RCA REVIEW

a technical journal

Published quarterly by
RCA LABORATORIES
in cooperation with all subsidiaries and divisions of
RADIO CORPORATION OF AMERICA

VOLUME XXVI MARCH 1965 NUMBER 1

CONTENTS

PAGE

Infrared Detectors 3
G. A. MORTON

Microwave Photomultipliers Using Transmission Dynodes 22
D. BLATTNER, H. JOHNSON, J. RUEDY, AND F. STERZER

A Mode Analysis of Quasi-Isotropic Antennas 42
O. M. WOODWARD

Application of Some Linear FM Results to Frequency-Diversity Waveforms 75
T. B. HOWARD

Helix Support Structure for Ultra-Wide-Band Traveling-Wave Tubes 106
E. F. BELOHOUBEK

Switching Analysis of Gate-Controlled Space-Charge-Limited Emission Processes in Semiconductors 118
R. C. WILLIAMS

Application of Content-Addressed Memory for Dynamic Storage Allocation 140
Y. CHU

RCA Technical Papers 153

Authors 156

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

RCA REVIEW

BOARD OF EDITORS

INFRARED DETECTORS

By

George A. Morton

RCA Electronic Components and Devices,
Princeton, N. J.

Summary—*In applications requiring great sensitivity and high speed of response, infrared photoconductive detectors have proved very effective. The photoconductor employed depends upon the wavelength of interest. For the very near infrared (long wave limit $\lambda_m = 1.8\ \mu$), germanium junction photodiodes are used. At somewhat longer wavelengths ($\lambda_m \sim 3.5\ \mu$), lead sulfide and indium arsenide have suitable responses. For still longer wavelengths ($\lambda_m \sim 7\mu$), PbTe, PbSe, and InSb are the most frequently used photoconductors. Except for relatively low sensitivity InSb junction devices, detectors employing these materials require liquid nitrogen cooling. Extrinsic photoconductors constituted of appropriately doped germanium are used for longer wavelengths. These can be made with a wide range of long-wave limits covering the spectrum to beyond 100 μ depending upon the activator employed. Such cells require considerable cooling to reduce thermal excitation of carriers. The sensitivity of these detectors depends upon their cooling, the level of background radiation, and the configuration of the detector. Methods of calculating the sensitivity, cooling requirements, and performance characteristics are discussed and illustrative examples given.*

Introduction

ALTHOUGH INFRARED radiation has been known since about 1800, when it was discovered by Sir William Herschel, very little effort went into the development of infrared detectors until the late 1930's and early 1940's. Since then the field has been very active and a very extensive technology has developed around the subject of infrared detectors.

Infrared radiation is that portion of the electromagnetic spectrum from just below the visible spectrum, i.e., 7500 Å, to the microwave radio region. Most of the black-body radiation from objects heated to temperatures normally encountered in terrestrial environments lies in the infrared portion of the spectrum. For example, an object at room temperature (300°K) has its peak of black-body radiation at about 10 μ. Because of this, infrared detectors are very effective devices for sensing the presence of distant objects that differ only slightly in temperature from their ambient. As a consequence these detectors are of considerable military importance. This fact has contributed enormously to their development, but will not be discussed further here. Infrared detectors are also useful in thermometric instruments, in determining certain chemical bindings by spectroscopic analysis, in plasma diagnostics, in astronomy, and in medicine. The

present discussion, however, is concerned with the physics of infrared detectors rather than their application.

There are two broad classes of detectors, namely, photon detectors and temperature detectors. In the latter, infrared radiation changes the temperature of the sensing element, and this change of temperature produces an observable electrical or physical change in the sensor. Photon detectors are those in which photons of the radiation produce directly an excitation in the sensing layer that can be observed electrically or otherwise. The present discussion is restricted to photon detectors that produce electrical response to infrared radiation.

Figure 1 shows the portion of the electromagnetic spectrum under consideration. Wavelength in microns is indicated along the center horizontal line. Below the line are indicated the principal detectors

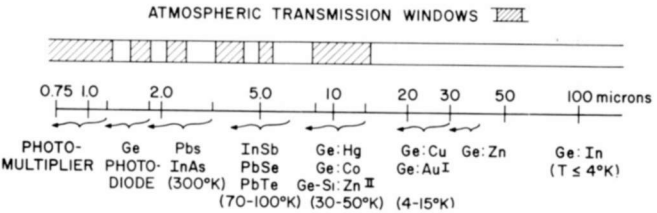


Fig. 1—Infrared spectrum with principal detectors.

used in the various portions of the spectrum. The shaded strips above the line give a rough indication of the spectral regions wherein normal atmosphere is relatively transparent. It will be noted that there are a number of windows in the region between the end of the visible spectrum and about 6 μ. Then there is a rather long opaque region and another window extending from about 8.5 to 13.5 μ. Beyond this the atmosphere is relatively opaque out virtually to the radio region with only a few small windows. The oxygen and nitrogen of the atmosphere are quite transparent for most of the infrared spectrum. The absorbers encountered in a normal atmosphere are primarily carbon dioxide and water vapor, particularly the latter. Since both of these gases decrease quite rapidly with altitude, the upper atmosphere is very much more transparent to infrared radiation.

Infrared Photomultipliers

The first detector named on the chart, sensitive in the region from 0.75 to about 1.2 μ, is the photomultiplier. An infrared-sensitive photomultiplier (type 7102) is shown diagrammatically in Figure 2.

This multiplier has a photocathode that emits electrons when it is exposed to infrared radiation in the spectral region indicated. These electrons are accelerated from the cathode and focused onto the surface of the first dynode of the multiplier. The dynode is coated with a layer that emits several electrons when bombarded by a primary electron. These electrons are accelerated onto a second dynode with a similar surface where each produces a number of secondary electrons. This

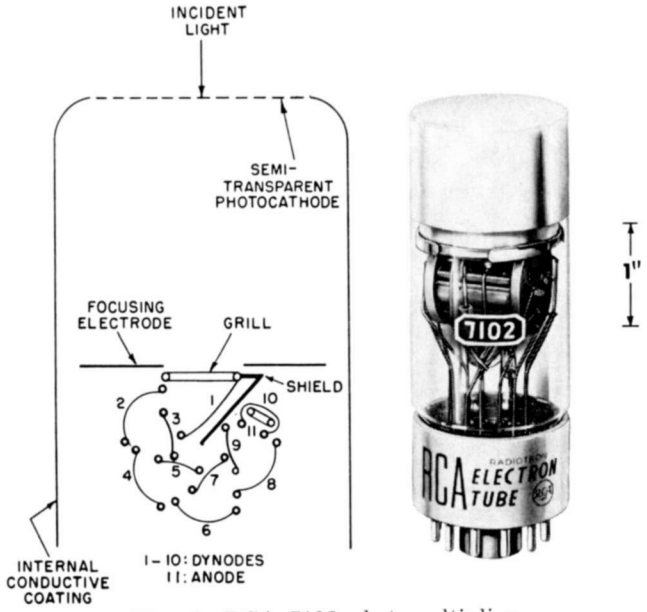


Fig. 2—RCA 7102 photomultiplier.

process is repeated for as many stages as is required to bring the electron current up to the desired level, at which point it is collected on the anode. If each dynode produces δ electrons for each primary, and there are k dynodes in cascade, the factor by which electrons from the cathode (usually referred to as the gain (G) of the tube) are multiplied is δ^k.

The most important element in this device is, of course, the photocathode. This infrared-sensitive cathode, which has been designated as an S-1 cathode, consists of a cesium oxide matrix containing silver, possibly in colloidal form, and a layer of cesium at the vacuum interface. The cathode is of interest because it is one of the oldest practical photocathodes known and the least understood of all of the cathodes in practical use. All efficient photoemitters are semiconductors

and generally take the form of a thin film of semiconductor treated on the vacuum interface with a material such as Cs to minimize the electron affinity, i.e., the potential difference between the vacuum potential and the conduction band of the semiconductor. This model can give a quantitative account of the performance of most photocathodes (e.g., all of the alkali antimonide cathodes); however, it does not suffice for the cesium–oxygen–silver cathode used in infrared-sensitive multipliers. The band-gap plus the electron affinity of cesium oxide has a long wavelength limit of about 0.4 or 0.5 μ. The silver incorporated in this cathode appears to play a decisive role as far as its infrared response is concerned. When examined in detail, the silver does not appear to behave simply as a donor-type dopant in the cesium oxide semiconductor. It has been proposed by W. E. Spicer that a possible explanation is that the photoemission is the result of photon excitation in the metallic silver and the emission of an electron from the silver into the cesium oxide and from thence over the electron affinity into the vacuum. Figure 3 shows an S-1 photocathode response.

There are broadly two ways of using a photomultiplier as a radiation detector. One is to chop the radiation with an appropriate shutter in the optical system (e.g., a rotating sectored disk) and observe the component of the output current having the chopping frequency. This can be done, for instance, with a phase-sensitive detector. The second method is to operate the multiplier with sufficient gain so that each electron produces a measurable pulse in the output and to count the rate of arrival of electron pulses. Actually, although pulse counting is the most sensitive way of detecting radiation with a photomultiplier, it is not practical with the type of multiplier illustrated when the cathode is maintained at room temperature. At room temperature, an S-1 cathode emits approximately 10^{-13} amp/cm^2 of thermionic emission. With the somewhat more than 10 cm^2 of cathode area of this multiplier, this means that there are some 2×10^7 electrons per second entering the structure. Such a high rate of pulse output saturates most counting equipment.

Cooling the cathode can greatly reduce the dark pulse rate. At 200° to 225°K (e.g., cooling with dry ice) the thermionic emission can be made negligible compared with other sources of dark pulses. Under these conditions photon counting becomes practical. If an integral pulse-height distribution curve, that is the number of pulses (n) that exceed a charge q during a given counting period, is plotted as a function of q, it will be found that over a fairly large range of q

$$n = n_0 e^{-\alpha q}.$$

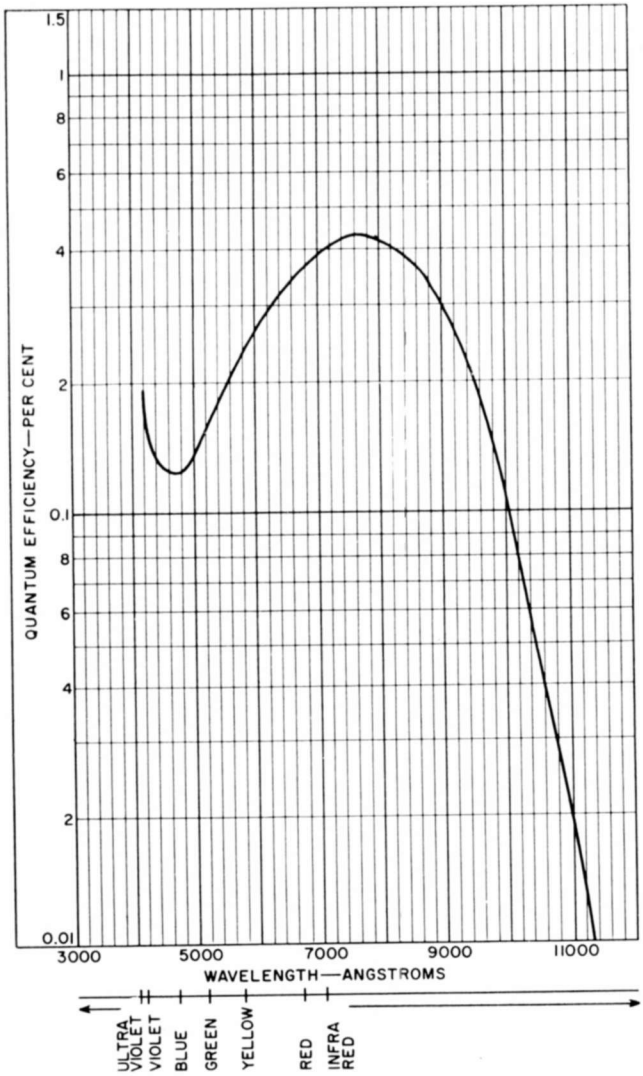


Fig. 3—Spectral response of S-1 photocathode.

This is illustrated in Figure 4 for dark pulses and a small amount of radiation on the cathode. At very small values of q, the pulse rate is very much higher than indicated. At large values of q, the dark pulse rate is larger than predicted. For optimum pulse counting, there should be a discriminator between the multiplier and the counter that transmits only pulses greater than about $0.4eG$ to eliminate the large

number of small spurious pulses. This will allow counting of 70 to 80% of photoelectrons from the cathode. The limit of sensitivity of this method of radiation detection is set by the statistical fluctuation in pulse rate. For a count of $\Delta n = n_T - n_D$, where n_T and n_D are the total and dark counts, respectively, in the period (e.g., 1 sec), a good approximation of the root-mean-square error is $2\sqrt{n_T + n_D}$. As an example, n_D might typically be 2000 counts per second and $2\sqrt{2n_D} \approx 130 \approx \Delta n$. If the quantum efficiency is 0.5% and the radiation has a

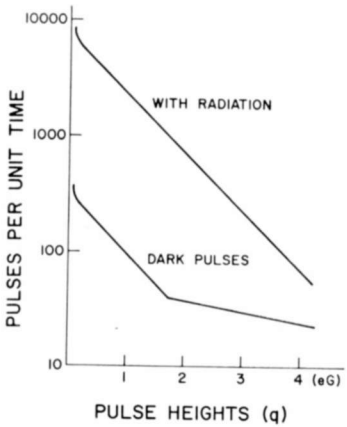


Fig. 4—Typical integral pulse height distribution curve for dark current and photoelectrons with a 7102 photomultiplier.

wavelength of 9000 Å, the equivalent of unity signal-to-noise ratio is obtained for about 3×10^{-17} watt.

The noise in dark current sets the limit of sensitivity of detection of radiation by the multiplier when current-measuring techniques are used. The r-m-s noise in the dark current is stated as equivalent to 1.7×10^{-12} watt at 8000 Å for this type of tube. At unity signal-to-noise ratio and with one-cycle-per-second bandwidth for the detector system, the lower limit of detectable radiation is therefore approximately 10^{-12} watt with the tubes uncooled. A second limit to the detection of a known radiating object results from the black-body radiation from the ambient. This limit applies to all detectors and is considered in more detail later.

Intermediate Infrared Detectors

In the 1 to 2 μ region of the spectrum, Ge junction diodes are the

most satisfactory detectors. Figure 5 illustrates the configuration of such a detector and also gives schematically the semiconductor band arrangement. They are prepared in much the same way as are diffused junction diodes used for nuclear radiation detection. A wafer of p-type Ge with a relatively small concentration of acceptor levels is mounted on a substrate and heated at a temperature somewhat below its melting point in an environment containing phosphorus. The phosphorus diffuses into the exposed surface. The diffusion is allowed to continue to a depth on the order of a micron. This forms an n-type layer on the p-type semiconductor. This junction is back biased. Radiation

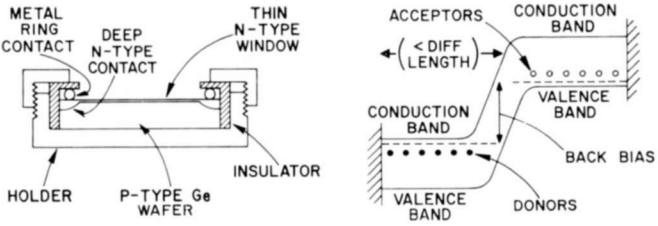


Fig. 5—Germanium photodiode.

entering the diode causes a hole–electron pair to be excited. If the excitation occurs in the n-type region, the minority carrier (hole) drifts to the junction and is drawn across the junction by the reverse-biased field. Similarly, absorption in the p-type region frees an electron which diffuses to the junction. Absorption of radiation either in the depletion layer of the junction or within a diffusion length on either side of the junction will free carriers and produce a signal current. Since the absorption coefficient for Ge, even quite close to its absorption edge, is of the order 10^4 cm^{-1}, it is relatively simple to have virtually all of the absorption occur in the useful portion of the detector.

One of the limits of sensitivity of this type of detector is the thermal excitation of carriers in the junction region. At room temperature, the noise resulting from the thermal excitation of carriers limits the sensitivity to a noise equivalent power of 10^{-12} watt per cycle bandwidth for a detector with an area of 5 mm^2. The noise varies exponentially with the reciprocal absolute temperature of the cell. Therefore, a small amount of cooling can eliminate this dark-current noise. Under these circumstances, the sensitivity of this cell is limited by black-body radiation noise from the object space into which the

detector looks and from such nonfundamental noise sources as leakage channels at the edges of the junction, contact noise, microplasma, breakdown, etc.

In the 2 to 3 μ region, PbS is by far the most widely used detector. This material is rather difficult to describe in terms of a semiconductor model. PbS has a band-gap of about 0.4 ev. However, as used, the material is in polycrystalline form and has been exposed to oxygen in its preparation, so that there are probably lanarkite barriers between the grains. The actual performance of the cell, particularly with respect to its variation with temperature, its noise characteristics, and its time constant, indicates that it is an array of barrier layer junctions. For example, its noise current output per unit frequency bandwidth varies inversely with the frequency. While giving quite good performance both at room temperature and at dry ice or liquid nitrogen temperature, these cells leave something to be desired both in their frequency response and their stability.

InAs, one of the III-V compounds having properties similar to Ge but with a considerably narrower band-gap (i.e., 0.33 ev) is now being developed as a detector for this portion of the spectrum. These detectors will undoubtedly be p-n junction detectors for room-temperature operation and will be quite similar to the Ge photodiode described above.

Going to somewhat longer wavelengths, InSb, PbTe, and PbSe are the detectors used in the neighborhood of 5 or 6 μ. InSb can be used as an ohmic intrinsic photoconductor or as a p-n junction photoconductor. As an ohmic photoconductor, even when operated at liquid nitrogen temperature, the resistance of the cell is rather low, which makes the electrical handling of the signal difficult. Junction-type InSb will operate both at room temperature, albeit with considerable thermal noise and consequently low sensitivity, and at liquid nitrogen temperature where its performance equals that of the ohmic photoconductor.

PbTe and PbSe both must be cooled to liquid nitrogen temperature for satisfactory operation. They are polycrystalline materials much like PbS and their frequency response and noise characteristics are very similar. They are, at present, easier to prepare than the InSb photodetectors, but their performance is definitely inferior.

Extrinsic Germanium Photoconductors

At present, photoconductor detectors for wavelengths longer than 7 μ do not use intrinsic semiconductors; rather, the long-wavelength response is due to the excitation of carriers from an impurity center to the conduction or valence band where they can move. Furthermore,

all of the detectors used at present are based on germanium. By selecting an impurity of the appropriate ionization energy, the detector can be made to have the spectral response required for the application at hand.

Before discussing the selection of specific activators to give desired spectral characteristics, let us consider briefly the mechanisms involved in photoconductivity. In general, the photoconductor takes the form of a small bar of semiconductor with ohmic contacts applied at each end. For satisfactory performance, the concentration of the activator atoms should not exceed approximately 10^{15} atoms/cm^3. Since

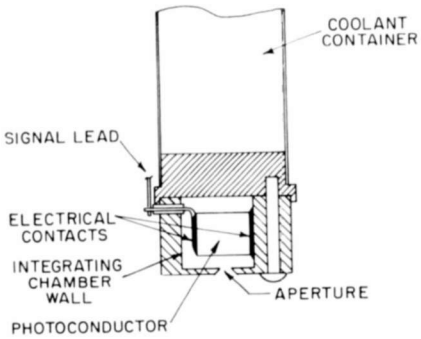


Fig. 6—Integrating chamber and photoconductor.

the photon absorption cross section of a typical impurity activator is of the order of 10^{-16} cm^2, the relatively low activator concentration means that the linear absorption coefficient of the material is between 1 and 0.1 cm^{-1}. It is undesirable from a number of standpoints to use a very large piece of photoconductor as the detector. Therefore, in order to absorb a reasonable fraction of the incident radiation, the extrinsic photoconductor element is generally mounted in an integrating chamber, as shown in Figure 6. The radiation enters through a small aperture in the wall of the chamber and, because of the high reflectivity of the integrating chamber walls, makes a number of traversals through the material until the radiation is absorbed. It is relatively easy to work out the relationships between the absorption properties of the semiconductor, its size, and the dimensions of the integrating chamber and aperture through which the radiation enters. It should be noted that here the effective area of the detector is the aperture in the integrating chamber wall and not the size of the photoconductive element.

Figure 7 illustrates schematically the energy-band scheme involved

in an extrinsic photoconductor. The activator impurity is shown as an acceptor, since all of the practical germanium detectors happen to employ p-type activator centers. The activator may be introduced into the germanium in the melt during the growth process or it may be diffused into a single-crystal bar of germanium. In order to obtain satisfactory performance, a small amount of compensator (n-type) impurity is also introduced into the semiconductor. The reason for this is two-fold.

First, if, for a given concentration of activator, there is no compensation, the Fermi level (E_f) lies halfway between the energy level

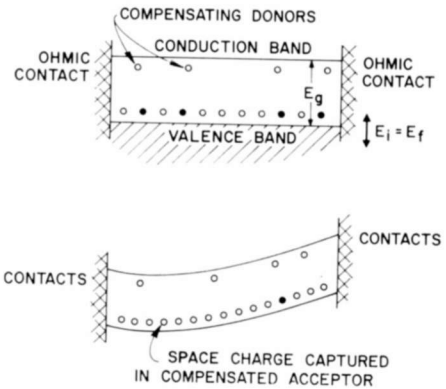


Fig. 7—Schematic energy-band diagram for an intrinsic photoconductor.

of the impurity (E_i) and the valence band. The variation of thermal excitation with cooling is an exponential function of E_f/kT. It is consequently proportional to $\exp(-E_i/2kT)$ since $E_f = E_i/2$. On the other hand, with the same impurity concentration, if there are a few centers compensated by electrons from the added donor impurity, the Fermi level will lie at the energy of the impurity, and the thermal excitation will be proportional to $\exp(-E_i/kT)$. In other words, the exponent is twice as large and, therefore, for a given thermal excitation, the required cooling is less.

The second and more important reason is electrical. To obtain greatest possible photoconductive gain, the element employs ohmic contacts. As the voltage is applied to the cell, carriers can be injected at one contact and leave at the other. A limit to the current flow is set by the charge distribution in the photoconductor. Where there are no compensated impurity centers, this space charge is the current flowing through the cell. However, carriers can be trapped in the

compensated centers giving a static space charge. Therefore, when an extrinsic element with some compensation is cooled to a very low temperature, the dark current is extremely low and consequently its resistance is essentially infinite. The action of these extrinsic photoconductors is therefore as follows: As soon as the voltage is applied to the element, charges enter from the contacts and are captured by the compensated center in sufficient number to form a space-charge block to further flow of current. Under these circumstances, there is a voltage gradient throughout the entire length of the cell. If radiation falls on the cell exciting a carrier into the valence band from one of the impurity centers, this carrier moves by virtue of the field and leaves the excited center from which it came with a residual charge on it. This upsets the potential distribution throughout the semiconductor and allows another carrier to enter from the positive contact. This carrier, in turn, moves through the photoconductor and, as it moves away from the electrode from which it comes, another carrier can enter. This continues until a carrier recombines on the site from which the original carrier was ejected. Thus, depending upon the lifetime of the carrier, the mobility, and the dimensions of the element, a single excitation may cause more than one carrier to be transported across the photoconductor.

Under these circumstances, the photoconductive gain can be greater than unity. Actually, with a very carefully prepared copper-activated germanium detector element, the mean free path of the carrier has been measured to be as large as 1.4 mm. A cell of this materiai with 0.5 mm between electrical contacts would have a photoconductive gain of almost 3. Obviously, the same reasoning applies to a thermally excited carrier, so that a photoconductor with high photoconductive gain will have a correspondingly high dark current. Furthermore, a cell with high photoconductive gain has a longer time constant or lower frequency response than one with lower photoconductive gain (neglecting trapping effects which, for the most part, appear to be unimportant to these germanium detectors).

Table I lists a number of impurities that have been investigated in germanium, together with their ionization energies. As has been pointed out earlier, one of the most important regions for detector operation is between 8.5 and 13.5μ. This would require an impurity level of about 0.09 ev.

The closest match to this is mercury with an ionization energy of 0.087 ev. This activator has been quite successful in long-wavelength detectors. Figure 8 shows the spectral response of mercury in germanium. Quite good photoconductive gain and low noise have been

Table I — Impurity Levels in Germanium

Activator	Acceptors			Donors	
Cu	0.26*	0.33†	0.04†		
Ag	0.09*	0.28*	0.13†		
Au	0.05*	0.20*	0.16†		0.05†
Zn		0.095†	0.035†		
Cd		0.16†	0.05†		
Hg		0.23†	0.087†		
Col. III			~0.01†		
Col. V					~0.01*
S					0.18*
Se				0.28*	0.14*
Te				0.30*	0.11*
Mn	0.37*	0.16†			
Fe	0.27*	0.35†			
Co	0.30*	0.25†			0.09†
Ni	0.30*	0.23†			

Values are in electron volts

* — measured from conduction band

† — measured from valence band

obtained with this material, but there are still some unanswered questions concerning the stability of mercury-doped germanium.

A second and rather different approach to obtaining material having exactly the required spectral response is that of employing an impurity atom in a Ge–Si alloy. Germanium and silicon will alloy in all proportions and the alloy can be grown into single crystals. As the concentration of silicon is increased, the band-gap increases monotonically. Similarly, the energy of an impurity in the band-gap increases with the silicon content. Photoconductors employing the second level of zinc

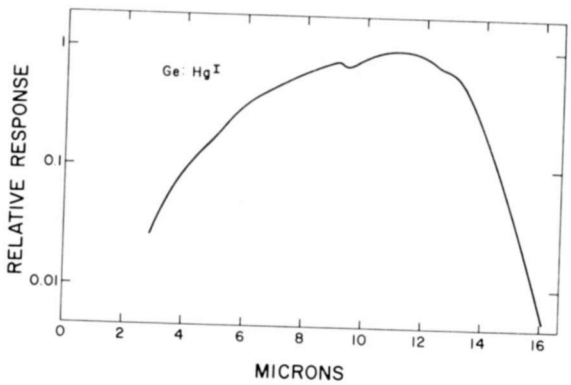


Fig. 8—Spectral response of a Ge:Hg^{I} photoconductor.

in germanium–silicon alloys have been investigated in some detail. It was found that an alloy containing about 6.1% of silicon in germanium and having the appropriate activator and compensator concentration could be made to have exactly the spectral response required to cover the radiation spectral band in question. One of the reasons why an exact long-wavelength limit should be sought is that the thermal generation–recombination noise is a very strong function of the long-wavelength limit. It was found possible to produce by alloy adjustment a photoconductor that, in a wide-angle-aperture detector cell, was background-noise limited at 50°K (a temperature that can be reached by pumping on liquid air) and that had a spectral response that closely fitted the atmospheric transmission window. Mercury-doped germanium required cooling to 35° or 40°K to accomplish the same job.

For response in the 20 to 30μ region, gold-doped and copper-doped germanium have given the best results. Copper has an ionization energy of about 0.04 ev, so that its long-wavelength limit is approximately 30μ. A detector with this activator must be cooled to about 10°K.

The use of gold as a long-wavelength activator is interesting. Gold has a donor level about 0.045 ev above the valence band. If a specimen of germanium is doped with a certain concentration of gold and then with a very shallow acceptor such as gallium, the electron from the donor level will drop into the gallium and leave the lowest gold level as a pseudo-acceptor. In principle, one would expect a pseudo-acceptor to have a much smaller cross section for recapturing a hole from the valence band than would a true acceptor such as copper. When a copper atom is ionized by an electron from the valence band being excited into a copper orbit, leaving a free hole, the copper level has a net charge of $-e$. This should produce a strong coulomb attraction for holes in the valence band. On the other hand, the gold pseudo-acceptor should be neutral after an electron has been excited into it from the valence band, and therefore would not have a coulomb attraction tending to recapture holes from the valence band. A difference in capture cross section of a factor of about 10^4 has been predicted from theoretical considerations, and appears to be at least partially verified by noise measurements. However, direct measurement of photoconductive gain comparing copper and gold fail to reveal this difference in lifetime. For a given activator concentration and compensation, the two activator atoms appear to have very nearly the same performance. Further work is required in this area.

For response around 40μ, zinc-activated germanium has proved quite effective. Detectors employing this photoconductor have been

used with considerable success in a number of scientific instruments.

Response out to 100 or 120μ can be obtained by using activators from either column III or column V of the periodic table. Arsenic, antimony, gallium, and indium have all been used as dopants.

Finally, photoconductive response has been obtained at extremely long wavelengths out even into the millimeter region by a somewhat different technique. The photoconductor used here was n-type indium antimonide cooled to a very low temperature. The impurity concentration was high enough to cause impurity banding and degeneracy; in other words, it was impossible to obtain carrier freeze-out irrespective of the amount of cooling. However, when the sample was immersed in a strong magnetic field, the degeneracy disappeared and the material behaved as a photoconductor. This work was done at the Royal Research Establishment in England by E. H. Putley. This appears to be a very interesting and effective way of covering the region of very long wavelengths where the infrared and radio wave spectra merge.

Detectivity

So far, sensitivity has been mentioned only in the most general terms. No attempt has been made to present a quantitative discussion of this aspect of photoconductors. However, the problem of noise and sensitivity must be analyzed in detail if an appraisal of the value of these devices is to be made.

Sensitivity is a measure of the limiting radiant power that these cells can detect. The threshold is determined by noise and random fluctuations in the output of the detector. Frequently, the limiting sensitivity of a detector is specified in terms of noise equivalent power (NEP). This is the radiant power in watts of infrared radiation on the cell that gives a signal-to-noise ratio of unity. The frequency bandwidth and the frequency at which the radiation is chopped must be specified. Also, the spectral content of the radiation must be given. The two most common specifications are for total radiation from a black body whose temperature is 500°K, or monochromatic radiation at the peak of the detector response.

In order to avoid the awkwardness resulting from the fact that the higher the sensitivity, the smaller the value of NEP, and also to give a figure that is independent of cell geometry, it is now customary to give the sensitivity in terms of detectivity D^*. Detectivity is defined by the following relation:

$$D^* = (\text{NEP})^{-1} A^{1/2} \Delta f^{1/2},$$

where A is the effective area of the detector and Δf is the frequency bandwidth.

For the specified NEP, the noise against which the signal is equated may be the result of a number of possible sources. The principal ones are (1) radiation (background) noise, (2) generation–recombination noise, (3) contact and surface noise, and (4) circuit noise.

Noise generated by the circuitry associated with a photoconductive detector is a very specialized subject. In the limit, this noise is due to the Johnson noise of the load resistor that couples the photoconductor to the input of the amplifier. Although this noise can be reduced by using cooled load resistors and very low-noise amplifiers, it cannot be eliminated. Obviously, the larger the signal that can be obtained from a photoconductor per unit radiation power incident on it, the lower the effect of the circuit noise. For this reason, considerable attention has been given to means of increasing the photoconductive gain by increasing the lifetime and mobility of these materials.

Contact and surface noise is not fundamental. Nevertheless, it can severely limit the performance of the photoconductive detector and therefore it must be minimized. This type of noise appears when the voltage between the contacts to the photoconductor becomes too large. It may be the result of minority carriers being injected by the contact or avalanche breakdown in the material or surface leakage over the surface of the photoconductor. The voltage at which this type of noise appears is quite sensitive to the way in which the contact has been made and to surface treatment (e.g., cleaning and etching). In general, this type of noise has frequency components that increase toward lower frequency and, for a given frequency band, is proportional to the reciprocal of its center frequency. For this reason, it is often termed $1/f$ noise.

Generation–recombination noise is fundamental. It is due to the thermal excitation of carriers. This excitation is random so that the current produced by carriers thus generated has superimposed on it a root-mean-square fluctuation that is proportional to the square root of the carriers excited. This noise current and the detectivity limit consequent from it can be determined as follows.

If n_A is the total acceptor concentration and n_D that of the compensating donors, then the net concentration of activators is $(n_A - n_D)$. The root-mean-square noise current $\{i_n\}_{rms}$ will be

$$\{i_n\}_{rms} = \frac{e\mu E}{L}\left(\frac{2N_c(n_A - n_D)}{n_D}\tau V\Delta f\right)^{1/2}\exp\left(-\frac{E_i}{2kT}\right),$$

But

$$E_i = \frac{h_c}{\lambda_m},$$

therefore

$$\{i_n\}_{rms} = \frac{e\mu E}{L}\left(\frac{2N_c(n_A - n_D)}{n_D}\tau V\Delta f\right)^{1/2}\exp\left(-\frac{h_c}{2k}\frac{1}{\lambda_m T}\right) \quad (1)$$

where E is the field in the photoconductor,
L is the distance between contacts,
τ is the lifetime,
V is the volume of the element,
Δf is the frequency band over which measurements are made,
N_c is the effective density of states in the valance band.

The other symbols have their conventional meaning.

If F is the radiant power incident on the detector and α_p the radiant cross section of each activator center, the signal current will be

$$i_s = \frac{e\tau\mu E}{L}(n_A - n_D)V\alpha_p F, \quad (2)$$

and the signal-to-noise ratio

$$\frac{S}{N} = \alpha_p F\left[\frac{n_D(n_A - n_D)}{2N_c}\frac{\tau V}{\Delta f}\right]^{1/2}\exp\left(\frac{h_c}{2k}\frac{1}{\lambda_m T}\right). \quad (3)$$

From this, it can be shown that

$$D^* = K_1\exp\left(K_2\frac{1}{\lambda_m T}\right) \quad (4)$$

where

$$K_1 = \left(\frac{\tau n_D\alpha_p}{2N_c}\right)^{1/2},$$

$$K_2 = \frac{hc}{2k},$$

where an integrating chamber is used. It should be noted that this noise varies exponentially with the ionization energy, E_i, of the impurities and with the reciprocal of the temperature, T, at which the cell is operated. The ionization energy E_i is, in turn, proportional to the reciprocal of the long-wavelength limit of the photoconductor. Therefore, the expression for noise can be put in terms of the temperature of the cell and its long-wavelength limit. Using this relationship and

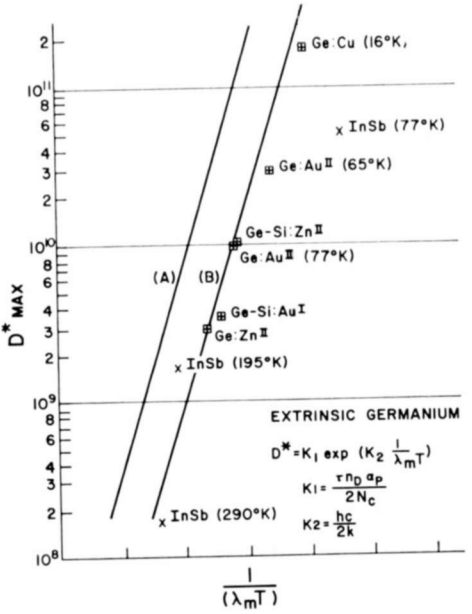


Fig. 9—Detectivity as a function of $1/(\lambda_m T)$.

the expression for the photocurrent in terms of the incident radiation, one can derive the general expression shown, with a coefficient containing semiconductor parameters and an exponential term involving the reciprocal of the long-wavelength limit and cell operating temperature. For a given class of photoconductors, the coefficient K_1 preceding the exponential term is relatively constant. It appears, for example, to apply to all of the impurity-activated germanium and germanium–silicon photoconductors that have been examined. Figure 9 gives a plot of this equation together with a number of experimental points for various impurity-activated germanium detectors that have been measured.

