two diode, but inherently balanced. Used mainly for measuring, rather than setting levels.

BARRIER GRID STORAGE TUBE AND ITS OPERATION*†

BY

A. S. JENSEN, J. P. SMITH, M. H. MESNER, AND L. E. FLORY

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

Summary—*Two versions of cathode-ray type of electron tubes to enable storage of video signals electrostatically upon an insulating target using a barrier grid or screen, have been designed and operated: the SDT using magnetic focus and deflection, and the STE using electrostatic focus and deflection. For any application, it is essential that their limitations and the functional relations between their characteristics be recognized. The inverse dependence of the fidelity with which the storage tube can reproduce a given signal, as measured by the cancellation ratio, upon the number of storage elements available on a given size target, is to be emphasized (see Equation (9) and experimental verification in Figure 17). However, there exists a maximum fidelity or a limiting cancellation ratio for which the difference between the input signal and its reproduction is just equal to the disturbance introduced by the tube. This indicates a corresponding minimum number of storage elements or amount of information to be stored, less than which no further improvement in fidelity can be realized.*

A differential method of measuring the characteristics of a storage tube is described and used. Though this method and nomenclature relating to such a subtraction or cancellation procedure is used, relationships are indicated between the characteristics described to those needed in the design of any arbitrary system involving storage of a signal.

The theory of the barrier grid target behavior is discussed. Tube data and operational limitations are presented, and it is shown that it is actually advantageous to use output amplifiers no wider in bandpass than is absolutely necessary to the overall system.

Storage times of up to 100 hours were observed with no evident distortion or decay.

INTRODUCTION

RECENTLY there has been evidenced an increasing interest in storage tubes.[1-3] In view of this fact, it seemed appropriate to describe a tube which, though still in an experimental stage

* Decimal Classification: R138 × R138.31.

† The work described in this paper was performed in whole, or in part, under Contract W28-003-sc-1541 between the U. S. Army Signal Corps Engineering Laboratories, Evans Signal Laboratory, Belmar, N. J. and Radio Corporation of America.

[1] A. V. Haeff, "A Memory Tube", *Electronics*, Vol. 20, pp. 80-83; September, 1947.

[2] J. A. Rajchman, "The Selectron—A Tube for Selective Electrostatic Storage", *Math. Tab. and Aids to Comp.*, Vol. II, pp. 359-361; October, 1947. (Abstract: *Proc. I.R.E.*, Vol. 35, p. 177; February, 1947.)

[3] R. A. McConnell, "Video Storage by Secondary Emission from Simple Mosaics", *Proc. I.R.E.*, Vol. 35, pp. 1258-1264; November, 1947.

and subject to further development beyond that outlined in this paper, may be of interest to system designers in applications requiring the storage and subsequent reproduction of video signals. There are many of these applications which are now only awaiting an appropriate storage device. For example, a reasonably short time delay (less than one second) could facilitate the solution to certain problems in television and standard audio broadcasting, electronic computer memory, frequency changing and multiplexing in communications, and in signal comparison, where either both signals are not available simultaneously or where it is desirable to make the comparison at an arbitrary phase relation. This last problem of signal comparison was uppermost in our minds during the development and testing of the barrier grid storage tube which is described, and the effect of this viewpoint will be felt in the presentation and in the nomenclature used. However, it will be pointed out that certain characteristics measured are practically directly convertible into characteristics needed in the design of other

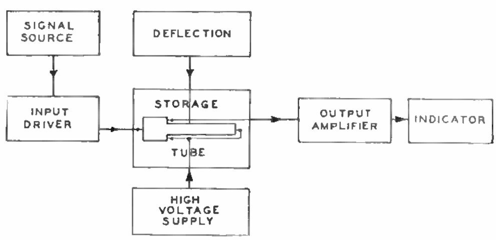


Fig. 1—Block diagram of storage tube cancellation circuits.

systems requiring storage. Likewise, since the storage times observed are of the order of days, the tube is also applicable to relatively long time storage problems.

In any of these applications a pertinent parameter of design is the fidelity with which a given signal may be reproduced and the functional relation of this fidelity to variables of the tube's operation. A critical method of measuring this fidelity is one in which a reproduced signal is compared with the original by subtracting one from the other and observing the difference. If the signals mutually cancel, reproduction is at highest fidelity, and the comparison of any residual signal to the useful output of the tube would be a measure of that fidelity. This measure, referred to as "cancellation ratio", is defined more specifically later on. It will be noticed that this method in one step compares reproduction fidelity in both amplitude and phase. The barrier grid storage tube by its design is particularly suited for this method of measurement since this subtraction or cancellation can be accomplished internally. Naturally, this fits the tube to that class of applications in which such an internal cancellation is desirable, insofar as it eliminates the need for the balanced circuits which would otherwise be required.

The procedure followed consists of impressing upon the tube on one scan a signal consisting of two square pulses whose amplitudes, polarities, and phases may be controlled, and on the succeeding scan two pulses, one of which, the "steady signal", is identical in amplitude, polarity and phase with one of the preceding, the second of which, the "variable signal", is different from the other of the preceding pulses only in polarity. On successive scans the steady signal remains as before, but the variable signal again changes in polarity only. The output from the steady signal then is a measure of the unfaithfulness of reproduction. The output of the variable signal is a measure of the output one would expect from the tube's simply storing and subsequently reproducing a desired signal. In the following, therefore, the output of the variable signal may be referred to now and then as "the desired signal."

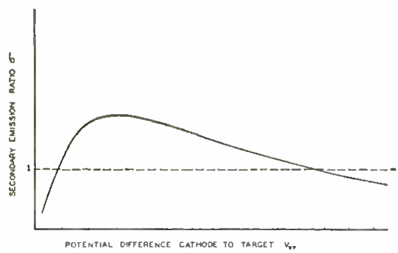


Fig. 2—Secondary emission ratio as a function of primary electron voltage.

The principle of electrostatic storage on an insulating surface has long been known and used in television pickup tubes, such as the iconoscope.[4] If an insulating surface is bombarded by an electron beam, the secondary emission ratio will vary with the energy of the bombarding electrons, according to the approximate curve shown in Figure 2. If the energy is such that the secondary emission ratio is greater than unity, then the potential of the target surface will change with respect to the electrode which collects the secondaries until the net number of secondaries leaving the target surface is exactly equal to the number of primaries arriving there. The surface potential, at which this action takes place, is known as the equilibrium potential. The remaining secondary electrons collect in the form of a space charge and rain back on the insulating surface, charging the unbom-

[4] V. K. Zworykin, G. A. Morton, and L. E. Flory, "Theory and Performance of the Iconoscope", *Proc. I.R.E.*, Vol. 25, pp. 1071-1092; August, 1937.

barded parts of the surface to a negative potential. Thus, a charge pattern is built up on the surface in the absence of any applied signal. The returning electrons, of course, partially neutralize any charges already on the surface and, thus, would make any comparison of signals from scan to scan impossible. Several ways have been attempted in the past to eliminate the redistribution effect, some of which are listed below:

1. Operation with a low energy beam where the secondary emission ratio is less than unity, as is done in the orthicon.[5]
2. Operation with a high energy beam where the secondary emission ratio also is less than unity.[6]
3. Maintaining the surface at a negative potential by a rain of electrons from a separate low energy source.[1]

Fig. 3—The barrier grid storage tubes: STE electrostatic tube in background, SDT magnetic tube in foreground.

4. Use of a grid or screen on or near the surface, operated at a potential preventing return of the electrons to the insulating surface.[6]

Each of these methods is adaptable to a particular use, the last one being chosen for the particular storage tube to be described. This tube consists essentially of an electron gun, an insulating target with a signal plate on the back and a fine mesh screen within a few mils of the front surface of the insulator, and a means of collecting the secondary electrons from the surface. The primary and secondary beams can be focused and deflected, either magnetically (SDT type) (Figure

[5] H. Iams and A. Rose, "Television Tubes Using Low Velocity Electron Beam Scanning", *Proc. I.R.E.*, Vol. 27, pp. 547-555, September, 1939.
[6] V. K. Zworykin and G. A. Morton, TELEVISION, John Wiley and Sons, New York, N. Y., 1940.

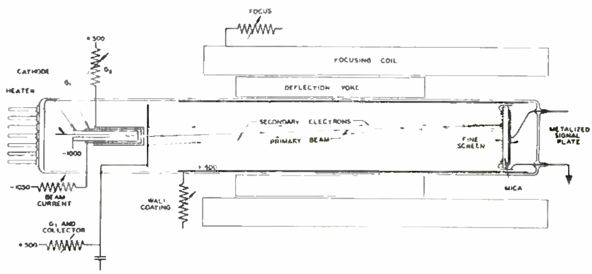


Fig. 4—Schematic diagram of SDT storage tube.

4), or electrostatically (STE type) (Figure 5). Means are provided to scan the insulating surface in a repetitive pattern, as for example, in a spiral or a "staircase" scan: In the spiral scan (Figure 6), since the angular velocity is usually constant, the linear scanning speed will vary from one end of the scan to the other. As is demonstrated later, the cancellation ratio will vary in a like manner. This is undesirable for experimental purposes, but may be used to an advantage in some applications.[7] The staircase scan (Figure 7) features constant scanning speed and constant interline spacing, both independently variable, which makes experiment simple and direct. Both scans use the target area equally efficiently. However, the spiral scan rejects the center of the target where the deflection disturbance is the least, and is, therefore, less desirable in this respect.

Since the target is an insulator, the only source of current to it is the primary beam, and the only drain of current from it is the sec-

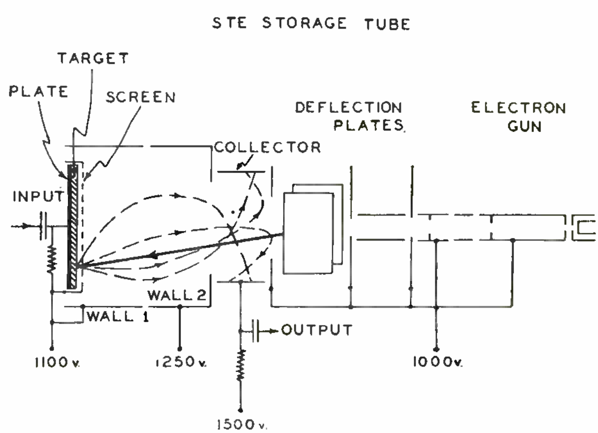


Fig. 5—Schematic diagram of STE storage tube.

[7] Calculations made by N. I. Korman, J. R. Ford, and L. Goldman, RCA Victor Division, Radio Corporation of America, Camden, N. J.

ondary beam. At equilibrium, these two must be equal, and any deposition or removal of charge on the surface will appear as a modulation of the secondary beam. However, since the energy of the primary electrons when they strike the dielectric is such that the secondary emission ratio is greater than unity (actually about two), those secondary electrons in excess of the number arriving in the primary beam must return to the target surface.

The barrier grid or screen functions as a virtual collector, so that the target equilibrium potential is established with respect to the screen and not to the actual collector electrode. At this potential a number of secondaries just equal to the number of arriving primaries are sufficiently energetic to penetrate the screen. These cannot return to the target, as appropriate fields outside the screen urge them away and toward the collector as the secondary beam. Meanwhile, the excess electrons are not sufficiently energetic to reach the screen, and

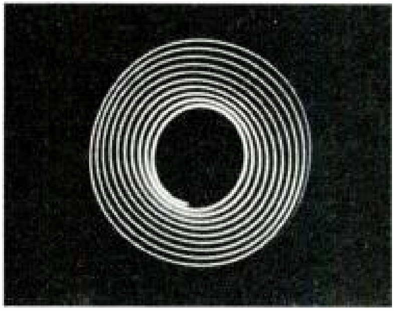


Fig. 6—Spiral scan used in SDT barrier grid storage tube.

are restricted in their motion by the close proximity of the screen to the dielectric surface. Thus, their redistribution to portions of the target not directly under the beam is considerably reduced.

When a signal is impressed upon the plate of the tube, the beam deposits on the insulating target a charge pattern, varying in intensity, that is a linear reproduction of the time variation of the impressed signal. If the surface is again scanned over the same path with no signal impressed upon the tube, the beam will remove the charge pattern, thus reading off a signal which is in polarity a mirror image of the original signal. Both during the writing and the reading, the signal will appear on the collector as a modulation of the secondary beam. In this operation, the tube has acted as a memory device, storing and subsequently reproducing a signal.

If, however, the same signal is impressed upon the tube on each successive scan, the beam will already have deposited the charge pattern necessary to match this signal variation. Therefore, that area

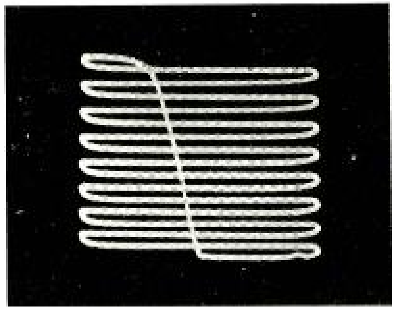


Fig. 7—Staircase scan used in STE barrier grid storage tube.

under the beam is instantaneously at equilibrium potential. No charge will be deposited on the target on succeeding scans, so that the secondary beam will be constant and unmodulated. Thus no signal will appear on the collector for steady input signals, constant in both amplitude and phase. However, any variation in the input signal will require deposition of charge by the beam. This will result in a modulation of the secondary beam and appear as a signal on the collector. In this fashion, steady signals are cancelled while varying signals are passed by the tube, the tube acting as an internal cancellation device (Figure 8).

An approximate alternate view of the internal cancellation operation considers the tube as a mixer. One signal is the presently impressed signal, the other is the charge pattern that has been deposited by the previous scan on the insulating target. Each modulates the

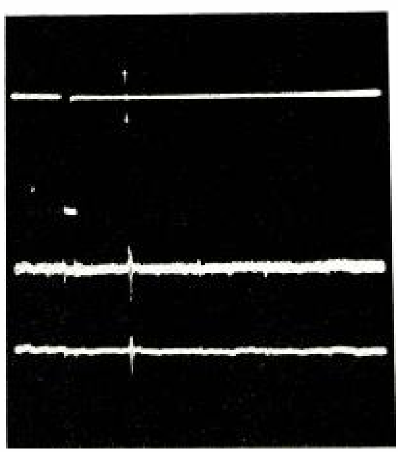


Fig. 8—Top line: Synthetic input signals used in tests. Variable signal is a pulse exactly like the steady pulse, but varying in amplitude from scan to scan. Middle line: Output without filter. Bottom line: Output filtered. Signals are in the same phase in each oscillogram.

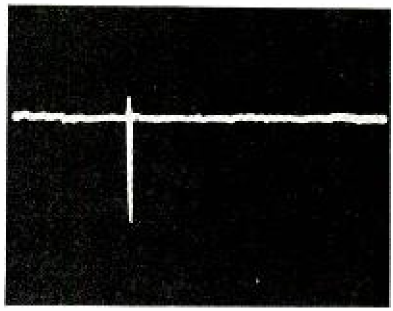


Fig. 9—Output signals, filtered, when output variable signal is a maximum, showing dynamic range. Signals are in the same phase as in Figure 8.

return beam with different polarity so that their mixture modulates the secondary beam with their difference. This indicates correctly that any part of the charge pattern that is not a faithful reproduction of the original signal will give rise to a residual signal.