Background Radiation Noise

Finally, there is the noise generated by the carriers that are excited by background radiation. In practical application, a photoconductive cell is exposed to background radiation coming through the optical system that images the target on the cell. Under normal conditions, this radiation is essentially black-body radiation corresponding to a temperature of approximately 300°K. The amount of radiation inci-

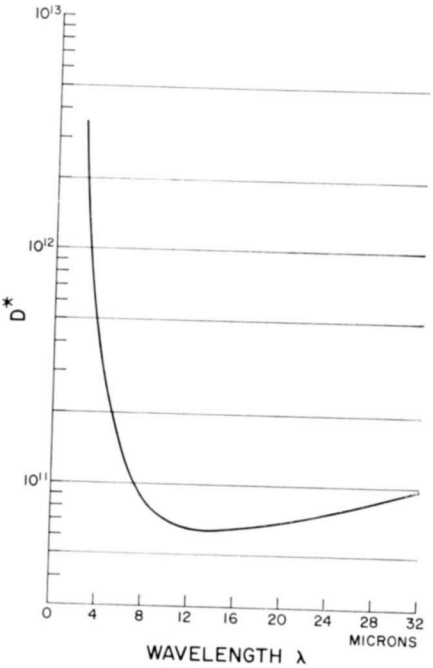


Fig. 10—Background-limited detectivity (for a 2π solid angle field of view).

dent on the cell depends upon the aperture angle of the optical system and the area of the cell. Like thermally excited carriers, the carriers excited by the background radiation have a statistical fluctuation that is proportional to the square root of the number of carriers excited. Therefore, we can calculate a limiting D^* for a detector knowing its long-wavelength limit and the spectral distribution of black-body radiation. The limiting $D^*(\pi)$ as a function of wavelength for a detector (having unity quantum efficiency) and a 2π solid angle of view is shown in Figure 10. If the angular aperture of the system, θ, is less than 180°, then

$$D^*(\theta) = D^*(\pi)\,(\sin\theta/2)^{-1}. \tag{5}$$

It is necessary in designing an infrared detector system to use the specified performance characteristics of the particular type of cell to be used. However, Equations (4) and (5) and the curves given in Figures 9 and 10 are very useful in making preliminary estimates of performance. For example, let us assume that a mercury-doped germanium detector with a long-wavelength cutoff (λ_m) of 15 μ, an effective area of 0.1 cm^2, and an angular field of $\theta = 58°$ is to be used in an optical system employing a parabolic mirror whose area is 300 cm^2. The object being viewed has an area of 1 square meter and is at a distance of $d = 10$ kilometers. If the atmospheric transmission γ is 25%, the detector system bandwidth 1 cycle/sec, and a signal-to-noise ratio of 5 is required, what is the minimum permissible temperature ΔT between the object and its environment?

From the curve in Figure 10 and Equation (5), the D* limit set by radiation background is 1.5×10^{11} cm watts^{-1}. With the aid of Figure 9 and taking into account the long-wave limit of 15 μ, it is found that the detector must be cooled below 36°K. From the cell area and frequency bandwidth, the detector has an NEP of 2.2×10^{-12} watt; therefore, the radiant flux density at the mirror must be approximately $J = 3.7 \times 10^{-14}$ watt/cm^2 for $S/N = 5$. The black-body radiation equations indicate the change in radiant power ΔW per degree temperature in the vicinity of room temperature is 5 watts/degree for a 1-square-meter object. Finally, the inverse square law relates radiation from the object and flux density at the mirror;

$$J = \frac{\gamma \Delta W}{\pi d^2} = \frac{0.25 \times 5 \times \Delta T}{\pi (10^6)^2}.$$

Therefore, $\Delta T = 0.1$°K.

From the foregoing discussion of extrinsic detectors, it is evident that a high degree of control of spectral response has been achieved. Improvements in this direction will be minor. It is expected that, as the field develops, there will be marked improvements in detectors as the result of reduction of nonfundamental noise due to contacts, channeling, etc. There may also be some further reduction in generation–recombination noise, but there is reason to believe that the improvement that can be made in this direction is rather limited. It is probable that the greatest advance will be made in improved quantum yield or responsivity. Initially, these improvements will be bounded by the gain–bandwidth limitation set by the relaxation time of the material, but eventually, as more complex detector structures are developed, this limit also may be exceeded.

MICROWAVE PHOTOMULTIPLIERS USING TRANSMISSION DYNODES*

By

D. Blattner, H. Johnson, J. Ruedy, and F. Sterzer

RCA Electronic Components and Devices,
Princeton, N. J.

Summary—*This paper describes a new type of photomultiplier tube that extends the low-noise gain of secondary-emission multiplication into the microwave range. The tube consists of a semitransparent photocathode sensitive to the required spectral region, a multistage structure of transmission secondary-emission dynodes, and a helical output coupler through which the electrons from the last dynode are focused.*

The experimental tubes described are designed for a center frequency of 1.5 gc and have a bandwidth of 1.0 gc. For visible light, the sensitivity of these tubes approaches that of an ideal detector. Calculations indicate that scaling of the multiplier design to frequencies as high as K_a band is possible.

Introduction

THE FUNDAMENTAL limitation on the frequency response of photomultipliers is the transit-time dispersion of the electrons as they traverse the multiplier structure. When the transit-time dispersion becomes a significant fraction of the period of the signal on the electron beam, the depth of modulation on the beam is reduced and the power output decreases. The two major causes of transit-time spread are (1) inequality of electron velocities, caused primarily by the variation of initial velocities with which the electrons are emitted and (2) inequality of path lengths of the electrons. (Transit-time spread due to time spread in the emission is in general negligible.[1,2])

Transit-time spread is difficult to minimize in multipliers that use conventional reflection-type dynodes, i.e., dynodes in which the secondary electrons are emitted from the same surface that is bombarded by the primary electrons. For minimum spread due to initial velocities, the voltage between dynodes should be as high as possible. In reflec-

* Sponsored by U.S. Army Electronics Laboratories, Fort Monmouth, New Jersey.

[1] M. H. Greenblatt, "On the Measurement of the Average Time Delay in Secondary Emission," *RCA Review*, Vol. 16, p. 52, March 1955.

[2] D. J. Blattner, H. C. Johnson, and F. Sterzer, "Upper Limit of Time Dispersion in Transmission Secondary Electron Emission from KCl Films," *Appl. Phys. Letters*, Vol. 4, p. 46, Feb. 1964.

tion-type dynodes, however, the current gain is a maximum when the primary electrons strike the emitting surface with velocities corresponding to only a few hundred volts; hence, minimum transit-time spread and high current gain are incompatible. Also, because the electron paths in reflection-type multipliers are generally curved, maintenance of equal path lengths for all electrons is inherently difficult.

As a result of the relatively large transit-time spread in multipliers using reflection-type dynodes, the frequency response of the fastest available electrostatically focused reflection-type multiplier is limited[3] to about 1 gc. This limit has been extended to about 6 gc in an experimental crossed-field reflection-type multiplier built by Miller and Wittwer.[4]

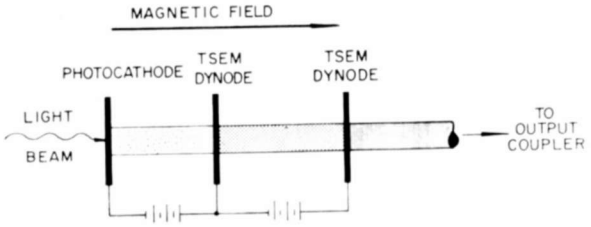


Fig. 1—Schematic diagram of two-dynode electron multiplier using transmission secondary-emission dynodes.

In this paper we describe microwave photomultipliers using *transmission* secondary-electron multiplication (TSEM) dynodes. In these dynodes, primary electrons are incident on one side of a film, and secondary electrons are emitted from the opposite side of the film. The maximum current gain of most TSEM dynodes occurs typically at several thousand volts, i.e., at voltages about an order of magnitude higher than those used for reflection-type dynodes. As a result, transit-time spread due to the spread in electron velocities can be made much smaller in TSEM multipliers than in the fastest reflection-type multipliers. Also, TSEM dynodes can be stacked in a parallel-plane geometry, as shown in Figure 1, so that all electrons will travel in nearly parallel straight lines of equal length. Measurements indicate that multipliers using TSEM films will be usable at frequencies up to 30 gc

[3] G. A. Morton, R. M. Matheson and M. H. Greenblatt, "Design of Photomultipliers for the Sub-Millimicrosecond Region," *IRE Trans. on Nuclear Science*, Vol. NS-5, p. 98, Dec. 1958.

[4] N. C. Wittwer, "Fast-Rise Time (Less Than 10^{-10} Secs.) Photomultiplier," paper presented at the twenty-second Conference on Electron Device Research, Cornell University, Ithaca, New York, June 1964.

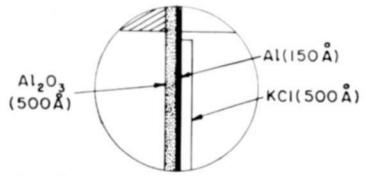


Fig. 2—Cross section of TSEM dynode.

— considerably higher than those achieved by the fastest multipliers using reflection-type multiplication.

TSEM Dynodes

The dynodes used in our experimental microwave photomultipliers consist of a sandwich of three films,[5] as shown in Figure 2—a supporting substrate of aluminum oxide (Al_2O_3) 500 Å thick, a 150 Å thick conducting film of aluminum, and a 500 Å thick film of the emitting material, potassium chloride (KCl). Dynodes in which the KCl film was replaced by magnesium oxide (MgO) were also tested.

The current gain of a typical KCl dynode is shown as a function of dynode voltage in Figure 3. A gain of 3.8 is obtained for a voltage of 4 kv, and a gain of 4.6 for a voltage of 6 kv.

The distribution of initial energies of secondary electrons from KCl dynodes varies from about 0 to 5 volts, and can be approximated by the following distribution function:

$$p(V_0) = 0.4\left(1 - \frac{V_0}{5}\right), \tag{1}$$

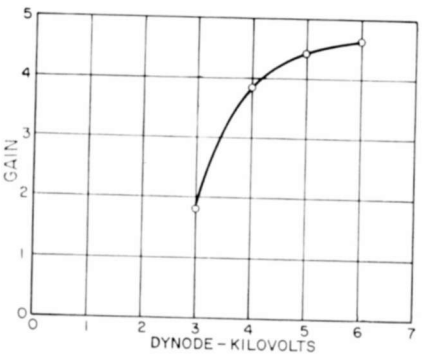


Fig. 3—Current gain of KCl transmission secondary-emission dynode as a function of dynode voltage.

[5] E. J. Sternglass, "High-Speed Electron Multiplication by Transmission Secondary Electron Emission," *Rev. Sci. Instr.*, Vol. 26, p. 1202, Dec. 1955.

where p is the fraction of electrons emitted per unit time with initial energies between V_0 and $(V_0 + \Delta V_0)$. The electron beam leaving the dynode also contains a small proportion of high-velocity electrons, most of which are probably primary electrons that are transmitted by the film. The ratio of secondary electrons to transmitted primary electrons (arbitrarily defined as electrons having energies greater than 50 volts) is about 30 at 3 kv, 17 at 4 kv, and 10 at 5 kv.[6]

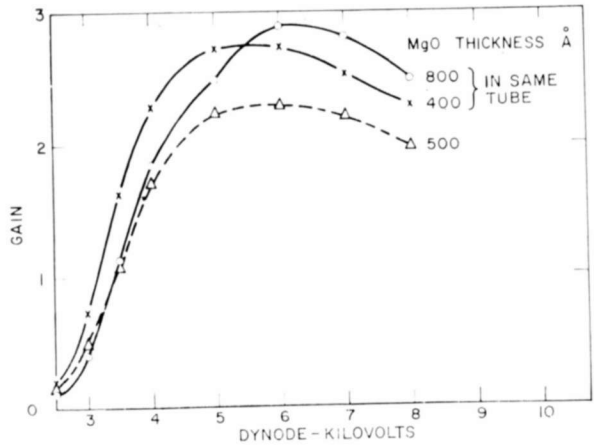


Fig. 4—Current gain of MgO transmission secondary-emission dynodes as a function of dynode voltage.

The current gain of dynodes using three different thicknesses of MgO film is shown in Figure 4. The maximum current gain of 2.8 for these dynodes is significantly lower than that attainable with KCl dynodes.

The results of life tests on both KCl and MgO dynodes are summarized in Table I. The life of KCl dynodes is seen to be a function of the total charge density incident upon the dynode; gain decreases by approximately 10% for every increment of 0.1 coulomb per square centimeter. The MgO dynode, on the other hand, showed no decrease in gain after incidence of a total charge density of 116 coulombs per square centimeter.

Output Couplers

The output coupler used in conventional reflection-type multipliers

[6] M. M. Wachtel, D. D. Doughty, and A. E. Anderson, *Advances in Electronics and Electron Physics*, Vol. 12, p. 64, Academic Press, New York, 1960.

is simply a load resistor in series with the electrode (anode) that collects the current emitted from the last dynode. The power delivered to such a load resistor is

$$P_R = \frac{m^2 I^2}{2} R \tag{2}$$

where m is the modulation index of the average current, I, flowing through the load resistor, R. The maximum value of load resistance that can be used is usually limited by bandwidth considerations; for bandwidths exceeding a few hundred megacycles, R is limited to, at most, a few hundred ohms.[7]

Table I—Life-Test Data for KCl and MgO Dynodes

Type of Dynode	Dynode Bombardment (amp/cm²)	Dynode Voltage (kv)	Time (hours)	Total Charge Density (coulomb/cm²)	Decrease in Gain (%)
KCl	3.2×10^{-6}	4	7	0.01	none
	0.4×10^{-6}	4	97	0.08	10
	0.4×10^{-6}	4	188	0.14	20
	40×10^{-6}	4	7.25	0.27	30
MgO	400×10^{-6}	5	80	116	none

In the TSEM photomultiplier, the current from the last dynode is focused through a traveling-wave-tube-type helix, and is then collected at a separate electrode (collector). The r-f output is taken directly from the helix rather than from a load resistor connected to the collector electrode. For the small currents used in microwave photomultipliers, the power output of the helical coupler is proportional to the square of beam current;[8] therefore, it is possible to define an effective output resistance $R_{eff} = P_{out}/(m^2 I^2/2)$. Helical couplers combine gigacycle bandwidth with effective output resistances of the order

[7] N. C. Wittwer, "Detection of Higher-Order Ruby Optical Maser Modes," *Appl. Phys. Letters*, Vol. 2, p. 194, 15 May 1963.

[8] B. J. McMurtry, "Microwave Phototube Design Considerations," *IEEE Trans. on Electron Devices*, Vol. ED-10, p. 219, July 1963.

of a megohm.[8,9] For a given current from the last dynode, therefore, the power output of a helical coupler is several orders of magnitude greater than that from a broad-band resistance coupler. Unlike resistance couplers, helical couplers are band-pass devices; however, they can be used for demodulating frequencies outside their pass band if they are operated in the r-f mixing mode, as discussed later. Helical output couplers are better suited for use in TSEM multipliers than resistive couplers because, as mentioned previously, the current drawn from high-gain TSEM dynodes must be kept small to ensure long life of the dynodes.

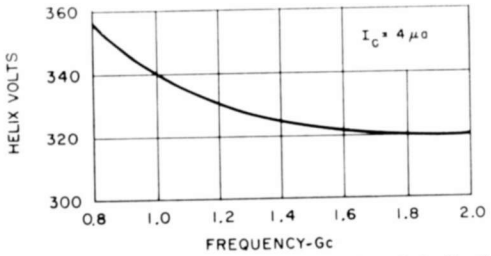


Fig. 5—Synchronous-velocity voltage of L-band helical output coupler as a function of frequency.

The helices in our photomultipliers are wound from 0.010-inch-diameter wire, are 12 inches long, have a pitch of 0.0192 inch and a mean diameter of 0.163 inch, and are supported in fluted bulbs of 7052 glass. The measured dielectric loading factor of the glass-supported helix varies from 0.96 at 1 gc to 0.92 at 2 gc. The synchronous-velocity helix voltage, shown as a function of frequency in Figure 5, varies from 340 volts at 1 gc to 320 volts at 2 gc. For a ratio of beam diameter to helix diameter of 0.8, the calculated beam-circuit interaction impedance of the helix ranges from 360 ohms at 1 gc to 60 ohms at 2 gc.[10,11]

Figure 6 shows the effective output resistance of the microwave phototube over the one-gigacycle pass band. The calculated values in this figure are based on McMurtry's Equation (15) corrected for helix loss.[8] The measured values were obtained by means of shot-noise measurements. The curves show that R_{eff} is of the order of one megohm.

---

[9] D. J. Blattner et al, "LASECONS: Microwave Phototubes with Transmission Photocathodes," *1963 IEEE Int. Conv. Record*, Pt. 3, p. 79, March 1963.

[10] J. R. Pierce, *Traveling-Wave Tubes*, D. Van Nostrand, New York, 1950.

[11] P. K. Tien, "Traveling-Wave-Tube Helix Impedance," *Proc. IRE*, Vol. 41, p. 1617, Nov. 1953.

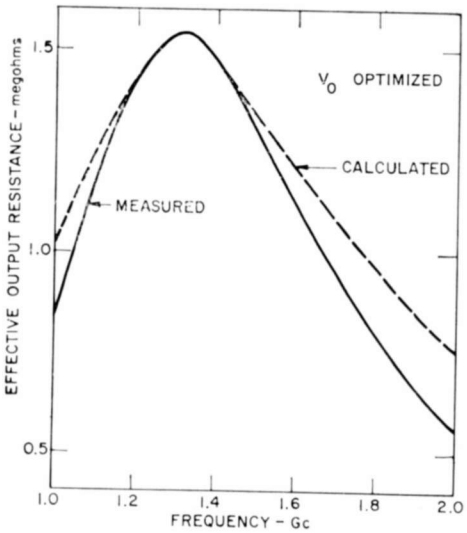


Fig. 6—Effective output resistance of L-band helical output coupler as a function of frequency.

Description of TSEM Microwave Phototube

Figure 7 is a schematic diagram of the complete TSEM microwave phototube. (This tube is also called a LASECON, a name derived from LASEr CONverter.) Incident light enters the tube axially through a large window and strikes a transmission-type photocathode. Photoelectrons emitted by the photocathode, which are bunched at the modulation frequency of the light, are amplified by one or more TSEM dynodes, and are then focused by means of an axial magnetic field through the coupling helix. This bunched beam of electrons excites an r-f wave on the helix. The wave is coupled through the glass

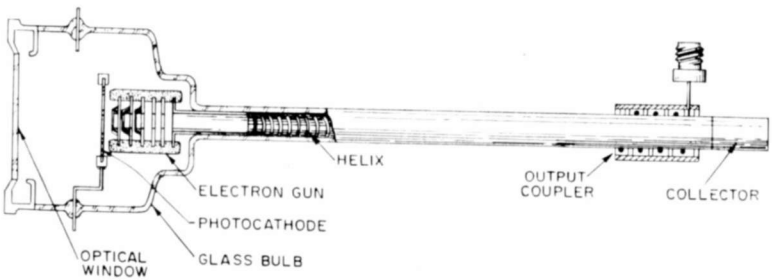


Fig. 7—Schematic diagram of TSEM microwave phototube.

envelope by a short length of helix connected to a coaxial connector. At present, focusing is accomplished in a solenoid.

The transmission-type photocathode used in these microwave phototubes can be made to provide any spectral response available in other phototubes. Present tubes have been made to provide S-1 and S-20 responses. The light flux required to produce a given photocurrent from the two different types of cathodes can be determined from photoresponse curves. The S-20 cathode is better for wavelengths below about 8000 Å; the S-1 cathode is superior for infrared radiation with wavelengths greater than 8000 Å.

Signal-to-Noise Ratio

The power signal-to-noise ratio at the output of a helix-type microwave photomultiplier can be written as follows:

$$\frac{S}{N} = \frac{P_s}{P_{sh} + P_{vel} + P_{part} + P_{th}}, \tag{3}$$

where

P_s = signal power,
P_{sh} = shot-noise power,
P_{vel} = velocity-fluctuation noise power,
P_{part} = partition-noise power,
P_{th} = thermal-noise power.

The signal power P_s is given by

$$P_s = \frac{1}{2} m^2 I_c^2 R_{eff} = \frac{1}{2} m^2 (\alpha_1 \alpha_2 \cdots \alpha_n)^2 I_0^2 R_{eff} \tag{4}$$

where

m = modulation index of the current entering the helix,
I_c = average collector current,
R_{eff} = effective output resistance of the helix (shown in Figure 6, for example). For large velocity distribution in the electron beam, R_{eff} is reduced as discussed below.
α_i = current gain of the i^{th} dynode,
I_0 = average cathode current,
n = number of dynodes.

Equation (4) assumes negligible electron-beam interception by the gun electrodes and the helix.

The shot-noise power P_{sh} is given by[12]

$$P_{sh} = 2eI_0BR_{eff}\,[\,(\alpha_1\,\alpha_2\cdots\alpha_n) + (\alpha_1\,\alpha_2\cdots\alpha_n{}^2) + (\alpha_1\,\alpha_2\cdots\alpha_{n-1}{}^2\,\alpha_n{}^2) + \cdots + (\alpha_1{}^2\,\alpha_2{}^2\cdots\alpha_n{}^2)\,] \tag{5}$$

where B is the coupler bandwidth and e is the electronic charge. Equation (5) was verified with a three-dynode microwave phototube operating at one gigacycle. The measured value of shot noise agreed with the value calculated from Equation (5) to well within measurement accuracy of ± 1 db.

The velocity-fluctuation noise power is negligible compared to the shot-noise power;[8] for a well-focused beam, the partition noise is generally also negligible.

Finally, the available thermal-noise power output P_{th} from the microwave phototube is given by

$$P_{th} = kTB, \tag{6}$$

where k is Boltzmann's constant and T is the absolute temperature of the output coupler.

When Equations (3), (4), (5), and (6) are combined, the signal-to-noise ratio can be written

$$\frac{S}{N} = \frac{\frac{1}{2}m^2\,(\alpha_1\cdots\alpha_n)^2 I_0{}^2R_{eff}}{2eI_0BR_{eff}\,[\,(\alpha_1\cdots\alpha_n) + (\alpha_1\cdots\alpha_n{}^2) + \cdots + (\alpha_1{}^2\cdots\alpha_n{}^2)\,] + kTB}. \tag{7}$$

If each dynode has the same gain, Equation (7) simplifies to

$$\frac{S}{N} = \frac{\frac{1}{2}m^2\,\alpha^{2n}\,I_0{}^2R_{eff}}{2eI_0BR_{eff}\,\frac{\alpha^n\,(\alpha^{n+1}-1)}{\alpha-1} + kTB}. \tag{8}$$

[12] K. R. Spangenberg, *Vacuum Tubes*, Ch. 12, McGraw-Hill, New York, 1948.

In tubes with sufficient current gain, the shot-noise power is very much larger than the thermal power; in this case the signal-to-noise ratio is given by

$$\frac{S}{N} = \frac{m^2 I_0}{4eB} \frac{\alpha - 1}{\alpha}. \tag{9}$$

If the dark current from the photocathode is negligible, as is often the case, the average cathode current I_0 is given by

$$I_0 = n_0 \eta e, \tag{10}$$

where n_0 is the average number of incident photons per second and η is the quantum efficiency of the photocathode. The shot-noise-limited signal-to-noise ratio from the photocathode can then be written

$$\frac{S}{N} = \frac{m^2 \eta n_0}{4B} \frac{\alpha - 1}{\alpha}. \tag{11}$$

This expression gives the signal-to-noise ratio for an ideal photodetector when $\eta\ (\alpha - 1)/\alpha$ equals unity; i.e., for a quantum efficiency of unity and infinite dynode current gain. In practice the maximum value of η is only about 0.2 (for an average S-20 cathode at a wavelength of about 4,000 Å). For KCl TSEM dynodes, α is typically 4, so under shot-noise-limited conditions and for $\eta = 0.2$, the signal-to-noise ratio of a TSEM tube is 8.2 db smaller than that of an ideal detector.

The number of photons per second required to obtain a given signal-to-noise ratio can be calculated from Equations (8) and (10). Figure 8 shows the photon flux required to obtain a signal-to-noise ratio of 10 db as a function of bandwidth for various numbers of dynodes. Figure 9 shows the signal-to-noise ratio as a function of photon flux for various numbers of dynodes at a fixed bandwidth of 100 mc.

The curves of Figures 8 and 9 show that microwave phototubes using TSEM dynodes require significantly fewer photons per second for a given signal-to-noise ratio than tubes without current multiplication. At narrow bandwidths (of the order of one megacycle) large numbers of dynodes are required to approach the shot-noise-limited signal-to-noise ratio; at bandwidths approaching one gigacycle, two or three dynodes are sufficient.

For shot-noise-limited operation (defined as $P_{sh} > 10P_{th}$), the

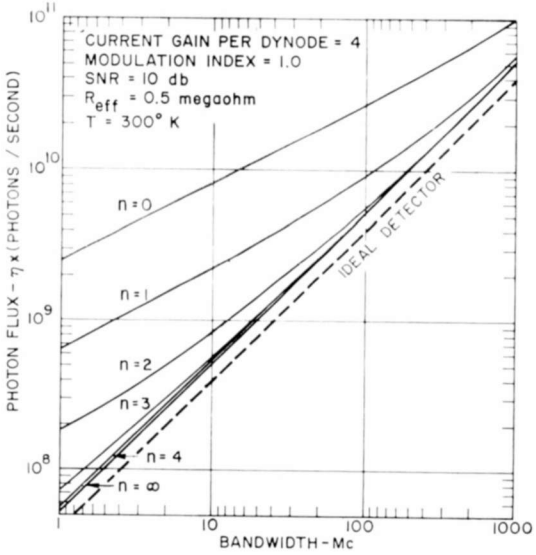


Fig. 8—Photon flux required for a signal-to-noise ratio of 10 db as a function of bandwidth.

minimum average beam current can be readily calculated from Equations (5) and (6). Table II gives the results of such calculations for tubes having one to four dynodes. The table also shows the total charge transported by the beams for every thousand hours of operation.

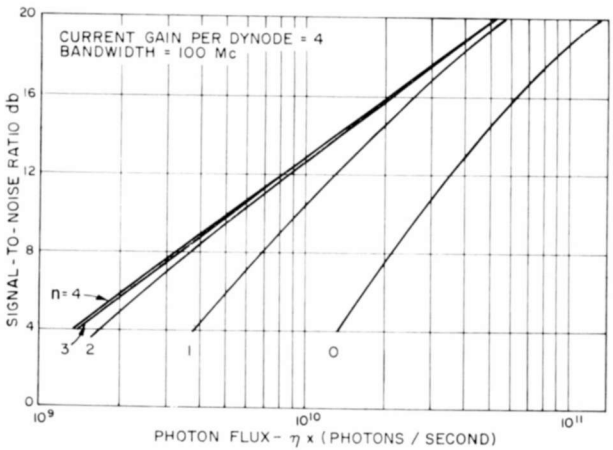


Fig. 9—Signal-to-noise ratio as a function of photon flux for various numbers of dynodes.

Table II is useful in estimating the life of KCl dynodes when the phototube is operated with just enough beam current to provide shot-noise-limited operation. In a tube with 4 dynodes, for example, the beam carries a charge of 1.4×10^{-3} coulomb for every thousand hours of operation. If the tube uses an L-band helical coupler, the cross-sectional area of the beam is approximately 0.12 cm^2; therefore the charge density per thousand hours is 1.17×10^{-2} coulomb/cm^2, and the charge density incident upon the last dynode is approximately 2.9×10^{-3} coulomb/cm^2 per thousand hours. Because a KCl dynode

Table II—Beam Current and Electronic Charge per Thousand Hours for $P_{sh} = 10\ P_{th}$ ($T = 300°K$, $\alpha = 4$, $R_{eff} = 1$ megohm)

Number of Dynodes	Average Beam Current (amperes)	Electronic charge per 1000 hours (coulombs/1,000 hours)
1	2.6×10^{-8}	9.4×10^{-2}
2	6.2×10^{-9}	2.2×10^{-2}
3	1.6×10^{-9}	5.8×10^{-3}
4	3.9×10^{-10}	1.4×10^{-3}

can handle about 8×10^{-2} coulomb/cm^2 with only a 10% decrease in gain (see Table I), more than 10,000 hours of operation can be expected.

For tubes using only one or two dynodes, or for tubes with helices designed for operation in the upper range of the microwave region, the life expectancy of KCl dynodes may be too short for many applications. In this case MgO dynodes should be considered for the last dynode stages; MgO dynodes can handle orders of magnitude more charge than KCl dynodes (as shown in Table I), but have the disadvantage of lower gain.

Transit-Time Dispersion

Dispersion between the Last Dynode and the End of the Helix

Because the electrons entering the helix have a spread of velocities, the electron beam debunches as it traverses the helix. The debunching effects in helix-type phototubes without TSEM are so small that they do not cause any significant deterioration of power output.[8] However, the width of the velocity distribution of electrons emitted from the TSEM dynodes is approximately an order of magnitude greater than

the width of the velocity distribution of photoelectrons, and the effect of this large velocity spread, though small, is probably not entirely negligible.

The transit time of an electron traveling from the last dynode to the end of the helix, t_h, can be written

$$t_h = t_0 - \frac{1}{\sqrt{2\frac{e}{m}}}\left[2d\frac{\sqrt{V_0}}{V} + \frac{l}{2}\frac{V_0}{V^{3/2}}\right], \qquad (12)$$

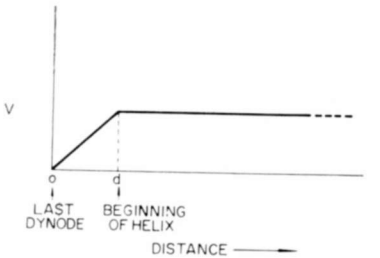


Fig. 10—Voltage distribution in the dynode-helix region assumed in derivation of Equation (12).

where

t_0 = transit time of an electron with zero initial velocity,

m = mass of electron,

V = helix voltage,

V_0 = emission voltage of electron,

l = helix length,

d = spacing between final dynode and the beginning of the helix.

Equation (12) was derived for the voltage distribution shown in Figure 10, with the assumption that the r-f voltages on the helix have negligible effect on the motion of the electrons.

The reduction in depth of modulation caused by any given velocity distribution can be calculated from Equation (12), as described in the Appendix. For tubes having a spacing, d, of 0.125 inch and a velocity distribution represented by Equation (1), the calculated reduction in depth of 1.5 gc modulation as a function of helix length

is shown in Figure 11. For a helix length of 12 inches, the modulation is reduced by 20 percent.

Using coupled-mode theory, McMurtry[8] has derived an approximate expression for the reduction of power output caused by a rectangular distribution of initial electron velocities. According to McMurtry, the fractional decrease in power output, $\Delta P/P$, is approximately given by

$$\frac{\Delta P}{P} = \left(\frac{N\Delta V}{2V}\right)^2, \tag{13}$$

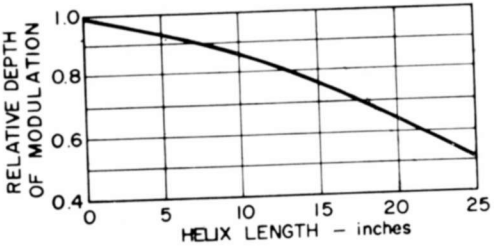


Fig. 11—Calculated reduction in depth of 1.5 gigacycle modulation of electron beam caused by velocity dispersion as a function of distance along the helix.

where

N = length of helix in slow-wave wavelengths,

ΔV = width of rectangular distribution of energies of emitted electrons,

V = helix voltage.

For the tubes used in our experiments ($V = 350$ volts, $N = 40$, $f = 1500$ mc), a power reduction of approximately 8% for $\Delta V = 5$ volts is calculated from Equation (13). This value is qualitatively consistent with the 20% reduction in modulation depth calculated from Equation (12) and the velocity distribution of Equation (1).

Figure 12 shows an experimental arrangement used for determination of the power reduction caused by the velocity spread of secondary electrons from TSEM dynodes. In this arrangement, the power output of a tube using a single TSEM dynode was compared with that of a tube using no dynode. The two tubes were made as similar as possible except for the inclusion of the TSEM dynode in the gun of one tube. The photocathodes of the tubes were illuminated by a He-Ne

gas laser operating at 6328 Å to produce the bunched electron beam. The laser was adjusted to operate in purely longitudinal modes. The mode spacing of the laser used was approximately 94 mc, and the tube voltages were optimized for maximum power output at the tenth difference frequency (about 935 mc). The output of the tube was amplified in a traveling-wave tube, heterodyned down to 30 mc, amplified in an intermediate-frequency amplifier, fed into a crystal detector,

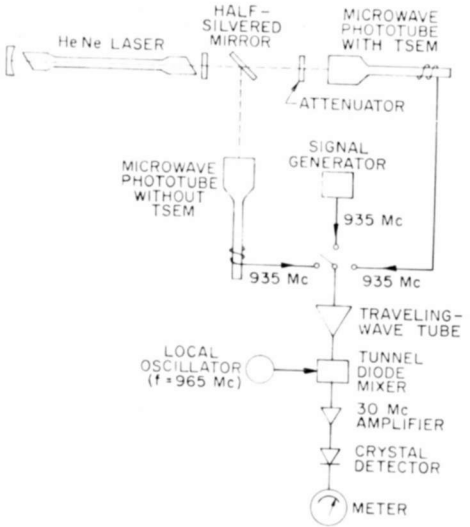


Fig. 12—Experimental arrangement used for measuring power reduction caused by velocity spread of secondary electrons from TSEM dynode.

and displayed on an output meter. All measurements were made under the following conditions:

(1) same collector current in both tubes (the light incident on the TSEM tube was attenuated with a neutral density filter to compensate for the TSEM gain),

(2) current interception on the helix less than 15%,

(3) constant laser excitation,

(4) same distances of the external output coupler from the electron gun in both tubes.

Figure 13 shows the power output at 935 mc as a function of the distance of the external output coupler from the gun for both tubes. For a very short length of helix, the power output of the TSEM tube is about 1 db higher than that of the tube without a dynode. This

disparity probably results from small differences in the construction, alignment, and focusing of the two tubes. For greater helix lengths the difference in power output between the two tubes becomes slightly less, as predicted by Equations (12) and (13). The scatter in the data is too large to support a definite conclusion that any deterioration of

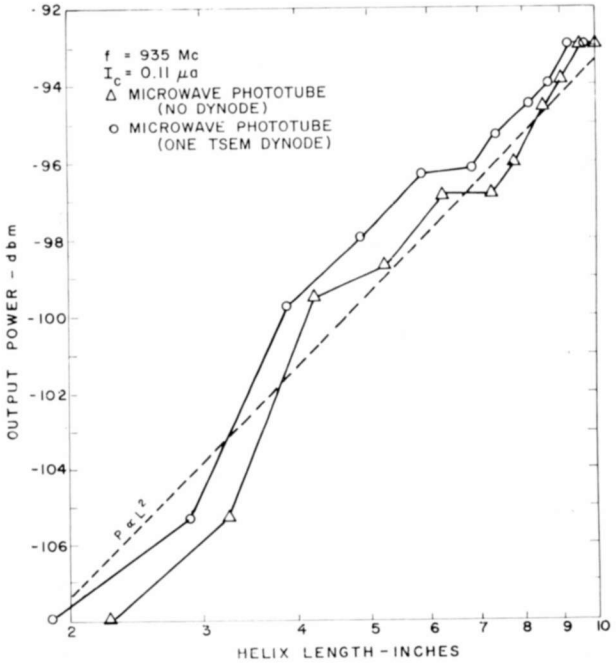


Fig. 13—Power output as a function of helix length for microwave phototubes with and without TSEM dynode.

power output is caused by velocity dispersion in the TSEM tube. However, if such deterioration does occur, the loss does not exceed one or two db. This amount of deterioration is negligible compared to the gain in power resulting from the multiplication of the beam current.

Determination of transit time dispersion was also made by comparing the shot-noise outputs of two microwave phototubes, one with a single dynode and one with no dynode. The spacing between the dynode and helix was unusually large (0.220 inch), so the transit-time dispersion in this region was accentuated. The measured frequency response of the TSEM tube was down by 3 db at 2100 mc, while calculations based on the Appendix and Equation (1) predict a drop in response of 4.3 db at 2100 mc. Thus it appears that the actual velocity

distribution of the secondary electrons is somewhat narrower than that described by Equation (1).

The time dispersion in the helical coupler is small regardless of the center frequency of the tube, because higher frequency tubes use shorter helices. Equation (12) indicates that the time dispersion in the helix is proportional to its length. However, the impedance of a helical coupler, neglecting losses, is proportional to its length in wavelengths. For example, an S-band (2-to-4 gc) coupler is only about half as long as an L-band (1-to-2 gc) coupler of the same impedance, and the ratio of the r-f period to the time dispersion is the same for both couplers.

Dispersion between the Photocathode and the Last Dynode

The transit time t_d of an electron traveling between two dynodes or between the photocathode and the first dynode can be written

$$t_d = s\sqrt{\frac{2m}{Ve}} - \frac{s}{V}\sqrt{\frac{2V_0 m}{e}}, \tag{14}$$

where s is the distance between the two electrodes, and V is the voltage between them, and where it is assumed that the electron travels in a straight line normal to the two electrodes.

The reduction in the depth of modulation as the electron beam traverses the multiplier structure can be calculated from Equation (14) with the aid of the Appendix and the velocity distribution of Equation (1). (The transit-time effects in the cathode–first-dynode region are generally negligible compared to the transit-time effects in the region between dynodes because of the large difference in initial velocity distribution between photoelectrons and secondary electrons.) For L-band tubes, these calculations predict negligible reduction in modulation.

The calculations were verified experimentally by use of a set-up similar to that shown in Figure 12. Instead of the tube having a single dynode, however, a tube with three dynodes was used. The helix length for each tube was the same. Within the experimental accuracy of ±1 db, no reduction of modulation was observed in the three-dynode tube at a frequency of 1000 mc.

To get maximum frequency response, the dynodes must be spaced as close as possible. A spacing of five millimeters, which can be achieved without difficulties, yields the following calculated reduction in the depth of modulation for a three-dynode tube: 9% at 5 gc; 29% at 30 gc.

R-F Mixing Operation

When the incident light beam is modulated at a frequency f_m that is outside the passband of the helix, the tube can still be used to demodulate the signal if an r-f pumping signal f_p is supplied at the input of the helix (see Figure 14), provided f_p and either $(f_m + f_p)$ or $|f_m - f_p|$, or both, are in the helix pass band. The r-f output from the helix contains f_p and one or both of the sidebands, so f_m can be recovered by passing the output into a crystal mixer. This novel technique greatly extends the useful frequency range of the TSEM micro-

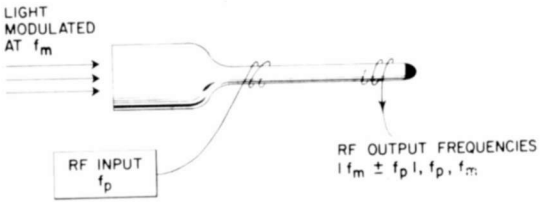


Fig. 14—R-F mixing mode of operation of microwave phototube.

wave phototube. For example, a tube having a helix that operates in the range from 1 to 2 gc can be used in this mode of operation to demodulate signals ranging from 0 to 4 gc.

For low currents, the power output at the sideband frequencies is proportional to the square of the beam current (see Figure 15), so an effective output resistance R_{mix} can be defined as follows:

$$P_{sb} = \frac{1}{2} m^2 I_0^2 R_{mix}, \tag{15}$$

where P_{sb} is the power at one side-band frequency. R_{mix} is a function of the r-f pump power; the optimum pump power is of the order of 0.1 to 1 milliwatt (see Figure 16). Our experiments show that R_{mix} is independent of signal frequency for frequencies ranging from audio to one thousand megacycles. For the L-band tube, the measured optimum value of R_{mix} in this range is about 0.1 megohm.

Conclusions

The use of transmission secondary-emission dynodes in a microwave phototube makes it possible to combine the high, noise-free gain of secondary-emission multipliers with the large bandwidth and high output resistance of a microwave helix. Tests of an L-band tube have

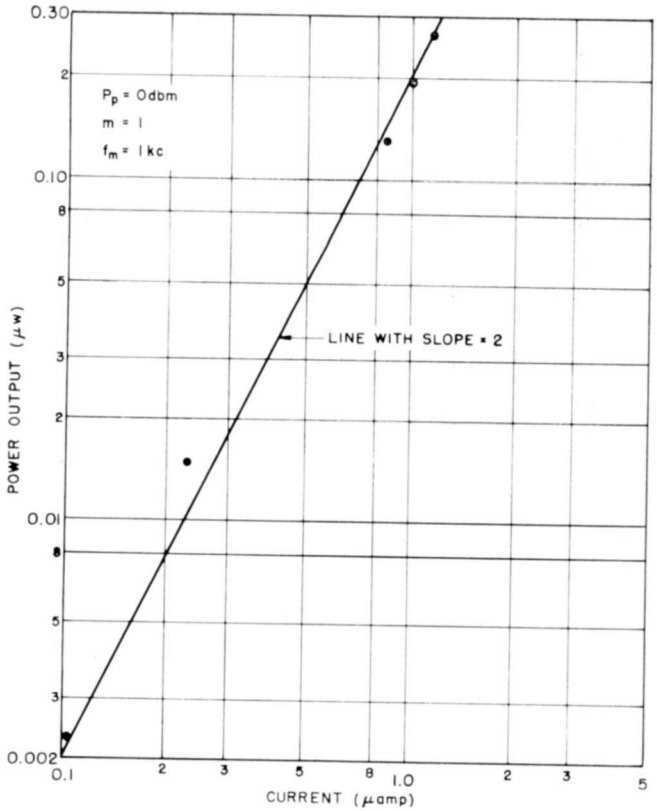


Fig. 15—Maximum power in one sideband versus average beam current, for r-f mixing operation of microwave phototube.

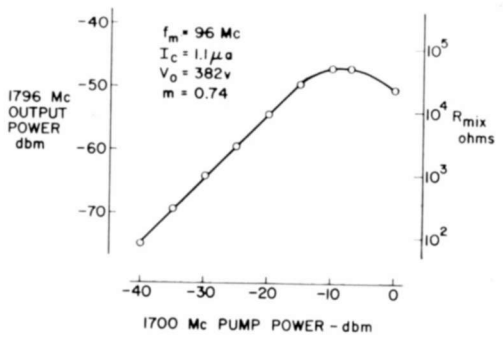


Fig. 16—Side-band power output versus pump power input for r-f mixing operation of microwave phototube.

demonstrated that there is no deterioration of modulation by the secondary-emission multiplication in the dynodes; this result confirms the calculation of expected performance. The noise is that predicted from the statistics of photo and secondary emission. Thus the sensitivity of the TSEM microwave phototube can approach the sensitivity of conventional low-frequency reflection-type photomultipliers.

Acknowledgments

The authors extend their thanks to G. A. Morton, A. H. Sommer, H. E. McCandless, and D. B. Wenner of RCA, and J. E. Creedon of USAEL, for discussions and assistance.

Appendix—A Simple Analysis of Modulation Loss Caused by Velocity Dispersion

Because of the spread in emission velocities of secondary electrons, the electrons reaching a given location at a given time can be considered to have been emitted at various times and, therefore, with various phases of the signal excitation (assumed to be sinusoidal). The difference in phase of excitation $\Delta\theta$ for electrons emitted at time intervals equal to Δt is given by

$$\Delta\theta = \omega\Delta t, \tag{16}$$

where ω is the signal frequency. The time interval Δt is determined from Equation (12) or (14).

The reduction in modulation depth caused by combining electrons with various phases of excitation at the location of interest is calculated by summation of vectors having amplitudes proportional to the number of electrons that have given values of $\Delta\theta$. The ratio of this vector sum to an algebraic sum of the same amplitudes defines the fraction of modulation depth remaining at the location corresponding to the value of Δt used.

Although the summation of modulation amplitude vectors can be performed with great accuracy by machine computation, a good estimate can be obtained by simple graphical addition. For example, the emission-voltage distribution of Equation (1) can be approximated by five vectors of relative amplitude 0.36, 0.28, 0.20, 0.12, and 0.04 (or 9:7:5:3:1), corresponding to $V_0 = 0.5$, 1.5, 2.5, 3.5, and 4.5 volts, respectively. These vectors, added at phase angles appropriate to their emission voltages and the assumed tube geometry, yield a resultant modulation depth that can be compared with unity (or 25) to determine the extent of the loss caused by the velocity dispersion.

A MODE ANALYSIS OF QUASI-ISOTROPIC ANTENNAS

By

O. M. Woodward

RCA Missile and Surface Radar Division,
Moorestown, N. J.

Summary—*A general approach is described for the analysis of antenna radiation characteristics from quasi-isotropic antennas. The theory is based on the superposition of symmetrical mode patterns produced from different types of decoupled ring arrays. In particular, a detailed study is made of electrically small radiators having omnidirectional radiation in one major plane. The calculations are presented in a number of graphs that are useful in the design of quasi-isotropic antennas for satellites or other spacecraft.*

Introduction

A COMMON PROBLEM that has faced the antenna engineer for some time is the design of a high-gain antenna having a directive beam. Today, however, he is frequently faced with the need for designing an antenna with the opposite characteristics; i.e., a low-gain antenna having radiation characteristics approaching those of an isotropic source. It has been shown that a truly isotropic antenna is impossible; that is, there is no antenna which for a given polarization will maintain constant field strength over the surface of the radiation sphere.[1] Subject to this basic limitation, however, it is possible to obtain radiation having definable ranges of amplitude variation and polarization characteristics over an appreciable portion of the sphere.

Communication antennas for satellites, missiles, and other spacecraft have become increasingly important during recent years. Considerations of stabilization, orbit trajectory, and spin have generally required that such antennas have patterns that are circularly symmetric with respect to the spin or reference axis of the space vehicle. Only patterns of this category are considered in this paper. The advantages of circular polarization over linear polarization for greater system optimization and overall gain in certain applications are now being extensively exploited. Furthermore, as the complexity of satellite equipment increases, the stringent weight, space, and reliability re-

[1] H. F. Mathis, "A Short Proof that an Isotropic Antenna is Impossible," *Proc. I.R.E.*, Vol. 39, p. 970, Aug. 1951; see also, H. F. Mathis, "On Isotropic Antennas," *Proc. I.R.E.*, Vol. 42, p. 1810, Dec. 1954.

quirements emphasize the importance of multi-port antennas capable of serving more than one transmitter or receiver without the need of remote switching.

The mode concept of pattern synthesis was started at RCA during the development of an electronic scanning antenna,[2] and subsequently led to the development of the multi-port antenna components for the NASA Relay communications satellite and other spacecraft antenna systems. It has proven to be a very useful tool, not only in establishing theoretical limitations on the radiation characteristics of familiar antennas, but also as a guide in deriving new types of antennas suitable for specific quasi-isotropic applications.

This study is confined to the theoretical radiation characteristics obtained from certain idealized antennas of small electric size.

Mode Analysis

The work is based on the theory that an antenna pattern having rotational symmetry with respect to a reference axis consists of the superposition of one or more independent, decoupled radiation mode types, each of which also has a symmetrical radiation pattern. In this connection, a mode is defined as radiation that is omnidirectional in magnitude in the plane normal to the reference axis, and that has an integral number (H) of cycles of phase variation throughout 360° in the plane. This phase variation may be either clockwise or counterclockwise. The mode-identifying integer may also be zero, i.e., the radiation may be constant in phase.

Such modes may be conceived as emanating from an infinite number of infinitesimal radiators (Hertz dipoles) arranged in a very small circle or ring coaxial to the reference axis. For any given mode, three orthogonal orientations of the radiators are possible, namely axial, tangential, and radial (Figure 1). It should be pointed out that as the radius, a, approaches zero, such a ring becomes a supergain radiator that is physically unrealizable. However, these concepts are quite useful as elementary building blocks in a theoretical study, first because they illustrate the ultimate limitations, and second because a number of practical antenna configurations have radiation characteristics that closely approximate these idealized patterns.

The radiation geometry can be most conveniently expressed in the usual θ, ϕ spherical coordinates with $\theta = 0°$ reference axis oriented normal to the plane of the rings (Figure 1). For convenience in this

[2] C. P. Clasen, J. B. Rankin, and O. M. Woodward, Jr., "A Radial-Waveguide Antenna and Multiple Amplifier System for Electronic Scanning," *RCA Review*, Vol. XXII, No. 3, p. 543, Sept. 1961.

paper, the reference axis is assumed to be vertical. At any point in the far-field region the radiation may be specified by orthogonal field components, F_θ and F_ϕ. F_θ is the component lying in the elevation plane containing the reference axis, and F_ϕ is the component normal to this plane.

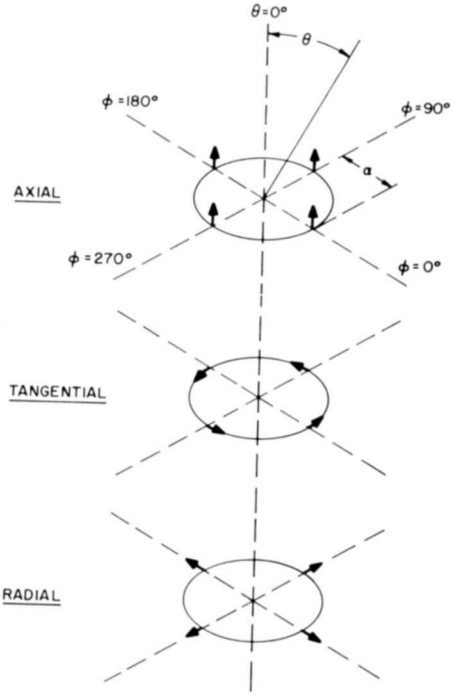


Fig. 1—Geometry of ring elements.

Chireix[3] has described the properties of rings with axial radiators. This work was extended by Knudsen in a number of extensive studies[4-6] that include the tangential and radial types as well. Expressions from

[3] H. Chireix, "Antennes a Rayonnement Zénithal Réduit," *L'Onde Elec.*, Vol. 15, p. 440, 1936.

[4] H. L. Knudsen, "The Field Radiated by a Ring Quasi-Array of an Infinite Number of Tangential or Radial Dipoles," *Proc. I.R.E.*, Vol. 41, p. 781, June 1953.

[5] H. L. Knudsen, "Radiation Resistance and Gain of Homogeneous Ring Quasi-Array," *Proc. I.R.E.*, Vol. 42, p. 686, April 1954.

[6] H. L. Knudsen, "Radiation from Ring Quasi-Arrays," *I.R.E. Trans. Antennas and Propagation*, Vol. AP-4, No. 3, p. 452, July 1956.

Table I—Relative Magnitudes of Elevation Patterns of an Infinite Number of Hertz Dipoles in a Ring whose Radius $a \rightarrow 0$ (from Knudsen).

Mode	Type	F_θ	F_ϕ
$H = 0$	Axial	$\sin\theta$	0
	Tangential	0	$\sin\theta$
	Radial	$\cos\theta \sin\theta$	0
$H = \pm1, \pm2, \pm3, \cdots$	Axial	$[\sin\theta]^{\lvert H\rvert+1}$	0
	Tangential	$\cos\theta\,[\sin\theta]^{\lvert H\rvert-1}$	$[\sin\theta]^{\lvert H\rvert-1}$
	Radial	$\cos\theta\,[\sin\theta]^{\lvert H\rvert-1}$	$[\sin\theta]^{\lvert H\rvert-1}$

these works for the relative elevation field patterns of rings having very small radii are summarized in Table I.

The elevation pattern characteristics may also be derived quite simply for the lower order modes from ring arrays of only four elements. This is evident since the patterns in the azimuth plane become nearly circular for this case as the ring radius, a, approaches zero. The normalized patterns for any ring radius in the elevation plane at $\phi = 0°$ are tabulated in column 1 of Table II with respect to

Table II—Relative Magnitudes and Phases of Patterns in Elevation Plane for Four Hertz Dipoles in Ring of Radius a $(k = 2\pi/\lambda)$.

Mode	Type	(1) Normalized patterns in elevation plane ($\phi = 0°$)		(2) $\lim a \rightarrow 0$	
		F_θ	F_ϕ	F_θ	F_ϕ
$H = 0$	Axial	$\sin\theta[1 + \cos(ka\sin\theta)]$	0	$\sin\theta$	0
	Tangential	0	$j\sin(ka\sin\theta)$	0	$j\sin\theta$
	Radial	$j\cos\theta\sin(ka\sin\theta)$	0	$j\sin\theta\cos\theta$	0
$H = \pm1$	Axial	$j\sin\theta\sin(ka\sin\theta)$	0	$j\sin^2\theta$	0
	Tangential	$\pm j\cos\theta$	$\cos(ka\sin\theta)$	$\pm j\cos\theta$	1.0
	Radial	$\cos\theta\cos(ka\sin\theta)$	$\mp j1.0$	$\cos\theta$	$\mp j1.0$

a reference element at the ring center that is in phase with the ring element located at $\phi = 0°$. The limiting field expressions as $a \rightarrow 0$ are given in column 2, and are seen to be identical in magnitude to the corresponding terms of Table I for $H = 1$. The positive sign of the mode number, H, is taken here to mean a phase increase that is proportional to the counterclockwise ϕ-angle as indicated in Figure 1.

Knudsen[6] has noted that only modes $H = \pm 1$ of the tangential and radial types can radiate in the directions $\theta = 0°$ and $\theta = 180°$, all other modes having a null in these directions. In addition, the tangential and radial patterns are identical in shape for $H = \pm 1$, differing only in their relative phasing (see column 2 of Table II). For any mode (except $H = 0$), either the tangential or the radial type is sufficient in a study of mode combinations radiating from small ring arrays (see Table I).

It may also be concluded that one (and only one) mode may be employed if the desired radiation pattern is required to be omnidirectional in the azimuth plane. This is evident from the fact that if two or more different modes are superimposed, their unequal phase variation with the angle ϕ will result in a variable amplitude characteristic about the reference axis. However, any combination of geometric types of a single mode (with the same sign of H) may be used without disturbing the omnidirectional properties in the azimuth plane.

Knudsen has also discussed the super-gain limitations of practical ring arrays and observed that the minimum circumference in wavelengths is approximately equal to the highest-order mode component. Hence, in this study of electrically small antennas (approximately one wavelength in circumference or less), only the lower-order modes ($H = 0$ and $H = \pm 1$) can be considered.

Mode Gain Factors

The gains with respect to a linearly polarized isotropic source are tabulated in Table III for various types of the two lower-order modes.

Referring to column 2 of Table II, the axial and tangential patterns of the $H = 0$ mode are identical to those of a single Hertz dipole having a gain of 1.5. The radial type for the $H = 1$ mode becomes a turnstile of orthogonal Hertz elements in the limit as the radius, a, approaches zero. Thus the gain for the F_θ and F_ϕ components is just half that of a single element, or 0.75. The tangential and radial types for $H = 1$ have the same gains, since their patterns are identical in magnitude.

The gain of the axial type for the $H = \pm 1$ mode is obtained by the usual method of integration for patterns having omnidirectional radia-

tion in the $\theta = 90°$ plane. Thus, on integrating the power density over the spherical surface for both an isotropic pattern and the above pattern, the gain is the ratio of the resultant summations;

$$\frac{\int_0^{\pi/2} \sin \theta d\theta}{\int_0^{\pi/2} \sin^5 \theta d\theta}$$

The gain for the radial type of the $H = 0$ mode is found in a similar manner. For this case the maximum field occurs at $\theta = 45°$ and $135°$.

Table III—Power Gain Relative to a Linearly Polarized Isotropic Source for the θ and ϕ Components

Mode	Type	G_θ	G_ϕ
$H = 0$	Axial	1.5	0
	Tangential	0	1.5
	Radial	1.875	0
$H = 1$	Axial	1.875	0
	Tangential	0.75	0.75
	Radial	0.75	0.75

General Radiation Characteristics for Combined Mode Types

Various radiation characteristics in the elevation plane may be obtained by changing the relative amplitudes and phases of superimposed rings of different geometric types. As pointed out earlier, however, these rings must have identical modes or phase progressions to preserve omnidirectional patterns in the plane of $\theta = 90°$.

All possible elevation patterns for the $H = \pm 1$ mode may be calculated from a combination of only two different types: (1) the axial, and (2) either the radial or the tangential, since the magnitudes of the latter two are identical as previously described. Although three different types are available for the $H = 0$ mode, only combinations of the axial and tangential types are considered here since these are sufficient for most of the practical designs of electrically small antennas.

The notation employed for the power division between the superimposed rings is

P_A = the fraction of the total input power to the axial ring,

$P_T = (1 - P_A)$ = the fraction of the total input power to the tangential ring.

Thus all of the different radiation characteristics may be defined in terms of only one power parameter, P_A. This is possible because the modes to be combined are decoupled.

The method chosen for defining the relative phasing is a phase comparison between the currents flowing in the superimposed rings at the same ϕ-angle. This approach has been found useful since it allows a simple physical identification between many types of practical antenna designs and the corresponding theoretical patterns. Hence, in the following calculations, β is defined as the phase of the tangential ring current at any ϕ-angle relative to the phase of the axial current at the same ϕ-angle. For the in-phase condition ($\beta = 0°$), the currents are assumed to flow in directions indicated in Figure 1.

Having derived the relative pattern shapes (Table II) and the gains (Table III) for the different mode types, the elevation field patterns for both the F_θ and F_ϕ components may now be written in terms of P_A and β for the various ring combinations.

The expressions for the axial and tangential combinations of the $H = 0$ mode relative to a linearly-polarized, isotropic source are

$$F_\theta = \sqrt{1.5\,P_A}\,\sin\theta + 0, \tag{1}$$

$$F_\phi = 0 + \sqrt{1.5\,(1 - P_A)}\,\sin\theta\,e^{j(\beta+90)}. \tag{2}$$

Corresponding expressions for the axial and tangential combinations of the $H = \pm 1$ mode are

$$F_\theta = e^{j90}\,[\sqrt{1.875\,P_A}\,\sin^2\theta \pm \sqrt{0.75\,(1 - P_A)}\,\cos\theta\,e^{j\beta}], \tag{3}$$

$$F_\phi = 0 + \sqrt{0.75\,(1 - P_A)}\,e^{j\beta}. \tag{4}$$

The above relationships are sufficient to determine the various radiation properties in the elevation plane of the two lower order modes, $H = 0$ and $H = 1$, in terms of P_A and β. These are considered separately in the following sections. The study has shown that the data for $\beta = 0°$ and $\beta = 90°$ is of greater practical importance at present. However, curves for the intermediate values of $\beta = 45°$ are also given for the sake of completeness and possible future use.

$H = 0$ Mode Combinations of the Axial and Tangential Types

As both the F_θ and F_ϕ patterns for the $H = 0$ mode (Equations (1) and (2)) are functions of sin θ only, then any combined relationship also has the same characteristic, and only a single normalized

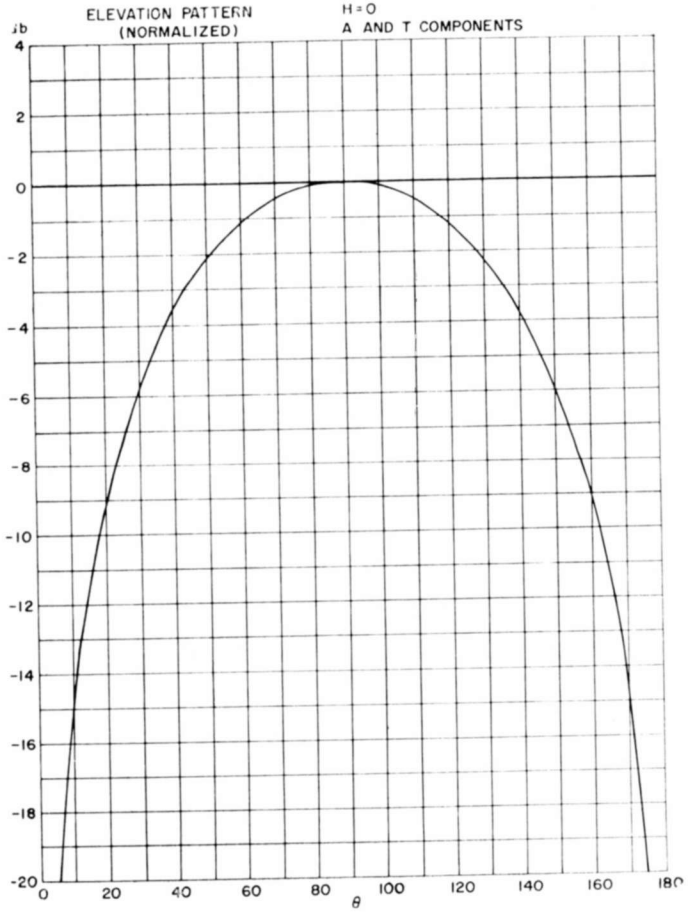


Fig. 2—The normalized elevation pattern for the $H = 0$ mode combinations of the axial and tangential types.

pattern in terms of θ (Figure 2) is needed to show the relative variations in the elevation plane for any of the radiation parameters. Therefore only the absolute magnitudes of the various radiation characteristics at the beam maximum ($\theta = 90°$) will be discussed.

The field intensities (in decibels) of the F_θ and F_ϕ patterns at $\theta = 90°$ relative to a linearly polarized isotropic source are given in

Figure 3 in terms of P_A. This data is independent of the parameter β, as neither Equations (1) or (2) involve both axial and tangential components simultaneously.

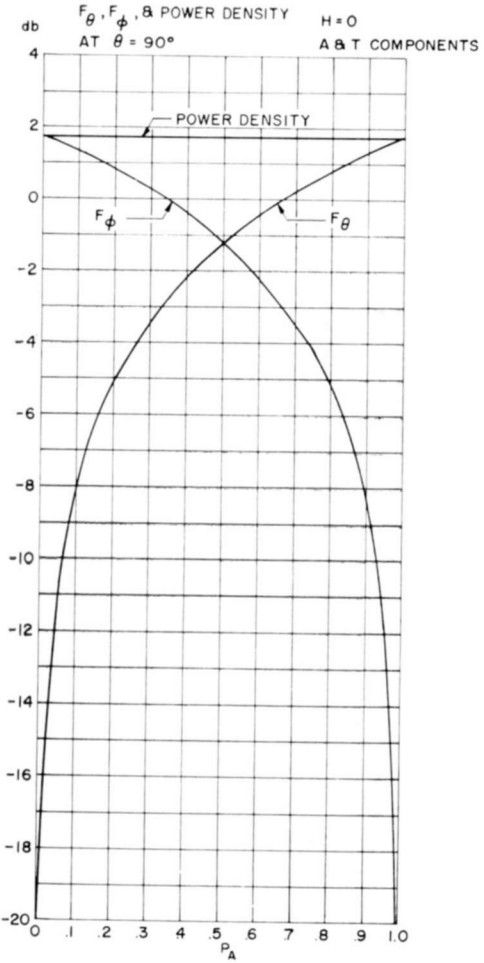


Fig. 3—The magnitudes (in db relative to a linearly-polarized isotropic source) of F_θ, F_ϕ, and power density at $\theta = 90°$ for the $H = 0$ mode combination of the axial and tangential types in terms of P_A.

Figure 3 also shows the power density radiated at $\theta = 90°$ as a function of P_A. The power density is equal to the sum of the powers radiated from any two orthogonal radiation components, either linearly polarized or circularly polarized. This radiation characteristic is useful

in applications such as polarization diversity reception. It is apparent from Equations (1) and (2) for the orthogonal F_θ and F_ϕ components that the power density is independent of P_A.

The maximum (F_{max}) and minimum (F_{min}) values of field intensity at $\theta = 90°$ may be found either by mathematical computations[7] or by graphical means from a polarization chart. As the latter process was employed in this work, a short description is given below.

This system[7] is especially well suited for plotting the pertinent polarization parameters on a single graph consisting of the ordinary

Table IV—Relationship between Impedance and Polarization Charts

Impedance Chart	Polarization Chart
Impedance magnitude	F_ϕ/F_θ
Impedance phase angle, Δ	phase of F_ϕ relative to F_θ ($\Delta = 0°$ for the right half of the great circle, 90° for the vertical axis, and 180° for the left half of the great circle)
Standing-Wave Ratio	Axial Ratio of F_{max}/F_{min}
Distance toward generator, α (in electrical degrees)	angular orientation of F_{max} with respect to F_θ

Carter[8] impedance chart with altered coordinates for representation of elliptically polarized characteristics. The illustrative example of Figure 4 indicates the geometric relationships of the polarization chart to the impedance chart, and the tabulation of Table IV gives the relationship between the two systems.

In Figure 4 the intersection on the polarization chart of the (F_ϕ/F_θ) curve and the phase angle curve (Δ) locates the point P. A circle is drawn through P as indicated. The lines joining the $\alpha = 90°$ point to P and to the point diametrically opposite P define the magnitudes of the F_θ and F_ϕ components. The intersection of the circle with the vertical axis defines the magnitudes of the F_{max} and F_{min} compo-

[7] V. H. Rumsey, "Part I—Transmission Between Elliptically Polarized Antennas," *Proc. I.R.E.*, Vol. 39, p. 535, May 1951 (see also papers following Part I).

[8] P. S. Carter, "Charts for Transmission-Line Measurements and Computations," *RCA Review*, Vol. III, p. 355, Jan. 1939.

nents. Thus, this polarization chart enables the F_{max} and F_{min} magnitudes to be determined in terms of the known F_θ and F_ϕ components.

Using this polarization chart, the F_{max} and F_{min} values at $\theta = 90°$ relative to a linearly polarized isotropic source are plotted (in decibels)

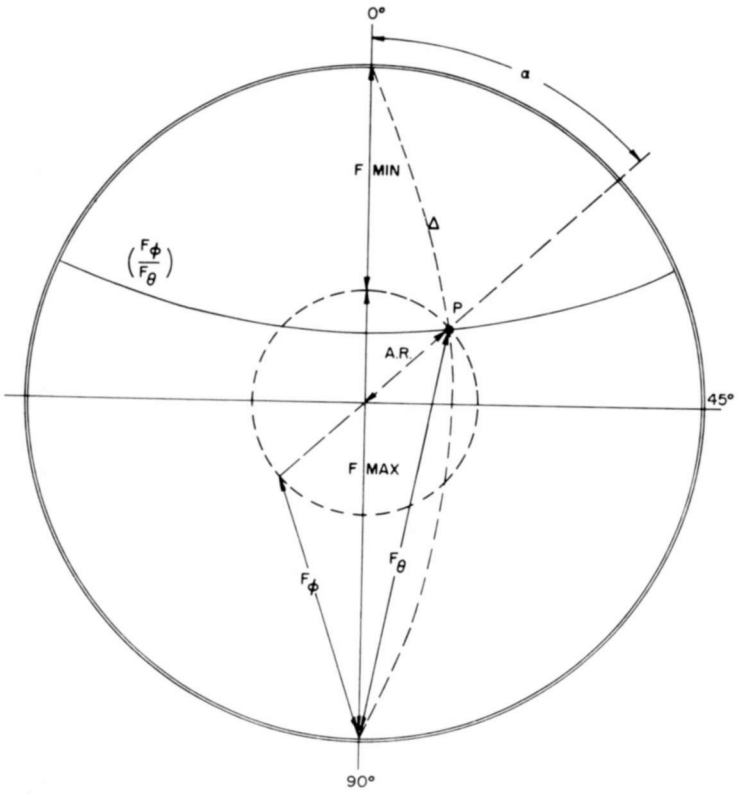


Fig. 4—Polarization chart showing the geometric representation of components for elliptical polarization.

as a function of P_A and β in Figure 5. It may be noted that if $\beta = 0°$, the F_{max} and F_{min} variations are identical to the F_θ and F_ϕ patterns of Figure 3. With the geometry and definitions employed here, the radiation is right-hand elliptically polarized[9] for all values of P_A other than 1.0 or 0. For $\beta = 90°$, the radiation is linearly polarized and the F_{max} curve of Figure 5 is the same as the power density curve of Figure 3.

[9] "Standards on Wave Propagation: Definitions of Terms, 1950," *Proc. I.R.E.*, Vol. 38, p. 1265, Nov. 1950.

The ratio of F_{max} to F_{min}, or axial ratio, is plotted in Figure 6 as a function of P_A and β as calculated from the data of Figure 5. Circular polarization can only be obtained for $P_A = 0.5$ and $\beta = 0°$. For $\beta = 90°$ the radiation is linearly polarized for all values of P_A and hence the axial ratio is infinite.

The elliptically polarized radiation with a major axis F_{max}, and minor axis F_{min}, may be resolved as two counter-rotating, circularly polarized wave components having field magnitudes of $1/2\ (F_{max} + F_{min})$ and $1/2\ (F_{max} - F_{min})$. Therefore the power absorbed by a

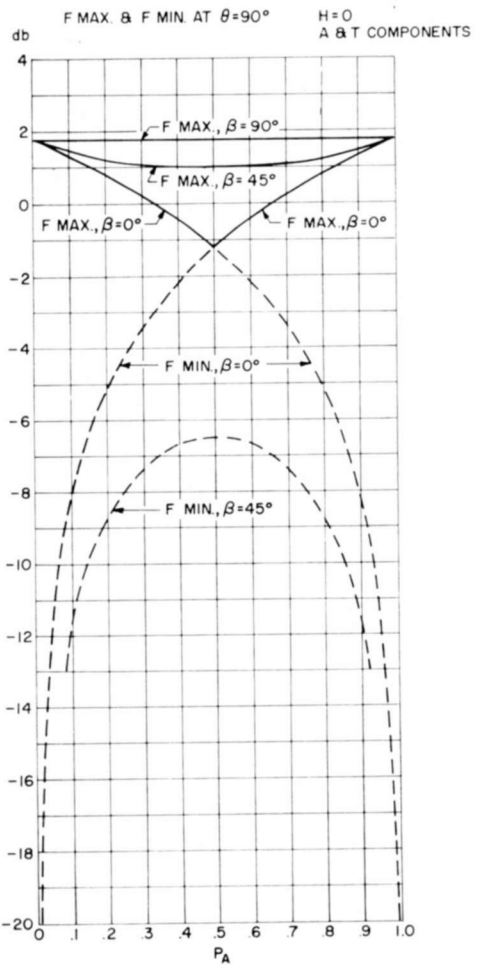


Fig. 5—The magnitudes (in db relative to a linearly-polarized isotropic source) of F_{max} and F_{min} at $\theta = 90°$ for the $H = 0$ mode combination of the axial and tangential types as a function of P_A and β.

circularly polarized receiving antenna (P_{cp}) is proportional to the sum of the squares of the orthogonal field magnitudes for each of the circularly polarized components, or

$$P_{cp} \propto \frac{1}{2} [F_{\max} \pm F_{\min}]^2. \tag{5}$$

The data of Figure 5 was used along with these relations to calculate the curves of Figure 7, which show the gain (in decibels) at $\theta = 90°$ relative to a circularly polarized isotropic source in terms of P_A and β when measured with right-hand (RH) and left-hand (LH) circularly polarized antennas. It is seen from the $\beta = 0°$ curves that

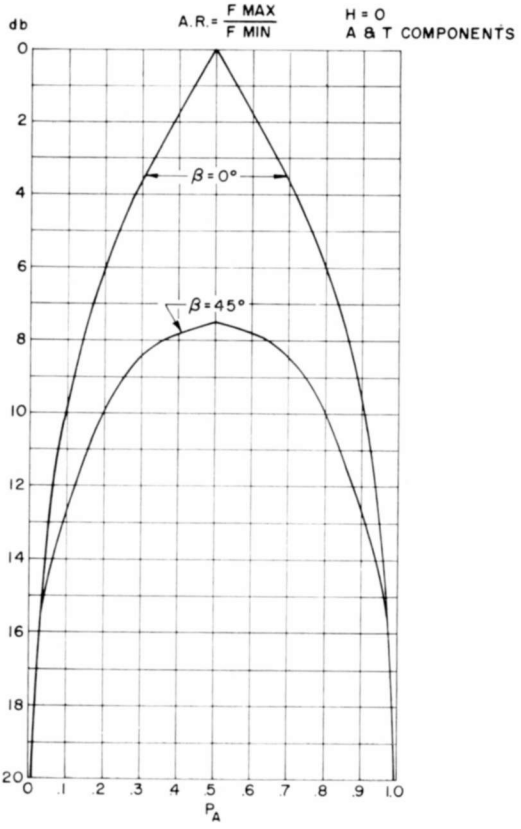


Fig. 6—The axial ratio characteristics (in db) at any θ angle for the $H = 0$ mode combinations of the axial and tangential types as a function of P_A and β.

the gain varies little for a wide variation in P_A. For linearly polarized radiation ($\beta = 90°$), the gain is down 3 decibels from the maximum gain, and is independent of P_A.