Target Behavior

The behavior of the target can be better understood by reference to Figure 10, in which is plotted the general relation between the energy of secondary electrons emitted from a surface and the number of secondaries emitted per unit energy interval. If M electrons in the primary beam strike the target, the area under this curve will be σM, the total number of secondaries emitted. Equilibrium will occur for the target surface at a potential of V_e with respect to the screen, for which the number of secondaries with sufficient energy (more than eV_e) to penetrate the screen is just equal to M. This is the area under the curve from V_e to infinity. The remaining $(\sigma - 1)M$ secondaries, the area under the curve from zero to V_e, will not have sufficient energy to penetrate the screen, and will be returned to the target by the field between the screen and the dielectric surface.

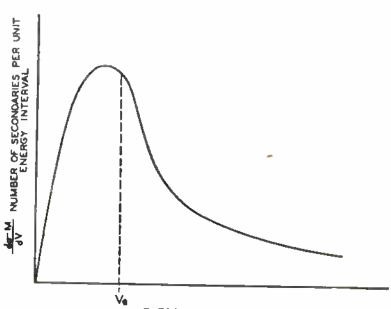


Fig. 10—Energy distribution of secondary electrons.

If, with the surface at equilibrium a few volts (V_e) positive to the screen, a signal is impressed upon the plate, the entire target will be swung capacitatively to a new potential. Now the number of secondaries that return to the target will be the area under the curve in Figure 10 from zero to this new potential. The net instantaneous current to the target will be the difference between these last areas per unit time, and the general curve is plotted in Figure 11. Note that in a restricted region around equilibrium the curve is essentially linear. This allows the tube to act as a more or less linear device to reproduce signal amplitudes. At the upper limit, for positive signals, the curve approaches the primary beam current as an asymptote. At the lower end, for negative signals, the curve is tangent to $(1-\sigma)$ times the beam current at a signal equal to $-V_e$. For this and more negative signals, all the secondaries will penetrate the screen and go to the collector.

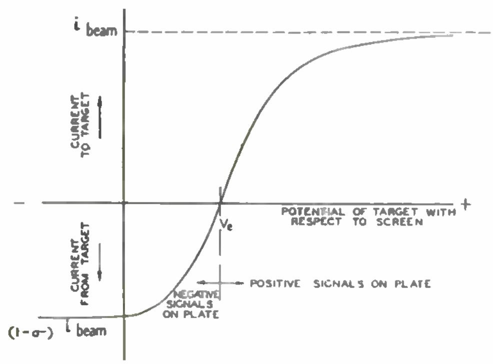


Fig. 11—Instantaneous current to target as a function of target potential with respect to the screen or barrier grid.

After the application of any signal and while the beam is on a particular region of the target surface, the instantaneous current to that portion of the dielectric obeys this curve. However, when the beam is scanning the target, it does not remain on any spot long enough to bring it entirely to equilibrium, that is, to discharge it completely to the equilibrium potential. The percent discharge effected per scan is called the "discharge factor".

The curve in Figure 11 points out that there is an essential difference between the responses to positive and negative signals, both in the manner of response and in the maximum value. As a result, the discharge factor for negative signals depends on both the signal amplitude and upon the beam current, whereas except for small signals, the discharge factor for positive signals depends on the beam current alone. This may, however, be chosen to give an acceptable discharge factor for both signs of reasonable signals. For both polarities the discharge factor is an inverse function of the capacitance per unit target area, the width of the beam and the scanning speed; and can, of course, never exceed unity. For the present mica targets, a discharge factor of 70 per cent has been measured for a beam current to the target of about 5 microamperes.

Signals

The external connections, shown in Figures 4 and 5, allow the tube to give an output signal, as described above, which is to a first approximation, the difference between the signal applied during scan I and that applied during scan II. Figure 12 shows in succession the input

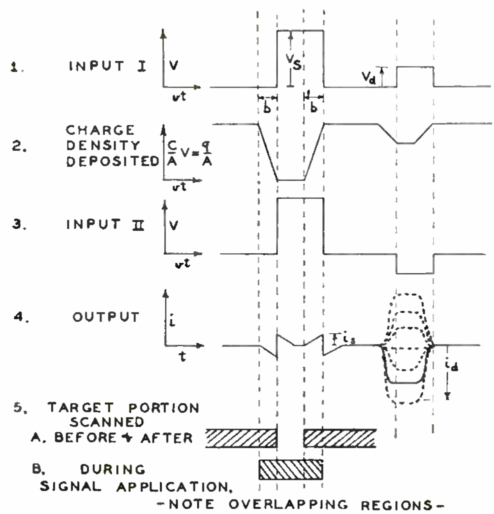


Fig. 12—Target behavior. In lines 1 to 3 the abscissa is the position (vt) of the beam spot along the scan line on the target. Line 4 is the output signal during scan II. The dotted lines indicate schematically the variable signal output for different variations in amplitude between scans. In line 5 is shown the actual target area scanned, with that area scanned during the signal removed to the side for clarity.

on scan I, the charge density deposited on scan I, the input signal on scan II, and the output signal on scan II. Note that the output is essentially the sum of the second and third lines of the figures.

Variable Signals

In general, a new signal (V_{sig}) on the plate will require the deposition of an amount of charge equal to the product of the capacitance per unit area of the target, the area of the target scanned by the beam during the signal, and the fraction of the signal discharged (fV_{sig}). The capacitance considered is that between the target surface and the signal plate. The net current to the target then will be:

$$i_{\mathrm{sig}} \propto f\ (\kappa/s)\ wv\ V_{\mathrm{sig}} \tag{1}$$

where wv is the area of the target scanned per unit time and κ/s is proportional to the capacitance per unit target area. It can be shown that, since the variable signal is changing in polarity from scan to scan, the effective signal, considering the effect of the discharge factor, is:

$$V_{\mathrm{eff}} = \left(\frac{2}{2-f} + \frac{b}{vt} \right) V_d. \tag{2}$$

For simple storage of a signal for a single scan previous to which the target was at equilibrium with no charge deposited at that portion of the target, the expression is:

$$V_{\mathrm{eff}} = (f + b/vt)\ V_d.$$

From Equations (1) and (2) the output variable signal is:

$$i_d \propto f\ (\kappa/s)\ wv \left(\frac{2}{2-f} + \frac{b}{vt} \right) V_d. \tag{3}$$

The first term in these expressions considers the simple charging of the scanning line of the target surface to the equilibrium potential as the spot moves along, while the second term is concerned only with the variation of the input signal with time. Hence, this latter remains of importance for very low scanning speeds (v approaching zero), and contributes the intercept ($R = 1$) in Figure 16.

In Figure 11 and in the text to this point, "the beam current" has referred to the current actually reaching the target and of that the portion actually returning to the collector. The screen, however, intercepts a portion of the primary beam from the gun before it reaches the target and a similar portion of the secondaries before they reach the collector. As a result, the a-c signal current will be considerably less than the d-c primary beam current from the gun. For a screen of 60 per cent transmission, the maximum modulation is only 36 per cent. The remainder goes to the collector as a direct current component, consisting of secondaries from the screen. Since this is an a-c system, however, this component may be neglected unless it is subject to a variation that would appear as a disturbance or noise, a spurious signal (q.v. below).

Residual Signals

Figure 12 shows the center of a steady signal completely cancelled. To obtain this, first, the dielectric target must have a sufficiently high product of resistivity and dielectric constant, such that an appreciable amount of charge cannot leak through the dielectric between scans. Second, there must be so little surface leakage across the dielectric, and the successive lines of the scan must be sufficiently spaced relative to the spot size that the beam cannot remove the charge that was deposited when it previously scanned a neighboring line. Either of these requires the deposition of additional charge on the next scan, and results in incomplete center cancellation. The latter results also in the appearance of a signal of opposite polarity at the time the portion of the charge is removed. This effect is usually called "interline crosstalk".

The spacing of the screen from the dielectric surface is determined by a not too critical compromise. If the spacing is too great, redistribution effects will shade the signals, introducing more interline crosstalk, and reduce the resolution. If the spacing is too small, whenever negative signals are applied to the plate, the very negative portion of the target surrounding the beam spot may, by a "coplanar grid effect", erect a potential barrier outside the screen, over which many of the secondaries cannot go. As a result they will be collected by the screen, and their absence from the secondary beam each scan will cause a positive signal to appear on the collector. It has been found that some few mils spacing of the screen is enough to prevent this coplanar grid effect. In a practical case, the use of a woven wire screen, whose thickness of weave provides a virtual spacing, is sufficient.

Considering the idealized signals in Figure 12, it can be observed (line 5) that the portion of the target scanned before the application of the signal overlaps that portion scanned during the application of the signal by just a beam width. This causes the charge pattern deposited (line 2) and hence the reproduced signal that would result from a simple storing on one scan and removal on a second scan, such as would be used for a simple memory problem as in a computer, to be shifted to an earlier phase by an amount proportional to the beam width. When the signals are compared from scan to scan, this shift in phase results in a residual signal output. Considering the internal subtractive procedure, the action is as follows: After a charge pattern has been laid down on the first scan, during each succeeding scan the beam will remove charge from the overlapping region before the application of the signal and replace it after the application of the signal. This transient removal and replacement of charge modulates the secondary beam and results in the residual uncancelled "spike" output for the steady signal input. The amount of charge involved depends upon the width of the beam and the discharge factor, and inversely upon the length of the target scanned during the signal rise time:

$$i_s \propto f\,(\kappa/s)\; wv\; V_s\,(b/vt). \tag{4}$$

The effectiveness of the tube as a cancellation device and the fidelity with which the tube can reproduce a signal may conveniently be measured by the "cancellation ratio", the ratio of the peak values

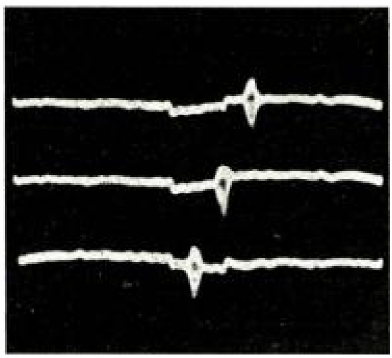


Fig. 13—Top line: Spike output from steady signal alongside of variable signal. Middle line: Variable signal coincident with one spike. Bottom line: Variable signal in center of steady signal.

(amplitudes) of the steady to the variable input signals for equal output signals:

$$R = V_s/V_d \qquad \text{when } i_s = i_d. \tag{5}$$

Since the variable signal will appear on the output with nearly the same amplitude, whether it is phased coincident with the center of a fixed signal or the spikes or not, except for very large values of the steady signals, Figure 13, this definition of cancellation ratio is practically independent of the phase of the variable signal. Thus the cancellation ratio may be calculated from (3) and (4) above:

$$R = [2/(2-f)]\;(vt/b) + 1. \tag{6}$$

Calculations[7] of spike output signals for more realistic wave shapes

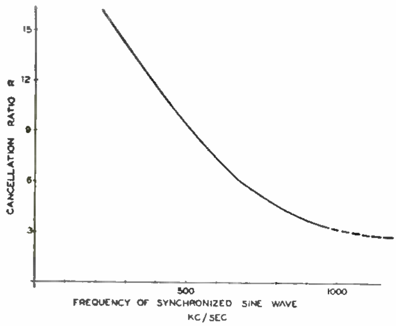


Fig. 14—Cancellation ratio as a function of frequency of a synchronized sine wave steady signal.

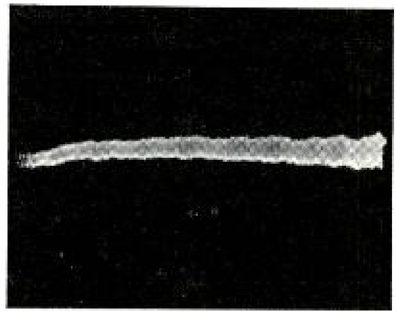


Fig. 15—Synchronized sine wave output from SDT tube using a spiral scan. The scan spirals inward so the scanning speed decreases from left to right and the cancellation ratio correspondingly decreases.

results in very complex integrals, but this same general trend prevails. It appears that the beam may be considered as a low pass filter whose frequency cut-off is roughly proportional to the ratio of the scanning velocity to the beam spot size. Thus to accurately cancel or reproduce signals of short rise times, the tube should either have a very fine spot or a rapidly moving spot. When a synchronized sine wave, whose phase is kept constant with respect to the start of the scan, is applied to the tube, a plot of the cancellation ratio as a function of the frequency of the sine wave is indicative of the operation of the writing beam as such a low pass filter (Figure 14). Likewise, the application of a synchronized sine wave signal to an SDT, using a spiral scan, shows qualitatively the relationship between cancellation ratio and scanning speed (Figure 15). Making use of the simplicity of control

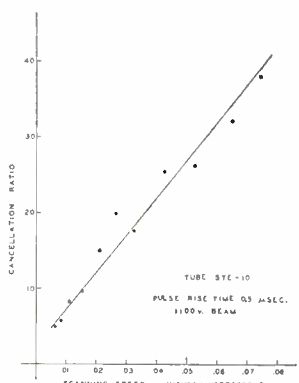


Fig. 16—Relation between cancellation ratio and scanning speed for a given pulse length.

of the scanning speed in an STE using a staircase scan one can measure quantitatively this relationship which is plotted from experimental data in Figure 16. This curve demonstrates essential agreement with Equation (6), which gives cancellation ratio as a linear function of scanning speed.

OPERATION

Another source of residual signal is input-output coupling. Unless normally careful external shielding is used and unless either the capacitance between the signal plate and collector is kept small by sufficient spacing or internal shields are used, an appreciable amount of the input signal will appear on the output by simple capacitative pickup. The screen acts as probably the most important internal shield as long as its impedance to ground is kept very low. It must have a short direct metallic lead out of the tube. If it has appreciable impedance to ground, not only does its shielding properties decrease, but as it swings with the signal it will modulate that portion of the primary beam which it intercepts, and which is normally a d-c component of the secondary beam. Since this is at least as large as the maximum a-c signal, modulation of it is serious. Measurements indicate, however, that in normal tubes this modulation can be negligible.

Disturbance

A number of factors individually contribute disturbance signals which may be viewed as a kind of noise and which represent a lower limit to the useful magnitude of the desired variable signal. A variable signal whose output is lower than the disturbance level is likely to be lost to an observer. The extent to which these contributions to the disturbance may be reduced depends largely on their character and source.

a. **Thermal noise:** Presently, tubes are operated at such beam currents that the output signals of all types are well above the noise and it is not a limiting factor. The worst disturbance is some five times the noise in amplitude. However, if smaller beam currents (with an appropriately smaller target capacitance per unit area to keep the discharge factor up) are attempted, to reduce the spot size further, the noise could be an important consideration.

b. **Deflection pickup:** In the STE type of tube, the collector must be properly internally shielded from the deflection plates to prevent pickup. The present design is successful in this respect.

c. **Deflection corners:** Target action theory[7] shows a second order signal that can arise as a result of change in curvature of the scanning pattern. This signal has not been observed and must be well below the noise.

d. **Deflection disturbance or shading:** The electric fields off the target surface must be designed such that the secondaries are collected uniformly from the surface. When this is not so, the resultant shading gives a signal that is synchronous with the frequencies in the scanning pattern. This is the most serious disturbance signal because both the secondaries from the target and those from the screen contribute. This means that there is available more than twice the current for modulation by this disturbance as there is for the desired signal. In the SDT tubes, this disturbance can be quite pronounced and control is difficult, since the same fields are used to focus both primaries and secondaries. In the STE, secondary and primary focus are separate, and this disturbance is more easily removed.

e. **Screen:** The successive interception of the beam by the screen wires generates a signal that is second in importance only to that of shading. If the beam does not extend for more than about three screen wires (this is usually the case), the signal resulting from the screen's intercepting the beam depends upon the secondary emission ratio of the screen wires and upon the ratio of the screen wire diameter to the beam spot length parallel to the direction of scanning.

The upper limit to the dynamic range of the variable signal is determined by its saturation value. Reference to Figure 11 will show the existence and limits of this saturation. A measure of this dynamic range is then the ratio of the maximum variable signal output (its saturation value) to the maximum disturbance output (the variable signal output's practical lower limit). This is called the "disturbance ratio".

$$D = i_{d\ \mathrm{max}}/i_{\mathrm{disturbance}}. \tag{7}$$

If the only contribution to disturbance is that from the screen,

$$D_s = b/(\sigma_s - 1)\ u. \tag{8}$$

Reducing the secondary emission ratio of the screen wires to unity gives the greatest promise for improvement of the disturbance ratio since it has been shown (Equation (6)) that the beam spot size must be small for good cancellation ratio and there are mechanical limitations on the fineness of the screen wire. Tubes with gold sputtered screens have shown disturbance ratios greater by a factor of two than those with stainless steel screens.

Figure of merit

From Equations (6) and (8) a figure of merit, some indications of the limitations of the tube, and means of improvement may be deduced.

$$(R-1)\ DN \propto [A/\delta\ (\sigma_s - 1)\ u]\ [(2/(2-f]\qquad (9)$$

where N is the number of pulses of rise and fall times t that can occur successively during the total scan, usually referred to as the number of storage elements on the target. Note that the three desired quantities, cancellation ratio, disturbance ratio, and number of elements available per tube, are so related that no one can be improved except at the expense of the others, or by enlarging the tube, or by causing the screen wire secondary emission ratio to approach unity. A finer spot,

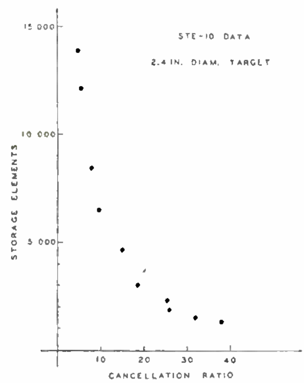


Fig. 17—Storage elements in an STE tube as a function of fidelity.

allowable if σ_s is reduced, would also allow δ, the interline spacing, to be reduced.

This product $(R-1)\ DN$ appears to be a convenient figure of merit for this tube and similar cancellation devices. The value for present two-inch target tubes is roughly $6 \cdot 10^5$ or greater.

In Figure 17 is plotted the same experimental data as in Figure 16, having calculated the number of pulses of 0.5 microsecond rise and fall times that can be placed on the 2.4 inch target of the tube with an interline spacing of 0.030 inch. This curve together with Equation (9), which it substantiates, indicates that for a given tube, wherein the spot size is essentially determined by the primary gun structure, the number of storage elements on the target is a function only of the cancellation ratio, being independent of the scanning speed or the pulse length. It follows then that these elements may be used to store information using pulses of *any* duration, the scanning speed varying inversely with the pulse length.

The storage area required per element (A/N) would be a figure of merit for the target construction; the present value for the usefully scanned portion of the target can be conservatively set approximately $6 \cdot 10^{-4}$ square inches per element for a cancellation ratio of 20.

Another figure of merit useful for some considerations has been suggested.[7] The "limiting cancellation ratio" can be defined as the ratio of the variable signal input to the steady signal input when the variable signal input is adjusted to give an output equal to the disturbance, and the steady signal is adjusted to give best overall performance from other considerations (e.g. sufficient discharge factor or linearity of response). This ignores the output dynamic range for cases where it is not important. Present STE tubes have limiting cancellation ratios of roughly 100 for which they should have sufficient area for about 600 elements (extrapolating Figure 17).

If two equal adjacent pulses are very close together, then the output signal from simple storage will not go to zero between them. but only to some finite value of amplitude (y). The ratio (Y/y) of the pulse amplitude (Y) to this finite value of the output signal (y) between the pulses can be taken as a measure of the resolution or the fidelity of reproduction. In fact, this is exactly the cancellation ratio defined above, $R = Y/y$. Another term, "percentage modulation" may be applied and defined as $P = (Y - y)/Y$ so that cancellation ratio may be related to "percentage modulation" such that $P = 1 - (1/R)$. For many applications such as television values of "percentage modulation" (P) as low as 5 per cent are useful. This would correspond to a cancellation ratio of only 1.05, which from Figure 17 extrapolated would indicate about $5 \cdot 10^5$ storage elements. The lowest percentage modulation, and therefore the greatest number of storage elements, that can be used is limited by the disturbance.

As a circuit element, the tube may be viewed in general as a high internal impedance generator, similar to ordinary electron tubes. Its output is essentially a current signal fixed in magnitude by the tube operation and characteristics. A reasonable figure would be 30 per cent modulation of a 3 microampere beam or an a-c signal of approximately 2 microamperes peak to peak. The output capacitance is about 20 micromicrofarads; the input capacitance approximately 400 micromicrofarads for a 2.4 inch target. The full 30 per cent modulation is attained for an input variable signal of 50 volts peak to peak. This may be summed up as a transconductance of 0.04 micromhos, from which the tube performance can be calculated in the usual manner.

Tube Data

Average characteristics for the latest STE type tubes are plotted in Figure 18. From these can be deduced the operating data. It is to be noticed that, similar to other vacuum tubes, there are different modes of operation possible depending upon whether or not the application permits saturation of certain signals. If the variable signal may be saturated for any large value of input the following is true. Data for SDT type tubes using mica dielectric about 0.8 mil thick, 230 mesh gold sputtered stainless steel woven wire screen spaced about 5 mils from the dielectric is $R = 20$, $D = 25$, $v = 0.020$ inch per microsecond, $t = 0.3$ microsecond, and $f = 80$ per cent. N can be calculated to be 1000. Limiting cancellation ratio is 50. For $t = 1$ microsecond.

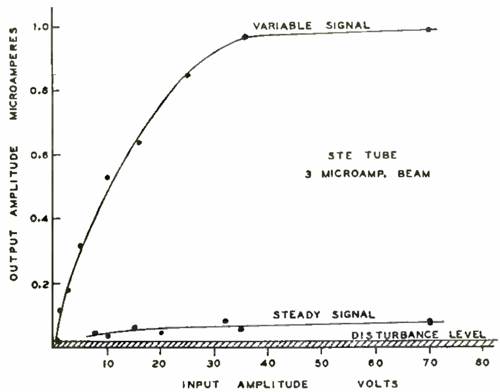


Fig. 18—Characteristic curves of operating storage tubes. Pulse length: 2 microseconds; scanning speed: 0.025 inch per microsecond; output filtered.

R becomes 45. STE electrostatic tubes use the same kind of target: 0.8 mil mica dielectric, 230 mesh gold sputtered stainless steel woven wire screen spaced about 5 mils from the dielectric; and their performance data is $R = 25$, $D = 25$, $v = 0.025$ inch per microsecond, $t = 0.3$ microsecond, and $f = 80$ per cent. Calculated N is 2000. Limiting cancellation ratio is 100. If saturation is not permissible, then reference to the characteristic curves in Figure 18 will show that for the STE tube, the cancellation ratio is about 12 while the limiting cancellation ratio is about 50. For this type of operation involving smaller input signals, the discharge factor is somewhat higher.

The SDT tube has a measured spot of 0.006 inch diameter for a beam current of 10 microamperes. The STE tube has a measured spot size of 0.008 inch diameter 8.5 inches from the main lens for a beam current of 10 microamperes and the screen 1000 volts above the cathode.

Storage Time

The present available apparatus does not allow the application of signals at repetitive rates less than 50 cycles per second. These slow repetition rates have given cancellation and disturbance ratios the same as for rates as high as 4000 cycles per second. This means that such signals are stored without appreciable change for at least 1/50th of a second. However, in a television test set in which the target is scanned in a standard television pattern and in which the output from the collector can be applied to the grid of a kinescope, so that the signals can be viewed at positions corresponding to the positions on the target from whence they came, signals that were impressed on the tube were observed to have negligible reduction or diffusion across the surface after 100 hours, during which time the beam was off. This tube had the same type of mica target as was described above.

Filtering

The data presented above are for the tube alone without the benefit of optimum aiding circuitry. The bandwidth of the amplifiers used in the measurements was 3 megacycles per second. By a judicious choice of the frequency response of the output amplifier, however, the performance of the tube as a cancellation device can be improved, since the spikes may contain frequency components roughly three times as high as the variable signal. A filter having a sharp cut-off just above the highest useful frequency can thus increase the cancellation ratio by attenuating the spikes. This is a true gain in a cancellation system; for other applications it simply indicates that the bandwidth of the system should be no greater than that required to pass the highest desired frequency. In addition, it was found that both disturbance and cancellation ratios could be improved by the introduction of a simple LC low pass filter, in this particular case having a half-value at 300 kilocycles per second, in the output circuit. This is shown in Figure 19 and also in Figure 8. The filter must have a fairly shallow low frequency cut-off to affect the screen disturbance, since the beam in scanning the screen crosses the wires at various angles. This means that there is generated not only the highest frequency due to scanning directly across individual wires, but the lower frequency components due to scanning the wires at more oblique angles. This spectrum unfortunately extends down into the region of useful signals and cannot be filtered out completely.

Concatenation

In an application requiring a cancellation device the use of two storage tubes in cascade has brought results which in many ways are gratifying, despite the added equipment and greater complexity from an operational standpoint. The concatenated set-up is made by feeding the conventional signal into the first storage tube as before, amplifying its output to a level to properly drive a second storage tube, then feeding this into the second tube.

This arrangement offers three distinct advantages. The disturbance contributed by the first tube constitutes a steady signal input to the second tube, which in turn cancels it. This means that the overall disturbance output is only that from the second tube alone. This tube may be carefully chosen so that this overall disturbance is a smaller than average amount. Secondly, since the output of each tube is essentially the first difference of its input signal, the output of the second tube is the second difference of the original input. Hence, the response to slowly varying signals is reduced, and the overall discharge factor

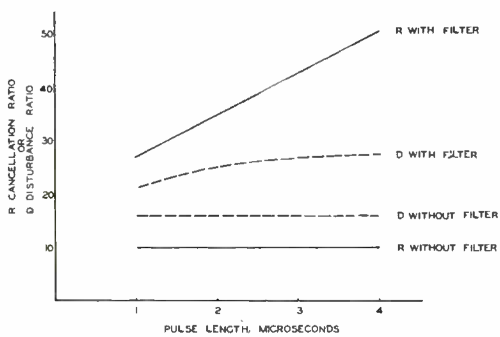


Fig. 19—Effect of pulse length and filtering on cancellation and disturbance ratios. Scanning speed 0.020 inch per microsecond in an SDT tube.

is greater than for either tube alone. Thirdly, for steady signal inputs to the first tube, the inputs to the second tube are the relatively smaller spikes. These are in turn attenuated by the second tube to give an overall cancellation ratio larger than either tube alone. This overall cancellation ratio is not as large as the product of the two tubes' cancellation ratios, however, since the spikes (see Figure 12, line 4) have rise times which are smaller than those in the original signal. The cancellation ratio of the second tube will be correspondingly smaller than that measured in the usual fashion. This is assuming the use of raw signals without filtering between tubes. Two SDT tubes, each of which showed an $R = 10$ singly, in cascade showed an $R = 60$. Variable signals that were lost in the disturbance of the first tube were readily visible in the output of the second tube.

Circuits

A quick survey of the stability of the associated circuits would show that there should be less than a pulse rise time jitter between the initial signal pulse and that which triggers the scanning pattern. Further, the deflecting circuits should allow the scanning raster to move only a small fraction of a line width. This would indicate deflection constant to one part in 1000. For both tubes, the spiral and staircase generators were fed from standard regulated supplies. The high voltage supplies were not regulated.

Conclusions

Besides being a direct measure of the characteristics of a storage tube as used in signal comparison problems, the described method of comparing input and output signals is of value in determining the general characteristics useful for the design of any system involving signal storage. In any storage system, the fidelity of reproduction of the stored signal is a primary consideration. This is measured by the cancellation ratio. Discharge factor is important in determining the writing and erasing requirements in any application. The disturbance ratio gives the output dynamic range while the limiting cancellation ratio is the input dynamic range. The limiting cancellation ratio may also be viewed as the greatest fidelity detectable since this is the fidelity for which the difference between the input and the reproduced signals is just equal to the disturbance introduced by the tube. The number of storage elements required by a signal is simply the amount of information that is contained in that signal. The number of storage elements then gives directly the number of discrete pulses which may be stored on the target with a given fidelity, as indicated by the cancellation ratio. Likewise, for more complex wave forms the number of storage elements can be taken as equal to the product of the bandwidth of the signal and a time t which is the duration of the signal which can be stored. Figure 17, which shows that the product of the cancellation ratio and the number of storage elements is a constant for any tube, than may be interpreted as indicating that the product of the bandwidth, the duration of the stored signal, and the fidelity is a constant. Thus for a given fidelity of reproduction the number of storage elements is fixed and for a signal of given bandwidth, the maximum duration is determined. Conversely for a certain desired duration of signal, the bandwidth (and hence the highest frequency) that could be stored is fixed by the same relation.

Acknowledgments

The barrier grid storage tube is an outgrowth of television pickup tube development. As such, it represents a result of the combined efforts of many people. Particular acknowledgment in this regard is due to Harley Iams, A. Rose, P. K. Weimer, and H. B. Law. The early work on the SDT tube was done by R. L. Snyder and S. V. Forgue. During the entire development, valuable assistance was rendered by P. G. Herkart and S. W. Dodge, under whose direction the tubes were fabricated, and by C. J. Busanovich and R. R. Goodrich, who made and processed some of the more intricate parts. The gun for the STE tube, a major contribution, was designed by F. H. Nicoll and D. W. Epstein. Considerable assistance with circuit problems, including the design of a power amplifier and the staircase scan generator, was given by J. M. Morgan. Finally, the authors wish to express their appreciation to V. K. Zworykin and I. Wolff, who directed the work, for their continued encouragement and valuable suggestions.