Summarizing the calculations for the axial and tangential combinations of the $H = 0$ mode, the data of Figures 3, 5, 6, and 7 give the relative gains at $\theta = 90°$ in terms of P_A and β for the various radiation parameters of F_θ, F_ϕ, power density, F_{max}, F_{min}, axial ratio, F_{RH}, and F_{LH}. Corresponding values at other θ-angles may be determined by utilizing the normalized curve of Figure 2.

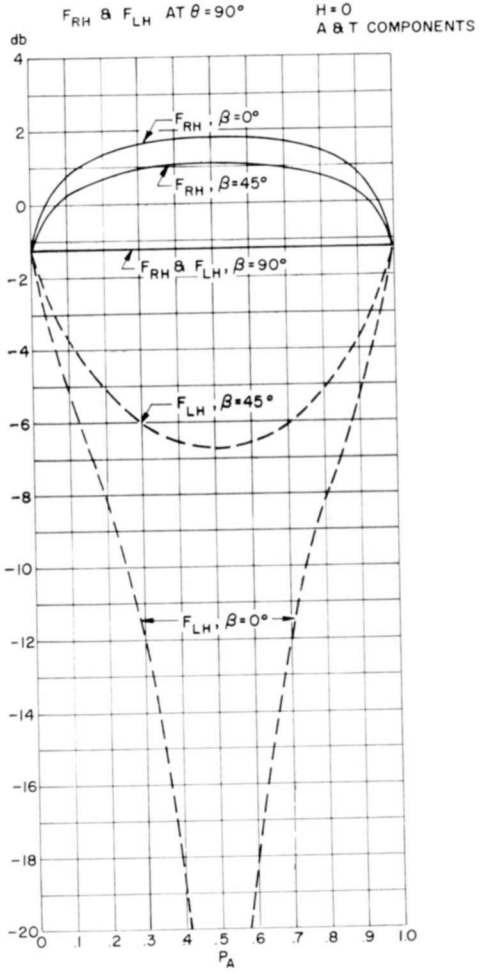


Fig. 7—The right-hand and left-hand circularly-polarized components at $\theta = 90°$ for the $H = 0$ mode combinations of the axial and tangential types as a function of P_A and β. The magnitudes (in db) are relative to a circularly polarized isotropic source.

The curves for $\beta = 45°$ and $90°$ in the above charts apply for both positive and negative values of β. It is seen from Equations (1) and (2) that the designated senses of circular polarization must be reversed for $90° < |\beta| \leq 180°$.

$H = 1$ Mode Combinations of the Axial and Tangential Types

Elevation patterns of F_θ and F_ϕ for the axial and tangential combinations of the $H = +1$ mode given in Equations (3) and (4) are

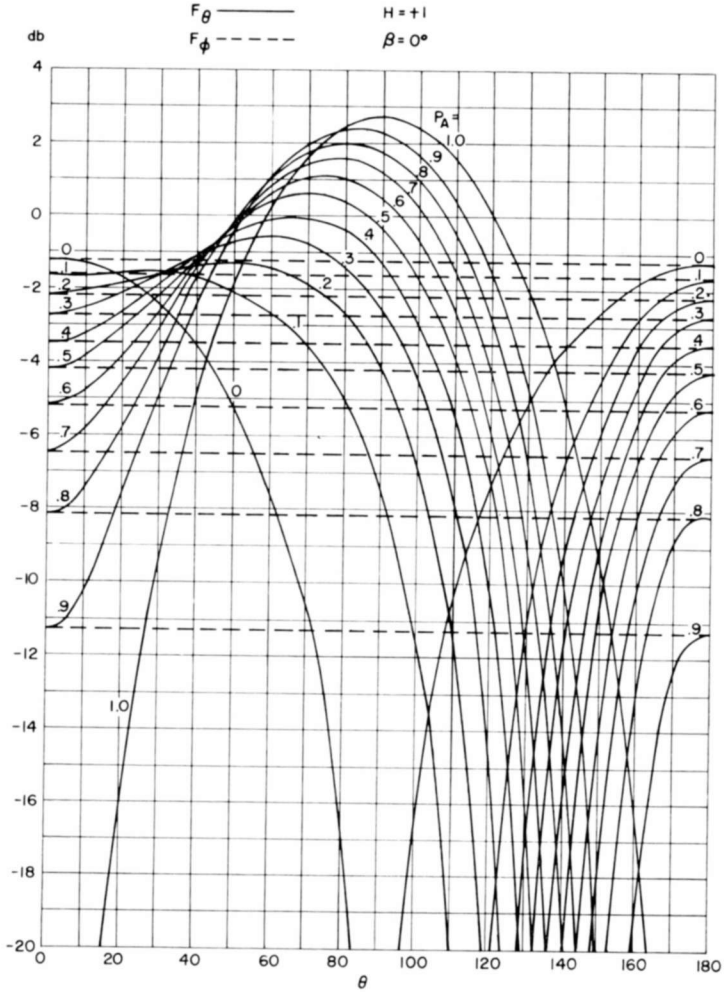


Fig. 8—F_θ and F_ϕ elevation patterns (in db relative to a linearly polarized isotropic source) for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 0°$ and P_A variable. These curves also represent the extremes of the field variation from F_{max} to F_{min} for $\beta = 0°$.

plotted in Figures 8, 9, and 10 against the angle θ in terms of the parameters P_A and β. Since the tangential type for $H = +1$ radiates circular polarization at $\theta = 0°$ and $\theta = 180°$, the F_θ and F_ϕ components are independent of β at these angles. From Equation (4), the magnitudes of the F_ϕ curves are functions of P_A only and are independent of the angle θ. Thus they appear as horizontal lines on the rectangular plots of Figures 8, 9, and 10. The F_θ curves, however, are functions of θ, P_A, and β (see Equation (3)). In Figure 8, for $\beta = 0°$ it is seen that the F_θ characteristics are unsymmetrical with respect to the

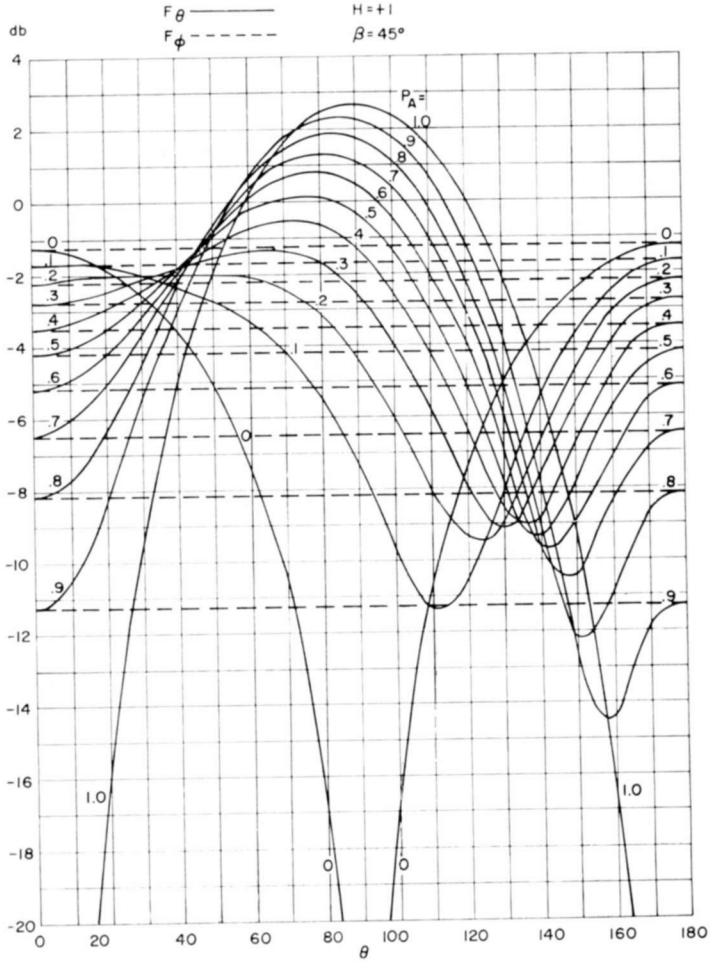


Fig. 9—F_θ and F_ϕ elevation patterns (in db relative to a linearly polarized isotropic source) for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 45°$ and P_A variable.

azimuth plane ($\theta = 90°$) for all values of P_A other than 1.0 and 0. This property is useful in improving hemispheric coverage for circular polarization. The curves of Figure 10 for $\beta = 90°$ exhibit perfect symmetry with respect to the azimuth plane.

Next, curves are presented for the maximum and minimum values of field intensity with respect to a linearly polarized isotropic source. As with the previous data given for $H = 0$, the F_{max} and F_{min} curves for $H = +1$ and $\beta = 0°$ are identical with the F_θ and F_ϕ curves of

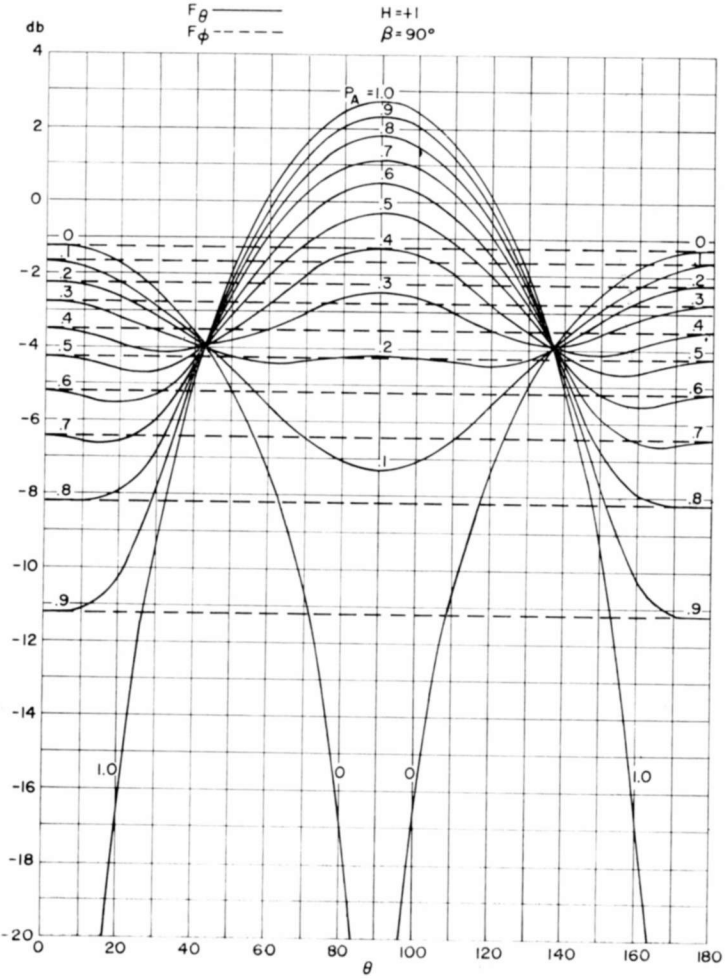


Fig. 10—F_θ and F_ϕ elevation patterns (in db relative to a linearly polarized isotropic source) for the $H = +1$ mode combination of the axial and tangential types with $\beta = 90°$ and P_A variable.

Figure 8. For this case, the solid and dashed lines represent only the extremes of the field variation from F_{max} to F_{min} as a function of the angle θ. Corresponding curves for $\beta = 45°$ and $\beta = 90°$ are given in Figures 11 and 12, respectively.

The axial ratio of F_{max} to F_{min} plotted against the angle θ is given in Figures 13, 14, and 15 for $\beta = 0°$, 45°, and 90°, respectively. For values of P_A other than 0 and 1.0, the curves of Figure 13, for $\beta = 0°$, are seen to rise first from 0 decibels (circularly polarized radiation) at

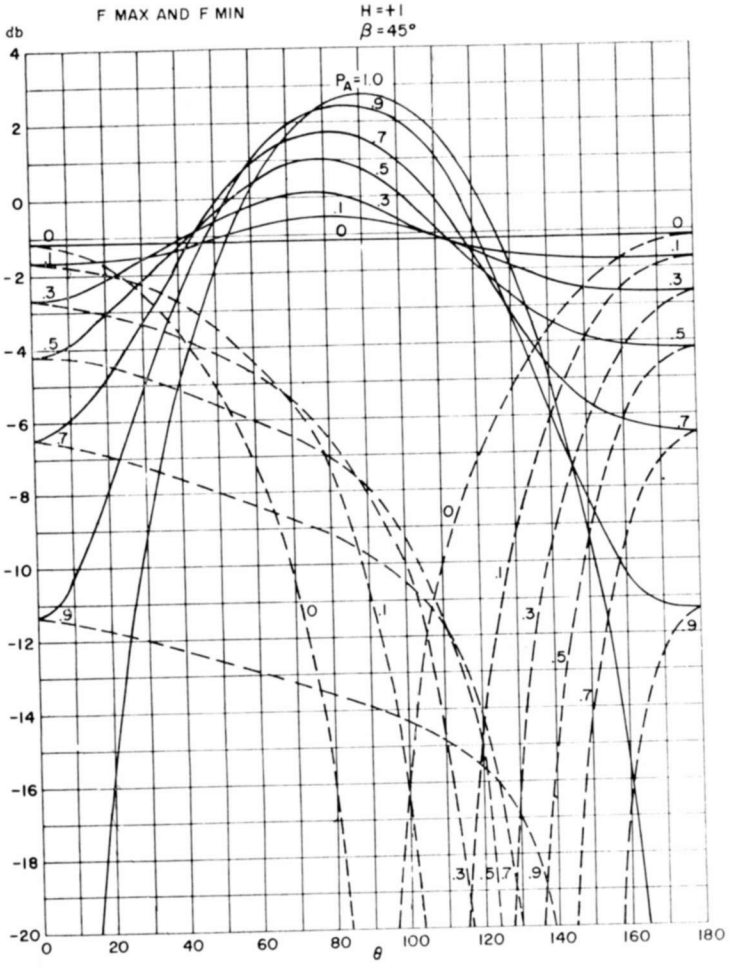


Fig. 11—F_{max} and F_{min} elevation patterns (in db relative to a linearly polarized isotropic source) for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 45°$ and P_A variable.

$\theta = 0°$ to some finite value (elliptical polarization of the same sense), then fall back again to 0 decibels before rising rapidly to a high axial ratio (linear polarization) at some value of θ which depends on P_A. At higher θ-angles, the axial ratio drops back to 0 decibels at $\theta = 180°$ (circular polarization of the opposite sense).

In contrast, the axial ratio curves of Figure 15, for $\beta = 90°$, show symmetry about the $\theta = 90°$ plane and also indicate that only linearly

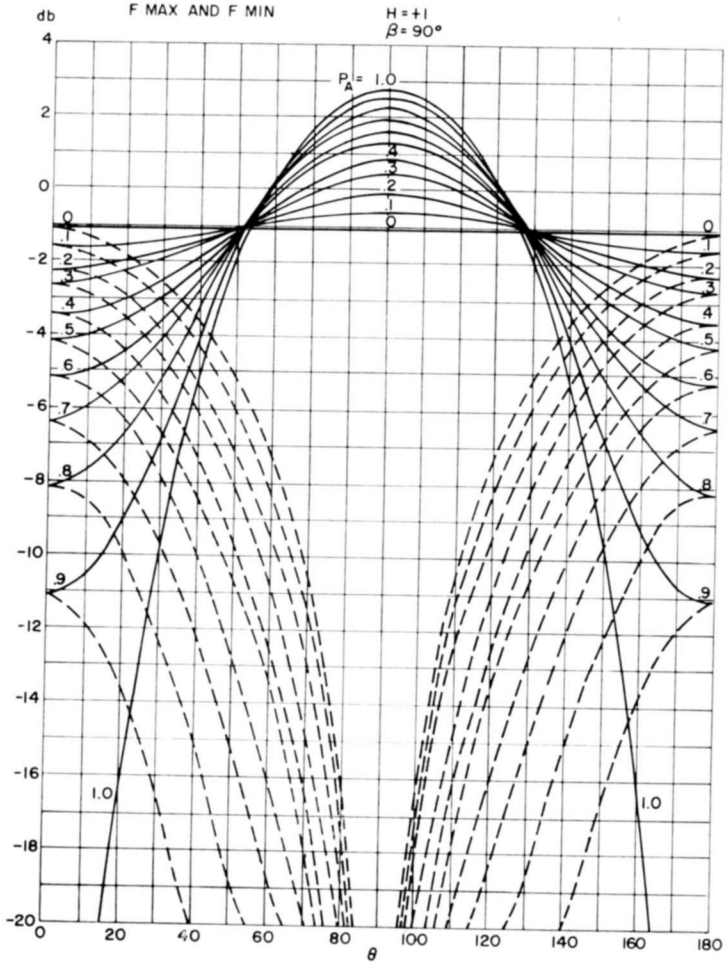


Fig. 12—F_{max} and F_{min} elevation patterns (in db relative to a linearly polarized isotropic source) for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 90°$ and P_A variable.

polarized radiation can be obtained in that plane regardless of the P_A value.

The F_{max} and F_{min} data of Figures 8, 11, and 12 are used in conjunction with Equation (5) to determine the gain (in decibels), rela-

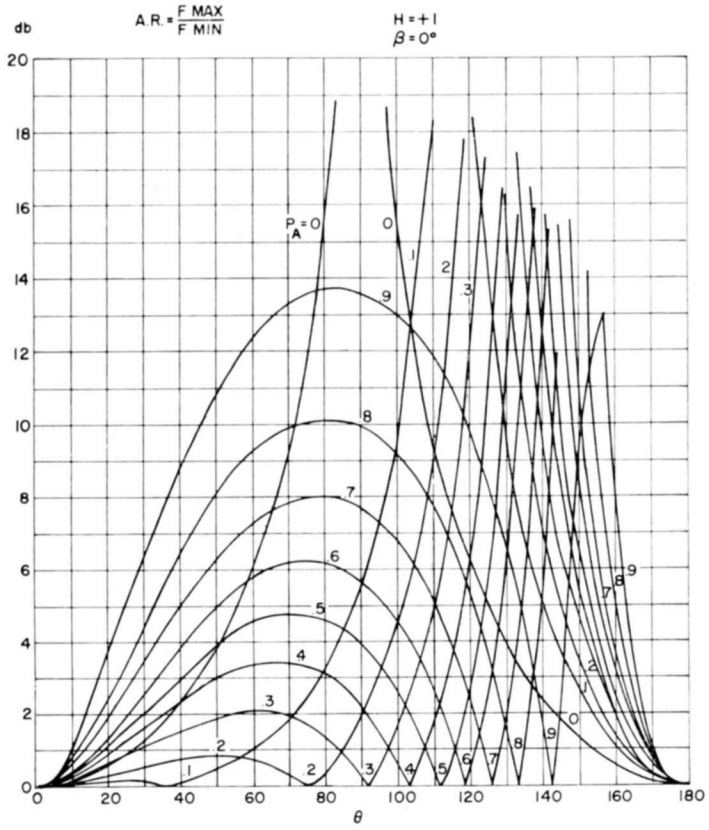


Fig. 13—The axial ratio characteristics (in db) plotted against the elevation angle, θ, for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 0°$ and P_A variable.

tive to a circularly polarized isotropic source, in terms of P_A and β as measured with circularly polarized antennas. Charts showing gain plotted against angle θ are given in Figures 16, 17, and 18 for $\beta = 0°$, 45°, and 90°, respectively. As mentioned previously, considerably more than a hemispheric coverage for a single sense of elliptical polarization can be obtained for $\beta = 0°$ and intermediate values of P_A (Figure 16).

The symmetry of the curves for $\beta = 90°$ (Figure 18) indicates equal values of RH and LH circular polarization, i.e., linear polarization at $\theta = 90°$.

Graphs for the power density versus angle θ for $\beta = 0°$, 45°, and

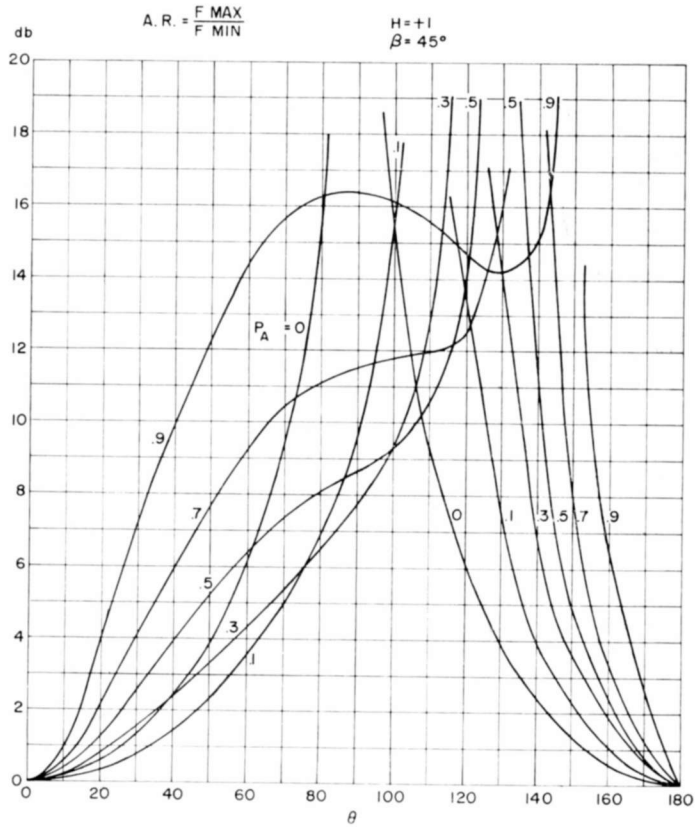


Fig. 14—The axial ratio characteristics (in db) plotted against the elevation angle, θ, for $H = +1$ mode combinations of the axial and tangential types with $\beta = 45°$ and P_A variable.

90° are given in Figures 19, 20, and 21, respectively. It is noted that nulls occur only at the poles ($\theta = 0°$ and 180°) for $P_A = 1.0$, and that the power density has little variation with θ for certain values of P_A and β.

This raises the interesting point of how close an electrically small antenna can approach a completely isotropic radiator in terms of power

density.* To answer this question, the maximum variation in power density was determined from Figures 19, 20, and 21 in terms of P_A and β, and the results plotted as the curves of Figure 22. It is seen that for $\beta = 90°$ and $P_A = 0.27$, the radiation is spherical within 0.6 decibel.

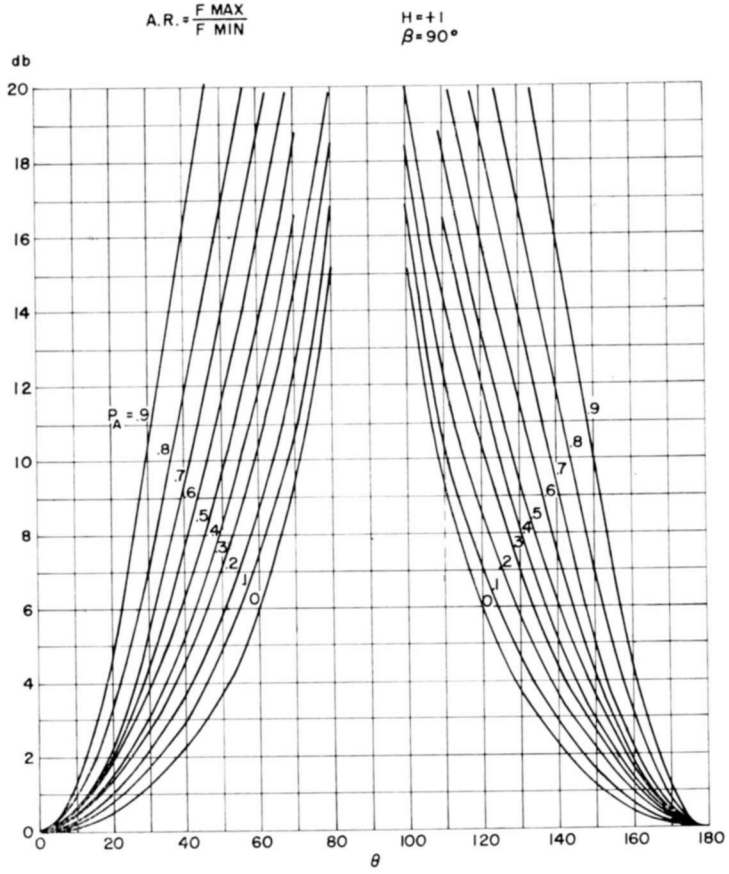


Fig. 15.—The axial ratio characteristics (in db) plotted against the elevation angle, θ, for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 90°$ and P_A variable.

Figures 23 to 27 are presented to summarize the foregoing data of the $H = +1$ mode in polar form to give a better overall picture of the significant changes in pattern shape with variation in the parameters.

* It is to be noted that this is not in violation of the impossibility of achieving a truly isotropic antenna since the references[1] apply to a specific polarization sense, not power density.

These charts enable the user to pinpoint ranges of parameters defining desired radiation characteristics quickly; then more detailed and accurate data may be found by reference to the previously described rectangular plots.

Figure 23 gives the F_θ and F_ϕ characteristics in terms of β and representative values of P_A. The gain reference is a linearly polarized isotropic source.

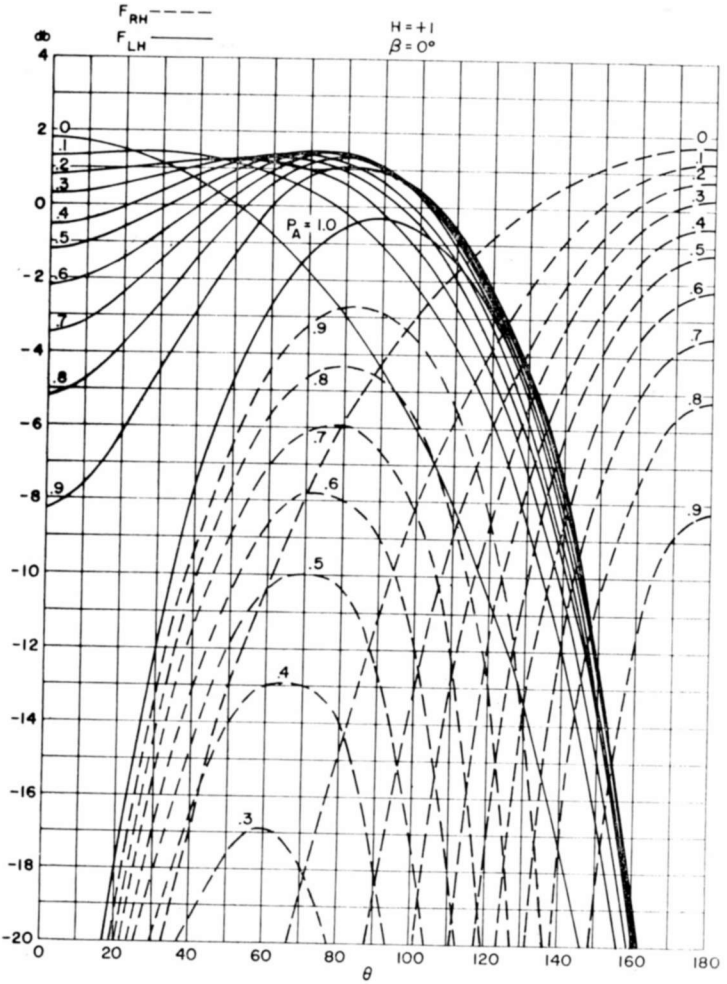


Fig. 16—Elevation patterns of the right-hand and left-hand circularly polarized components for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 0°$ and P_A variable. The magnitudes (in db) are relative to a circularly-polarized isotropic source.

Figure 24 presents the F_{max} to F_{min} variations in a similar manner. The shaded areas represent the region in which the field intensity varies. Thus, for example, the plot for $P_A = 0.3$ and $\beta = 0°$ indicates that the radiation is circularly polarized at $\theta = 0°$, 90°, and 180° ($F_{max} = F_{min}$), elliptically polarized in the upper hemisphere, and becomes linearly polarized at $\theta \simeq 130°$ in the lower hemisphere.

The axial ratio in decibels of F_{max} to F_{min} is given in the next set

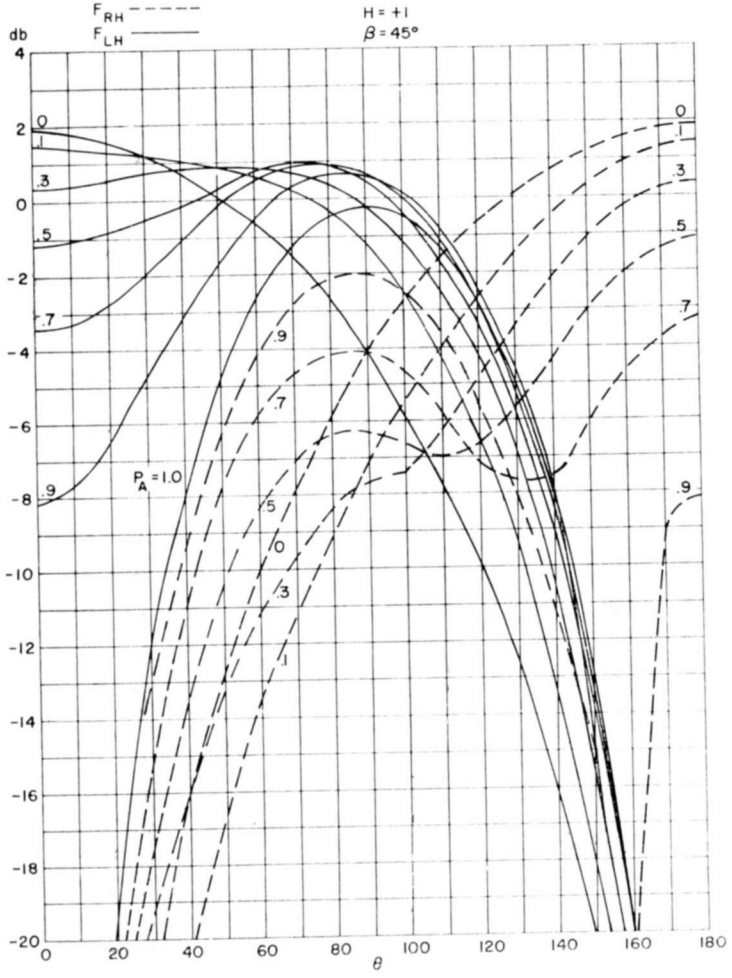


Fig. 17—Elevation patterns of the right-hand and left-hand circularly polarized components for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 45°$ and P_A variable. The magnitudes (in db) are relative to a circularly-polarized isotropic source.

of patterns (Figure 25). From Figures 24 and 25, it is seen that for $H = +1$, the radiation is linearly polarized

(1) at $\theta = 90°$ for $P_A = 0$ and any value of β;
(2) at $\theta = 90°$ for $\beta = 90°$ and any value of P_A;
(3) at all directions for $P_A = 1.0$; and
(4) at some one direction in the range of $180° > \theta > 90°$ for $1.0 > P_A > 0$ and $90° > \beta \geq 0°$.

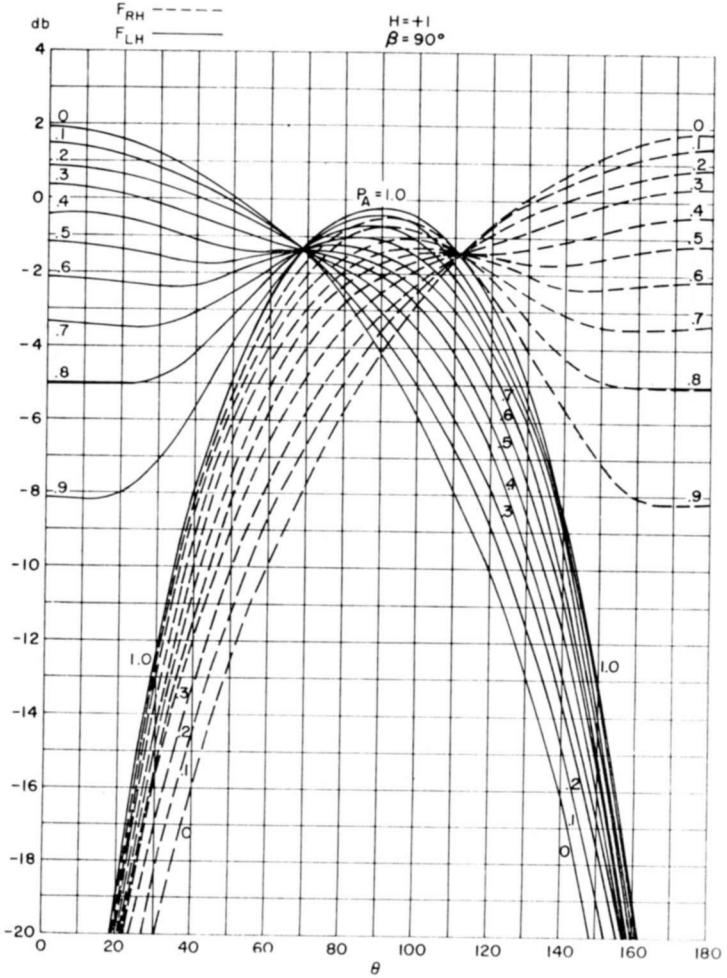


Fig. 18—Elevation patterns of the right-hand and left-hand circularly polarized components for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 90°$ and P_A variable. The magnitudes (in db) are relative to a circularly-polarized isotropic source.

Circular polarization can be obtained only at the poles ($\theta = 0°$ and $\theta = 180°$) and at one other θ-angle for $\beta = 0°$ and $1.0 > P_A > 0$.

The data of Figure 26 gives the RH and LH circularly polarized patterns with reference to a circularly polarized source. This set of curves has been found particularly useful in satellite-to-ground links requiring the maximum signal transmission over wide "look" angles.

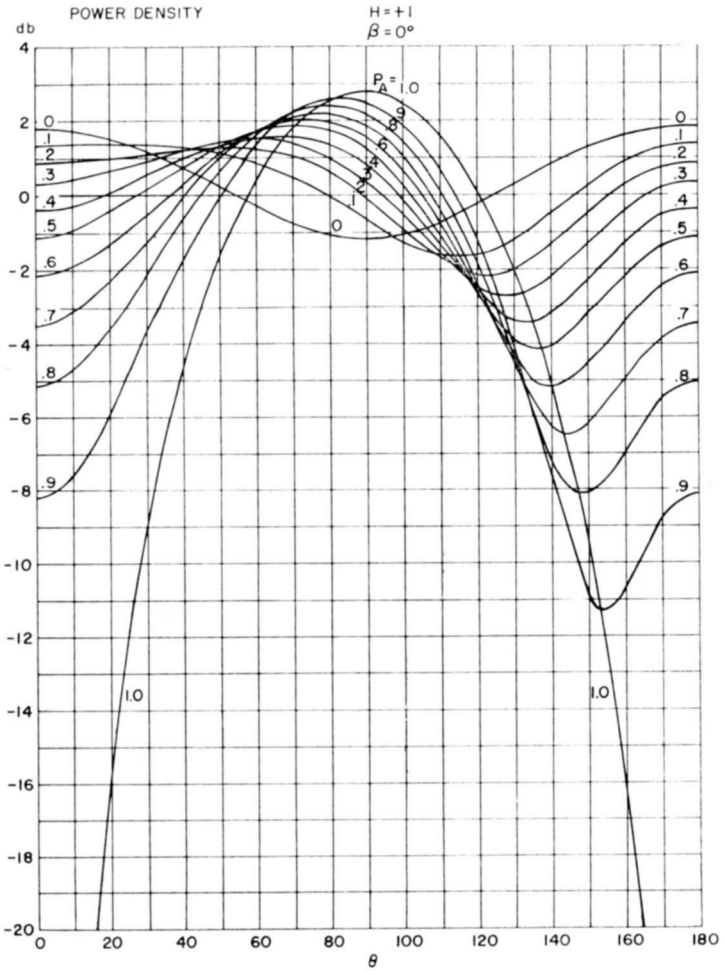


Fig. 19—Power density (in db relative to an isotropic source) plotted against the elevation angle, θ, for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 0°$ and P_A variable.

It is seen that the greatest hemispherical coverage is obtained for $\beta = 0°$ and intermediate values of P_A.