Symbols

A	Useful area of target; $A = \delta\lambda$
b	Beam width parallel to scan.
C/A	Capacitance per unit area of target (target surface to plate).
D	Disturbance ratio; $i_{d\ max}/i_{disturbance\ max}$.
e	Electronic charge.
f	Discharge factor; percent discharge per scan.
i_d	Output from variable signal; a-c component of secondary beam.
i_s	Output from steady signal; a-c component of secondary beam.
i_{sig}	Output from any new signal on the scan during which the signal first appeared.
M	Number of primary electrons bombarding target during a convenient time interval.
n	Pulse repetition rate.
N	Number of pulses of rise and fall times t that can occur successively during the total scan; $\lambda = 2v\,tN$; total number of "elements" available in tube.
P	Percentage modulation $(Y - y)/Y$.
q/A	Charge density deposited on target.
R	Cancellation ratio; V_s/V_d for $i_s = i_d$.
s	Thickness of dielectric target.
t	Rise time of input pulse.
u	Diameter of screen wires.
v	Scanning speed.
V_d	Input variable signal amplitude.
V_{eff}	Effective input variable signal, considering the effect of discharge factor.

V_e Equilibrium potential of target surface with respect to screen.

V_{kt} Potential difference between cathode and target, determining bombarding electron energy.

V_m eV_m is the energy of the most numerous secondaries from the dielectric surface.

V_s Input steady signal amplitude.

V_{sig} Amplitude of any new signal.

w Beam width perpendicular to scan.

y Amplitude of output between two adjacent pulses.

Y Output amplitude of a single pulse.

δ Separation, center to center, of scan lines. Interline spacing.

κ Dielectric constant of target insulator.

λ Total length of scan.

σ Secondary emission ratio of dielectric.

σ_s Secondary emission ratio of screen wires.

RADIO-FREQUENCY PERFORMANCE OF SOME RECEIVING TUBES IN TELEVISION CIRCUITS*†

BY

ROBERT M. COHEN

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

Summary—*Several types of receiving tubes may be used to advantage in television receivers designed to tune all thirteen channels. This paper discusses the performance of these tube types in radio-frequency amplifier, mixer, and local oscillator applications. Both push-pull "balanced" circuits and single-ended "unbalanced" circuits are discussed. Data are presented for over-all gain, noise, image rejection, and, to a lesser extent, on oscillator frequency stability. These data are taken at two representative channels in the television band: Channel No. 4 (66 to 72 megacycles) and Channel No. 11 (198 to 204 megacycles).*

INTRODUCTION

THE design of television tuners involves consideration of a number of factors which as yet are not fully evaluated due to lack of sufficient field experience. These factors are briefly described before discussing the performance of tubes in specific tuner designs.

An essential requirement of a tuning unit is that the local oscillator produce low radiation. A figure of 0.01 microwatts has been suggested[1] as the maximum tolerable amount of local oscillator radiation that will cause no appreciable radiation interference with receivers fifty feet apart operating in the field-strength region of about 500 microvolts per meter. This requirement is difficult to achieve in practice and necessitates the use of a radio-frequency stage to reduce the transfer of oscillator voltage from the mixer circuit to the antenna. In addition, enclosure of the entire tuning unit in a shield is advisable to prevent direct radiation by the oscillator tank circuit.

The significance of good signal-to-noise ratio is another important consideration in tuner design and performance. There is some evidence that a signal input voltage between ten to twenty times the root-mean-square value of equivalent noise is needed to produce a picture having barely satisfactory entertainment value when the receiver is in a

* Decimal Classification: R262 × R593.6.

† This paper was presented in abbreviated form at the Winter Technical Meeting of the I.R.E. in New York City in March 1947, and at the I.R.E. Chicago Engineering Conference in April 1947.

[1] E. W. Herold, "Local Oscillator Radiation and Its Effect on Television Picture Contrast", *RCA REVIEW*, Vol. VII, No. 1, pp. 32-53, March, 1946.

noise-free location. The better the signal-to-noise ratio of a receiver, the more capable it is of satisfactory performance in suburban areas and the less critical the antenna design. In most television tuners, the main portion of the random noise is attributed to the first amplifier tube.[2,3] Tubes which are space-charge limited produce an equivalent noise voltage, referred to the signal grid, which is a function of transconductance. High-transconductance triodes have low noise factors and are valuable as radio-frequency amplifiers at high frequencies.

A third important factor is accurate termination of the antenna system in order to prevent standing waves on the transmission line. It is complicated by the wide frequency range required of a thirteen-channel system. Proper termination of the antenna must be accomplished at the receiver terminals of the transmission line because so-called "wide-range" television antennas present a widely varying impedance at the antenna terminals of the transmission cable. If the antenna is mismatched and the receiver mismatch is in the order of 10 per cent at the receiver terminals, noticeable loss of high-frequency response in high-fidelity receivers may result. If the mismatch is large and the transmission line is long, double images or ghosts may result. Many designers believe that the elimination of the adjustably-tuned antenna circuit is economically justified and terminate the 300-ohm antenna cable at the input terminals of the first amplifier tube with resistance loading to provide the proper impedance and a fixed inductance to provide a broadly tuned resonant circuit. This loading, of course, adversely affects the noise performance of the receiver, limiting its effective sensitivity. It may also adversely affect the receiver in other ways which are discussed in a later paragraph.

Test Methods

On the basis of the foregoing considerations, several tuning units were constructed representative of current engineering practice. The minimum bandwidth of the radio-frequency circuits was six megacycles in any channel. Measurements of each unit were made on Channel No. 4 and Channel No. 11 (See Table 2). These are, respectively, the middle channels of the low-frequency and high-frequency television bands. A block diagram of the equipment employed in making these measurements is shown in Figure 1.

---

[2] B. J. Thompson, D. O. North, and W. A. Harris, "Fluctuations in Space-Charge-Limited Currents at Moderately High Frequency", *RCA REVIEW*, Vol. IV, No. 3, pp. 269-285, January, 1940.

[3] W. R. Ferris and D. O. North, "Fluctuations Induced in Vacuum Tube Grids at High Frequencies", *Proc. I.R.E.*, Vol. 29, No. 2, pp. 49-50, February, 1941.

Noise measurements were made using a high-gain intermediate-frequency amplifier which had a bandwidth of 4.5 megacycles and was terminated with a square-law vacuum-tube voltmeter. The signal generator employed has a balanced output terminated with a 150-ohm resistor in each conductor. In order to measure the equivalent noise voltage referred to the antenna terminals of the receiver, it is necessary to reduce the output of the signal generator to zero. The value indicated on the meter is proportional to the square of the amplified noise voltage. An unmodulated carrier signal is then applied from the signal generator and its value is adjusted to provide an output indication that is twice that of the noise. The equivalent noise voltage is then equal to the value indicated on the signal generator. The ratio of this equivalent noise voltage to that produced by the thermal agitation noise in the 300-ohm antenna circuit (5.0 microvolts when the band-

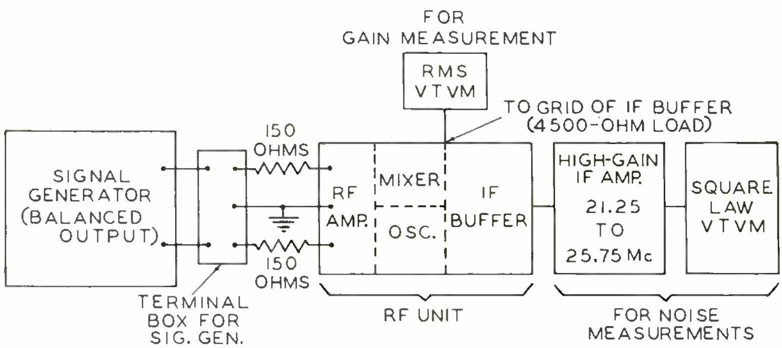


Fig. 1—Block diagram of test set-up.

width is 4.5 megacycles) is expressed in decibels as the noise figure. Gain figures are obtained by dividing the output voltage, measured using a 4500-ohm load impedance at the grid of the first intermediate-frequency amplifier, by the value of signal input voltage as indicated on the signal generator.

The possibility of including channel switches in these experimental tuning units was given consideration. However, channel switches and their connecting leads or continuously variable inductors introduce additional circuit reactances which would vary with each design.

Balanced Push-Pull Circuits

The circuit shown in Figure 2 is of the balanced type consisting of a neutralized push-pull 6J6 radio-frequency amplifier followed by a 6J6 mixer with push-pull grids and parallel plates. The local oscillator is also a 6J6 connected in push-pull in a conventional Hartley circuit.

The advantages of this arrangement can be summarized as follows:

1. Common cathode connection for push-pull operation cancels the inductive reactance of the cathode lead thereby reducing degeneration.
2. Push-pull mixing allows rejection of certain spurious responses.
3. The fixed-tuned extremely broad input circuit affords good antenna match and reduces cost of channel switch by eliminating two tuned circuits.
4. Push-pull circuits do not require switching at high-current points of circuit.

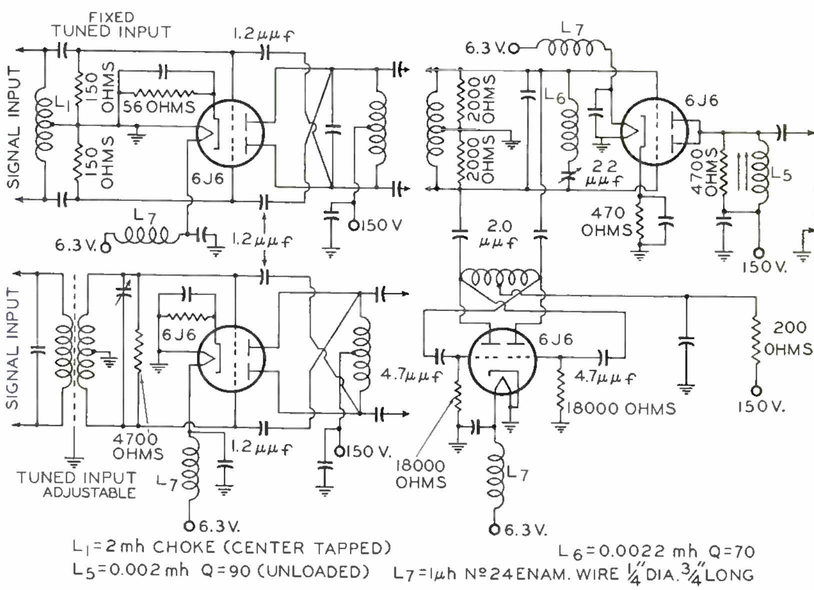


Fig. 2—Push-pull radio-frequency amplifier, oscillator, and balanced-input mixer.

5. Uniform performance from 44 to 216 megacycles is obtained.
6. The absence of radio-frequency current in ground return leads permits greater flexibility of mechanical design.

The disadvantages of this circuit arrangement are:

1. The circuit requires an intermediate-frequency trap in mixer grid circuit.
2. The circuit may require a narrow-band mixer plate circuit so that the trap in the mixer grid will be effective.
3. Minimum of six switch-contact points per channel is needed.
4. There is a large amount of local oscillator radiation when the radio-frequency amplifier is unbalanced.

5. An unbalanced (push-push) component of local oscillator voltage at the mixer grids is readily passed through the capacitance of the radio-frequency stage to the antenna lead causing additional radiation.

The alternate adjustably-tuned input arrangement given in Figure 2 (see also Table 3) offers improved signal-to-noise ratios but at the expense of requiring additional tuned circuits. These circuits add to the complexity of the selector switches and also increase the possibility of antenna mismatch and the attendant reflections. In addition, neutralization of the radio-frequency amplifier is more critical.

The over-all measured gain (i.e., ratio of intermediate-frequency grid voltage to signal generator voltage) of the tuning unit shown in Figure 2 is 60. The distribution of the over-all gain is approximately: radio-frequency amplifier, seven to ten; mixer, six to eight. The equivalent noise is approximately 13 decibels above the theoretical minimum value attributed to the thermal effect of a 300-ohm resistor. The image-rejection figure for this system is 35 decibels.

When the tuned-input circuit is used, the gain is increased to approximately 120. The equivalent noise is reduced to 6 decibels and the image rejection is improved to 45 decibels.

A computation of gain from the grid of the untuned radio-frequency amplifier stage to the grid of the mixer is given approximately by

$$\text{Gain} = \frac{g_m}{2\pi\Delta f\sqrt{C_1 C_o}}$$

where C_o is the plate-to-plate output capacitance of the radio-frequency double triode and C_1 is the grid-to-grid capacitance of the mixer input. The bandwidth, Δf, is measured at the 3-decibel attenuation points. The transconductance, g_m, which should be used is that for one-half of the radio-frequency double triode. Assuming $\Delta f = 6$ megacycles, $C_o = 6.5$ micromicrofarads, $C_1 = 7.5$ micromicrofarads and $g_m = 5300$ micromhos, the grid-to-grid gain is 20, and the gain from the signal generator to the mixer grid is approximately 10.

The mixer circuit gain, from the grid-to-grid of the balanced mixer to the grid of the single-ended intermediate-frequency amplifier tube may be expressed as

$$\text{Gain} = g_c R_L$$

where R_L is the impedance of the mixer plate circuit (4500 ohms) and g_c is the conversion transconductance of each half of the 6J6. Assuming the latter to be one-quarter of the transconductance, or 1325 micromhos, the gain is 6.

The over-all gain is the product of the individual stage gains and is thus equal to 60, which is in exact agreement with the measured values.

The equivalent thermal-agitation noise voltage for the 300-ohm dummy antenna is 5.0 microvolts for a bandwidth of 4.5 megacycles. The measured value of equivalent noise voltage is 21 microvolts. Thus, the noise figure is

$$20 \operatorname{Log}_{10} \frac{E \text{ measured}}{E\ 300\Omega} = 20 \operatorname{Log}_{10} \frac{21}{5} = 13 \text{ decibels}$$

It was found that the particular tuned input circuit used, increased the signal voltage and the antenna noise voltage equally, and by a factor of 2. Since the noise attributable to the tube does not materially increase, one would expect the signal-to-noise ratio to be improved by a factor of nearly two, making the noise figure for the tuned input circuit approximately 7 decibels above the thermal noise of the 300-ohm antenna. The 6-decibel value actually measured is in good agreement with the predicted value.

UNBALANCED CIRCUIT

A radio-frequency unit* utilizing a single-ended 6AG5 mixer driven by a pentode radio-frequency amplifier having an unusual type of balanced input circuit is given in Figure 3 (see also Table 4). The input impedance of the radio-frequency amplifier is somewhat higher than that of a grounded-grid amplifier but, of course, lower than that of the conventional grounded-cathode amplifier. If the input circuit is properly balanced, it is easily shown that the load presented to the circuit by the tube is approximately $\frac{2}{g_k}$ when transit-time loading is neglected. The advantages of this circuit are:

1. Single-ended switching circuits require only three switch contacts for each station selected.
2. Local oscillator radiation is lower than that provided by the circuit shown in Figure 2 and is not critical as to balance.
3. The broad, fixed-tuned input circuit (substantially non-reactive termination) aids in preventing reflections.

---

* The input circuit is a development of R. Romero, Industry Service Laboratory, RCA Laboratories Division, New York, N. Y.

The disadvantages are:

1. Pentode tubes produce more noise than triodes.
2. Lower gain is obtained than with push-pull twin triodes.
3. The transconductance available per tube may be less than that obtainable with twin triodes when the triodes are used in parallel.
4. The unbalanced mixer-grid circuits permit reception of more spurious responses than do push-pull mixers.
5. Non-uniform performance from 44 to 216 megacycles is obtained with available pentodes.

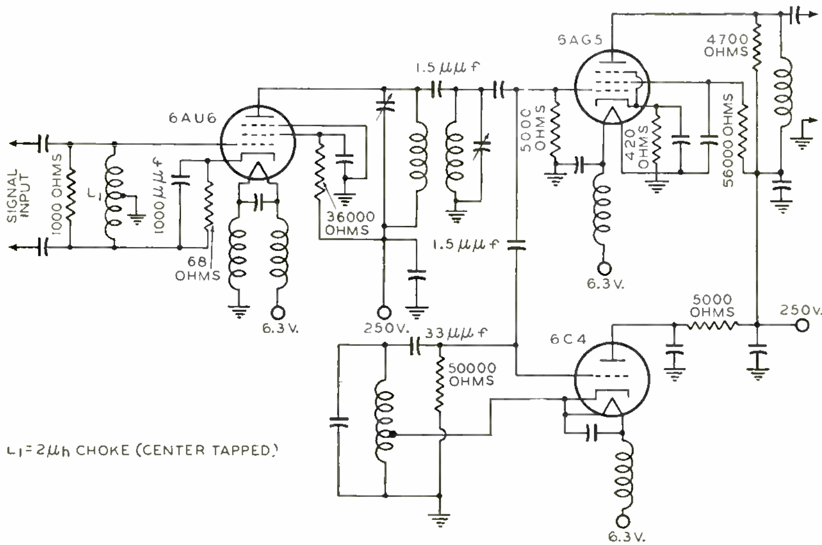


Fig. 3—Radio-frequency unit, unbalanced circuit.

The over-all measured gain of this system, when a 6AU6 is used as the radio-frequency amplifier, a 6AG5 as the mixer, and a 6C4 as the local oscillator, is 30 at Channel No. 4 and 25 at Channel No. 11. The noise figure is approximately 20 decibels at both frequencies. The image-rejection values for this circuit are 35 decibels at Channel No. 4 and 30 decibels at Channel No. 11. The 6BA6 may also be used in this circuit as a radio-frequency amplifier but, due to its lower transconductance will give slightly lower gain.

The gain of the 6AU6 radio-frequency amplifier (i.e., ratio of grid-to-ground voltage on the mixer to grid-to-cathode voltage of the radio-frequency amplifier) is calculated below. Even though there is signal impedance between cathode and ground, the full value of g_m is used because the input signal is applied between grid and cathode, and not between grid and ground.

$$\text{Gain (for the 6AU6)} = \frac{g_m}{2\pi \Delta f \sqrt{C_o C_i}} = 10$$

where g_m, the 6AU6 mutual conductance, is equal to 5200 micromhos, C_o is the plate-to-ground output capacitance of 14 micromicrofarads, C_i is the grid-to-ground mixer input capacitance of 14 micromicrofarads, and Δf is the bandwidth, 6 megacycles, to the 3-decibel attenuation points.

Because there is an antenna-termination loss of 6 decibels with the matched-impedance input, the gain from the signal generator to the plate circuit of the radio-frequency amplifier is 5.

The gain from the grid of the 6AG5 mixer to the grid of the intermediate-frequency amplifier is given approximately by

$$\text{Gain} = g_c R_L = 1250 \times 10^{-6} \times 4500 \doteq 5.6$$

where R_L is the impedance of the mixer plate circuit (4500 ohms) and g_c is the conversion transconductance of the 6AG5, assuming the latter to be one-quarter of the transconductance or 1250 micromhos.

The over-all gain is $5 \times 5.6 = 28$ which checks reasonably well with the measured values given above.

The measured value of equivalent noise voltage is approximately 45 microvolts or 20 decibels above the thermal noise of a 300-ohm antenna.

Grounded-Grid Radio-Frequency Amplifier and Cathode-Coupled Mixer

A circuit using a twin triode as a grounded-grid radio-frequency amplifier working into a second twin triode used as a cathode-coupled mixer[4] is given in Figure 4 (see also Table 5). The antenna matching transformer used in this investigation required three changes in tap positions in the low-frequency bands and two in the high-frequency ones. The advantages of this circuit are:

1. It requires fewer switch contacts than double-ended circuits.
2. This circuit has less oscillator radiation than any of the others.
3. Oscillator injection is accomplished by means of a separate grid thereby requiring less critical excitation adjustment.
4. The fixed-tuned input circuit affords non-critical antenna match.
5. Low noise from triode tubes results in higher signal-to-noise ratio.
6. Uniform performance from 44 to 216 megacycles is obtained.

[4] G. C. Sziklai and A. C. Schroeder, "Cathode-Coupled Wide Band Amplifiers", *Proc. I.R.E.*, Vol. 33, No. 10, pp. 701-709, October, 1945.

The disadvantages are:

1. Lower gain for each stage than that obtained from push-pull twin triodes.
2. Input circuit requires more additional switch contacts than circuit shown in Figure 3.

The gain of this circuit is uniform over the frequency range of the television band and is approximately 15. The noise figure is 8 decibels at Channel No. 4 and 9 decibels at Channel No. 11. Image rejection is approximately 40 decibels at both channels.

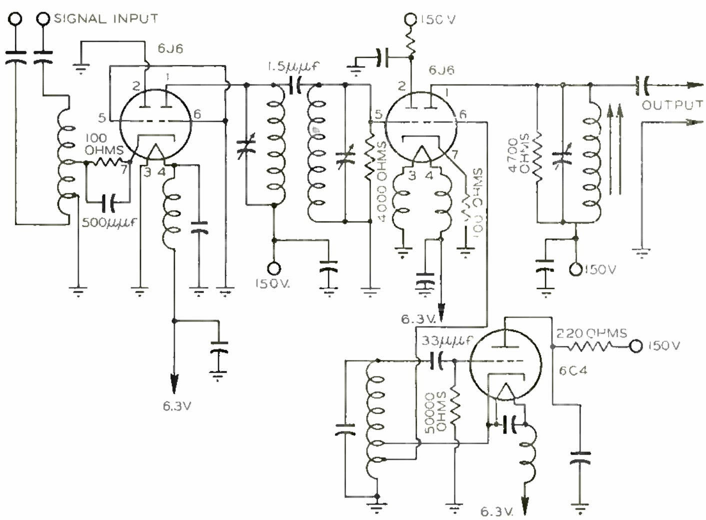


Fig. 4—Radio-frequency unit with grounded-grid triode radio-frequency amplifier, twin-triode cathode-coupled mixer, and triode oscillator.

The use of both halves of the 6J6 in parallel as a grounded-grid radio-frequency amplifier is not desirable because the tube does not have sufficient plate-to-cathode shielding for this application. The 6J4 is recommended for use in this circuit in television receivers in which higher tube cost is justified. When the 6J4 is substituted for the 6J6 radio-frequency amplifier, the gain is increased to 30 and the noise figure is improved to 6 decibels in both channels. The image rejection is not measurably changed.

The calculated gain of the grounded-grid radio frequency amplifier measured from the antenna to the grid of the mixer is calculated as shown below[5]

[5] M. C. Jones, "Grounded Grid Radio-Frequency Voltage Amplifiers", *Proc. I.R.E.*, Vol. 32, No. 7, pp. 423-429, July, 1944.

$$\text{Gain} = \frac{1}{4\pi\Delta f\,\sqrt{C_o\,C_i}}\sqrt{\frac{g_m}{R_{ant}}} = 3.7.$$

In this equation Δf, the bandwidth measured at the 3-decibel attenuation points, is 6 megacycles, C_o is the radio-frequency amplifier plate-to-ground output capacitance of 14 micromicrofarads, C_i is the mixer grid-to-ground input capacitance of 16 micromicrofarads, R_{ant} is the antenna transmission line impedance of 300 ohms, and g_m, the transconductance of one section of the 6J6, is 5300 micromhos. It was not possible to use both halves of the 6J6 in a grounded-grid circuit because of excessive regeneration.

The calculated gain of the cathode-coupled twin-triode mixer from the mixer grid to the intermediate-frequency amplifier grid is approximately[4]

$$\text{Gain} = \frac{g_c\,R_L}{2} = 3.0$$

where g_c, the conversion transconductance of one tube section, is 1325 micromhos, assuming g_c equal to one-quarter the 6J6 transconductance of 5300 micromhos; and R_L is the mixer plate circuit impedance of 4500 ohms.

The gain from the antenna to the intermediate-frequency amplifier grid is the product of the individual stage gains or 11.2.

This value of 11.2 is lower than the measured value of 15 for the system. Applying bias voltage to the radio-frequency amplifier to produce a known variation in transconductance resulted in a larger reduction in gain than the calculated theoretical reduction which should be proportional to the square-root of the reduction in transconductance. This reduction in gain indicates that regeneration is present in the radio-frequency amplifier, although not to such an extent as to produce instability.

Oscillator Design

The problems of local oscillator design are similar in all of the circuits described. Type 6J6 works well as a push-pull oscillator in circuits using push-pull mixers or in applications where high power output is necessary in order to excite the mixer tube sufficiently. Type 6C4 is suitable for use in single-ended circuits having efficient oscillator injection. In circuits where the cathode is tapped on the oscillator grid coil, radio-frequency chokes should be placed in series with the oscillator heater leads and the heater should be bypassed to the cathode to reduce microphonics. Ceramic sockets or mica-filled rubber-wafer sockets are recommended for use with the oscillator tube to reduce frequency drift during warm-up. When the tubes are used in oscillator circuits similar to those described in this paper, the warm-up drift expressed as a change of capacitance across the oscillator tank circuit is approximately 0.025 micromicrofarads for type 6J6 and 0.009 micromicrofarads for type 6C4.

INTERFERENCE EFFECTS WITH FIXED-TUNED-INPUT CIRCUITS

Several of the circuits described in this paper employ a very broad fixed-tuned antenna-input circuit. Although the use of such circuits reduces antenna matching problems and reduces the cost of the receiver, interference effects may be severe. The use of the input circuit of Figure 3 is beneficial in reducing the interference effect. Strong signals on the grids of the radio-frequency amplifier will cause the tube to generate undesired harmonics. If the plate circuit of the radio-frequency amplifier resonates at a frequency which is a multiple of the interfering signal, strong interference will result. For example, a receiver tuned to the upper-frequency channels would be subject to second-harmonic interference generated by strong frequency-modulated signals operating in the 88-to-108-megacycle band. The push-pull radio-frequency amplifier circuit will not generate strong even-order harmonics, but will be susceptible to strong signals whose frequency is one-third that of the desired signal. One of the important advantages of the grounded-grid radio-frequency amplifier is that it is less susceptible to harmonic generation because of the degeneration introduced by the impedance in the cathode circuit.

(Tables on following pages)

Table 1—Summary of Test Results

Circuit	Measured Gain* Channel #4	Measured Gain* Channel #11	Image Rejection (decibels) Channel #4	Image Rejection (decibels) Channel #11	Noise Figure (decibels) Channel #4	Noise Figure (decibels) Channel #11
Figure #2						
6J6 Fixed-tuned Input	60	60	35	35	13	13
6J6 Adjustably Tuned Input.	120	120	45	45	6	6
Figure #3						
6AU6 Grid-Cathode Input	30	25	35	30	20	20
Figure #4						
Grounded-Grid Radio-Frequency Amplifier and Cathode-Coupled Mixer 1/2 6J6 Radio-Frequency Amplifier	15	15	40	40	8	9
Figure #4						
With 6J4 Radio-Frequency Amplifier	30	30	40	40	6	6

* Measured gain figures are obtained by dividing the mixer intermediate-frequency output voltage developed across a 4500-ohm load, by the value of signal input voltage as indicated on the signal generator using a 300-ohm dummy antenna. (See Figure 1).

Table 2—Television Channel Assignments

Channel No.	Channel Frequency (Megacycles)	Oscillator Frequency (Megacycles)
1*	44-50	71
2	54-60	81
3	60-66	87
4	66-72	93
5	76-82	103
6	82-88	109
7	174-180	201
8	180-186	207
9	186-192	213
10	192-198	219
11	198-204	225
12	204-210	231
13	210-216	237

Picture carrier is placed 1.25 megacycles above low-frequency edge of band. Sound carrier is placed 5.75 megacycles above low-frequency edge of band. Thus, in Channel No. 1, picture carrier is 45.25 megacycles and sound carrier is 49.75 megacycles.

(* Will probably be assigned by Federal Communications Commission to services other than television.)

Table 3—Operating Conditions for Circuit of Figure 2

	Radio-Frequency Amplifier 6J6	Mixer 6J6	Oscillator 6J6	
Plate Supply Voltage	150	150	150	volts
Plate Current	10*	9#	12*	milliamperes
Grid Voltage	— 1.2	— 4.5	— 19*	volts
Grid Current	...	0.1**	1.1	milliamperes

* Per triode unit.
** Per triode unit with oscillator on.
\# Plates in parallel.

Table 4—Operating Conditions for Circuit of Figure 3

	Radio-Frequency Amplifier 6AU6	Mixer 6AG5	Oscillator 6C4	
Plate and Grid — No. 2 Supply Voltage	250	250	250	volts
Plate Current	11	6	11	milliamperes
Grid-No. 2 Voltage	95	150	...	volts
Grid-No. 2 Current	4.3	1.8	...	milliamperes
Grid-No. 1 Voltage	...	— 4.25	— 25	volts
Grid-No. 1 Current	...	20	500	microamperes

Table 5—Operating Conditions for Circuit of Figure 4

	Radio-Frequency Amplifier 6J6	Mixer 6J6	Oscillator 6C4	
Plate Supply Voltage	150	150	150	volts
Plate Current	9.5#	10.2* 10.5**	11	milliamperes
Grid Voltage	— 1	— 2	— 10	volts
Grid Current	0	# #	0.2	milliamperes

* Input triode unit.
** Output triode unit.
\# Per triode unit.
\#\# Not measured.

STEREOSCOPIC VIEWING OF CATHODE-RAY TUBE PRESENTATIONS*

By

H. A. IAMS, R. L. BURTNER, AND C. H. CHANDLER

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

Summary—This paper describes two methods of presenting three dimensional information on cathode ray tubes. Possible applications are considered and examples given showing both oscillograms and radar presentations in three dimensions. A discussion of the effect of flicker on stereoscopic viewing is included.

CATHODE-RAY OSCILLOGRAPHY

THE CATHODE-RAY oscilloscope has been a useful analytical tool for a number of years. Its application however has been somewhat limited because it is able to display information of two variables only. Various attempts to overcome this deficiency have resulted in multi-beam cathode-ray tubes, electronic switching and beam intensity variation. Another method of attack depending on the synthesis of a three dimensional presentation by stereoscopic means has had very little experimental investigation until recently.[1]

Recently the desirability of a simple three-dimensional indicator became more apparent with the development of radar sets which produced information in the form of azimuth, elevation, and range data. Since cathode-ray tubes are universally used as radar indicators full use of the information requires that at least two tubes be used. Azimuth vs. range (type B indication) may be shown on one tube while either azimuth vs. elevation (type C indication) or range vs. elevation (type E indication) may be shown on the other. The interpretation of such indicators is difficult and usually requires at least two operators. It was with these considerations in mind that the following work was done.