The last set of curves (Figure 27) shows the power density characteristics.

General characteristics noted on all five figures are

(1) complete symmetry for all values of P_A is obtained with respect to the $\theta = 90°$ plane for $\beta = 90°$, and the greatest asymmetry is exhibited for $\beta = 0°$;

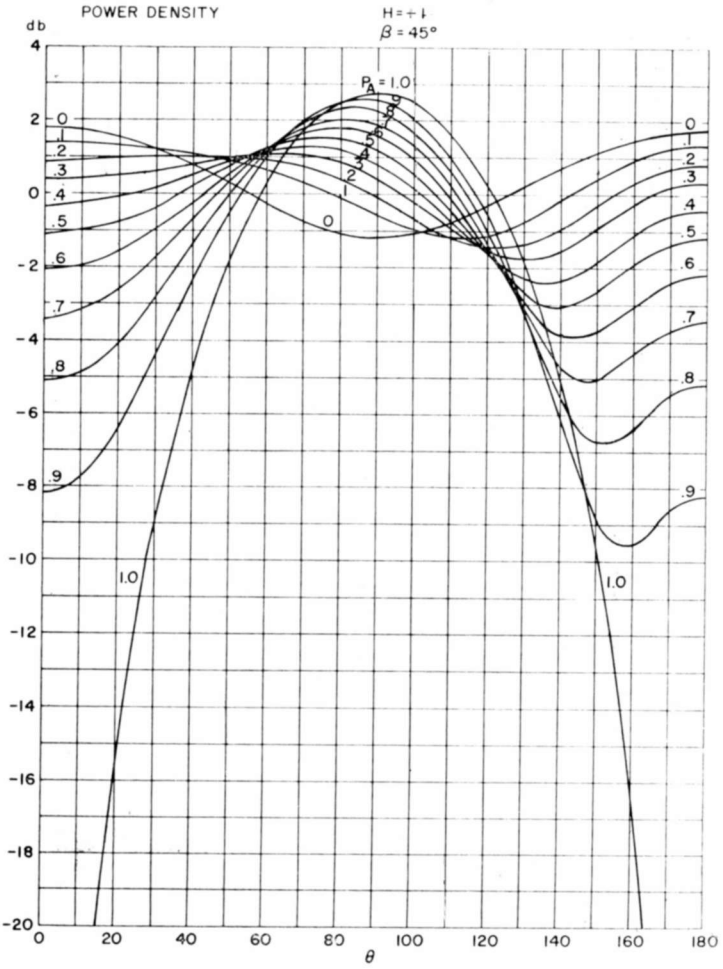


Fig. 20—Power density (in db relative to an isotropic source) plotted against the elevation angle, θ, for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 45°$ and P_A variable.

(2) curves for $P_A = 0$ and $P_A = 1.0$ are independent of β;

(3) there are relatively rapid changes in pattern shape near the maximum and minimum values of P_A and particularly in the range from 0.9 to 1.0.

In these and all foregoing charts for the $H = +1$ mode, the curves for $\beta = 45°$ and $\beta = 90°$ apply to both positive and negative angles.

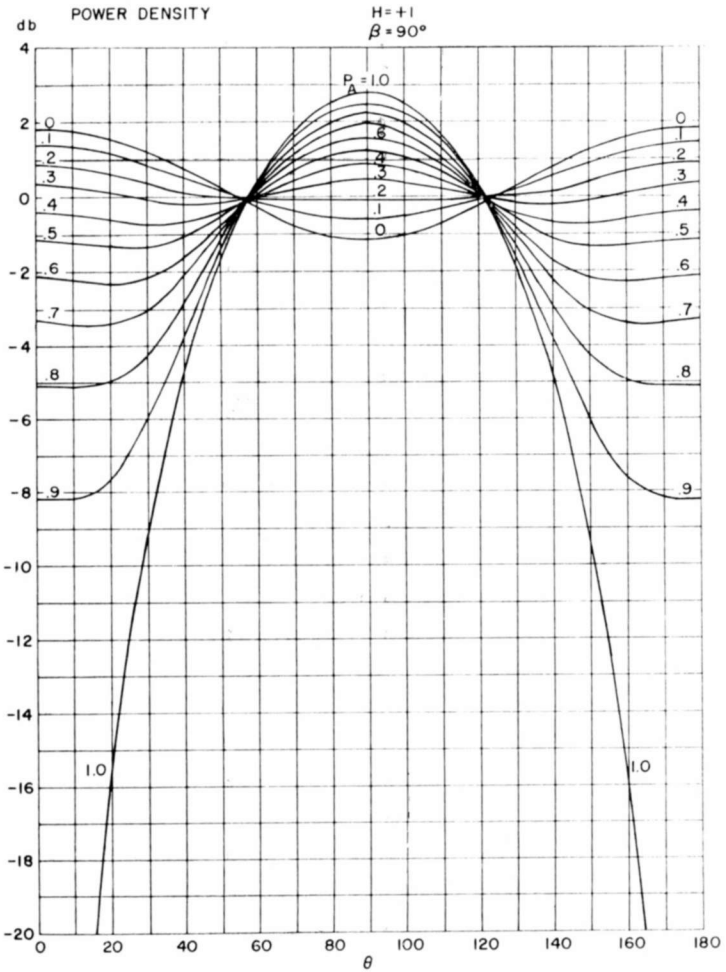


Fig. 21—Power density (in db relative to an isotropic source) plotted against the elevation angle, θ, for the $H = +1$ mode combinations of the axial and tangential types with $\beta = 90°$ and P_A variable.

For $90° < |\beta| \leq 180°$, the θ-scale must be inverted on all charts and the designated senses of circular polarization must be interchanged. This is apparent from an inspection of Equations (3) and (4).

As previously described, the radial and tangential modes are identical in magnitude but are in phase quadrature (see Table II). Hence, if an $H = +1$ mode combination of axial and radial types is employed,

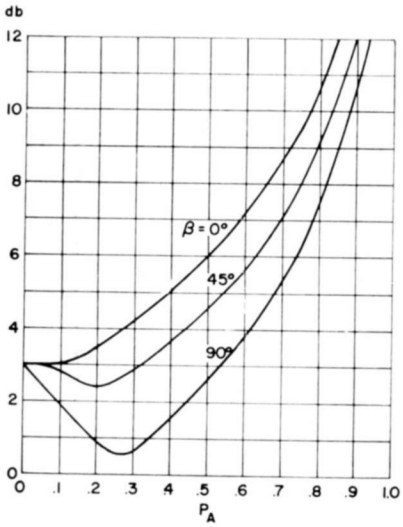


Fig. 22—Maximum variation in power density as a function of P_A and β for the $H = 1$ mode combinations of the axial and tangential types.

90° must be added to the phase angle between the axial and radial types to obtain the β-values given in the charts.

If the progressive phasing for the rings is changed from $H = +1$ (counterclockwise) to $H = -1$ (clockwise), it is apparent that the antenna is in effect merely turned upside down. Thus the radiation characteristics are unchanged and only the θ-scale need be inverted on all charts.

The key drawing of Figure 28 provides a graphical summary of the terminology and definitions to aid in using the charts for both $H = 0$ mode and $H = 1$ mode combinations.

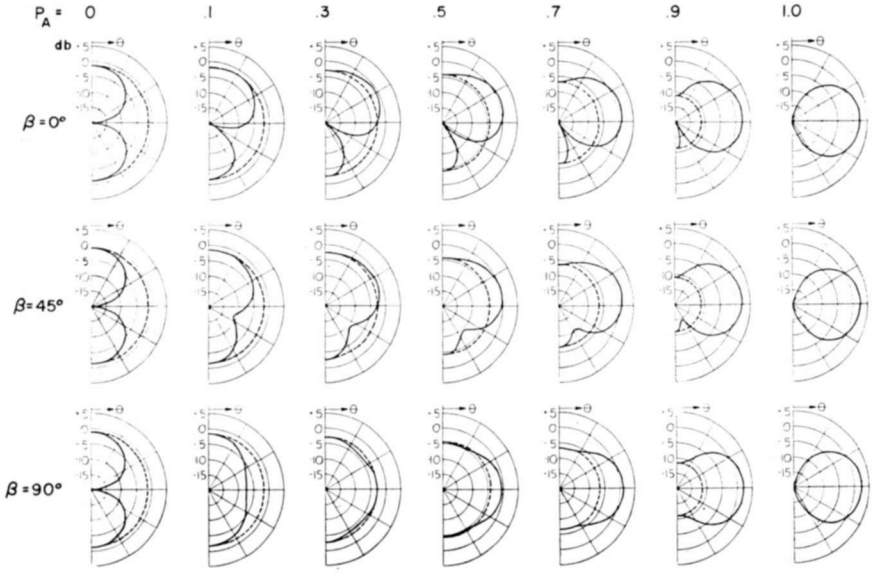


Fig. 23—Polar elevation patterns (in db relative to a linearly polarized isotropic source) of the F_θ (solid lines) and F_ϕ (dashed lines) components for the $H = +1$ mode combinations of the axial and tangential types.

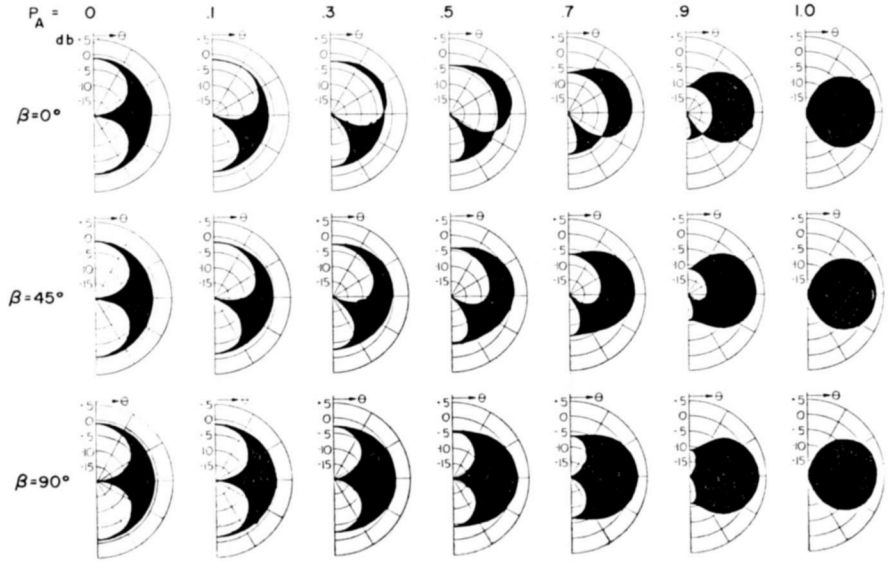


Fig. 24—Polar elevation patterns (in db relative to a linearly polarized isotropic source) of the F_{max} and F_{min} components for the $H = +1$ mode combinations of the axial and tangential types. The shaded areas represent the region in which the magnitude varies from maximum to minimum.

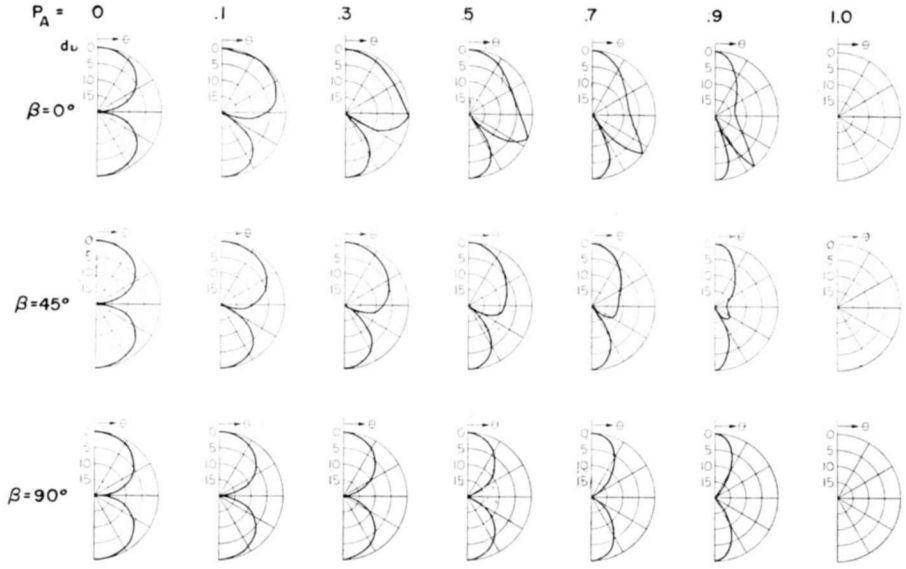


Fig. 25—The axial ratio characteristics, F_{max}/F_{min}, (in db) plotted against the elevation angle, θ, for the $H = +1$ mode combinations of the axial and tangential types.

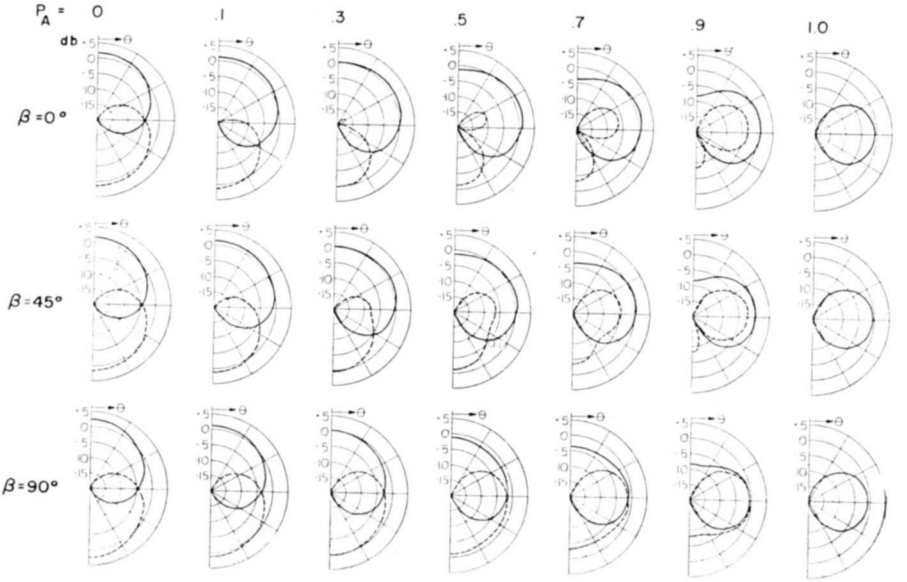


Fig. 26—Polar elevation patterns of the right-hand (dashed lines) and left-hand (solid lines) circularly polarized components for the $H = +1$ mode combination of the axial and tangential types. The magnitudes (in db) are relative to a circularly polarized isotropic source.

Conclusions

The radiation pattern from a quasi-isotropic antenna may be synthesized by the superposition of circularly symmetric radiation modes emanating from idealized ring arrays. Each mode has an omnidirectional pattern in one major plane with an unique far-field phase progression in that plane. The mode pattern in the orthogonal major plane is a function of the element orientations in the ring array and the excitation of the elements around the ring.

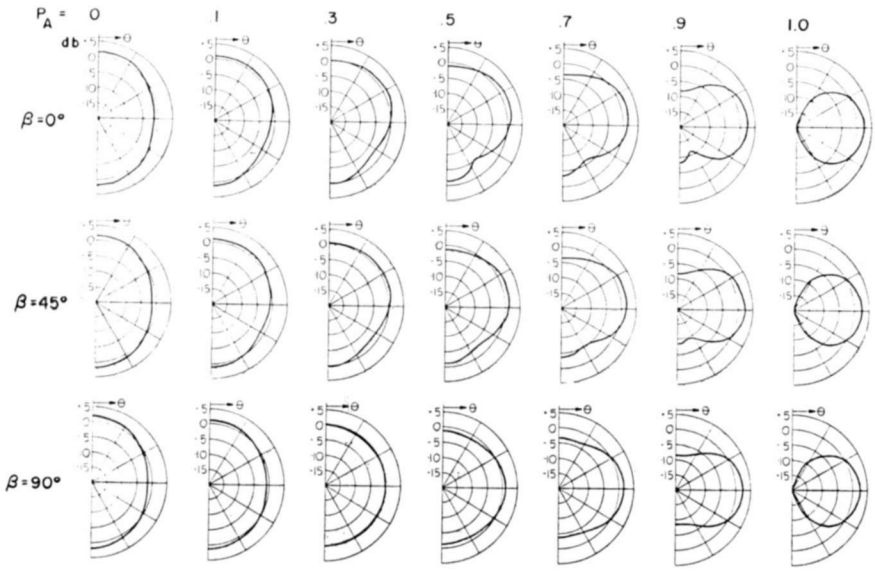
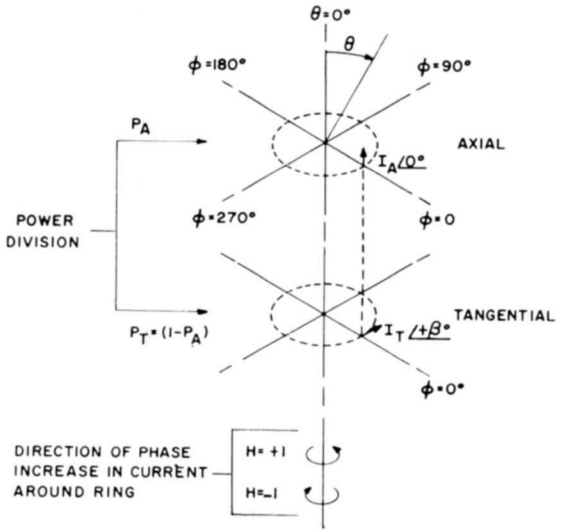


Fig. 27—Polar elevation patterns of the power density (in db relative to an isotropic source) for the $H = +1$ mode combinations of the axial and tangential types.

The number of mode components forming the radiation pattern from a practical antenna of small electrical dimensions is limited by supergain effects. Further restrictions are placed on the mode components if the antenna pattern is omnidirectional in one major plane. A simple method is described for determining all possible radiation patterns from quasi-isotropic antennas of this kind. The calculated data is presented in a number of graphs in terms of the various radiation parameters. These graphs are useful in the analysis and design of quasi-isotropic radiators, such as spacecraft antennas, which require specific amplitude and polarization characteristics over the radiation sphere.



β = phase of current in tangential element at any azimuth angle with respect to current in axial element at the same azimuth angle.

For $|\beta| \leq 90°$, use charts as labeled for $H = 0$ and $H = +1$.

For $90° < |\beta| \leq 180°$, interchange senses of circular polarization for $H = 0$ and $H = +1$, and invert θ scale for $H = +1$.

Invert θ scale for $H = -1$.

Fig. 28—Key showing a graphical summary of the terminology and definitions employed.

Acknowledgment

The author is indebted to W. C. Wilkinson,* J. Epstein,* J. B. Rankin,** and R. W. Klopfenstein*** for helpful comments and suggestions during many discussions on this work, and to Mrs. V. R. LeBoutillier who made all of the graphs and drawings.

* Antenna Skill Center, RCA Missile and Surface Radar Division, Moorestown, N. J.

** Formerly at RCA Missile and Surface Radar Division, Moorestown, N. J.; now at M.I.T. Lincoln Laboratory, Lexington, Mass.

*** RCA Laboratories, Princeton, N. J.

APPLICATION OF SOME LINEAR FM RESULTS TO FREQUENCY-DIVERSITY WAVEFORMS

By

Thomas B. Howard

RCA Missile and Surface Radar Division,
Moorestown, N. J.

Summary—*The resolution characteristics for certain simple pulse-burst, frequency-diversity waveforms are determined by finding the response of the matched filter to the doppler-shifted signal spectrum. Considered as a function of both time and doppler frequency, the complex filter response is essentially Woodward's autocorrelation function for a combined time and frequency shift. Its squared magnitude, the signal ambiguity function, provides a theoretical measure of the ability of the radar to resolve targets in range and radial velocity. It is concluded that a waveform possesses fairly good resolution properties when the transmission times for its subcarriers are specified by a parabolic curve in frequency. The repetition period of the resultant pulse train is staggered and the pulse train has pulse-to-pulse frequency shifting. The waveform power spectrum is tapered in order to improve the zero-doppler response. On the basis of the spectrum's taper, the curve for subcarrier timing is then modified by the principle of stationary phase to obtain limited control of ambiguity in the presence of large doppler shifts. It is believed that the approach and results are useful in furnishing physical insight and in illustrating waveform performance and limitations.*

Introduction

THE TRANSMISSION of successive pulses of different frequencies, called frequency diversity, can be used with high-energy pulse-burst waveforms to extend radar signal bandwidths beyond 100 mc. The matched receiver of the frequency-diversity radar achieves large bandwidth conveniently by employing several frequency channels in parallel as shown in Figure 1. Each receiver channel incorporates a band-pass filter, matched to a transmitted subpulse at one of the radar subcarrier frequencies, followed by a decoding delay line. The channel outputs are amplitude weighted, decoded in phase, and added coherently at i-f or r-f to form the matched-receiver output. For waveforms comprising repeating subcarriers there will be some duplication of the functional elements between the band-pass filter bank and the summing bus shown in Figure 1.

This paper is devoted to the problem of timing the subcarriers of the transmitted waveform so that high resolution is achieved in range and doppler simultaneously. Theoretical performance in range and

doppler for a number of simple waveforms is described together with a convenient method of analysis. Since complex waveforms can often be considered as the combination of several simple waveforms, the analytical results can be applied by superposition to more complicated timing arrangements.

When the transmitted waveform is doppler shifted by ϕ cycles per second and applied at the input of a filter matched to the non-doppler-shifted transmitted waveform, the squared magnitude of the resultant time response is essentially the Woodward ambiguity function,[1,2]

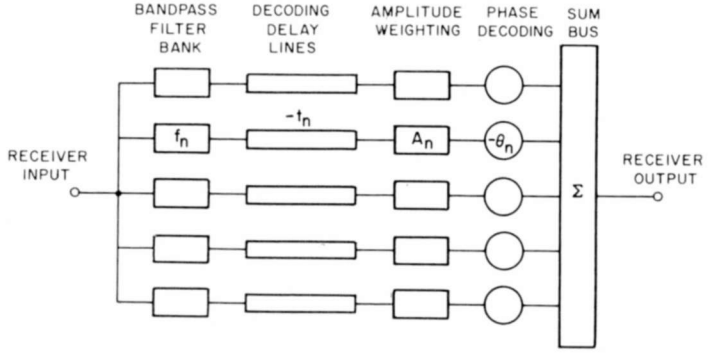


Fig. 1—Matched receiver for pulse-burst frequency-diversity waveform.

$|\chi(t,\phi)|^2$. The resolution characteristics of the waveform can be indicated graphically by a three-dimensional diagram of either $|\chi|$ or $|\chi|^2$ plotted vertically on a base plane with time as one axis and doppler as the second as shown in Figure 2. If the ambiguity surface has a single sharp central peak at the origin ($t = \phi = 0$) and yet maintains a relatively low amplitude elsewhere in a region of interest, the waveform permits high simultaneous resolution in range and doppler. The waveform corresponding to the $|\chi|$ plot of Figure 2 has this desirable property if interest is confined to a central doppler strip along the t-axis. In this study, the amplitude of the matched-filter response considered as a function of both time, t, and doppler, ϕ, is presented as the measure of waveform performance.

---

[1] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, pp. 120, 27, 101, and 119, McGraw-Hill Book Co., Inc., New York, N. Y., 1953.

[2] J. R. Klauder, "Design of Radar Signals Having Both High Range and High Velocity Resolution," *Bell Sys. Tech. Jour.*, Vol. 39, p. 809, July 1960.

Waveform Description

As indicated in Figure 3(a), the frequency-diversity waveform considered in this analysis consists of N phase-related subpulses, one at each subcarrier frequency, where N is taken as an odd integer for convenience. Subpulses of amplitude A_n and width τ, occur at times t_n with a total duration for the pulse sequence of T seconds. Figure

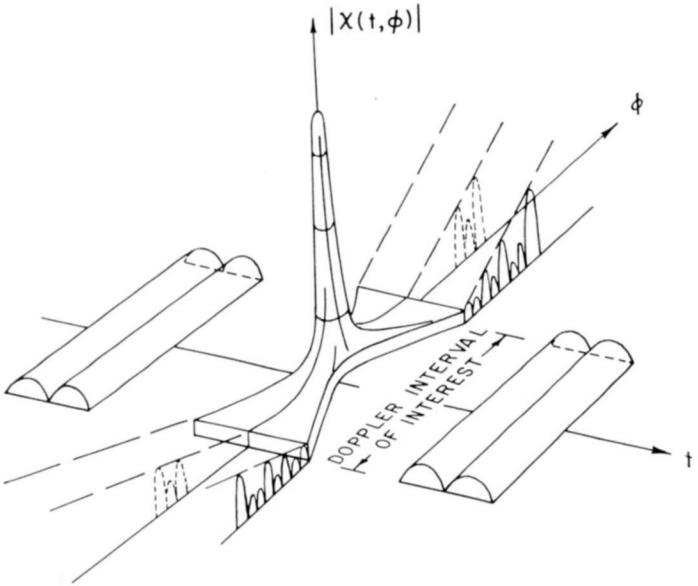


Fig. 2—$|\chi|$ surface that is well behaved along t-axis.

3(b) shows the subcarrier frequencies equally spaced Δ cycles per second apart, where Δ is made equal to the reciprocal of the pulse width τ in order to eliminate gaps in the waveform's power spectrum. The nth subcarrier frequency, f_n, is offset by $n\Delta$ cycles per second from the central high-frequency carrier, f_0, so that the total bandwidth, W, bracketing N subcarriers is $N\Delta$. The subcarrier timing function, $T_D(f)$, illustrated in Figure 3(c) is selected so that it is single-valued over the frequency range $(-W/2, W/2)$ and extends over a time interval of length T. In addition to amplitude weighting and subcarrier timing, subcarrier phase coding angles θ_n are included to provide more generality to the results.

In this treatment, three continuous functions $A(f)$, $T_D(f)$, and $\theta(f)$ are employed to describe the waveform. When they are sampled at discrete frequencies $n\Delta$, these functions specify, respectively, sub-

carrier amplitude weighting, subcarrier timing, and phase coding as denoted by

$$A(n\Delta) = A_n; \quad T_D(n\Delta) = t_n; \quad \theta(n\Delta) = \theta_n. \tag{1}$$

$A(f)$ is an even, real low-frequency function used to provide spectral

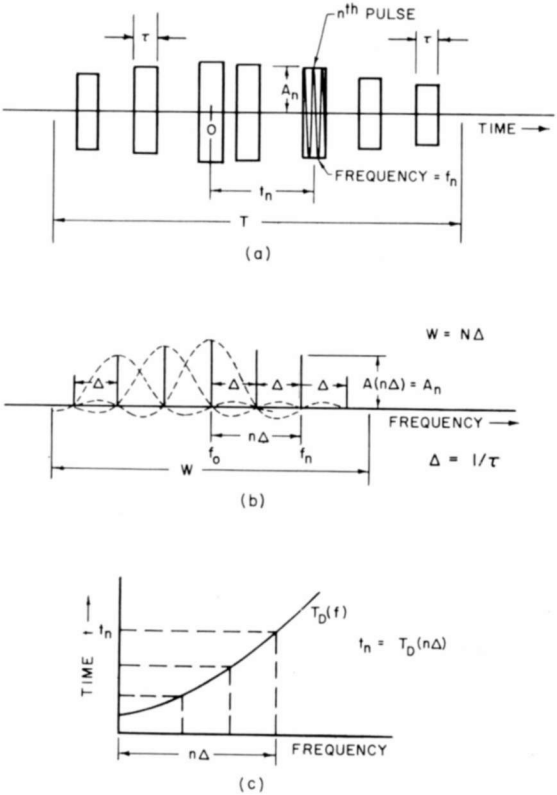


Fig. 3—Subcarrier amplitude weighting and timing.

taper and truncation and is zero everywhere except in the frequency range $(-W/2, W/2)$.

The composite transmitted waveform is now represented mathematically by $\psi_T(t)$, a complex high-frequency function;[1]

$$\psi_T(t) = \sum_{n=-\infty}^{\infty} A_n p(t - t_n) \exp\{j[2\pi f_n(t - t_n) + \theta_n]\}, \tag{2}$$

where $f_n = f_0 + n\Delta$. $p(t)$ specifies the subpulse envelope shape, taken

as rectangular and of width τ and defined by

$$p(t) = \text{rect}\,\frac{t}{\tau} = \begin{cases} 1 & |t| < \dfrac{1}{2}\tau \\ 0 & |t| > \dfrac{1}{2}\tau. \end{cases} \tag{3}$$

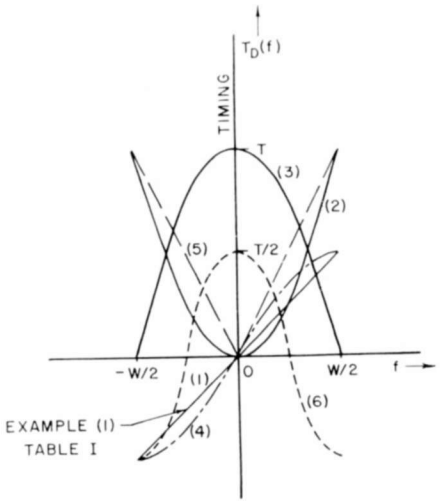


Fig. 4—Subcarrier timing curves, $T_D(f)$.

Since N is odd, the definition of $A(f)$ and the selection of W make $A_n = 0$ when $|n| > (N-1)/2$. Consequently, all terms of the infinite sum indicated are zero except for the N terms in the interval

$$-\frac{N-1}{2} \leq n \leq \frac{N-1}{2}.$$

To study the effect of subcarrier timing, the timing curves shown in Figure 4 are considered in the calculation of the matched filter response. Over the frequency interval $(-W/2, W/2)$, the timing curves are

(1) linear,
(2) parabolic,
(3) half-cycle cosine,
(4) half-cycle sine,
(5) linear slopes of opposite sign,
(6) full-cycle cosine.

These specific shapes were chosen because they permit simple determination of central ambiguity without use of a computer. In a practical case when the duty cycle for the pulse train is low, pulse positions can be time-shifted slightly off the curves to prevent time overlap, if this is necessary to achieve high transmitter efficiency.

Matched-Filter Response

To establish a performance criterion for the various waveforms considered, a general expression for the matched-filter response is found as a function of time and doppler frequency, ϕ. If the spectrum of the transmitted waveform is designated as $\Psi_T(f)$, the corresponding matched filter is defined by $\Psi_T^*(f)$. The output response $\psi_0(t,\phi)$ of the matched filter to the doppler-shifted transmitted waveform, with spectrum $\Psi_T(f-\phi)$, is then given by the inverse Fourier transform;

$$\psi_0(t,\phi) = \int_{-\infty}^{\infty} \Psi_T(f-\phi)\,\Psi_T^*(f)\,\exp\,(j2\pi ft)\,df. \tag{4}$$

Since the impulse response of the matched filter with the frequency function $\Psi_T^*(f)$ is $\Psi_T^*(-t)$ and the doppler-shifted waveform in the time domain is $\psi_T(t)\exp(j2\pi\phi t)$, an equivalent expression for the output function is

$$\psi_0(t,\phi) = \{\psi_T(t)\,\exp\,(j2\pi\phi t)\} \bigstar \{\psi_T^*(-t)\}, \tag{5}$$

where $\bigstar$ denotes convolution. When the complex high-frequency function given by Equation (2) is used in Equation (5), the convolution results in the following double summation

$$\psi_0(t,\phi) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \exp j\{\pi[f_n + f_m + \phi]\,[t-(t_n - t_m)]$$
$$+ 2\pi\phi t_n + \theta_n - \theta_m\}\,A_n A_m \rho[t-(t_n-t_m),\,(n-m)\Delta + \phi], \tag{6}$$

where

$$\rho(t,\phi) \equiv \int_{-\infty}^{\infty} p\left(x+\frac{t}{2}\right) p^*\left(x-\frac{t}{2}\right) \exp\,(j2\pi\phi x)\,dx, \tag{7}$$

and where the A's, θ's and t's are as defined in Equation (1).

In Equation (6), because of amplitude truncation, all but N^2 terms

are zero. It is noted that $\rho(t,\phi)$ as defined by Equation (7) can be used to specify the ambiguity diagram for each subpulse considered separately. Since it is expressed in the convenient symmetrical form used in Reference (2), $\rho(t,\phi)$ is real whenever $p(t)$ is selected as an even time function.

For purposes of investigating the detailed behavior of ambiguity in a central region of the $t-\phi$ plane, it is assumed that the relative doppler shift, ϕ, remains a small fraction of the nominal bandwidth ($\tau^{-1}=\Delta$) of each matched receiver channel. Under these conditions, the spectrum of each subpulse of the received waveform always remains approximately centered in its matching channel. Central ambiguity depends primarily on the N component responses found by pairing each receiver channel with the subpulse that it matches except for the slight doppler offset. Because of the decoding delay lines, the N contributing responses have time-coincident envelopes centered at $t=0$. The central response of the matched receiver, which is defined as the superposition of these N component responses, is formed to examine the behavior of central ambiguity. In the nomenclature of Rihaczek,[3] the immediate discussion is concerned with determination of "proximal" resolution properties as indicated by the shape of the "central surface." These properties are investigated by a study of the central response of the matched receiver.

An expression for the central response is found from Equation (6) by removing the high-frequency carrier and then summing low-frequency terms corresponding to $m=n$. These are the N terms along the principal diagonal of the representative matrix of the N^2 elemental responses. Under the assumptions made in this paper, the remaining $N(N-1)$ terms of the summation comprise the side responses, which contribute to secondary ambiguities; these are discussed later. If the carrier in the presence of doppler shift is taken[2] as $[f_0+(\phi/2)]$, the low-frequency version of the central response is simply

$$u_0(t,\phi)=\rho(t,\phi)\sum_{n=-\infty}^{\infty}A_n^2\exp(j2\pi\phi t_n)\exp(j2\pi n\Delta t), \tag{8}$$

where use is made of the fact that $f_n=(f_0+n\Delta)$. For rectangular subpulses, $p(t)$, Equation (7) defining $\rho(t,\phi)$ becomes

$$\rho(t,\phi)=\operatorname{rect}\left(\frac{t}{2\tau}\right)\int_{-x_0}^{x_0}\exp(j2\pi\phi x)\,dx,$$

[3] A. W. Rihaczek, "Radar Resolution Properties of Pulse Trains," *Proc. IEEE*, Vol. 52, p. 153, Feb. 1964.

where $x_0 = (\tau - |t|)/2$;

$$\rho(t,\phi) = \text{rect}\left(\frac{t}{2\tau}\right)\frac{\sin \pi\phi(\tau - |t|)}{\pi\phi}. \tag{9}$$

The symmetrical limits of integration result from the truncating effects of the product of the two displaced rectangular functions in x considered for both positive and negative time.

To permit application of published results pertaining to continuous types of waveforms, the complex Fourier series in Equation (8) is next replaced[4] by an equivalent periodic function. The new periodic function consists of the superposition of identical complex time functions that repeat in an infinite train with period Δ^{-1} seconds. By this substitution, concepts and analytical techniques described in References (5), (6), and (7) on continuous waveforms become applicable to each function in the train. Since $\tau = \Delta^{-1}$, it can be seen from Equation (9) that $\rho(t,\phi)$ never exceeds a width of $2/\Delta$. Consequently, because of the presence of $\rho(t,\phi)$ as a factor in Equation (8), the central response is confined to the time interval $(-\Delta^{-1}, \Delta^{-1})$. Preparatory to replacing the Fourier summation of Equation (8) by an explicit time train, the function $G(f,\phi)$ is introduced as

$$G(f,\phi) \equiv A^2(f) \exp\{j2\pi\phi T_D(f)\}, \tag{10}$$

with inverse transform,

$$g(t,\phi) = \int_{-\infty}^{\infty} G(f,\phi) \exp(j2\pi ft)\, df. \tag{11}$$

By use of Equations (1) and (10), Equation (8) can be rewritten in the alternate form of the Fourier Series

---

[4] J. L. Allen, "Phased Array Studies," M.I.T., TR No. 228, AD 249470, Aug. 1960. (See p. 166 for analogy. The finite Fourier series for the array function is given also by the superposition of line source functions in an infinite train.)