Normal three dimensional vision may be explained on the basis of focus, aspect, and convergence, each of which contributes its share to

* Decimal Classification: R371.5 × R537.131.

[1] O. H. Schmitt, "Cathode Ray Presentation of Three Dimensional Data", *Proc. I.R.E.*, Vol. 35, No. 2, p. 174, February, 1947; *Jour. Appl. Phys.*, Vol. 18, No. 9, pp. 819-829, September, 1947. C. Berkeley, "Three-Dimensional Representation on Cathode Ray Tubes", *Proc. I.R.E.*, Vol. 35, No. 2, p. 178, February, 1947.

the overall depth perception of an individual. Focusing is accomplished by altering the shape of the lens of the eye. The magnitude and direction of this alteration produce a certain sense of depth. The eyes are physically separated so that each gets a slightly different view or aspect of a given object which further contributes to the depth perception. Finally the eyes must converge if each is to see the same object at the same time. The last of these, convergence, is probably the most important factor in judging the distance of nearby objects and is the only one used in the device to be described.

Figure 1 is a diagram of the optical arrangement of the three dimensional oscilloscope and Figure 2 is a photograph of the device with the optical elements removed to show the two cathode-ray tubes.

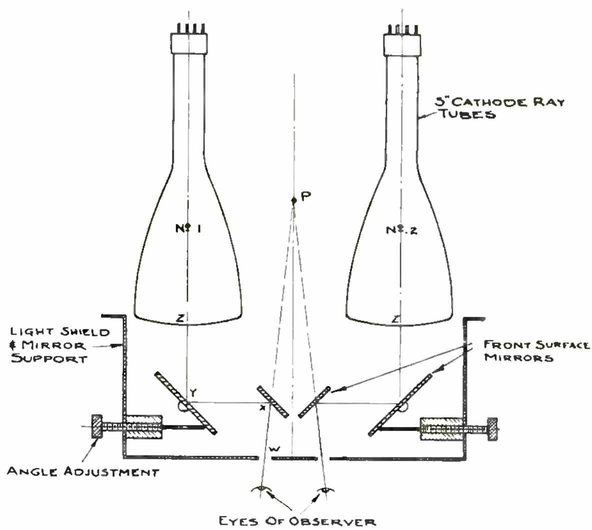


Fig. 1—Arrangement of the stereo-oscilloscope.

If a point of light is produced in the center of each tube and the tubes viewed as shown in Figure 1 the left eye of the observer sees the point of light on tube #1 and the right eye sees the point of light on tube #2. When the mirror angles are properly adjusted the effect is as though a single point of light were present at point *P*. It is obvious that if the spot on tube #1 is shifted to the right and the spot on tube #2 is shifted to the left the eyes will converge more sharply and the apparent point *P* will appear to be closer to the observer. By this means the point *P* may be moved toward and away from the observer by applying proper centering or shifting voltages to the cathode-ray tubes.

Figure 3 shows the geometry of such a system with the mirrors omitted for the sake of simplicity. L represents one-half the inter-ocular distance of an average person and is approximately $1\frac{1}{4}$ inches. The line a-a represents a screen on which two spots of light l and l' appear and P represents the apparent point of light when one eye sees l and the other l'. It is of interest to see how D varies as the distance between spots is varied. From the figure the following relations may be written:

$$D = \frac{S}{\tan \alpha} \qquad \text{and} \qquad \tan \alpha = \frac{L - S}{C}$$

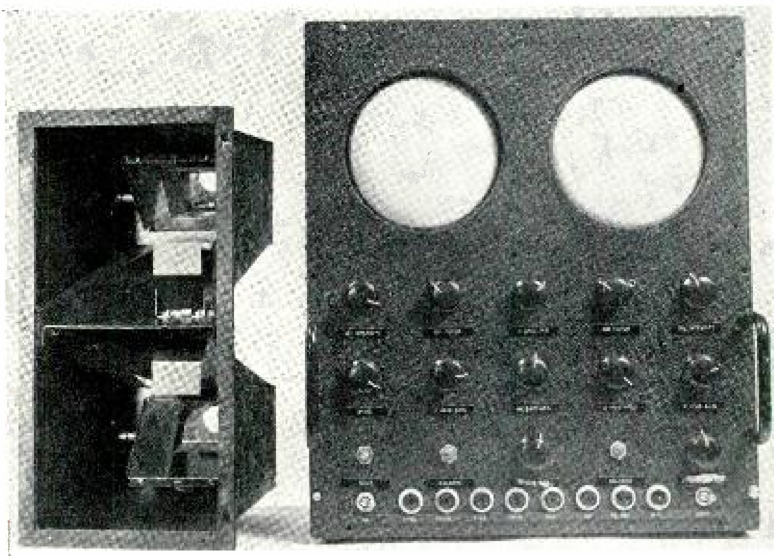


Fig. 2—Stereo-oscilloscope, showing cathode-ray tubes.

Therefore

$$D = \frac{SC}{L - S}.$$

This equation is plotted in Figure 4 and is a hyperbola which is asymptotic to the lines $S = L$ and $D = -C$. The curve shows that the production of an undistorted third dimension requires non-linear spot displacements on the cathode ray tubes. Usually, however, it is desirable to produce a third dimensional scale which is about the same length as the other two scales and therefore only a small part of the curve of Figure 4 need be used. For small excursions of the spots toward and away from each other, the curve is seen to be approximately linear particularly if the lower portion near $S = 0$ is used. By differentiating Equation (1) the slope of the curve at any point may be found.

$$\frac{dD}{dS} = \frac{CL}{(L-S)^2}$$

at $S = 0$ the slope m is:

$$m = \frac{C}{L}.$$

This tangent is plotted in Figure 4 and indicates that linear individual spot excursions up to $\pm .2\,L$ about $S = 0$ will produce negligible distortion in the third dimension. Locating the midpoint of the spot

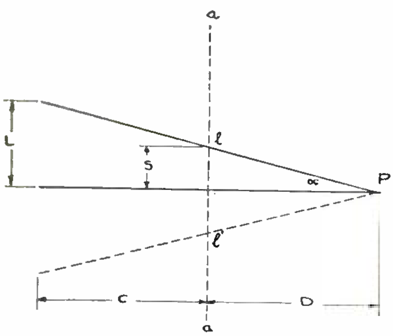


Fig. 3—Geometry of steroscopic viewing system.

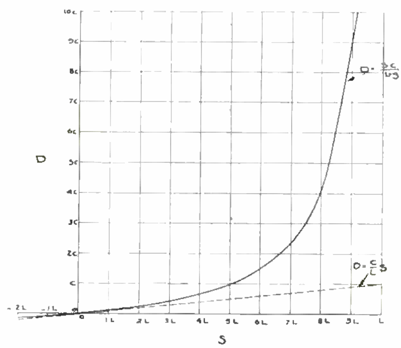


Fig. 4—Apparent distance of fused image as a function of spot separation.

excursions at $S = 0$ has the added advantage of making distance $wxyz$ (Figure 1) equal to wxP. This means that the eyes must focus at about the same distance at which they converge as is normally the case in viewing any object. Letting $c = 16$ inches and limiting S to values between $-.2\,L$ and $+.2\,L$ a third dimensional scale 6.4 inches long is produced which is adequate for most applications since the other two scales are limited by the size of the tubes to something less than 5 inches.

Figure 5 shows a circuit suitable for producing a full three dimensional scan. A voltage applied to terminals x-x moves the two cathode ray tube spots parallel to the X axis and a voltage applied to terminals y-y moves the spots parallel to the Y axis. It should be noted that these voltages move the spots in the same direction keeping them always the same distance apart so that when they are viewed through the optical system of Figure 1 they appear as a single spot moving in the XY plane.

A voltage applied to the terminal z-z is fed to a triode phase splitter which is coupled to corresponding plates on the cathode ray tubes. The voltages generated by the phase splitter are of equal magnitude and opposite polarity causing the spots on the cathode ray tubes to move in opposite directions. When seen through the optical system the effect is that of a single spot of light which moves toward and away from the observer parallel to the Z axis. It is possible then to move the spot anywhere within a prescribed volume of space simply by introducing the proper voltages to the three sets of terminals. This implies that a volume may be scanned rapidly by electronic means.

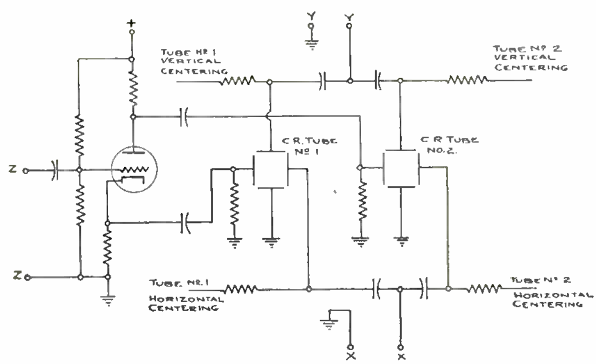


Fig. 5—Circuit for three-dimensional scan.

The device described above has many applications. Figure 6 is a stereo-photograph of a helix which was produced when a linear sweep voltage was introduced on the x-x terminals and sine and cosine voltages of equal frequency were introduced on the y-y and z-z terminals. Possible oscillographic applications might include the observation of grid voltage, plate voltage and plate current relations for a vacuum tube, the observation of filter characteristics in the complex plane, or observation of the relations between pressure, volume and temperature in a mechanical engine.

Radar Presentations

As mentioned previously the device can be applied to radar as a method of presenting three-dimensional information. This application was tested using a laboratory K-band system which was capable of linearly scanning 40 degrees in azimuth about 30 times per second and sinusoidally scanning as much as 30 degrees in elevation about once per second with a symmetrical beam having a half power width of 1 degree. The pulse repetition rate was 4,000 per second. Objects were seldom seen at distances greater than about 5 miles because of terrain characteristics and the relatively low power used (10 kilowatts peak). Because a complete scan required approximately a second, long persistence cathode ray tubes were used. This gave the indicator volume the long persistence characteristics necessary to produce a continuous three dimensional picture.

Figure 7 is a photograph of the stereo indicator tubes showing radar information in three dimensions. The indicator was connected to give a type *C* indication (azimuth vs. elevation) with range shown as the third or depth dimension. The indications seen are from ground objects extending to a range of about two miles. The smeared indication at the bottom is a group of trees about 300 yards from the trans-

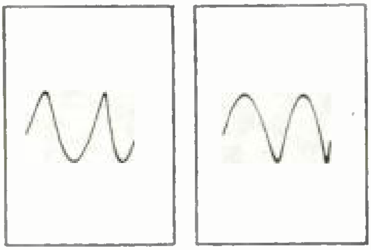


Fig. 6—Photograph of stereo-oscillogram of helix.

Fig. 7—Photograph of stereoscopic radar indications.

mitter, and the distant indications at the top are built up areas in Princeton, New Jersey about two miles distant along the top of a hill.

When the indicator mentioned above was viewed, it gave the impression of a three-dimensional model of the scene. By changing the wiring of the viewing apparatus the apparent position of the observer could be moved to any of the six sides of the figure (depending upon whether azimuth, elevation or range was shown stereoscopically). Many observers thought the pictures were too small and dim for commercial use, and some had trouble in fusing the two pictures into a single image.

There is no doubt that these difficulties could have been reduced by the use of known techniques and greater care in adjusting the optical system. However, instead of exhausting the possibilities of this method it was considered advisable to test another which offered the possibility of providing more pictures per second, more light, and pictures of identical proportions for the two eyes to view.

Alternate Fan Beam Scanner

To obtain these advantages it was proposed that the pictures for the two eyes should be presented alternately on a single cathode-ray tube, that the observer should wear polaroid glasses with crossed polarizations, and that the picture appropriate to each eye should be selected by moving polaroid filters in front of the tube. This, of course, raised the question of how many alternate pictures per eye per second it is necessary to have in order to retain stereoscopic vision.

Tests were made to determine the accuracy with which an index could be reset at the apparent position of a stereoscopic pair of spots located ten feet from the eyes. When the spots were illuminated continuously each of three observers repeated his settings within about ± 1 per cent of the distance; because of differences in interocular distance the apparent position differed slightly between observers. When the pictures were shown alternately to the two eyes, speeds above 40 pictures per eye per second were essentially equivalent to continuous illumination. At lower speeds flicker was noticed, particularly when the spots were bright. At 20 pictures per eye per second flicker was tolerable, and ability to judge distance unimpaired. Eight pictures per eye per second flickered objectionably, but there was no great difficulty in fusing the spots into a single image. The accuracy with which the index could be reset was about ± 2 per cent. Speeds much lower than this gave the impression of two independent flashing lights. From these tests it was concluded that about 20 pictures per eye per second would be needed for a satisfactory system.

It seems unreasonable to use a pencil beam scanning space to provide information so rapidly. Aside from the technical problem of building the scanning antenna, the speed of light is such that one could scan only 232 beam widths in 1/40 second, if a range of 10 miles and one pulse on the target per scan are assumed. A radar presentation with $\sqrt{232} = 15$ lines resolution would not be highly regarded.

Fortunately, it is not necessary to scan with a pencil beam. The V-beam height finder developed during the war used two fan beams, narrow in azimuth and wide in elevation. One of the fans was located in a vertical plane, and the echoes obtained with it gave the true azimuth and range of an object. The second beam was in a plane tilted somewhat from vertical, so that the azimuth position at which an echo was received was a function of the height of the reflecting object. A comparison of the location of the two echoes from a single object allowed the height to be determined.

It appeared that this principle could be extended to the use of two fan beams, tilted opposite directions from vertical, and scanning alternately. The time for one sweep can be made 1/40 second or less, and the resulting radar pictures can be viewed stereoscopically to give the impression of a three dimensional model viewed from above.

Figure 8 is a photograph of a scanner built to cover 90 degrees of azimuth (less 8 degrees of switching time) with a beam having a half-power width of 1.9 degrees in azimuth and 11 degrees in elevation. In one revolution there is one sweep from each of the four antennas. Alternate antennas are tilted plus and minus 10 degrees from the vertical so that the fan beam from one antenna illuminates an elevated

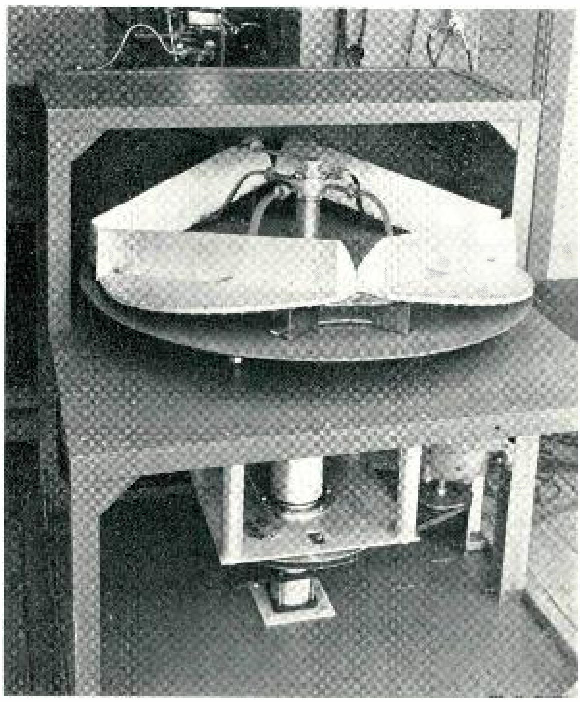


Fig. 8—Scanning antenna for alternate-fan-beam stereoscopic radar system.

object sooner than a ground object at the same bearing and the fan beam from the next antenna illuminates the elevated object later than the ground object.

Figure 9 is a photograph of the viewing device. The pictures from all the antennas are presented on a single-cathode-ray tube, the screen of which is viewed through an inclined mirror. A drum carrying four polaroid filters revolves around the mirror, and when the observer wears suitably polarized glasses the eye which sees the screen is determined by the filter which is interposed. The drum is driven by a selsyn to keep it in step with the motion of the scanning antennas.

On the cathode ray tube the scanning pattern is preferably of the type shown in Figure 10. This differs from a quadrant of a Plan-Position-Indicator in that lines of constant range are straight, rather than arcs of a circle. This change is needed to enable the eyes to fuse the two spots into a single image. In a display of this type the apparent elevation of a reflecting object is very nearly independent of range, as long as its elevation angle is less than 20 degrees, as viewed from the transmitter.

Fig. 9—Viewing device for alternate-fan-beam stereoscopic radar system.

A scanning and viewing system as described above was tested, using a medium power radar transmitter on a wavelength of $1\frac{1}{4}$ centimeters. A two-place Aeronca plane flying at an elevation between 600 and 800 feet and a range of one mile was seen clearly, and the apparent position of its image was well above objects on the ground. At a scanning speed of 18 pictures per eye per second the flicker did not seem objectionable. The illumination and size of the presentation were considerably greater than in the previous test using a lower scanning rate and two long-persistence cathode-ray viewing tubes.

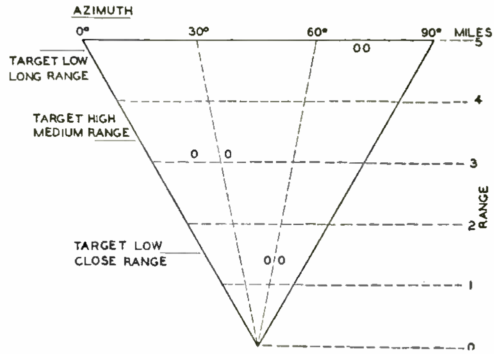


Fig. 10—Scanning pattern for alternate-fan-beam stereoscopic radar system.

Although the performance of the alternate-fan-beam system was better than that of the pencil-beam system in most respects, the range appeared to be less. This is to be expected, because the act of spreading the power into a fan resulted in loss of antenna gain which was not compensated by the reduction of scanning loss.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to the men who gave such substantial assistance in carrying out the project. I. Wolff made many valuable suggestions; H. B. DeVore designed the radar system on which the tests were made; and W. C. Wilkinson assisted in the testing and adjustment of the fan-beam antennas.

FREE SPACE MICROWAVE PROPAGATION*

By

A. L. Hammerschmidt

Engineering Development, National Broadcasting Company, Inc.,
New York, N. Y.

Summary—*Nomographic charts have been developed which facilitate the calculation of the performance of microwave relay equipment. One chart shows the relationship of frequency, distance, and path attenuation between isotropic antennas. A second chart shows the relationship between diameter, frequency, and gain of a parabolic reflector antenna with respect to an isotropic antenna. An example demonstrates the application of these charts to a transmission problem, and a table indicating calculated transmission characteristics of a typical microwave television relay equipment for various antenna combinations is included.*

Introduction

Excellent methods of computing the transmission characteristics of microwave relay systems have been covered in previously published papers.[1,2] However, application of these methods to a particular relay circuit is often laborious and time-consuming. This paper describes recently developed nomographic charts which relate the major factors encountered in free-space propagation. Use of these charts greatly simplifies the computations involved in determining the transmission characteristics of a microwave relay system.

Nomographic Charts

Two charts are included in this paper. Chart A shows the relationship of frequency, distance, and path attenuation between isotropic (hypothetical omnidirectional) antennas. This chart was developed from the free space formula (Appendix A). The derivation of the formula used to construct the chart is indicated in Appendix B.

Chart B shows the relationship between diameter, frequency, and gain of a parabolic reflector antenna referred to an isotropic antenna. A parabola efficiency of 65 per cent, i.e., the effective area equals 65

* Decimal Classification: R113.737.

[1] H. T. Friis, "A Note on a Simple Transmission Formula", *Proc. I.R.E.*, Vol. 34, No. 5, pp. 254-255, May, 1946.

[2] C. W. Hansell, "Radio Relay Systems Development by the Radio Corporation of America", *Proc. I.R.E.*, Vol. 33, No. 3, pp. 156-168, March, 1945.

per cent of the aperture area, was assumed in the preparation of Chart B. This efficiency is readily obtainable in practice. The derivation of the formula used to construct this chart is indicated in Appendix C.

Application of Charts

An example of the use of these charts applied to a transmission problem involving typical microwave television relay equipment follows. In this example, it is required to determine the line-of-sight distance which will provide a 40-decibel signal-to-noise ratio transmission circuit. It is assumed that parabolic reflector antennas four feet in diameter are utilized at each end of the circuit.

Transmitter carrier frequency	7,000 megacycles
Power applied to transmitting antenna terminals	100 milliwatts
Receiver noise factor	20 decibels

The receiver noise power due to thermal agitation for a band of modulation frequencies five megacycles wide is (Appendix D)

$$P_m = (0.8 \times 10^{-20})\,(10 \times 10^{6}) = 8 \times 10^{-14} \text{ watts.} \qquad (1)$$

Assuming the receiver noise factor is 20 decibels (100 times) above thermal, the receiver noise power (P_n) due to all equipment causes is then

$$P_n = (100)\,(P_m) \quad \text{or} \quad P_n = (100)\,(8 \times 10^{-14}) = 8 \times 10^{-12} \text{ watts.} \qquad (2)$$

The permissible path attenuation for a one-to-one signal-to-noise ratio is

$$10 \log \frac{P_t}{P_n} = 10 \log \frac{100 \times 10^{-3}}{8 \times 10^{-12}} = 101 \text{ decibels.} \qquad (3)$$

Since a 40-decibel signal-to-noise ratio is specified, Equation (3) above becomes

$$101 \text{ decibels} - 40 \text{ decibels} = 61 \text{ decibels.} \qquad (4)$$

From Chart B, it is found that the gain of a four-foot parabola at a frequency of 7,000 megacycles is 37.25 decibels. The total gain due to the receiving antenna and transmitting antenna is then

$$(2)\,(37.25 \text{ decibels}) = 74.5 \text{ decibels.} \qquad (5)$$

The improvement factor due to the FM method of transmission for unity frequency deviation assuming the television signal is symmetrical is (Appendix E)

$$20 \log \frac{\sqrt{3}\, f_d}{f_a} = 20 \log \sqrt{3} = 4.8 \text{ decibels.} \tag{6}$$

The maximum permissible signal attenuation is found by adding Equations (4), (5), and (6) above.

$$61 + 74.5 + 4.8 = 140.3 \text{ decibels.} \tag{7}$$

Placing a straight edge on the 7,000 megacycles and 140.3 decibel points of Chart A, a distance of 22 miles is obtained.

Table I demonstrates the use of the nomographs for calculating the free-space transmission characteristics of the above-described equipment for several combinations of transmitting and receiving antennas. In each case, the maximum permissible path attenuation is obtained by adding Equations (4) and (6) above.

A signal-to-noise ratio of 40 decibels was chosen for these examples as a reasonable compromise for a single-hop television relay circuit. Considerations other than good engineering practices may dictate a poorer signal-to-noise ratio. For example, if a 34-decibel signal-to-noise ratio is tolerable, the distances indicated in Table I will be doubled.

Table I—Calculated Performance
Microwave Television Relay Equipment

Antennas Used	Total Gain of Antennas (decibels)	Maximum Permissible Path Attenuation (decibels)	Free Space Range 40 decibels signal-to-noise (miles)
Two 4 foot parabolas	74.5	140.3	22
4 foot and 6 foot parabola	77.9	143.7	32.5
Two 6 foot parabolas	81.4	147.2	49
Two horn* antennas	47.2	113	1 (Approx.)
One horn* and one 4 foot parabola	60.8	126.6	4.5
One horn* and one 6 foot parabola	64.2	130	6.8

* Horn Antenna: Aperture 11.68 × 8.84 inches
Aperture area = 103.25 square inches
Approx. effective area = 51.6 square inches or 333 square centimeters (Appendix A)

$$\text{Power Gain} = \frac{(4\pi)\,(\text{Effective Area})}{\lambda^2} = 227 \text{ or } 23.6 \text{ decibels}$$

Additional Considerations

It should be emphasized that the term signal-to-noise ratio as used in this paper refers to radio frequency root-mean-square values. An improvement factor due to FM has been included in the noise figure since the examples of this paper are based on an equipment utilizing the frequency-modulated system of transmission. In order to evaluate the overall performance of an amplitude-modulated system, it is necessary to include a factor related to the degree of modulation of the radio-frequency carrier.

In a practical relay circuit, other factors which may alter the calculated results include:

1. Effect of the ground reflected signal which, depending upon its phase, may decrease or increase the intensity of the received signal. This effect may be minimized by the proper choice of antenna heights,[2] or under certain circumstances by utilizing extremely directive antennas.

2. Obstructions in the path or the path clearance including obstructions between the receiving antenna and transmitting antenna. Aircraft, for example, may constitute transient obstructions.

3. Attenuation due to rain drops.[3] This effect is usually negligible at 1,000 megacycles but may become appreciable at 10,000 megacycles.

4. Fading due to anomalous propagation.[4,5]

Appendix

A. *Free Space Transmission Formula*[1]

The free space transmission formula is

$$\frac{P_r}{P_t} = \frac{A_r A_t}{d^2 \lambda^2} \tag{1}$$

[3] Sloan D. Robertson and Archie P. King, "The Effect of Rain Upon the Propagation of Waves in the 1- and 3-Centimeter Regions", *Proc. I.R.E.*, Vol. 34, No. 4, pp. 178P-180P, April, 1946.

[4] Edwin G. Schneider, "Radar", *Proc. I.R.E.*, Vol. 34, No. 8, pp. 528-578, August, 1946.

[5] M. Katzin, R. W. Bauchman, and W. Binnian, "3- and 9-Centimeter Propagation in Low Level Ocean Ducts", *Proc. I.R.E.*, Vol. 35, No. 9, pp. 891-905, Sept., 1947.

where P_t = power fed to input terminals of transmitting antenna
P_r = power available at output terminals of receiving antenna
} Same Units

A_t = effective area of transmitting antenna
A_r = effective area of receiving antenna
d = distance between antennas
λ = wavelength
} Same order of Units

Typical antennas have the following effective areas[1]*:*

1. Effective area of isotropic $= \dfrac{\lambda^2}{4\pi}$
2. Effective area of half-wave dipole $= 0.1305\lambda^2$
3. Effective area of parabola $\approx$ 65 per cent of aperture area
4. Effective area of optimum horn $\approx$ 50 per cent of aperture area

B. *Path Attenuation Between Isotropic Antennas*

With isotropic antennas at the receiving and transmitting locations, Equation (1) becomes

$$\frac{P_t}{P_r} = 4.543 \times 10^3 f^2 d^2 \qquad (2)$$

where f = frequency in megacycles
d = distance between antennas in miles

The path attenuation is obtained by rearranging (2) above.

Path Attenuation (decibels) $= 10 \log [(4.543 \times 10^3)(f^2)] + 10 \log d^2$

Equation (3) was used in the construction of Chart A. (3)

C. *Parabolic Reflector Gain*

The power gain of an antenna equals the effective area of that antenna divided by the effective area of the antenna to which it is compared (usually an isotropic). The gain of a half-wave dipole referred to an isotropic is therefore 2.15 decibels. Likewise, the power gain of a parabolic reflector antenna referred to an isotropic is given by

$$\text{Power Gain} = \frac{4\pi A}{\lambda^2} \qquad (4)$$

where A is the effective area of the parabola and is in the same units as λ. Equation (4) may be reduced to

$$\text{Power Gain} = 2.649 \times 10^{-5} f^2 r^2 \qquad (5)$$

where f = frequency in megacycles

r = radius of reflector in feet

Rearranging Equation (5)

$$\text{Power Gain (decibels)} = 10 \log [(2.649 \times 10^{-5})(f^2)] + 10 \log r^2 \qquad (6)$$

Equation (6) was used in the construction of Chart B.

D. *Receiver Noise Factor*[2]

The noise power due to the thermal agitation is

$$P_m = (0.8 \times 10^{-20})\,(B) \text{ watts} \qquad (7)$$

where B is the effective bandwidth of the relay system in cycles per second (normally twice the band of modulation frequencies employed). The receiver noise power due to all equipment causes will usually be between 10 and 100 times the noise power due to thermal agitation as indicated by Equation (7).

E. *FM Improvement Factor*

The signal-to-noise ratio improvement due to the use of FM is[6]

$$\text{FM Improvement Factor (RMS volts)} = \sqrt{3}\,\frac{f_d}{f_a}$$

where f_d = peak frequency deviation of carrier

f_a = highest modulating frequency

The above formula is valid for symmetrical modulation of the FM carrier. The asymmetrical character of a television signal may, therefore, alter the improvement factor somewhat.

---

[6] Murray G. Crosby, "Frequency Modulation Noise Characteristics", *Proc. I.R.E.*, Vol. 25, No. 4, pp. 472-514, April, 1937.