[5] J. R. Klauder, A. C. Price, S. Darlington, and W. J. Albersheim, "The Theory and Design of Chirp Radars," *Bell Sys. Tech. Jour.*, Vol. 39, p. 745, July 1960.

[6] C. E. Cook, "Pulse Compression—Key to More Efficient Radar Transmission," *Proc. I.R.E.*, Vol. 48, p. 310, March 1960.

[7] E. L. Key, E. N. Fowle, and R. D. Haggarty, "A Method of Design Signals of Large Time–Bandwidth Product," *I.R.E. Conv. Record*, Part 4, p. 146, March 1961.

$$u_0(t,\phi) = \rho(t,\phi) \sum_{n=-\infty}^{\infty} A^2(n\Delta) \exp\left[j2\pi\phi T_D(n\Delta)\right] \exp\left(j2\pi n\Delta t\right)$$
$$= \rho(t,\phi) \sum_{n=-\infty}^{\infty} G(n\Delta,\phi) \exp\left(j2\pi n\Delta t\right). \tag{12}$$

The desired equivalent form for the low-frequency version of the central response is obtained by application of Poisson's sum formula[8] which states that if $v(t)$ and $V(f)$ are a Fourier transform pair,[1] then

$$\sum_{n=-\infty}^{\infty} V(n\Delta) \exp\left(j2\pi n\Delta t\right) = \frac{1}{\Delta} \sum_{k=-\infty}^{\infty} v\left(t - \frac{k}{\Delta}\right). \tag{13}$$

Using Equation (13), the central response, Equation (12), becomes

$$u_0(t,\phi) = \frac{1}{\Delta} \rho(t,\phi) \sum_{k=-\infty}^{\infty} g\left(t - \frac{k}{\Delta}, \phi\right), \tag{14}$$

where the summation on the right is the expression for the repeating train.

In what follows, it is assumed that $T \gg \tau$, and use is made of the previous assumption that $|\phi| \ll 1/\tau = \Delta$. The repeating train represented by the summation now contains essentially all the doppler information since, with the latter assumption,

$$\rho(t,\phi) \approx \rho(t,0) = \text{rect}\left(\frac{t}{2\tau}\right)(\tau - |t|).$$

Because $A(f)$ is truncated, the individual time functions in the train must overlap. But when N is large and $|\phi|$ is not excessive, the degree of overlap is small, even between adjacent time functions. The behavior of the central response can, therefore, be investigated by examning the behavior of any one of the individual functions, for instance, $g(t,\phi)$ itself, found by taking the inverse transform of Equation (10) for each waveform example. Hence, $g(t,\phi)$ serves as a pilot function useful in the selection and control of $T_D(f)$ to shape central ambiguity.

It is noted that the behavior of the central response as represented by Equations (8) and (14) does not depend on the phase coding as specified by $\theta(f)$. However, phase coding can affect the side responses

[8] A. Papoulis, *The Fourier Integral and Its Applications*, p. 47, McGraw-Hill Book Co., Inc., New York, N. Y., June 1962.

and, if altered on a waveform-to-waveform basis, can also affect the central response for second-time-around echoes.

Examples of Central Response

The case of equal subcarrier amplitude weighting is studied by taking

$$A(f) = A^2(f) = \text{rect}\left(\frac{f}{W}\right) = \begin{cases} 1, & |f| < \dfrac{W}{2} \\ 0, & |f| > \dfrac{W}{2}. \end{cases} \tag{15}$$

Applying the inversion formula given by Equation (11), $g(t,\phi)$ is found as

$$g(t,\phi) = \int_{-\infty}^{\infty} \text{rect}\left(\frac{f}{W}\right) \exp\{j2\pi[\phi T_D(f) + ft]\}df. \tag{16}$$

Now for all subcarrier timing curves, $T_D(f)$,

$$g(t,0) = W \text{ sinc } Wt, \tag{17}$$

where $\text{sinc } x = (\sin \pi x)/(\pi x)$. The function to be examined, $g(t,\phi)/W$, is listed in Table I for the six subcarrier timing curves shown in Figure 4 in the equal-amplitude-weighting case.

Relation of Timing Examples (1) and (2) to Linear F-M Results

Table I shows that $g(t,\phi)$ of example (1) exhibits ambiguity in t and ϕ similar to that possessed by a linear FM pulse, i.e., doppler produces a time shift in the matched-filter response giving rise to the familiar[1] correlation between time and frequency as shown by the relation

$$\frac{1}{W} g(t,\phi) = \frac{1}{W} g\left(t + \phi \frac{T}{W}, 0\right). \tag{18}$$

It is the purpose of this discussion, however, to emphasize relationships existing between example (2), corresponding to a parabolic subcarrier timing curve, and certain other linear FM results.

Table I

Example	$T_D(f)$	$\frac{1}{W} g(t,\phi)$
(1) (linear)	$\frac{T}{W} f$	$\operatorname{sinc} W\left(t + \frac{\phi}{W} T\right)$
(2) (parabolic)	$\frac{4T}{W^2} f^2$	$\frac{1}{4\sqrt{T\phi}} [Z(V_2) - Z(V_1)] \exp\left\{\frac{-j\pi (Wt)^2}{8T\phi}\right\}$ where $V_2, V_1 = \frac{W}{2\sqrt{T\phi}}\left(t \pm \frac{4T\phi}{W}\right)$
(3) (half-cycle cosine)	$T \cos \pi \frac{f}{W}$	$\sum_{m=-\infty}^{\infty} J_m(2\pi T\phi) \operatorname{sinc} W\left(t + \frac{m}{2W}\right) \exp\left(\frac{jm\pi}{2}\right)$
(4) (half-cycle sine)	$\frac{T}{2} \sin \pi \frac{f}{W}$	$\sum_{m=-\infty}^{\infty} J_m(\pi T\phi) \operatorname{sinc} W\left(t + \frac{m}{2W}\right)$
(5) (linear slopes of opposite sign)	$\frac{2T}{W} \lvert f \rvert$	$\frac{1}{2} \operatorname{sinc} \frac{W}{2} (t+\alpha) \exp\left\{j \frac{\pi W}{2} (t+\alpha)\right\} + \frac{1}{2} \operatorname{sinc} \frac{W}{2} (t-\alpha) \exp\left\{-j \frac{\pi W}{2} (t-\alpha)\right\}$ where $\alpha = \frac{2T\phi}{W}$
(6) (full-cycle cosine)	$\frac{T}{2} \cos 2\pi \frac{f}{W}$	$\sum_{m=-\infty}^{\infty} \exp\left(\frac{jm\pi}{2}\right) J_m(\pi T\phi) \operatorname{sinc} W\left(t + \frac{m}{W}\right)$

$Z(V) = C(V) + jS(V)$ is the complex Fresnel integral[5,9] and $J_m(\theta)$ is a Bessel function[9] of the first kind.

[9] E. Jahnke and F. Emde, *Table of Functions*, Dover Publications, Inc., New York, N. Y., 1945.

By use of Equation (11), $g(t,\phi)$ can be expressed as

$$g(t,\phi) = \int_{-W/2}^{W/2} \exp\left[j\beta(f)\right] \exp\left(j2\pi ft\right) df, \tag{19}$$

where

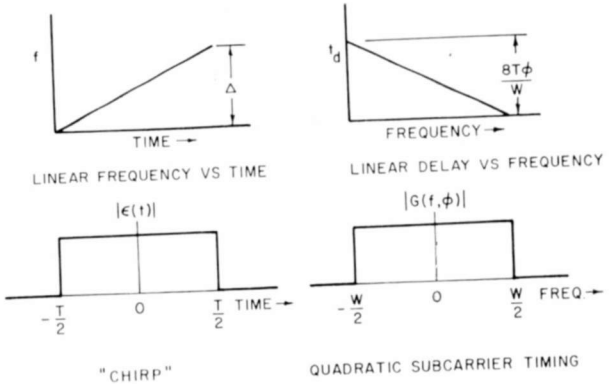
$$\beta(f) = 2\pi\phi T_D(f) = 2\pi\phi\left[\frac{4T}{W^2} f^2\right].$$

Since the envelope delay associated with the phase function $\beta(f)$ is given by

$$t_d = -\frac{1}{2\pi}\frac{d\beta}{df} = -\frac{8T\phi f}{W^2}, \tag{20}$$

the response function, Equation (19), can be considered to represent the impulse response of a rectangular band-pass filter of bandwidth W possessing an envelope delay that changes linearly by $8T\phi/W$ across the band of width W. Figure 5 shows the analogy between the quadratic subcarrier timing characteristic and the "chirp" of Reference (5). By use of the relationships in Figure 5, $|g(t,\phi)|$ can be plotted by reinterpreting curves of the type presented in References (5) and (6). Thus, $|g(t,\phi)|$ corresponds to the spectral amplitude curves for a rectangular pulse of linear FM presented in the above references. (Reference (5) also explains a time-domain reinterpretation of its spectral curves in discussing passive generation of the long f-m transmitter signal.)

The central response as given by Equations (8) and (14) is normalized[1] by dividing by $2E = N\tau = W\tau/\Delta$ (see Appendix I). Figures 6 and 7 show the corresponding normalized functions $|g(t,\phi)|/W$ and $|\rho(t,\phi)|/\tau$ for specific values of $|\phi|$ when $|\phi| \ll \tau^{-1} = \Delta$. The figures reveal the approximate behavior of the magnitude of the central response when $|\phi| < N/8T$. The inequality $|\phi| < N/8T$ defines a central doppler strip along the time axis. Within this strip, as can be seen from the figures, little time overlap exists between the adjacent functions of the repeating train in Equation (14). The magnitude of the sum of individual time functions is therefore closely approximated by the sum of their individual magnitudes. The plots illustrate that the choice of τ equal to Δ^{-1} ensures that $\rho(t,\phi)$ [$\simeq \rho(t,0)$] will in large measure suppress undesirable repetitions of the summation in Equation (14) when the indicated product is formed. Central ambiguity as measured by the absolute value of the product exhibits



$$\epsilon(t) = \text{rect}\left(\frac{t}{T}\right) \exp[j\psi(t)]$$

$$\psi(t) = 2\pi \int f\, dt$$

$$G(f,\phi) = \text{rect}\left(\frac{f}{W}\right) \exp[j\beta(f)]$$

$$\beta(f) = -2\pi \int t_d\, df$$

KLAUDER'S NOTATION	PRESENT NOTATION
Δ	$\frac{8T\phi}{W}$
T	W
t	f
$f_0 - f$	t
$k = \frac{\Delta}{T}$	$\frac{8T\phi}{W^2}$

Fig. 5—Analogy between "chirp" and quadratic subcarrier timing.

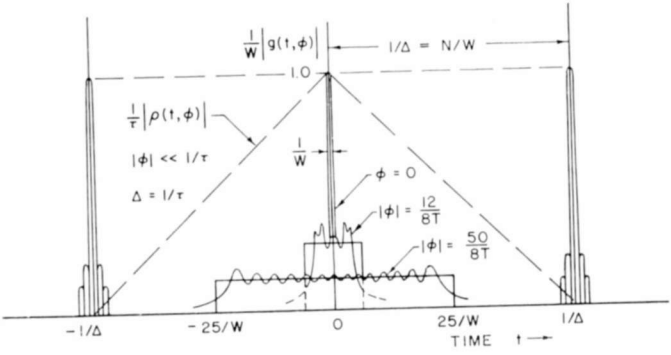


Fig. 6—Central ambiguity for waveform with quadratic subcarrier timing, $N = 51$.

desirable behavior in the sense that it is highly peaked near $t = \phi = 0$ and yet maintains a relatively low level elsewhere as long as $|\phi| < N/8T$. As shown, when $|\phi| > 1/T$, the level and width (in the time dimension) of $|g(t,\phi)|/W$ are respectively about $1/\sqrt{8T|\phi|}$ and $D = 8T|\phi|/W$. At $|\phi| = N/8T$, $|g(t,\phi)|/W$ reaches a level of $1/\sqrt{N}$ and a width of $N/W = \Delta^{-1} = \tau$. Figure 8 shows $|g(t,\phi)|/W$ more accurately for three doppler shifts.

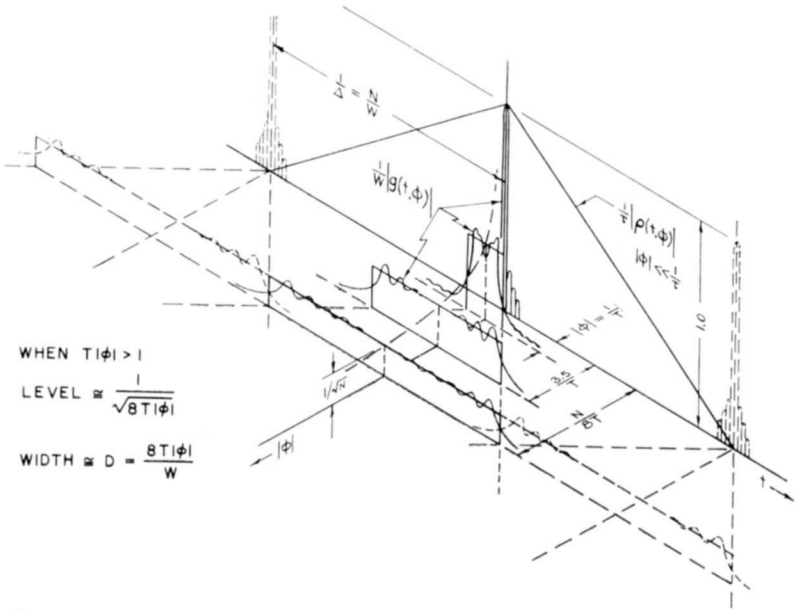


Fig. 7—Central ambiguity for waveform with quadratic subcarrier timing, $N = 51$.

When $|\phi| > N/8T$, overlapping and interaction of the superposed functions must be taken into account. As discussed later, performance now slowly deteriorates as the envelope peaks of the resultant function build up in a complex manner with increasing doppler. Figure 9 illustrates the behavior of the central response in the doppler dimension with the envelope deterioration caused by overlapping. For each doppler, the figure shows the approximate maximum in a vertical cut parallel to the time axis.

The superposed functions of the train in Equation (14) are analogous to the aliased spectrum[10] in equispaced time-sampling analysis,

[10] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*, p. 117, Dover Publications, Inc., New York, N. Y., 1958.

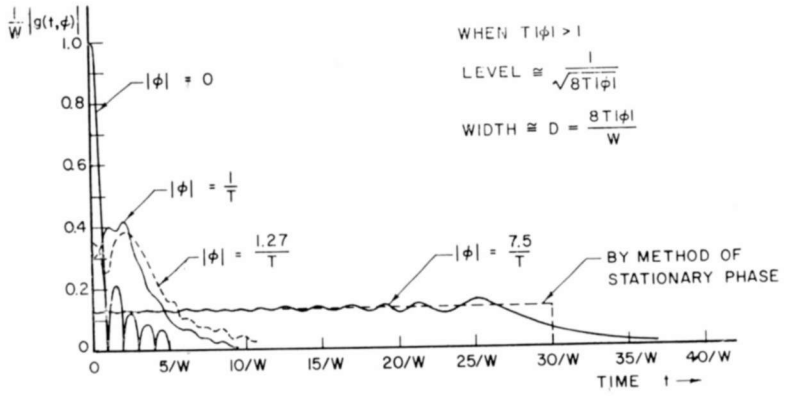


Fig. 8—Central ambiguity for waveform with quadratic subcarrier timing, $T_D(f) = (4T/W^2)f^2$.

except that here sampling is in the frequency domain. To prevent overlapping, the sampling theorem requires sample spacing, Δ, to be less than $1/D$, where D is the duration of $g(t,\phi)$.

Minimum Level of Central Response

The N triangular pulses of subcarrier in the real high-frequency time function corresponding to the low-frequency complex function, Equation (8), are almost orthogonal as indicated by a small correlation coefficient (see Appendix II). Thus the energy in the central response tends to remain fairly constant and independent of the relative phases of its N component pulses. The maximum factor of variation possible for the total energy in Equation (8) is less than two, and approaches two as N becomes very large. Consequently, amplitude peaks in time

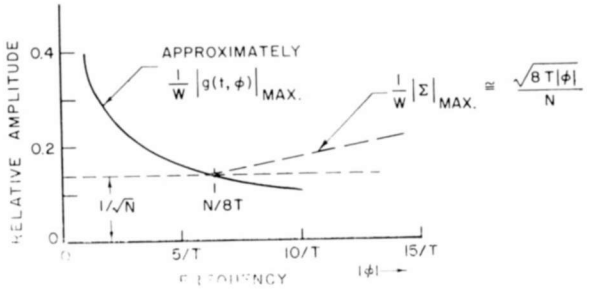


Fig. 9—Effects of overlapping on envelope of central response.

are minimized by equal distribution of energy over the width $2\tau = 2N/W$ of the time-coincident pulses. This amounts to spreading the zero-doppler energy by a factor of about $2N$ in time to achieve a maximum amplitude reduction ratio of about $1/\sqrt{2N}$. The foregoing reasoning illustrates that by adjustment of a single-valued $T_D(f)$, it is impossible to achieve a voltage level lower than about $1/\sqrt{2N}$ in the normalized central response, if the level must be maintained for all points in time. Thus, at $|\phi| = N/8T$, the waveform of example (2), by maintaining a level of $1/\sqrt{N}$, begins to approach the theoretical limit. A much lower level for any doppler $|\phi| \ll \Delta$ would appear to require more subpulse spectral overlap which is achieved by making $\tau < \Delta^{-1}$ at the cost of increased side response. It could, of course, also be achieved for particular dopplers by employing each subcarrier more than one time. Either of these methods can make the amount of energy in the central response more sensitive to doppler shift.

Overlapping in Central Response

As mentioned above, for the timing of example (2) considered in the absence of spectral taper, overlapping and interaction of the summed functions must be taken into account when $|\phi| > N/8T$. As overlapping increases with increasing doppler, performance deteriorates, since, although the magnitude of individual functions varies inversely with the square root of doppler, the number of significant contributing functions at any point in time increases directly with doppler. Envelope build-up is shown in Figure 9. The magnitude of the sum of individual functions has many closely spaced peaks and nulls in the time dimension when the number of interacting functions is small. For example at $|\phi| = N/4T$, overlapping requires consideration of two functions of amplitude $1/\sqrt{2N}$ at all points in time. These functions have a difference frequency of $W/2$ which manifests itself in the resultant as $N/2$ envelope peaks of amplitude $\sqrt{2/N}$ in the period $1/\Delta$. Although the complex behavior of the superposed functions has not been investigated in detail, some general comments can be made. The envelope peaks tend to decrease in frequency in the time dimension as the number of contributing functions increases with doppler. The normalized peaks have a maximum amplitude of about $\sqrt{8T|\phi|}/N$. At $|\phi| = N^2/8T$, the peaks reach unit amplitude, there being N contributing functions (properly phased) of amplitude $1/N$. In the time dimension these unit peaks occur at, $t = \pm 1/(2\Delta)$, $\pm 3/(2\Delta)$, etc., as is most easily seen from the Fourier series in Equation (8). Multiplication by $|\rho(t,\phi)|/\tau$ reduces the peaks at $t = \pm 1/(2\Delta)$ by about 6 db.

CENTRAL AMBIGUITY FOR TIMING EXAMPLES (3), (4), AND (6)

Figure 10 illustrates $|g(t,\phi)|/W$ corresponding to the cosine subcarrier timing curve shown in Figure 4 for example (3). In this case, the expression for $g(t,\phi)$ is quite similar to expressions encountered in paired-echo distortion analyses.[5] The real and imaginary parts of $g(t,\phi)$ are each determined by a single Bessel coefficient at appropriate stations in time separated by W^{-1}. This fact is useful in making approximate plots. The behavior for small doppler frequencies is almost the same as that for timing curve (2) of Figure 4. For large

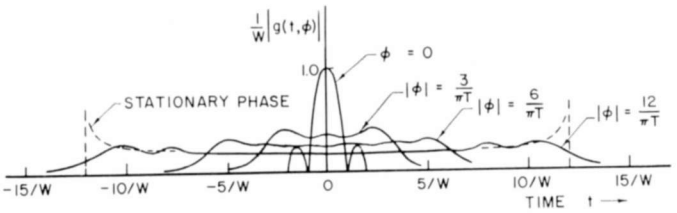


Fig. 10—Central ambiguity for timing example (3), $T_D(f) = T\cos\pi(f/W)$.

doppler shifts, the plotted results were checked by the principle of stationary phase,[7,8] which gives the approximation

$$\frac{1}{W}|g(t,\phi)| \approx \frac{1}{\pi\sqrt{T|\phi|}}\left[1-\left(\frac{Wt}{\pi T\phi}\right)^2\right]^{-\frac{1}{4}}, \qquad |\phi| \gg \frac{1}{T},\ |t| < \frac{\pi T|\phi|}{W}. \tag{21}$$

For the timing of example (4), $g(t,\phi)$ is a real function. Figure 11 shows that $g(t,\phi)/W$ exhibits ambiguity somewhat similar to $g(t,\phi)/W$ of example (1); that is, a doppler shift causes a time shift of the main lobe. However, in this case there is also a fall-off in peak amplitude accompanied by a pronounced change in shape for large doppler.

Approximate curves of $g(t,\phi)/W$ for the timing curve of example (6) are shown in Figure 12. Doppler shift creates a symmetrical envelope response with pronounced undesirable ripples in the time dimension.

TAPERING THE WAVEFORM SPECTRUM

As shown by Equation (17) for the case of equal subcarrier ampli-

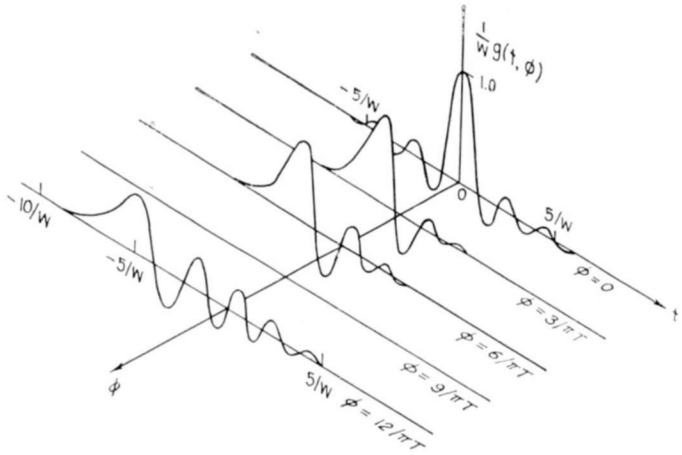


Fig. 11—Central ambiguity for timing example (4), $T_D(f) = (T/2) \sin \pi (f/W)$.

tude weighting, $g(t,0)$ is a $(\sin x)/x$ function resulting in undesirable high-level sidelobes. Referring to Figure 8, the first sidelobe shown has a normalized peak amplitude of 22% (—13.2 db). The magnitude and slow fall-off of sidelobes can severely limit range resolution capability when the dynamic range of received signals is large. In order to improve the shape of $|g(t,0)|$, a spectral taper factor is applied to the rectangular frequency function of Equation (15). For example, if

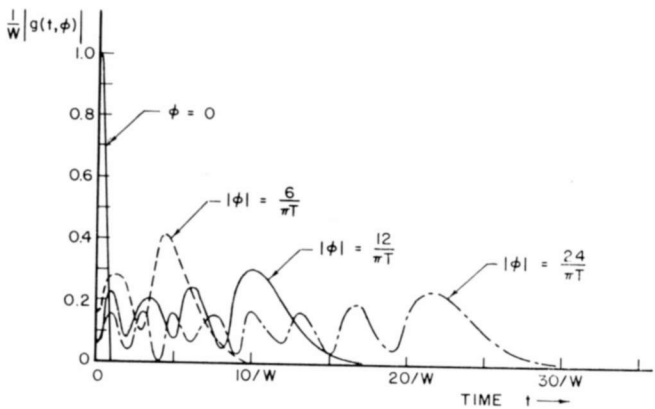


Fig. 12—Central ambiguity for timing example (6), $T_D(f) = (T/2) \cos 2\pi (f/W)$.

$$A^2(f) = \left[1 + 2F_1 \cos 2\pi \frac{f}{W} \right] \operatorname{rect}\left(\frac{f}{W} \right), \tag{22}$$

then, for all subcarrier timing curves $T_D(f)$,

$$\frac{1}{W} g(t, 0) = \operatorname{sinc} Wt + F_1 \operatorname{sinc} W\left(t - \frac{1}{W} \right) + F_1 \operatorname{sinc} W\left(t + \frac{1}{W} \right). \tag{23}$$

(With the above taper factor, $2E$ remains equal to $N\tau = W\tau/\Delta$. See Appendix I.)

Now, when $F_1 = 0.42$, all the sidelobes of $g(t, 0)$ are suppressed to a level 40 db or more below peak signal.[5] For the linear timing of example (1), doppler shift causes the same time shift as before spectral taper was applied so that, as before,

$$\frac{1}{W} g(t, \phi) = \frac{1}{W} g\left(t + \dot{\phi} \frac{1}{W}, 0 \right). \tag{18}$$

The approximate effects of doppler shift for the parabolic timing curve of example (2) are determined when $|\phi| \gg \tau^{-1}$ by employing the principle of stationary phase. For each ϕ, $\beta(f) \equiv 2\pi\phi T_D(f)$ is a dispersive phase characteristic conforming to the assumptions of Reference (7). The assumptions are essentially

(a) The group delay function, $-\frac{1}{2\pi} \frac{d\beta(f)}{df}$, has a large slope, $\left| \frac{1}{2\pi} \frac{d^2\beta(f)}{df^2} \right|$, at every frequency in the band of width W.

(b) There is only one frequency corresponding to each delay of the group delay function, $-\frac{1}{2\pi} \frac{d\beta(f)}{df}$, over the band of width W.

Under these conditions, by the principle of stationary phase,

$$|g(t,\phi)| \approx \left[\frac{A^2(f)}{\sqrt{\left| \frac{1}{2\pi} \frac{d^2\beta(f)}{df^2} \right|}} \right]_{f=\lambda} ;$$

at time

$$t = -\frac{1}{2\pi}\frac{d\beta(f)}{df}\Bigg]_{f=\lambda}. \tag{24}$$

Since $A(f)$ is even and since, for example (2), $\beta(f) = 2\pi\phi[(4T/W^2)f^2]$, Equations (24) and (22) yield

$$\frac{1}{W}|g(t,\phi)| \approx \frac{1}{\sqrt{8T|\phi|}} A^2\left(\frac{Wt}{D}\right)$$

$$\approx \frac{1}{\sqrt{8T|\phi|}}\left[1 + 2F_1 \cos 2\pi\frac{t}{D}\right] \operatorname{rect}\frac{t}{D}, \tag{25}$$

where $D = 8T|\phi|/W$ is the approximate time width of $|g(t,\phi|$ for each $|\phi| \gg T^{-1}$. Thus, for large dopplers $|g(t,\phi)|$ assumes a taper similar to that of $A^2(f)$ and has the same functional form. (This result follows more directly from an example in Reference (7) and could also have been obtained from a general theorem in Reference (5) applied to the large time–bandwidth product case.)

Equation (24) suggests the possibility of controlling the shape of $|g(t,\phi)|$ for large dopplers, $|\phi| \gg T^{-1}$, by control of $T_D(f)$. The energy associated with $|g(t,\phi)|$ is of course fixed by the choice of $A^2(f)$ and does not depend on $T_D(f)$ (nor on doppler as is shown in Equation (29)). Again using the methods of Reference (7), the energy in $|g(t,\phi)|$ is now distributed more uniformly in time for purposes of reducing its peak amplitude. The new shape of $|g(t,\phi)|$ is approximately rectangular. $|g(t,\phi)|$ as approximated by Equation (24) will remain constant in time (flat-topped) if the new timing curve satisfies

$$\frac{d^2\beta}{df^2} = K\,[A^2(f)]^2 = KA^4(f), \tag{26}$$

where K is a real nonzero constant. A timing curve that satisfies Equation (26) when $F_1 = 0.42$ is

$$T_D(f) = \frac{1}{2\pi\phi}\beta(f)$$

$$= 0.66\left(\frac{4T}{W^2}f^2\right) - 0.17T\cos 2\pi\frac{f}{W} - 0.01T\cos 4\pi\frac{f}{W}. \tag{27}$$

The resultant approximate width of $|g(t,\phi)|$ in time is obtained as the magnitude change in delay of $-\frac{1}{2\pi}\frac{d\beta(f)}{df}$ in the frequency interval $(-W/2, W/2)$. Then, corresponding to the new timing curve,

$$\frac{1}{W}|g(t,\phi)| \approx \frac{1.43}{\sqrt{8T|\phi|}}\,\mathrm{rect}\left(\frac{t}{0.66D}\right), \tag{28}$$

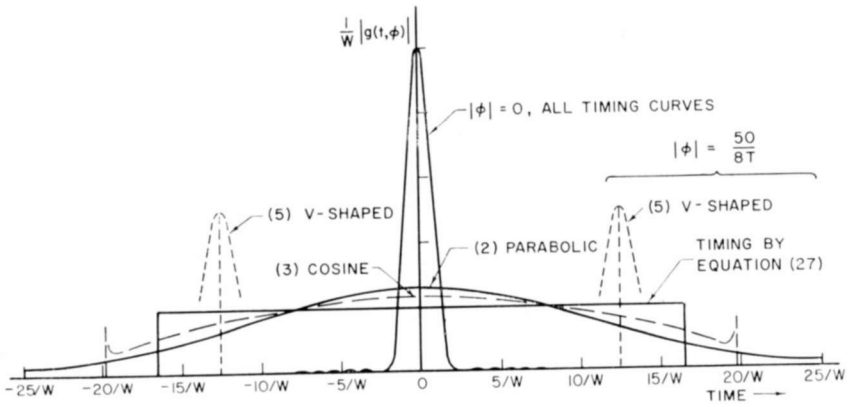


Fig. 13—Tapered case, comparison of central ambiguity for timing examples (2), (3), (5), and by Equation (27).

where $D = 8T|\phi|/W$. The energy of $|g(t,\phi)|$ as approximated by Equation (28) is checked by Parseval's formula;

$$\int_{-\infty}^{\infty}|g(t,\phi)|^2\,dt = \int_{-\infty}^{\infty}A^1(f)\,df,$$

$$\left(\frac{1.43W}{\sqrt{8T|\phi|}}\right)^2(0.66D) = (1+2F_1^2)\,W, \tag{29}$$

$$1.35W = 1.35W.$$

A comparison of Equations (28) and (25) shows that the approximate $|g(t,\phi)|/W$ corresponding to the new timing curve, Equation (27), has been reduced in peak amplitude by 22% and that its width to end points has been diminished by 34%. The approximate $|g(t,\phi)|/W$ for timing example (2) (parabolic) and for the timing of Equation (27) are plotted in Figure 13 for $\phi = 0$ and $|\phi| = 50/8T$.

Figure 13 also shows the approximate $|g(t,\phi)|/W$ for two other subcarrier timing curves employed with the same tapered spectrum, Equation (22) with $F_1 = 0.42$. The two other timing examples are example (3), cosine, and example (5), V-shaped plotted, also for $\phi = 0$ and $|\phi| = 50/8T$.

Figure 14 compares the timing curves of example (2), example (3) (inverted), and that given by Equation (27) (shifted).

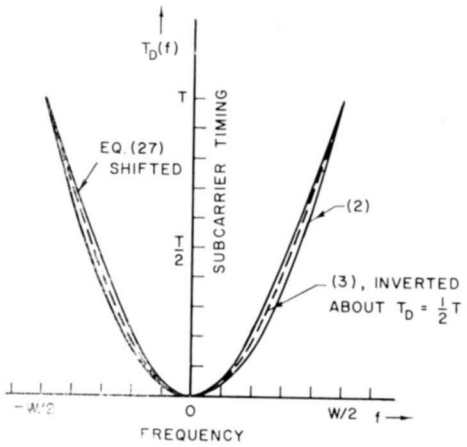


Fig. 14—$T_D(f)$ for examples (2), (3), and Equation (27).

Comparison of Central Ambiguity for Various Timing Examples

Shaping of central ambiguity for large doppler shifts, $|\phi| \gg T^{-1}$, is accomplished by controlling the slope of the group delay function, $-\frac{1}{2\pi}\frac{d\beta(f)}{df}$, as discussed in the preceding section. Since $\beta(f) = 2\pi\phi T_D(f)$, the group delay function is proportional to the derivative of the subcarrier timing curve, $T_D(f)$. Consequently, for the V-shaped subcarrier timing curve, example (5), two discrete values of delay (relative) result. Referring to Figure 4, it can be seen that the derivative of curve (5) takes on equal but opposite values in each half of the frequency band. The group delay function is therefore represented by a constant time delay over one half the band but by an equal constant time advance over the second half. When this fact is correlated with the corresponding plot of Figure 13 for $|\phi| = 50/8T$, two peaks are seen to exist in the central ambiguity—one delayed and the second advanced accordingly. For the other cases, the response amplitude for

$|\phi| = 50/8T$ is determined by the magnitude of the second derivative of the timing curve as shown by Equation (24) and by the definition of $\beta(f)$. For example, in the center of the frequency band (see Figure 14), the curve given by Equation (27) has the highest second derivative, the cosine curve (3) the next highest, and the parabolic curve (2) the smallest second derivative. This is reflected by the corresponding plots in Figure 13 in a region centered at $t = 0$. The relative amplitudes are in inverse order to the second-derivative magnitudes. At the exact center of the band, the V-shaped curve (5) has an infinite second derivative, a fact which tends to clear the region near $t = 0$ in Figure 13.

The energy associated with all of the responses plotted in Figure 13 is the same. Since by Equation (14), $g(t,\phi)$ repeats itself every Δ^{-1}, the energy in the central response tends to remain fairly constant as stated previously when $\tau = \Delta^{-1}$. However, Figure 13 and Equation (14) illustrate that the energy in the central response corresponding to certain dopplers $|\phi| \gg T^{-1}$ can be substantially reduced by making τ less than Δ^{-1} to reduce the width of $\rho(t,\phi)$. In particular, for the case of the V-shaped curve of example (5), the factor $\rho(t,\phi)$ of the central response, Equation (14), can be made to reduce the peaks shown in Figure 13. Also, over a central doppler strip repetitions of $g(t,\phi)$ are eliminated by selecting τ small compared with Δ^{-1}. Thus, by selection of the relation between τ and Δ and by shaping of the subcarrier timing curve, the energy in the central response can be reduced over selected doppler regions. By this process, ambiguity is transferred from the ϕ-axis to other regions of the $t-\phi$ plane. Making $\tau\Delta$ less than unity will increase the side response levels resulting in more ambiguity along the t-axis.

Second-Time-Around Echo Rejection

If, in the absence of doppler shift, the phase coding angles are set at zero for the transmitted waveform and its matched filter, then $g(t,0)$ corresponding to a phase-coded second-time-around echo can be shown to be the inverse transform of

$$G(f,0) = A^2(f)\, e^{j\theta(f)}. \tag{30}$$

When $\theta(f) = \pi(\tau/W)f^2$, the delay function, $-\dfrac{1}{2\pi}\dfrac{d\theta(f)}{df}$, decreases linearly by $\tau = \Delta^{-1}$ over the frequency interval $(-W/2, W/2)$. In the non-tapered case, the second-time-around echo is dispersed and

reduced in amplitude by the factor $1/\sqrt{N}$ with respect to a desired echo with the same energy. $g(t,0)$ for this dispersed echo is identical to $g(0, N/8T)$ as found from Table I for timing example (2). Thus, neglecting doppler shift, an offset in the phase coding angles of

$$\theta_n = \theta(n\Delta) = \pi \frac{\tau}{W} (n\Delta)^2 = \pi \frac{n^2}{N} \tag{31}$$

on a waveform-to-waveform basis will reduce second-time-around echoes by the factor $1/\sqrt{N}$ without requiring alteration of the delay code. This is true for all subcarrier timing curves $T_D(f)$.