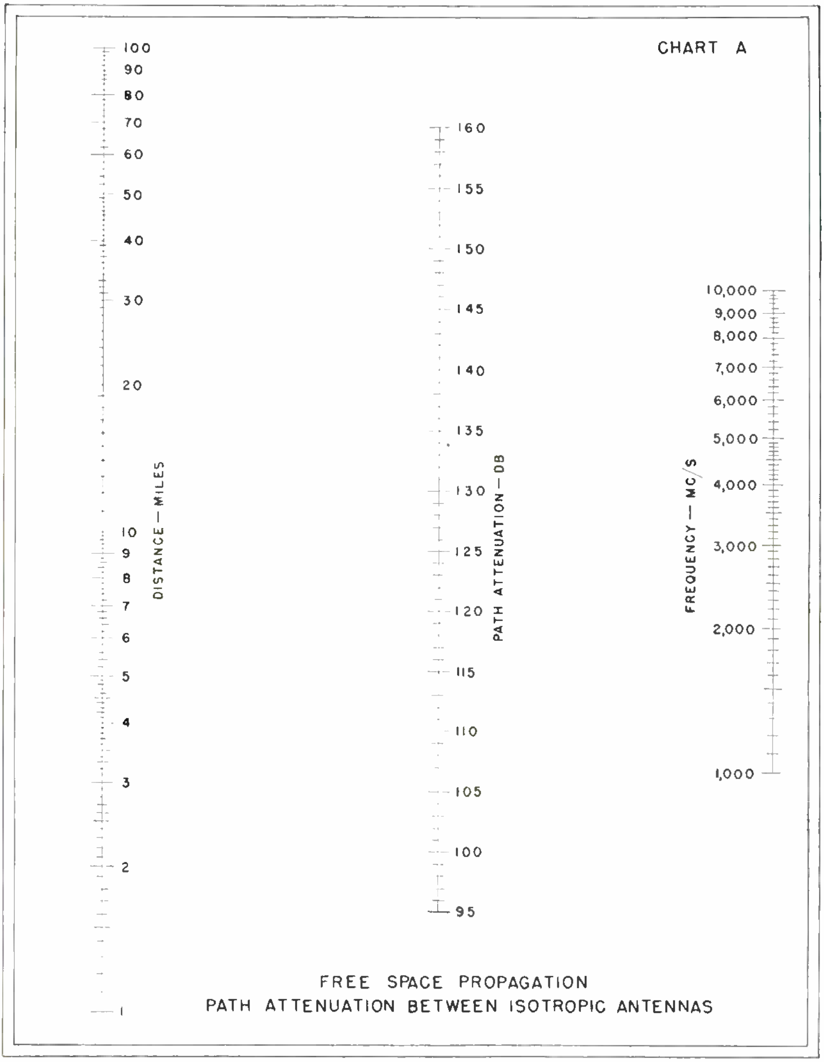
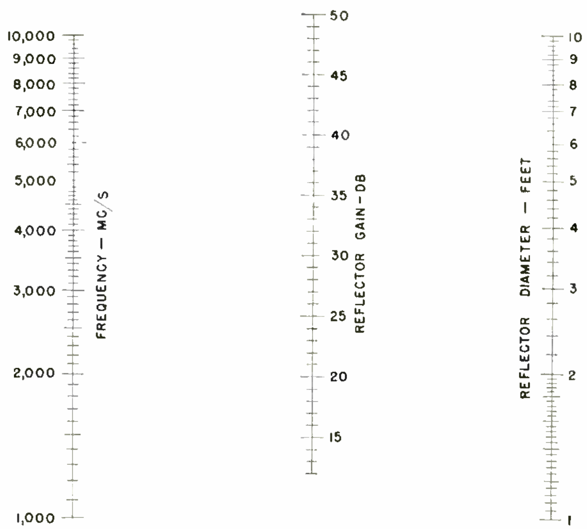


CHART A
100
90
80
70
60
50
40
30
20
10
9
8
7
6
5
4
3
2
1
DISTANCE — MILES
160
155
150
145
140
135
130
125
120
115
110
105
100
95
PATH ATTENUATION — DB
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
FREQUENCY — MC/S
FREE SPACE PROPAGATION
PATH ATTENUATION BETWEEN ISOTROPIC ANTENNAS

CHART B



PARABOLIC ANTENNA GAIN
REFERRED TO ISOTROPIC ANTENNA
ASSUMED EFFICIENCY 65%

RCA TECHNICAL PAPERS†

Fourth Quarter, 1947

Any requests for copies of papers listed herein should be addressed to the publication to which credited.

"An Analysis of FM Servicing", N. L. Chalfin, *RCA Rad. Serv. News* (October-November) 1947

"Audio Noise Reduction Circuits", H. F. Olson, *Electronics* (December) 1947

"The Behavior of 'Magnetic' Electron Multipliers as a Function of Frequency", L. Malter, *Proc. I.R.E.* (October) 1947

"Blower Requirements for RCA Forced-Air-Cooled Tubes", *RCA Application Note AN-125*, RCA Tube Department, Harrison, N. J. (November 15) 1947

"A Comparison of Lateral and Vertical Disc Recording Systems", R. A. Lynn, *Broad. Eng. Jour.* (December) 1947

"Compensation of Frequency Drift", *RCA Application Note AN-122*, RCA Tube Department, Harrison, N. J. (October 15) .. 1947

"Dynamic Symmetry and Acoustic Room Design", M. Rettinger, *Audio Engineering* (October) 1947

"The Electron and Light Microscopy of Beggiatoa", F. H. Johnson and R. F. Baker, *Jour. Cellular and Comparative Physiology* (October) 1947

"Electronic Metal Detector", P. A. Greenmeyer, *Electronic Ind.* (November) 1947

ELEMENTS OF ACOUSTICAL ENGINEERING (Second Ed.), H. F. Olson, D. Van Nostrand Co., Inc., New York, N. Y. 1947

"Force at the Stylus Tip While Cutting Lacquer Disk-Recording Blanks", H. E. Roys, *Proc. I.R.E.* (November) 1947

"Frequency Allocations", P. F. Siling, *RCA REVIEW* (December) 1947

THE FUTURE OF TELEVISION, (Rev. Ed.), O. E. Dunlap, Jr., Harper and Bros., New York, N. Y................. 1947

"Improving Candy Inspection Methods", P. A. Greenmeyer, *Manufacturing Confectioner* (October) 1947

"Induction Effects in Terrestrial Magnetism; Part III—Electric Modes", W. M. Elsasser, *Phys. Rev.* (November 1) 1947

"Interlocked Scanning for Network Television", J. R. DeBaun, *RCA REVIEW* (December) 1947

† Report all corrections or additions to RCA REVIEW, Radio Corporation of America, RCA Laboratories Division, Princeton, N. J.

"Intermodulation Distortion Analysis as Applied to Disk Recording and Reproducing Equipment", H. E. Roys, *Proc. I.R.E.* (October) 1947

"Lightweight Recorders for 35- and 16-mm Film", M. E. Collins, *Jour. Soc. Mot. Pic. Eng.* (November) 1947

"Measured Characteristics of the Pylon Antenna", O. O. Fiet, *Broadcast News* (December) 1947

"The Motion of Electrons Subject to Forces Transverse to a Uniform Magnetic Field", P. K. Weimer and A. Rose, *Proc. I.R.E.* (November) 1947

"New Television Field-Pickup Equipment Employing the Image Orthicon", J. H. Roe, *Proc. I.R.E.* (December) 1947

"Performance Characteristics of Long-Persistence Cathode-Ray Tube Screens; Their Measurement and Control", R. E. Johnson and A. E. Hardy, *RCA REVIEW* (December) 1947

"Radio—Scientific Developments and Television in the U. S.", SECTION OF THE TEN EVENTFUL YEARS 1937-1946, Encyclopedia Britannica 1947

"Radio—1946", G. L. Beers, SECTION OF AMERICAN YEARBOOK .. 1947

"Radiophoto Standards", S. H. Simpson, Jr. and R. E. Hammond, *RCA REVIEW* (December) 1947

"Ratio Detectors for FM Receivers", S. W. Seeley, *FM & Tele.* (December) 1947

"Receiver Microphonics Caused by Heater-Cathode Capacitance Variations", *RCA Application Note AN-123*, RCA Tube Department, Harrison, N. J. (October 15) 1947

"Recent Developments in Radiotelegraph Transmitters for Shore Stations", J. F. McDonald, *RCA REVIEW* (December) 1947

"Recorder and Timer for Short Intervals", W. H. Bliss, *Electronics* (November) 1947

"Slot Antennas", N. E. Lindenblad, *Proc. I.R.E.* (December) .. 1947

"Small-Signal Analysis of Traveling-Wave Tube", C. I. Shulman and M. S. Heagy, *RCA REVIEW* (December) 1947

"Storage Orthicon and Its Application to Teleran", S. V. Forgue, *RCA REVIEW* (December) 1947

"Suppression of Arc-Over, Corona, and High-Voltage Leakage in the 5TP4 Kinescope", *RCA Application Note AN-124*, RCA Tube Department, Harrison, N. J. (November 15) 1947

"Technical Aspects of Television Studio Operation", R. W. Clark and H. C. Gronberg, *RCA REVIEW* (December) 1947

"TELERAN—Part II, First Experimental Installation", D. H. Ewing, H. J. Schrader and R. W. K. Smith, *RCA REVIEW* (December) 1947

"Television Field Equipment", J. H. Roe and N. S. Bean, *FM & Tele.* (November) 1947

"Television Field Equipment", J. R. Smith, *FM & Tele.* (December) 1947

NOTE—Omissions or errors in these listings will be corrected in the yearly index.

AUTHORS

RICHARD BURTNER received his B. S. degree in Electrical Engineering from Purdue University in 1942. From 1939 to 1942 he was a member of the technical staff of radio station WBAA, Lafayette, Indiana. In 1942 he joined the Cathode-Ray Tube Department of the RCA Victor Division, Lancaster, Pa., and in 1943 was transferred to RCA Laboratories Division, Princeton, N. J. where he was engaged in research on rapid scanners and microwave radar systems. During 1946, he conducted research on television distribution systems for the Telicon Corporation of New York City and since January 1947, has been employed in the Aerophysics Laboratories of the Goodyear Aircraft Corporation. Mr. Burtner is a Member of Sigma Xi and the Institute of Radio Engineers.

WENDELL L. CARLSON was graduated from the Bliss Electric School in 1918. Since 1918 he has been engaged in radio receiver development, first at the Washington Navy Yard (1918 to 1923); next, with the General Electric Company at Schenectady (1923 to 1930); then, with the RCA Manufacturing Company at Camden, N. J. (1930 to 1941); and since 1941, he has been with RCA Laboratories Division at Princeton, N. J. Mr. Carlson is an Associate Member of the Institute of Radio Engineers.

CHARLES H. CHANDLER received the B.A. degree in physics and mathematics from the College of Wooster in 1940, and the M.Sc. degree in physics from the Ohio State University in 1946. He was a graduate assistant in the Physics Department of the University while taking graduate work during the academic year 1940-1941. In 1941 he was commissioned in the Signal Corps, where his assignments included radar maintenance and communications frequency-monitoring overseas, and administration of radar research and development projects at Camp Evans, N. J. After leaving the Army in 1946, he joined RCA Laboratories Division, in Princeton, N. J., working on microwave applications. Mr. Chandler is a Member of the Institute of Radio Engineers.

ROBERT M. COHEN joined the Research and Engineering Dept., Radiotron Division, RCA Manufacturing Company in May 1940. He has specialized in receiving-tube application engineering. At present he is assigned to the development of new receiving tubes in the Applications Engineering Laboratory, Tube Department, RCA Victor Division at Harrison, N. J. He is a Member of the Institute of Radio Engineers.

LESLIE E. FLORY received his B.S. degree in Electrical Engineering from the University of Kansas in 1930. In the same year he joined RCA Manufacturing Company at Camden, N. J., conducting research on television tubes and related electronic problems. In 1942, he transferred to RCA Laboratories Division at Princeton, N. J., continuing work on electronic tubes and special circuit applications. Mr. Flory is a Member of Sigma Xi and a Senior Member of Institute of Radio Engineers.

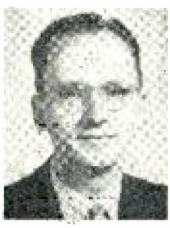
A. L. HAMMERSCHMIDT received the B.E.E. degree from Ohio State University in 1938. Between 1934 and 1938 he was employed as an operator of WBNS, Columbus, Ohio. From 1938 to 1941 he was Technical Supervisor of the Ohio State University Radio Station WOSU. In 1941 he joined the Engineering Department of NBC in the capacity of Television Engineer, later transferring to the Development Group as Development Engineer. Mr. Hammerschmidt is a member of Eta Kappa Nu and a Senior Member of the Institute of Radio Engineers.

WILLIAM D. HERSHBERGER received the A.B. degree in 1927 from Goshen College, the A.M. degree in Physics in 1930 from George Washington University and the Ph.D. degree in Electrical Engineering from the University of Pennsylvania in 1937. From 1927 to 1931 he worked on supersonics at the Naval Research Laboratory and from 1931 to 1936 at the Signal Corps Laboratories, where he was responsible for the work on target detection with radio waves. He was a Harrison Fellow at the University of Pennsylvania in 1936-1937. He joined the Radio Corporation of America in 1937. Since 1942 he has been at RCA Laboratories Division, Princeton, N. J., where he has been associated, first, with the work on radio altimeters and airborne radar and, more recently, on physical studies at microwave frequencies. Dr. Hershberger is a Member of the American Physical Society, Sigma Xi, and the American Association for the Advancement of Science, and a Senior Member of the Institute of Radio Engineers.

HARLEY IAMS graduated from Stanford University in 1927 with a B.A. degree in mechanical engineering. From 1927 until 1930, he was associated with the Westinghouse Electric Company working on facsimile transmission and television. In 1931, he joined the research laboratories of the RCA Manufacturing Co. transferring in 1942 to RCA Laboratories Division, Princeton, N. J., where he has been engaged in the development of television, infrared, centimeter-wave and radar tubes and apparatus. In 1947, he joined the North American Aviation Co. in Los Angeles, Calif. Mr. Iams is a Member of Sigma Xi and Phi Beta Kappa, and a Senior Member of the Institute of Radio Engineers.

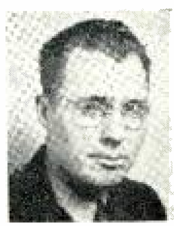
ARTHUR S. JENSEN received the B.S. degree in 1938, the M.S. degree in 1939, and the Ph.D. degree in 1941 in Physics from the University of Pennsylvania where he was a research fellow from 1939 to 1941. After a short period as a research physicist at the Naval Research Laboratory, he reported for active duty with the U. S. Naval Reserve in August, 1941. From then until 1946 he served as an officer-instructor in physics and aviation physics at the U. S. Naval Academy with the rank of lieutenant commander. At the end of 1945, he joined RCA Laboratories Division. Dr. Jensen is a Member of the Institute of Radio Engineers, the American Physical Society, the American Association of Physics Teachers, the American Association for the Advancement of Science, Pi Mu Epsilon, and Sigma Xi.

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, transferring to RCA Laboratories Division at that time, with which Division he is now associated. Dr. Luck is a member of the American Physical Society.



MAX H. MESNER attended Kansas City Junior College and the University of Missouri, graduating in 1940 with a B. S. degree in E. E. From 1940 to 1942, he was employed by RCA Manufacturing Co., Camden, N. J., as a radio engineer working on airborne radar equipment. Since 1942, he has been with RCA Laboratories Division, Princeton, N. J., as a research engineer working principally with storage and computor devices. Mr. Mesner is a Member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, Association for Computing Machinery, and an Associate Member of the Institute of Radio Engineers.

LOWELL E. NORTON received his B.S. degree in electrical engineering in 1932, and the M.S. degree in communications in 1935 from the University of Minnesota. He entered the services of the Research Division of RCA Manufacturing Company at Camden, N. J., in 1935 and since 1942 has been at RCA Laboratories Division in Princeton, N. J. Mr. Norton is a Member of Sigma Xi.



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.

JOHN P. SMITH received the B. S. degree in Electrical Engineering from Texas A. and M. College in 1927 and in the same year joined the General Electric Co. in Schenectady, N. Y. From 1927 to 1930, he was engaged in the Radio Department working on carrier current test and later television development. From 1930 to 1940, he continued television development with RCA Manufacturing Co., Camden, N. J., specializing in electronic synchronizing generators. During the war he worked on various defense projects at RCA Laboratories Division, Princeton, N. J. At present he is engaged in television research at the Laboratories. Mr. Smith is a member of Sigma Xi and a Senior 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. He is now located at RCA Laboratories Division in Princeton, N. J., engaged in television research and development. Mr. Wendt is a Member of Alpha Chi, 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.

RCA