Side Responses

The side responses and their possible interaction with the central response have been neglected in the above discussion of the central response. A complete description of waveform behavior requires consideration of all N^2 terms in Equation (6). Wherever time overlap occurs between any of the terms of the complete response, it is necessary to superpose all overlapping terms in correct relative phase. As pointed out previously, the N^2 terms include $N(N\text{-}1)$ terms contributing to side responses in addition to the N terms of the central response. Side responses exist even when $|\phi| \ll \Delta$ because each received rectangular subpulse excites not only the receiver channel centered at its own subcarrier, but also the remaining N-1 channels centered at the other subcarriers. The side-response amplitude is used as the measure of distal[3] ambiguity. Side-response performance can depend heavily on the ratio τ/T, which can affect the amount of time overlap in its contributing terms, as well as the overlap of side terms on the central response. Contributing terms correspond to pulses having widths of 2τ and extending approximately over the time interval $(-T, T)$. Zero-doppler side-response behavior is directly related to the many fine-grain irregularities that exist in the transmitted power spectrum—irregularities brought about by subpulse spectra having relative phases that change rapidly with frequency. (The zero-doppler cut is the waveform's autocorrelation function and, consequently, is the inverse transform of the waveform's power spectrum.)

A complete description of theoretical performance for a final waveform would perhaps be obtained by computer as a last step in waveform synthesis. In the early stages of synthesis, however, it is expedient and illuminating to break the side responses into groups and to consider each group separately and independent of the central response.

When $\tau \ll T$ and when N is not too large, the approximate behavior of the complete response is determined with relative ease since, under these conditions, the interaction of superposed central and side-term groups has a secondary effect. The most significant of the side-response groups are now considered separately for two of the subcarrier timing curves of Figure 4.

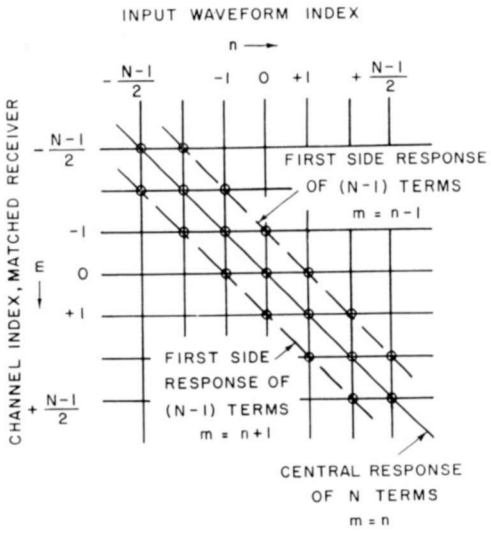


Fig. 15—Matched receiver input–output matrix (shown for $N = 7$).

First Pair of Side Responses—Nontapered Spectrum

It is convenient to consider the side responses in pairs, the *first pair* consisting of two summations, each having $N - 1$ terms found by letting $m = (n + 1)$, $(n - 1)$ in Equation (6). The first pair corresponds to the pair of diagonals adjacent to the principal diagonal of the matched receiver input–output matrix as illustrated in Figure 15. When $|\phi| \ll \Delta = \tau^{-1}$, it is permissible to neglect ϕ in the argument of ρ in Equation (6). Under these conditions, the shapes of the individual pulse envelopes, corresponding to the individual terms of the first side response, are shown as $|n| = 1$ in Figure 16. These shapes are the same as those shown in Reference (11) in a three-dimensional plot of the ambiguity diagram for a single rectangular pulse. ($|n| = 1$ corresponds to the vertical planes, $|\phi| = \tau^{-1}$, of that reference.)

[11] S. Appelbaum and P. E. Howells, "Waveform Design for Tomorrow's Radars," *Space/Aeronautics*, Vol. 32, p. 186, Oct. 1959.

For the linear timing of example (1), the $N - 1$ terms of each summation represent $N - 1$ time-coincident pulses. Consequently two expressions result, each of which is quite similar to that given in Equation (8) for the central response. Therefore, Poisson's sum formula can be used to form two product functions, each similar to the one in Equation (14). When $\theta_n = 0$, the summations so formed lead in each case to repeating high-amplitude peaks, a superposition

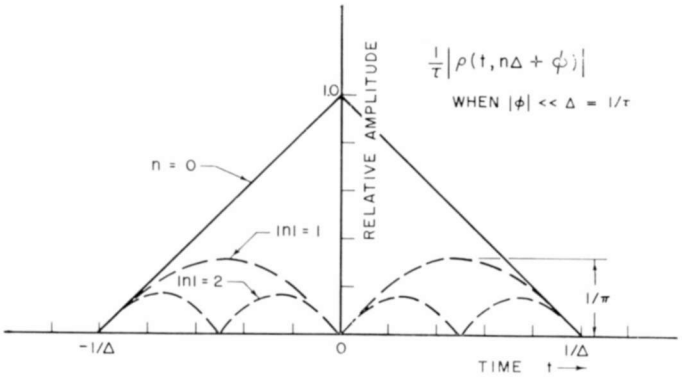


Fig. 16—Shapes of pulses contributing to central ($n = 0$) and side ($n = 2, 3, \cdots$) responses.

of sinc functions analogous to that encountered in the central response for timing example (1). Fortunately, when doppler is small, these peaks are suppressed by the first factor $\rho\,[t \pm (T/N), \mp \Delta]$ of the product. Sinc function peaks at $t = \pm T/N$, for example, fall at the nulls of $\rho\,[t \pm (T/N), \mp \Delta]$. (See Figure 16 for $|n| = 1$.) In a large doppler case ($|\phi|$ comparable to $N/2T$), the amplitude peaks shift in time away from the nulls. Under these conditions, it has been found convenient to employ phase coding,

$$\theta_n = \frac{\pi}{N-1}\,\frac{n^3}{3}, \tag{32}$$

to disperse the peaks of the first side response, reducing them in amplitude. Equation (32) was developed by a procedure similar to that used in arriving at Equation (31). Table II summarizes the side-response behavior for timing example (1).

For the timing of example (2), the $2(N - 1)$ pulses contributing

Table II—Peak Amplitudes for First Pair of Side Responses

$\lvert\phi\rvert$	θ_n	Peak of First Side Response
0	0	$\dfrac{1}{\pi N}$
$\dfrac{N}{2T}$	0	$\dfrac{N-1}{\pi N} \approx \dfrac{1}{\pi}$
All Dopplers	Eq. (32)	$\dfrac{\sqrt{N-1}}{\pi N} \approx \dfrac{1}{\pi\sqrt{N}}$

to the first pair of side responses occur in $N - 1$ time-coincident pairs. If $\tau < 4T/N^2$, the $N - 1$ pairs of pulses contributing to the first side responses are spread out in time, so that one pulse pair at most occurs at any instant. In this case the peak amplitude for the first pair of side responses is $2/(\pi N)$ maximum. A method for locating first side responses in time and frequency is illustrated in Figure 17. The time locations of the first side responses with respect to the central response are shown in Figure 18. As long as $|\phi| \ll \tau^{-1}$, the maximum side-response level remains essentially independent of doppler shift. Points

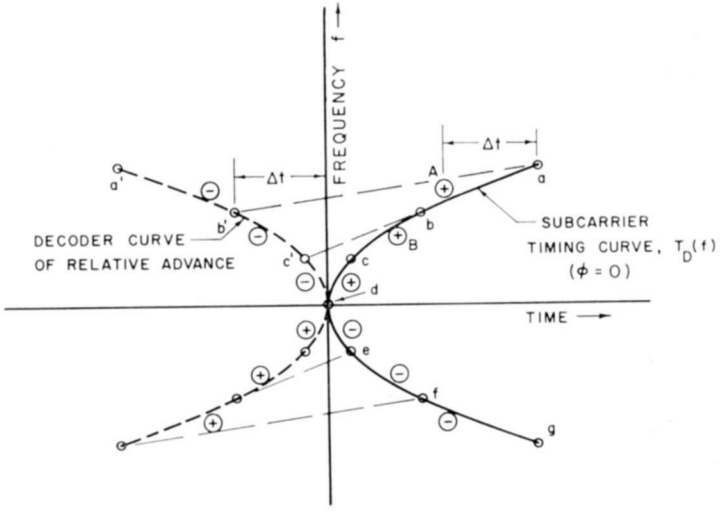


Fig. 17—Locating first side responses in time and frequency ($\phi = 0$).

a, b, and c in Figure 17, represent the locations of subpulses of a received waveform in time and in frequency. Points a′, b′, and c′ represent the corresponding locations of the subpulses of the image of the transmitted waveform and, hence, are the locations of the subpulses of the matched-receiver impulse response. When the subpulse at point a excites the receiver channel at the next lower subcarrier, the time advance corresponding to point b′ is applied to the delay time of point a. A side response occurs at point A. Similarly, points b and c′ result in a side response at point B, etc. The $N - 1$ points marked by the plus signs result because (except for the one at point g) each

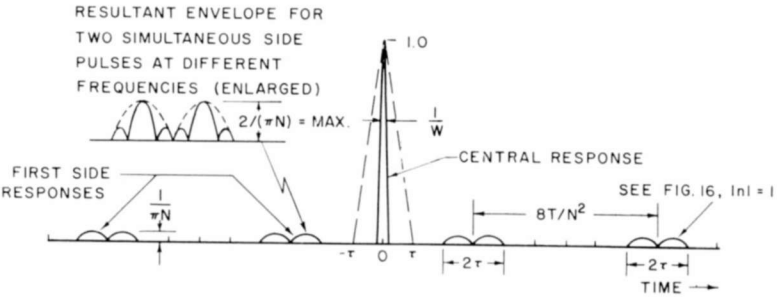


Fig. 18—Central and first side responses, zero-doppler cut.

received subpulse excites the receiver channel at the next lower subcarrier frequency. The $N - 1$ points marked by minus signs result because (except for the one at point a) each received subpulse also excites the channel at the next higher subcarrier frequency. The points marked "plus" and "minus" fall on two straight lines.

Conclusions

Signal ambiguity has been presented for several frequency-diversity waveforms in terms of the matched receiver output envelope. The discussion has emphasized envelope shapes in a central region near $t = \phi = 0$. Side responses were treated briefly for the case of linear and quadratic subcarrier timing.

Under the assumptions of this paper, waveforms with subcarriers timed according to curves (2) and (3) of Figure 4 (i.e., curves with parabolic and cosine (half-cycle) shapes) perform well in a central strip along the t-axis. The width of this strip in doppler is proportional

to N/T. Outside of this central strip, performance deteriorates in a complex manner as peaks in the output envelope slowly build up with increasing $|\phi|$. The build-up occurs because of overlapping effects that are analogous in the time domain to those encountered in the aliased spectrum in sampling analysis. Neglecting side responses, ambiguity is also confined to a central strip along the ϕ-axis whose width in the time dimension is that of the subpulse autocorrelation function.

It has been demonstrated that the principle of stationary phase can be employed to provide limited control of the shape of ambiguity profiles in the t-dimension when $|\phi| \gg T^{-1}$. This is accomplished by reshaping the subcarrier timing curve after spectral taper has been chosen to shape the zero-doppler envelope response. A final timing curve should probably take cognizance of the shape of the envelope of the subpulse autocorrelation function (or, in a practical case, the shape of the envelope response of a band-pass filter tailored to the subpulse). For example, stationary phase can be applied to make $|g(t,\phi)|$ sag in its middle portion in the region where the subpulse autocorrelation function has its highest amplitude.

It is suggested that by superposition the foregoing results can be applied to more complex waveforms when the complex waveforms are comprised of a number of simple waveforms of the type described.

Acknowledgment

The author is greatly indebted to T. Murakami for his many helpful suggestions and for his invaluable assistance in the preparation of the figures.

Appendix I—Energy in Transmitted Waveform

Evaluation of Equation (5) at $t = \phi = 0$ results in

$$\psi_0(0,0) = \int_{-\infty}^{\infty} |\psi_T(\tau)|^2 \, d\tau = 2E, \tag{33}$$

where E is the energy contained in the real physical transmitted waveform of N subpulses. If the N subpulses are orthogonal, E is equal to the sum of the energies of the individual subpulses. It then follows from the waveform description that

$$\psi_0(0,0) = 2E = \sum_{-(N-1)/2}^{(N-1)/2} A_n^2 \tau. \tag{34}$$

Subpulses are of course orthogonal when they do not overlap in time and also in time-coincident cases when $\tau = \Delta^{-1}$. It is possible, however, to have nonorthogonal subpulses when two overlap in time by an amount equal to a nonintegral number of cycles of their difference frequency. For reasons of simplicity, this possibility is ignored.

If none of the side responses in Equation (6) contribute to $\psi_0(0,0)$, then $\psi_0(0,0)$ is equivalent to Equations (8) and (14) evaluated at $t = \phi = 0$. Equation (34) can be obtained by equating Equations (6) and (8) for $t = \phi = 0$. By similar use of Equation (6) with Equation (14), $\psi_0(0,0)$ is also obtained as

$$\psi_0(0,0) = 2E = \frac{\tau}{\Delta} g(0,0). \tag{35}$$

Since, by Equation (23), $g(0,0) = W$, it follows that $2E = \tau W/\Delta = N\tau$ in both the tapered and nontapered ($F_1 = 0$) cases. Normalization[1] of Equations (6), (8), and (14) is accomplished by dividing by $\psi_0(0,0) = 2E$.

It is noted that Equation (6) is equivalent to $\exp(j2\pi f_0 t)\ (2E)\ \chi(-t, -\phi)$ where $\chi(t,\phi)$ is the combined time and frequency autocorrelation function of Reference (1) and where $\chi(0,0) = 1$.

Appendix II

The purpose of this discussion is to illustrate that the normalized correlation coefficient for any two component pulses of the central response is small. The two pulses are expressed as

$$e_1(t) = q(t) \cos [2\pi(f_1 + m\Delta)t + \theta] \tag{36}$$

and

$$e_2(t) = q(t) \cos 2\pi f_1 t. \tag{37}$$

In Equation (36), m is an integer, $m \neq 0$, and θ is an arbitrary phase angle.

The pulses have triangular envelopes given by

$$q(t) = \tau \left[1 - \frac{|t|}{\tau} \right] \text{rect}\left(\frac{t}{2\tau} \right), \tag{38}$$

where $\tau = \Delta^{-1}$. When $f_1 \gg \Delta = \tau^{-1}$, the two pulses have approximately the same energy U given by

$$U = \int e_1^2(t)\,dt = \int e_2^2(t)\,dt = \frac{1}{2}\int q^2(t)\,dt = \frac{1}{3}\tau^3. \quad (39)$$

The normalized correlation coefficient $\rho(m,\theta)$ is defined as

$$\rho(m,\theta) = \frac{1}{U}\int e_1(t)\,e_2(t)\,dt. \quad (40)$$

Substituting from the preceding four equations into this definition and evaluating the integral,

$$\rho(m,\theta) \approx \frac{3}{2(m\pi)^2}\cos\theta. \quad (41)$$

Equation (41) illustrates that $|\rho|$ is a maximum when $m = 1$ and when θ is equal to zero or 180°. Consequently the normalized correlation coefficient for any two component pulses is small,

$$|\rho| \leq \frac{3}{2\pi^2} \approx 0.152. \quad (42)$$

HELIX SUPPORT STRUCTURE FOR ULTRA-WIDE-BAND TRAVELING-WAVE TUBES

By

E. F. Belohoubek

RCA Electronic Components and Devices,
Princeton, N. J.

Summary—*A new helix support structure consisting of half-moon-shaped ceramic rods is described. The structure has an anomalous dispersion that permits an increase in the operating bandwidth of traveling-wave tubes. It also provides an input impedance suitable for wide-band matching to a coaxial coupler.*

Introduction

THE GENERAL TREND in many present-day applications for medium-power helix traveling-wave tubes is toward greater bandwidth. For the bandwidth of a traveling-wave tube to be increased much beyond an octave, the slow-wave circuit and the coupler must be specially tailored for wide-band operation. The main requirements for the slow-wave circuit of such a tube are as follows:

(a) A dispersion suitable for wide-band operation;

(b) A high interaction impedance;

(c) A constant, and preferably low, transverse impedance so that broad-band matching of couplers can be achieved;

(d) A rugged support structure that provides good heat conduction to the outside environment;

(e) Small r-f losses;

(f) Small outer diameter of vacuum envelope so that lightweight focusing assemblies can be used.

In most of today's helix tubes, the support structures consist of three ceramic rods that are held in intimate contact with the helix by brazing, glazing, or compression loading. If this supporting structure is designed properly, the dispersion of the helix can be shaped to provide both maximum bandwidth and an input impedance suitable for broad-band matching to a coaxial coupler. This paper presents the results of cold-test measurements of the dispersion of various helix assemblies and of the matching characteristics of associated wide-band coaxial couplers.

Helix Support Structure

Conventional helix-type slow-wave circuits are all somewhat dispersive, i.e., their group and phase velocities are unequal, and the interaction between the electron beam and the r-f circuit is limited to a specific frequency range. The first step toward increasing the

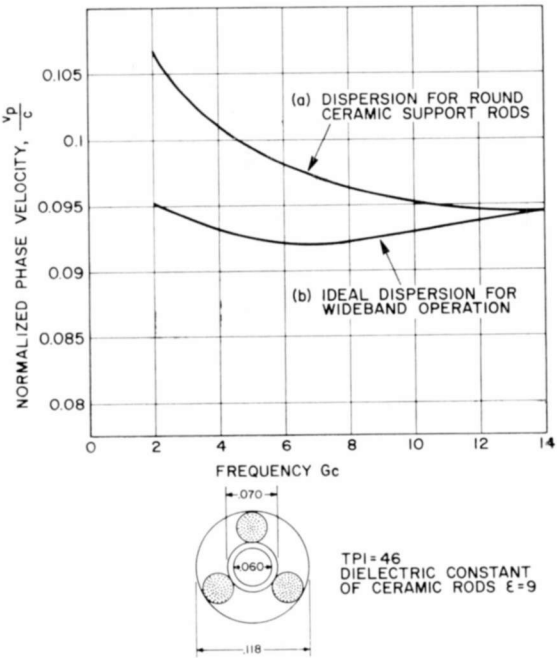


Fig. 1—Comparison of the dispersion for a typical helix assembly supported by round ceramic rods with the ideal dispersion required for wideband operation.

bandwidth of a traveling-wave tube is to reduce the dispersion of the slow-wave circuit. However, even if a fully dispersionless circuit could be realized, the resulting tube would still have a rather limited bandwidth because of the increase in beam wavelength and the corresponding decrease in gain per unit length at lower frequencies. For truly wide-band operation, it is necessary to find a dispersion that provides good interaction at high and low frequencies and reduces the tube gain in the band center.

The shape of the dispersion can be influenced strongly by the type of helix support structure used. Curve (a) of Figure 1 shows the dispersion of a typical helix assembly in which the support structure consists of three round ceramic rods enclosed by a metallic barrel. The

small-signal gain corresponding to this dispersion was computed for the design parameters listed in Table I, from curves of Birdsall and Brewer.[1] The dielectric and shield reduction factors for the interaction impedance of the helix were computed from curves of Tien[2] and McMurtry[3] on the basis of the measured dispersion of the helix assembly. The small-signal gain per unit length is given by curve (a) in Figure 2. The gain decreases rather rapidly at the lower end of the frequency band, and the usable bandwidth is restricted to about an octave.

Table I

Mean helix diameter, $2a$	0.065 inch
Helix wire diameter, $2s$	0.005 inch
Shield diameter, $2a_s$	0.118 inch
Beam diameter, $2b$	0.0325 inch
Beam voltage, V_0	2.5 kilovolts
Beam current, I_0	40 milliamperes
Turns per inch	46 for helix assembly having round support rods 45 for helix assembly having optimum anomalous dispersion

Inspection of the gain-parameter curves as a function of the synchronism parameter b_0, as published by Birdsall and Brewer, indicates that the operating bandwidth can be broadened substantially by decreasing the phase velocity of the helix in the middle and lower regions of the frequency band. In this way, the gain is enhanced at low frequencies, reduced in the center of the band, and left unchanged at the high-frequency end. Curve (b) in Figure 1 shows the ideal dispersion for producing this broad-banding effect. The corresponding gain per unit length is shown by curve (b) in Figure 2; a comparison of this curve with curve (a) shows the improvement that can be obtained. A significant increase in over-all bandwidth is obtained without loss

[1] C. K. Birdsall and G. R. Brewer, "Traveling-Wave Tube Propagation Constants for Finite Values of C," Hughes Aircraft Company, Technical Memorandum #331, Oct. 1953.

[2] P. K. Tien, "Traveling-Wave Tube Helix Impedance," *Proc. IRE*, Vol. 41, p. 1617, Nov. 1953.

[3] B. J. McMurtry, "Fundamental Interaction Impedance of a Helix Surrounded by a Dielectric and a Metal Shield," *Trans. IRE PGED*, Vol. ED-9, p. 210, March 1962.

in gain at the high-frequency end of the band. The variation in the small-signal gain remains within 20% over a frequency ratio of 3.5 to 1; for the same gain variation, the frequency ratio for a conventional helix assembly supported by round ceramic rods is 2.5 to 1.

Cold-test evaluations of several helix support structures were made to find one that would provide the ideal dispersion curve given in Figure 1. Because the r-f field extends further away from the helix

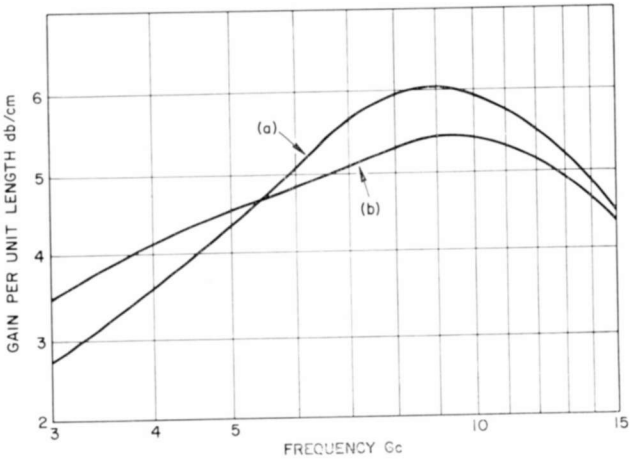


Fig. 2—Gain parameter as a function of frequency (a) for a conventional helix assembly supported by round ceramic rods and (b) for a helix assembly that provides the ideal anomalous dispersion.

at lower frequencies, a selective decrease in phase velocity at low frequencies can be achieved by use of the half-moon-shaped ceramic rods shown on the left in Figure 3. Rods of this type produce negligible additional dielectric loading at very high frequencies where the r-f fields are tightly bound to the helix, but cause a strong decrease in the phase velocity at low frequencies. Thus they produce an anomalous dispersion, as shown by curve (a) in Figure 3. An even closer approximation to the ideal dispersion curve is obtained by either breaking the corners of the half-moon-shaped rods or by using a double curvature, as indicated in Figure 3(b). Thus, a relatively simple modification of the helix support structure permits a substantial increase in the operating bandwidth for helix-type traveling-wave tubes.

In most metal–ceramic tubes, a compression technique is employed to secure the helix within the vacuum envelope. The standard method

is to elastically deform the outer shield into triangular shape before inserting the helix assembly with its three ceramic support rods. This method is not applicable to the new support structure because the half-moon-shaped ceramic rods cover nearly the entire inner surface of the helix shield. In the compression technique employed to accom-

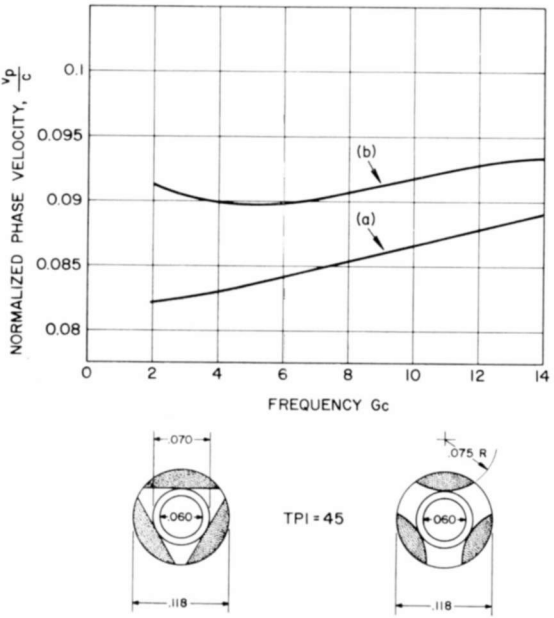


Fig. 3—Two examples of helix supports that produce the anomalous dispersion necessary for wide-band operation. Curve (a) shows the dispersion for a helix supported by half-moon-shaped ceramic rods (sketch at lower left) and curve (b) shows the dispersion when double-curved support rods are used (sketch at lower right).

modate the half-moon-shaped ceramic support rods, the outer vacuum envelope consists of a thin-wall tantalum tubing into which the helix, including the support rods, is inserted with minimum clearance. Subsequently, soft iron shims acting as pole pieces for the periodic permanent focusing magnets are copper-brazed to the tantalum tubing in vacuum. Because of the difference in the thermal expansion coefficients of tantalum and Armco-iron, the helix assembly is under heavy compression after the brazing cycle. This technique provides very good heat conduction from the helix to the outside of the tube. An added advantage of this technique is that the focusing shims may be placed very close to the beam; thus good focusing can be obtained with focusing magnets of minimum size and weight.

Coaxial Coupler

The type of coupler best suited for very wide-band operation is a direct coaxial connection. Such a coupler requires little space, has a low insertion loss, and can be matched to a helix over a very wide frequency range.

For proper design of the coaxial coupler, the input impedance of the helix must be known. Theoretical values that provide a first-order approximation for the transverse helix impedance have been given by Mathers and Kino.[4] This transverse impedance, not to be confused with the interaction impedance of the helix, is defined as the ratio of twice the power propagating along the helix to the square of the helix current. For helices having a closely spaced outer shield, the helix current is nearly equal to the shield current, and the transverse impedance approaches the input impedance of the helix as it appears to a coaxial line connected to one end of the helix while the other end is terminated by a matched load. The calculated values are only approximate, however, because they are based on solutions for a sheath helix rather than a tape helix, and the dielectric loading by the support rods is not taken into account.

To determine the input impedance more accurately, a series of different helix assemblies were cold-tested. The measurement setup consists of an X-band slotted line that has an opening in the bottom of the waveguide into which the helix assembly with its shield is inserted, as shown in Figure 4. The helix assemblies are scaled up in size by a factor of 2.5 for this measurement to ease the assembly and handling problems and to improve the measurement accuracy. A thin-wall tubing mounted inside the slotted waveguide forms a wire-above-ground transmission line that is connected tangentially to the helix. A small section is cut from the shield of the helix assembly over one third of its circumference at the input end to provide a smooth transition from the helix to the wire-above-ground line. The input impedance of the helix is determined on the basis of VSWR measurements made through the slot in the waveguide.

Figure 5 shows a Smith-chart plot of the input impedance of a conventional (round-rod-supported) helix assembly as a function of frequency. The reference plane for this measurement was chosen as the point where the wire-above-ground line penetrates the outer shield of the helix assembly.

To obtain the best possible match to a 50-ohm coaxial line by use

---

[4] G. C. Mathers and G. S. Kino, "Some Properties of a Sheath Helix with a Center Conductor or External Shield," Technical Report #65, Stanford University, Electronic Research Laboratory, June 1953.

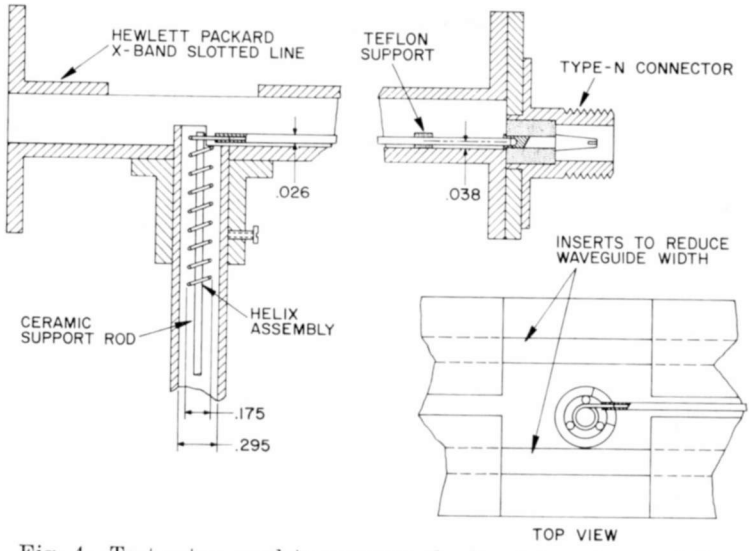


Fig. 4—Test setup used to measure the input impedance of the helix assembly.

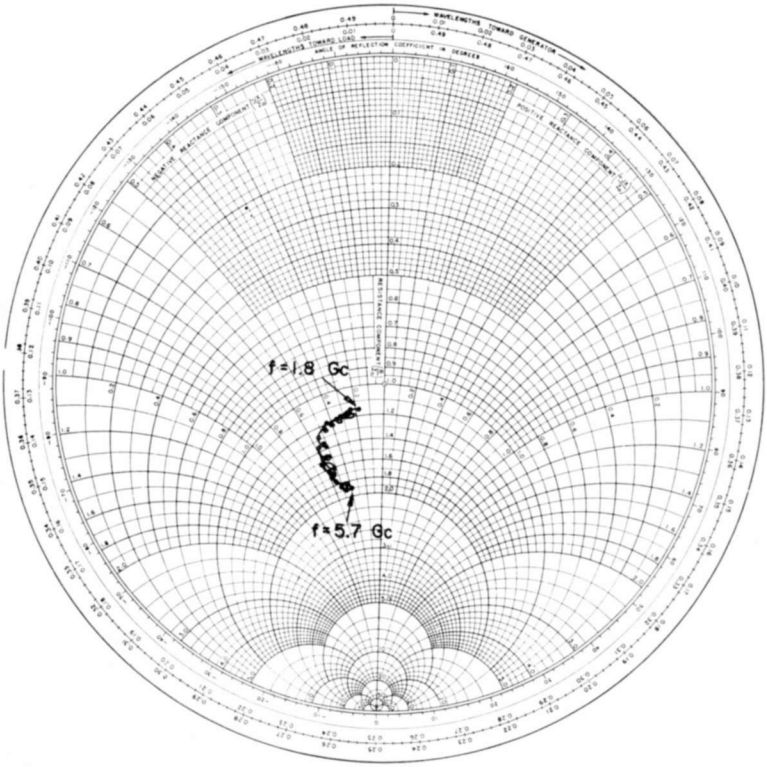


Fig. 5—Smith-Chart representation of the helix input impedance for an assembly having round support rods and a straight outer shield.

of a transformer section, the resistive part of the helix impedance should be close to 50 ohms and should show a minimum variation as a function of frequency; the reactive portion of the impedance should be very small. The latter requirement can be fulfilled if the reference plane in the Smith chart is shifted somewhat toward the load. The

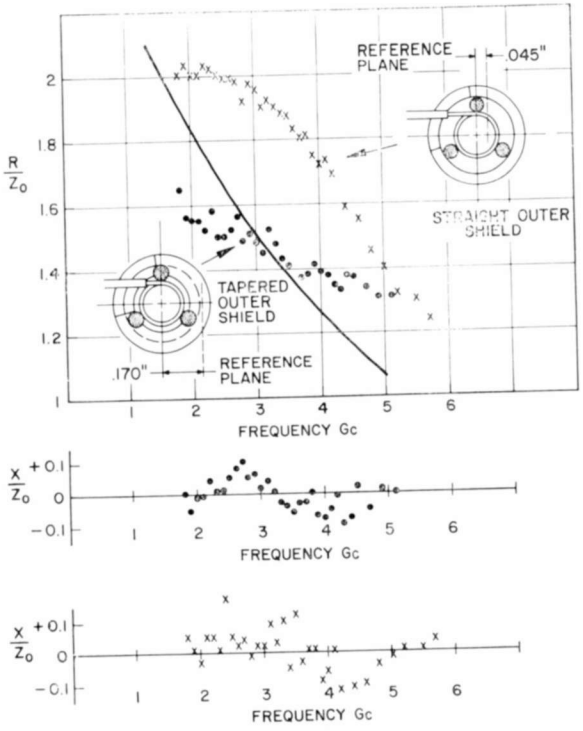


Fig. 6—Resistive and capacitive components of the helix input impedance as a function of frequency (helix TPI = 18.4; $Z_0 = 104$ ohms). X's and dots indicate measured values for straight and tapered outer shields, respectively; solid curve is calculated input resistance for a straight outer shield.

resistive and reactive parts of the resulting impedance are shown by the crosses in Figure 6. This figure also shows the theoretical transverse impedance for the helix assembly, as obtained from the curves of Mathers and Kino. The poor agreement between the measured and theoretical impedance illustrates the need for a separate impedance measurement if a wide-band coupler that provides a truly low VSWR is to be designed.

Two common techniques may be used to improve the shape of the helix input-impedance curve for wide-band matching: either the outer

shield *or* the helix pitch can be tapered in the region close to the coupler. As shown by Figure 7(a) and the dotted curves in Figure 6, these tapers not only decrease the input impedance but also reduce its variation as a function of frequency. In these examples, the helix pitch is changed by a ratio of 2.4 to 1, and the shield-to-helix diameter ratio is tapered from 1.7 down to 1.3.

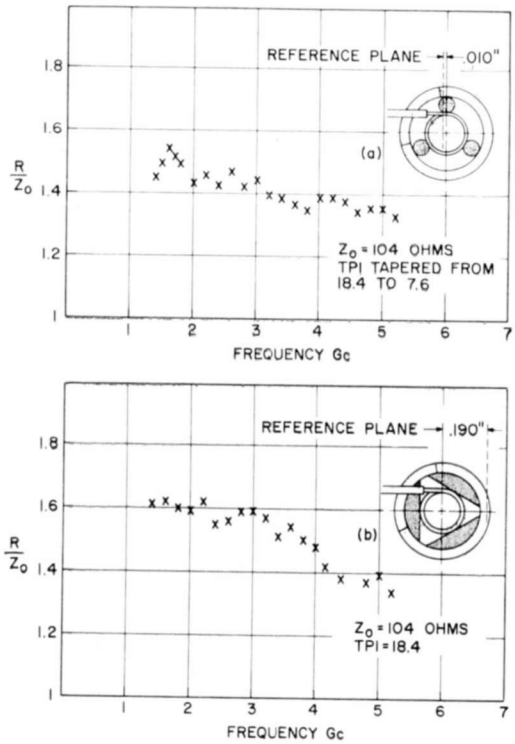


Fig. 7—Resistive component of the helix input impedance as a function of frequency (a) for a tapered helix supported by round ceramic rods and (b) for a uniform helix supported by half-moon-shaped ceramic rods.

For comparison, the input impedance of a uniform helix assembly having half-moon-shaped support rods is shown in Figure 7(b). In this case, the selective dielectric loading not only permits wide-band interaction with the electron beam but also decreases the input impedance more at the low-frequency end than at high frequencies; as a result, the impedance curve is fairly constant over a wide frequency range. If the reference plane is properly chosen, the normalized reactive component X/Z_0 is below 0.1, and can be neglected. The helix assembly using half-moon-shaped support rods can, therefore, be

matched to a coaxial line over a wide frequency band without the need of a taper either in the shield or in the helix pitch.

A cross section of a coaxial coupler for a wide-band helix assembly is shown in Figure 8. The coupler uses a two-step quarter-wavelength transformer to reduce the helix input impedance to 50 ohms. The ceramic window is placed at the end of the transformer section in the

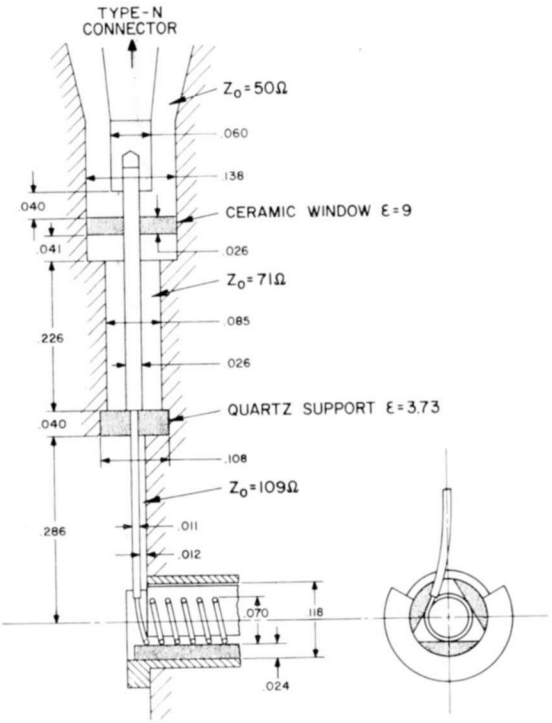


Fig. 8—Coaxial coupler for a wide-band helix assembly.

50-ohm line so that its transverse dimensions can be kept small. A short taper section joins the window to a type-N connector (not shown in the figure).

The calculation of the characteristic impedance for the transformer sections is based on the measured input impedance of the scaled helix assembly. For a mean helix input impedance of 155 ohms, the corresponding transformer impedances are 71 ohms for the output section and 109 ohms for the section connected to the helix. Based on tables of Young,[5] the maximum VSWR of the transformer within the band

[5] L. Young, "Tables for Cascaded Homogenous Quarter-Wave Transformers," *Trans. IRE PGMTT*, Vol. 7, p. 233, April 1959.

from 5 to 12 gigacycles is 1.3 to 1. The length of the 109-ohm transformer section must be shortened because of the shift in the reference plane which, as determined from cold-test measurements, is necessary to minimize the reactive component of the input impedance. Similarly, the length of the 71-ohm transformer section is adjusted to effect some cancellation of the corner capacitances at the quartz support disk. The dimensions of the coaxial window are calculated from curves by Moats[6]

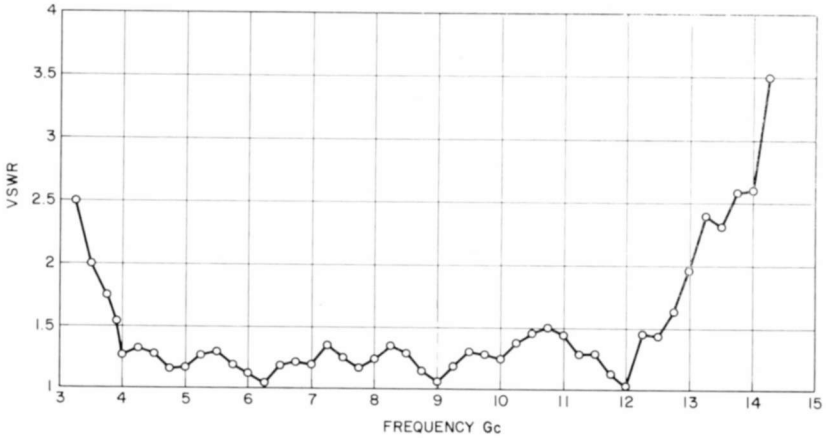


Fig. 9—Measured input VSWR of the coaxial coupler shown in Figure 8. (The measured values include the reflections from the type-N connector.)

for a minimum VSWR of 1.1 to 1 over the frequency band from 0 to 12 gigacycles. Figure 9 shows the measured input VSWR of an actual coupler assembly. The VSWR, including the reflections from the coaxial window and the type-N connector, remains below 1.5 to 1 over the frequency range from 3.9 to 12.6 gigacycles. The insertion loss of the coupler varies from 0.2 to 0.3 db over the same frequency range. An extension of the bandwidth to ratios of 4 to 1 and more is possible if a somewhat higher coupler VSWR can be tolerated.

Conclusions

Cold-test measurements were performed on various helix support structures and their suitability for ultra-wide-band operation was investigated. A support structure consisting of three half-moon-shaped

---

[6] R. R. Moats, "Design of Broadband Ceramic Coaxial Output Windows for Microwave Power Tubes," *Sylvania Technologist*, Vol. 11, p. 86, July 1958.

ceramic support rods, compression loaded by the outer metallic envelope, showed very wide bandwidth capabilities. Gain calculations based on the anomalous dispersion of this helix assembly indicated a bandwidth ratio of 3.5 to 1 for a 20% variation in small-signal gain. The wide operating bandwidth is implemented by an input impedance well suited for matching to a coaxial line with a minimum VSWR. Additional advantages of this type of helix assembly are its good heat-dissipation capability and the very small outer diameter of the vacuum envelope, which permits the weight and size of the periodic-magnet focusing structure to be held to a minimum.

ACKNOWLEDGMENT

The support of J. Klatskin in the performance of the cold-test measurements is gratefully acknowledged.

SWITCHING ANALYSIS OF GATE-CONTROLLED SPACE-CHARGE-LIMITED EMISSION PROCESSES IN SEMICONDUCTORS

By

Roscoe C. Williams*

Summary—*The switching behavior of gate-controlled p-i-n structures has been examined by dividing the switching time into two phases, I and II. Phase I is a constant-current phase where gate reverse current flows due to an applied reverse bias. Phase II is concerned with reverse gate current that is a pure diffusion current that decreases monotonically to zero.*

It is shown that Phase II current is many orders of magnitude less than Phase I current. Thus, one concludes that the switching time is essentially the duration of Phase I.

Space-charge neutrality considerations show that termination of electron flow into the i region results in hole-flow cessation, which means current flow has stopped since it consists of two-carrier flow.

It is shown that electron flow termination can be achieved by (a) appropriate location of the gate with respect to the n-i (N^+-N) junction; (b) suitable choice of reverse bias applied to gate. Correct selection of these two parameters yields microsecond switching times.

Introduction

THE ANALYSIS of gate-controlled space-charge-limited emission process has been set forth by the author[1] using:

(a) Injection parameters to obtain closed-form solutions for the p-i-n diode.

(b) Green's functions to obtain closed-form solutions for gate-controlled p-i-n structures.

The switching analysis of Reference (1) is based on the fact that current flow in p-i-n structures is obtained through two-carrier flow. Charge neutrality is preserved throughout the i region (high-resistivity N or P type material) by injecting an electron at the n-i junction for every hole injected at the p-i junction.

If the flow of electrons into the i region can be terminated in some manner, charge neutrality alone forces the termination of the injection of holes into the i region, resulting in the cessation of current flow. Since the gate[1,2] controls the current that flows through the device,

---

* Formerly with RCA Laboratories, Princeton, N. J.; now at the University of British Columbia, Vancouver, B. C., Canada.

[1] R. C. Williams, "Analysis of Gate-Controlled Space-Charge-Limited Emission Processes in Semiconductors," *RCA Review*, Vol. XXV, p. 262, June 1964.

[2] H. N. Yu, "The Chargistor, A New Class of Semiconductor Devices," *IBM Jour. Research and Development*, Vol. 5, No. 4, p. 328, Oct. 1961.

it will be shown that by switching the bias on the gate from forward to reverse the electron flow into the i region can be terminated within a fairly short interval of time (microseconds or less), resulting in a current cessation time (switching time) of the same order of magnitude.

It was shown in Reference (1) that the transit time of a hole across an i region four diffusion lengths long is of the order of 10 microseconds. This long transit time leads to operating frequencies of the order of 10^5 cycles per second when the device is used as an oscillator.[2] The fact that short switching times are feasible is entirely due to the reasons related to charge neutrality stated above.

The switching behavior of gate-controlled p-i-n structures will be examined by dividing the switching time into two phases, I and II. Phase I begins when the gate bias is rapidly switched from forward to reverse bias. Because of the large number of excess holes in the i region (high-resistivity N-type material), reverse current flows out of the gate as soon as the reverse bias is applied.

The fact that reverse current flows at all is related to the reverse transient characteristic of a P-N junction diode due to minority carrier storage.[3-13] In a simple p-n junction, the minority carriers (holes) are stored in the n region when the junction is forward biased; when

---

[3] B. R. Gossick, "Effect of Transient Time on Germanium Rectifier Behavior," *Phys. Rev.*, Vol. 91, p. 1011, Aug. 1953.

[4] R. G. Shulman and M. E. McMahon, "Recover Currents in Germanium p-n Junction Diodes," *Jour. Appl. Phys.*, Vol. 24, p. 1267, Oct. 1953.

[5] R. H. Kingston, "Switching Time in Junction Diodes and Junction Transistors," *Proc. I.R.E.*, Vol. 42, p. 829, May 1954.

[6] E. L. Steele, "Charge Storage in Junction Diodes," *Jour. Appl. Phys.*, Vol. 25, p. 1148, Sept. 1954.

[7] B. Lax and S. F. Neustadter, "Transient Response of a p-n Junction," *Jour. Appl. Phys.*, Vol. 25, p. 1148, Sept. 1954.

[8] S. R. Lederhandler and L. J. Giacoletto, "Measurement of Minority Carrier Lifetime and Surface Effects in Junction Diodes," *Proc. I.R.E.*, Vol. 43, p. 478, April 1955.

[9] J. Henderson and J. R. Tillman, "Minority Carrier Storage in Semiconductor Diodes," *Proc. I.E.E.*, Vol. 104B, p. 318, Jan. 1957.

[10] J. Halpern and R. H. Rediker, "Out Diffusion as a Technique for the Production of Diodes and Transistors," *Proc. I.R.E.*, Vol. 46, p. 1068, June 1958.

[11] W. H. Ko, "The Reverse Transient Behaviour of Semiconductor Junction Diodes," *I.R.E. Trans. on Electron Devices*, Vol. ED-8, p. 123, March 1961.

[12] D. P. Kennedy, "Reverse Transient Characteristics of a p-n Junction Diode Due to Minority Carrier Storage," *I.R.E. Trans. on Electron Devices*, Vol. ED-9, p. 174, March 1962.

[13] S. Y. Muto and S. Wang, "Switching Response of Graded-Base p-n Junction Diodes," *I.R.E. Trans. on Electron Devices*, Vol. ED-9, p. 183, Mar. 1962.

a reverse bias is applied, these stored holes give rise to reverse current. This phenomenon has been examined by many authors.[3-13] Kingston[5] and Lax and Neustadter[7] first analyzed the switching response of a uniform base diode; their work led to more detailed investigations involving models of finite dimensions with graded bases[12,13] and ohmic contacts with arbitrary recombination velocity.

These investigations assumed that initially there was a steady-state current flow in the forward direction that established a certain hole distribution throughout the n region of the p-n junction. When the bias is instantaneously switched from forward to reverse, a hole density persists at the junction for a finite length of time, i.e., it does

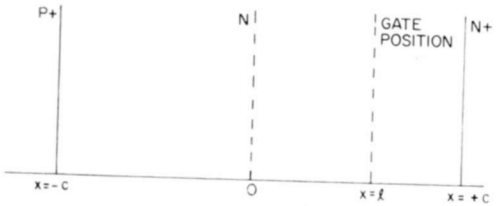


Fig. 1

not vanish instantaneously. This means that the junction voltage changes monotonically from a forward to a reverse bias even though the bias has been switched abruptly. This causes a large initial current to flow in the reverse direction immediately after switching, and it continues to flow, determined entirely by the external applied reverse-bias and external resistances, until the hole density at the junction vanishes. This is called the constant current phase.[12,13]

Phase I is the constant current phase for gate-controlled p-i-n structures. The minority-carrier (hole) distribution throughout the i region (high-resistivity N-type material) is determined by[1,14] hole densities at the junction interfaces located at $x = \pm c$ (see Figure 1), the hole lifetime, and length of the i region.

If the triode[1,2] is operated so that the gate is cut off before Phase I begins, the device will operate as a constant-current device with a large hole density at the gate. When the gate bias is reversed, Phase I begins and perists until the hole density at the gate vanishes.

As soon as this occurs, the voltage at the gate becomes reverse-

[14] D. A. Kleinman, "The Forward Characteristics of the p-i-n Diode," *Bell Syst. Tech. Jour.*, Vol. 35, p. 685, May 1956.

biased and the reverse current decreases in magnitude. Phase II starts the *instant* the hole density at the gate vanishes because the voltage at the gate becomes zero, which causes the potential difference between the gate and n-i (N+-N) junction to become zero, biasing off the n-i junction and reducing the electric field between it and the gate to zero.

With the n-i junction biased off, the electron supply is choked off; consequently, all the electrons available to maintain charge neutrality during all of Phase II are contained within the i region at the beginning of Phase II. Any holes injected into the i region after Phase II begins would violate charge neutrality, hence, the p-i (P+-N) junction, must also become biased off.

Since the field has vanished between the gate and n-i junction, drift current no longer flows in that region. Additional holes other than those present at the n-i junction are not required to maintain charge neutrality there since electron injection has ceased; consequently, hole flow toward the n-i junction stops and hole flow toward the gate begins.

The absence of a field between gate and n-i junction means that the diffusion gradient alone is responsible for reverse current flow toward the gate. This reverse diffusion current will be much smaller than the current that was flowing in the device prior to and during Phase I, since that current was primarily drift current.

The switching sequence, in view of the above facts, is as follows:

Initial State: Forward current flowing through the device with gate cutoff because potential at the gate exceeds voltage (forward bias) applied to it.

Phase I: Forward gate bias is reversed abruptly, reverse current flows out of the gate, and a diminished forward current continues to flow between gate and n-i junction. The current that reaches the n-i junction is the difference between the current that flowed initially and the reverse current at the gate. Phase I ends when the hole density at the gate vanishes.

Phase II: Operates as described above. Forward current ceases to flow in the device, and reverse current that is orders of magnitude less than the Phase I forward current starts to flow between the gate and the n-i junction. Phase II reverse current is not constant, but decreases monotonically to zero.

Phase I Analysis

The differential equation that describes the behavior of the hole density throughout the intrinsic region is derived in a manner similar to that in References (1) and (14). The following symbols are used:

I_n = number of electrons/cm^2/sec,

I_p = number of holes/cm^2/sec,

$p(x,t)$ = hole density,

$\beta = 1/(kT)$, where k = Boltzmann's constant and T the absolute temperature,

E = electric field,

D_p = diffusion constant for holes, in cm^2/sec,

D_n = diffusion constant for electrons, in cm^2/sec,

τ = hole lifetime,

I_{pr} = hole component of the reverse current,

p_N = thermal equilibrium hole density,

$\gamma = \dfrac{I_p(-c)}{I}$, the p-i junction injection parameter,

$\xi = \dfrac{I_n(c)}{I}$, the n-i junction injection parameter,

I_{ps} = saturation current density for the p-i junction.

The equations of particle flow are

$$I_p = -D_p \frac{dp}{dx} + \beta D_p p E, \tag{1}$$

$$I_n = D_n \frac{dp}{dx} + \beta D_n p E. \tag{2}$$

We have set $n = p$ in Equation (2) as in References (1) and (14).

The equations of continuity for electrons and holes, sometimes referred to as recombination laws,[15] are

[15] M. A. Lampert and A. Rose, "Volume-Controlled Two-Carrier Currents in Solids: The Injected Plasma Case," *Phys. Rev.*, Vol. 121, p. 26, 1 Jan. 1961.

$$-\frac{dI_p}{dx}=\frac{p}{\tau}+\frac{\partial p}{\partial t}, \tag{3}$$

$$\frac{dI_n}{dx}=\frac{p}{\tau}+\frac{\partial p}{\partial t}. \tag{4}$$

Differentiating Equations (1) and (2) with respect to x yields

$$-\frac{dI_p}{dx}=D_p\frac{d^2p}{dx^2}-\beta D_p\frac{d}{dx}(pE)=\frac{p}{\tau}+\frac{\partial p}{\partial t}, \tag{5}$$

$$\frac{dI_n}{dx}=D_n\frac{d^2p}{dx^2}-\beta D_n\frac{d}{dx}(pE)=\frac{p}{\tau}+\frac{\partial p}{\partial t}. \tag{6}$$

In order that the term involving $(d/dx)(pE)$ can be eliminated, multiply Equation (5) by $b=D_n/D_p$ and add it to Equation (6); the result is

$$\frac{d^2p}{dx^2}-\frac{p}{L^2}=\frac{\tau}{L^2}\frac{\partial p}{\partial t} \tag{7}$$

where $L^2=2D_n\tau/(b+1)$. Equation (7) is the equation that governs the spatial and temporal behavior of the hole density.

The boundary conditions that the hole density $p(x,t)$ must satisfy at the p-i and n-i junctions and at the gate are

The hole density must be continuous at the gate, thus

$$p(x<l,t)\Big|_{x=l}=p(x>l,t)\Big|_{x=l} \tag{B.C.1}$$

The second boundary condition will be stated here and proved later.

$$\frac{d}{dx}p(x<l,t)\Big|_{x=l}=\frac{d}{dx}p(x>l,t)\Big|_{x=l}-\frac{I_{pr}}{D_pp_N}S(t) \tag{B.C.2}$$

where $S(t)$ is the Heaviside unit-step function;

$$S(t)=0, \qquad t<0,$$

$$S(t)=1, \qquad t>0.$$

The proof of Boundary Condition (B.C.2) is contained in Appendix A.

This condition is based on the assumption that reverse current that flows during Phase I, $J_r = qI_r$ is at least an order of magnitude less than the current that flows between the p-i junction and the gate, $J = pI(x < l,t)$. It is shown in Appendix A that this assumption has the consequence that the product pE can be treated as constant throughout the device.

This means that E can become large and $p(x,t)$ can become small as long as they do it in such a way that pE remains constant. When reverse current starts to flow out of the gate at the beginning of Phase I, the current flowing through the device consists of diffusion and drift components. It has been shown in Reference (1) that the drift component is much larger than the diffusion component, in fact orders of magnitude larger. It has also been shown that the field at the gate is much larger than it is at the n-i junction, which causes those holes in the drift component of the current to move by the gate much faster than those in the diffusion component. Since the reverse current flows out of the gate by diffusion, the reverse current is due to the difference in the diffusion currents approaching and leaving the gate, since the drift component is moving too quickly to participate in the much slower diffusion process.

Because the diffusion component is such a small part of the total current flowing through the device, this switching mechanism can never be fast (millimicroseconds) unless the reverse current at the gate affects the drift component as well as the diffusion component. This would lead to a discontinuity in the field and Boundary Condition (B.C. 2) would no longer hold.

The other two boundary conditions are derived in Reference (1) and, in effect, constitute the basis of the solution of the p-i-n diode problem. It should be pointed out that without Equations (12) and (13) in Reference (1), i.e., the hole concentrations in terms of injection parameters and particle density flow, this analysis would not be possible. At $x = -c$,

$$p(-c,t) = \frac{\sqrt{1-\gamma}}{R}\left[\frac{I}{I_{ps}}\right]^{1/2} \qquad \text{(B.C.3)}$$

which holds until $p(l,t) = 0$ since I is constant. At $x = +c$,

$$p(+c,t) = \sqrt{1-\xi}\left[\frac{I - I_r}{I_{ps}}\right]^{1/2}, \qquad \text{(B.C.4)}$$

since I and I_r are both constant during Phase I.

Let $P(x,s)$ be the Laplace transform of $p(x,t)$, i.e.,

$$P(x,s) = \int_0^\infty e^{-st}\, p(x,t)\, dt. \tag{8}$$

Then the transformed equations of Equations (7) and Boundary Conditions (B.C.1) through (B.C.4) are

$$\frac{d^2P(x,s)}{dx^2} - \frac{K^2}{L^2} P(x,s) = \frac{\tau}{L^2} p(x,0) \tag{9}$$

$$P(x<l,s)\Big|_{x=l} = P(x>l,s)\Big|_{x=l} \tag{B.C.5}$$

$$\frac{d}{dx} P(x<l,s)\Big|_{x=l} = \frac{d}{dx} P(x>l,s)\Big|_{x=l} - \frac{I_{pr}}{sD_p p_N} \tag{B.C.6}$$

$$P(-c,s) = \frac{\sqrt{1-\gamma}}{Rs}\left[\frac{I}{I_{ps}}\right]^{1/2} \tag{B.C.7}$$

$$P(+c,s) = \frac{\sqrt{1-\xi}}{s}\left[\frac{I-I_r}{I_{ps}}\right]^{1/2} \tag{B.C.8}$$

where

$$K^2 = 1 + \tau s. \tag{10}$$

The solution to Equation (9) is

$$P(x<l,s) = A_1 \sinh\frac{K}{L}(c-x) + A_2 \sinh\frac{K}{L}(c+x) + \frac{1}{s} p(x,0) \tag{11}$$

$$P(x>l,s) = B_1 \sinh\frac{K}{L}(c-x) + B_2 \sinh\frac{K}{L}(c+x) + \frac{1}{s} p(x,0). \tag{12}$$

Application of Boundary Conditions (B.C.5) and (B.C.6) yields two equations in the four unknowns A_1, A_2, B_1, and B_2. These constants are evaluated (see Appendix B) using the remaining boundary conditions. Let

$$A = \sqrt{1-\xi}\left[\frac{I}{I_{ps}}\right]^{1/2}\left[1-\left(1-\frac{I_r}{I}\right)^{1/2}\right],$$

then the constants are

$$A_1 = 0,$$

$$A_2 = -\frac{A}{s \sinh K\lambda} - \frac{I_{p_1} L \sinh \dfrac{K}{L}(c-l)}{p_N s K D_p \sinh K\lambda}$$

$$B_1 = -\frac{I_{p_1} L \sinh \dfrac{K}{L}(c-l)}{p_N s K D_p \sinh K\lambda},$$

$$B_2 = -\frac{A}{s \sinh K\lambda}.$$

The transformed hole density in $-c < x < c$ is

$$p(x<l,s) = -\frac{A \sinh \dfrac{K}{L}(c+x)}{s \sinh K\lambda}$$

$$-\frac{I_{p_1} L}{D_p p_N} \frac{\sinh \dfrac{K}{L}(c \mp L) \sinh \dfrac{K}{L}(c \pm x)}{sK \sinh K\lambda} + \frac{1}{s} p(x,0); \qquad (13)$$

where the upper set of signs apply to the region $-c < x < l$, and the lower set of signs to the region $l < x < c$. The Inversion Theorem[16] applied to Equation (13) yields, for the Phase I hole density $p_1(x,t)$,

$$p_1(x<l,t) = p(x,0) - A\left[\frac{\sinh\left(\dfrac{c+x}{L}\right)}{\sinh \lambda} + \frac{2}{\lambda}\sum_{n=0}^{\infty} \frac{\dfrac{(-1)^n n\pi}{\lambda}}{1+\dfrac{\pi^2 n^2}{\lambda^2}} \sin\left[\frac{n\pi}{2}\left(1+\frac{x}{c}\right)\right] \exp\left\{-\left(1+\frac{\pi^2 n^2}{\lambda^2}\right)\frac{t}{\tau}\right\}\right]$$

[16] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*, p. 239-271, Clarendon Press, Oxford, England (second edition).

$$-\frac{I_{pr}L}{D_p p_N}\left[\frac{\sinh\left(\frac{c\mp l}{L}\right)\sinh\left(\frac{c\pm x}{L}\right)}{\sinh\lambda}+\frac{2}{\lambda}\sum_{n=0}^{\infty}\frac{(-1)^n}{1+\frac{\pi^2n^2}{\lambda^2}}\right.$$

$$\left.\sin\left[\frac{n\pi}{2}\left(1\mp\frac{l}{c}\right)\right]\sin\left[\frac{n\pi}{2}\left(1\pm\frac{x}{c}\right)\right]\exp\left\{-\left(1+\frac{\pi^2n^2}{\lambda^2}\right)\frac{t}{\tau}\right\}\right]; \tag{14}$$

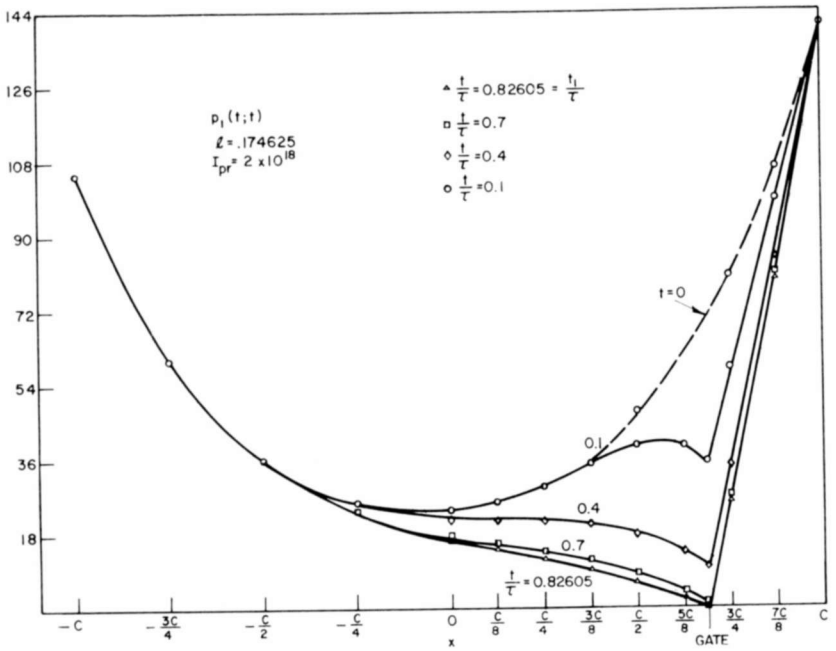


Fig. 2—Plot of $p_1(x,t)$ given by Equation (14) with the gate positioned at $l = (11/16)c = 0.175$ cm ($c = 0.254$ cm) and a reverse hole particle density $I_{pr} = 2 \times 10^{18}$ particles/cm²/sec. $p_1(x,t)$ is plotted for five different time intervals. The hole lifetime $\tau = 200$ microseconds.

where again the upper set of signs applies to the region $-c < x < l$, and the lower set of signs to the region $l < x < c$.

Figure 2 is a plot of $p_1(x,t)$, i.e., Equation (14), with the gate positioned at $l = (11/16)c = 0.175$ cm when $c = 0.254$ cm and a reverse hole particle density $I_{pr} = 2 \times 10^{18}$ particles/cm²/sec. Since the current flowing through the device has a particle density/sec of $I = 10^{21}$, $I_{pr}/I = 2 \times 10^{-3}$ so that $I_{pr} \ll I$ and Boundary Condition (B.C.2)

holds. Figure 2 has five different hole distributions plotted up for five different times: $t/\tau = 0$, 0.1, 0.4, 0.7, and 0.826. τ is the hole lifetime. These curves show that $p_1(l,t)$ vanishes when $t/\tau = 0.826$. The duration of Phase I is thus $t_1 = 0.826\tau$.

Figure 3 is a plot of Equation (14), with the gate positioned at $l = c/2 = 0.127$ cm when $c = 0.254$ cm and with the same reverse

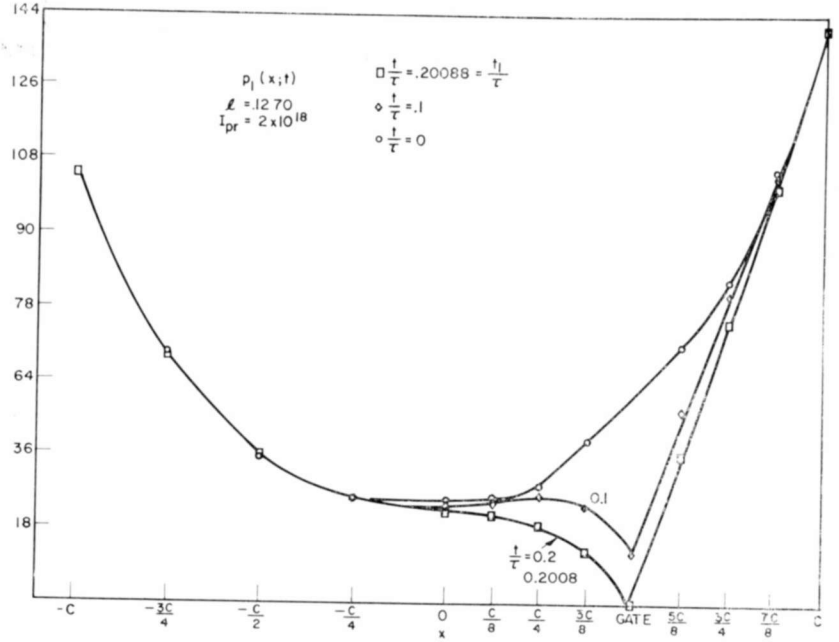


Fig. 3—Plot of Equation (14) with the gate positioned at $l = c/2 = 0.127$ cm ($c = 0.254$ cm). The reverse current is the same as in Figure 2. $p_1(x,t)$ is plotted for four different time intervals. The hole lifetime $\tau = 200$ microseconds.

current flowing out of the gate as in Figure 2. These curves show that $p_1(l,t)$ vanishes when $t/\tau = 0.2$. The duration of Phase I when the gate is placed at $l = c/2$ is four times shorter than when the gate is placed at $l = (11/16)c$.

The reason for this difference is that the hole density at $l = c/2$ is less than that at $l = (11/16)c$. It should be noted that the closer the gate is to the n-i junction at $x = +c$, the higher the hole density must be to preserve charge neutrality for the incoming electrons. Farther away from the n-i junction, recombination enters into play making it easier to deplete the region beneath the gate.

Figure 4 shows the hole distribution with the gate located at

$l = c/2$ and with a reverse particle density $I_{pr} = 2 \times 10^{19}$ particles/cm^2/sec. I_{pr} is thus a factor of 50 less than I. These curves show that $p_1(l,t) = 0$ when $t/\tau = 2.02 \times 10^{-3}$. Hence, the duration of Phase I, $t_1 = 2.02 \times 10^{-3}\tau$ for $I_{pr} = 2 \times 10^{19}$ particles/cm^2/sec.

The relative maximum located at $x \approx (3/8)c$ and the shoulder at

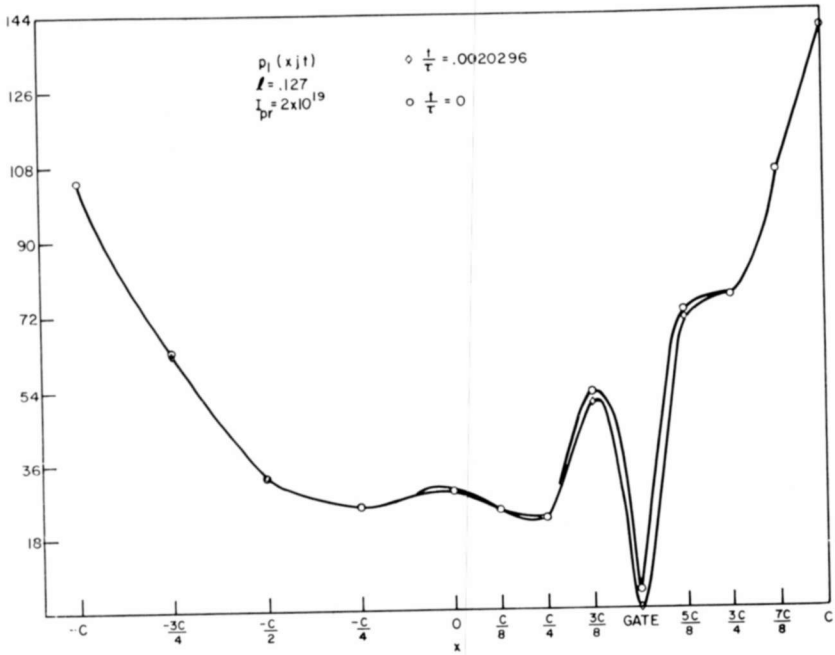


Fig. 4—Plot of Equation (14) with the gate positioned at $l = c/2 = 0.127$ cm ($c = 0.254$ cm) and a reverse hole particle density $I_{pr} = 2 \times 10^{19}$ particles/cm^2/sec. The hole lifetime $\tau = 200$ microseconds.

$x \approx (5/8)c$ is due to the fact that the holes have left the gate so rapidly that the field, which goes inversely as the hole density,[1] has increased just as rapidly, thus sweeping out all electrons at the gate not required for charge neutrality.

These electrons cause the pair density to build up to a relative maximum at $x \approx (3/8)c$. Similarly the holes swept away from the gate cause the shoulder at $x \approx (5/8)c$.

PHASE-II ANALYSIS

The dynamics of Phase II has been described in the Introduction.

The boundary conditions that the hole density must satisfy are

$$p_2(x<l,t_1)\Big|_{x=l} = p_2(x>l,t_1)\Big|_{x=l} \qquad \text{(B.C.9)}$$

Note that t_1 is the time when Phase I ends.

$$p_2(x=l,t') = 0 \qquad t_1 < t' < \infty \qquad \text{(B.C.10)}$$

$$p_2(x=c,t') = p_N \qquad t_1 < t' < \infty \qquad \text{(B.C.11)}$$

$$p_2(x=-c,t') = p_N \qquad t_1 < t' < \infty \qquad \text{(B.C.12)}$$

$$\begin{aligned} p_2(x<l,t) &= p_1(x<l,t) \\ p_2(x>l,t) &= p_1(x<l,t) \end{aligned} \qquad \text{(B.C.13)}$$

Boundary Condition (B.C.9) simply states that the hole density is continuous at the gate. Boundary Condition (B.C.10) states that the hole density is zero at the gate and remains zero once Phase I ends.

Boundary Condition (B.C.11) is a consequence of the fact that since the potential at the gate is zero, the null potential difference between gate and n-i junction means the forward-bias on the n-i junction vanishes, resulting in $p_c \to p_N$. Reference (1) shows that when $V_c \to 0$, $p_c \to p_N$. Boundary Condition (B.C.12) follows from (B.C.11). Since the n-i junction is shut off, space-charge neutrality requires that the p-i junction must also shut off. Reference (1) also shows that when $V_{-c} \to 0$, $p_{-c} \to p_N$. Boundary Condition (B.C.13) simply states that the hole density throughout the intrinsic region at the end of Phase I is equal to that at the beginning of Phase II.

The hole density throughout the intrinsic region, except at the gate, satisfies Equation (7). The Laplace transform of this equation with respect to time for Phase-II operation is

$$\frac{d^2P_2(x,s)}{dx^2} - \frac{K^2}{L^2}P_2(x,s) = -\frac{\tau}{L^2}p_2(x,t_1) = -\frac{\tau}{L^2}p_1(x,t_1), \qquad (15)$$

since Boundary Condition (B.C.13) states that the hole density at the end of Phase I is equal to that at the beginning of Phase II.

A general solution to Equation (15) is

$$P_2(x<l,s) = A_1 \sinh\frac{K}{L}(c-x) + A_2 \sinh\frac{K}{L}(c+x)$$

$$-\frac{\tau}{KL}\int_l^x p_1(\xi<l,t_1)\sinh\frac{K}{L}(x-\xi)\,d\xi \qquad \text{(16a)}$$

$$P_2(x > l,s) = B_1 \sinh \frac{K}{L}(c - x) + B_2 \sinh \frac{K}{L}(c + x)$$

$$-\frac{\tau}{KL}\int_l^x p_1(\xi > l,t_1) \sinh \frac{K}{L}(x - \xi)\,d\xi. \tag{16b}$$

Since the hole densities used in the Phase-I analysis were really the dimensionless hole density given by $p(x,t)/p_N$, where p_N is the thermal equilibrium hole density in the intrinsic region, Boundary Conditions (B.C.11) and (B.C.12) become

$$p_2(x = \pm c,t') = 1,$$

$$t_1 < t' < \infty.$$

Taking the Laplace transform of Boundary Conditions (B.C.10) and (B.C.11) (with the right-hand side equal to unity) yields

$$P_2(x = l,s) = 0, \tag{B.C.14}$$

$$P_2(x = c,s) = \frac{1}{s}. \tag{B.C.15}$$

Applying these boundary conditions to $P_2(x > l,s)$ given in Equation (16) permits evaluation of B_1 and B_2; thus

$$P_2(x > l,s) = \left[\frac{\sinh \frac{K}{L}(c + x)}{\sinh K\lambda} - \frac{\sinh \frac{K}{L}(c + l) \sinh \frac{K}{L}(c - x)}{\sinh K\lambda \sinh \frac{K}{L}(c - l)}\right]\left[\frac{1}{s}\right.$$

$$\left. + \frac{\tau}{KL}\int_l^x p_1(\xi > l,t_1) \sinh \frac{K}{L}(c - \xi)\,d\xi\right]$$

$$-\frac{\tau}{KL}\int_l^x p_1(\xi > l,t_1) \sinh \frac{K}{L}(x - \xi)\,d\xi. \tag{17}$$

Consider now the integral

$$J(c,l) = \int_l^c p_1(\xi > l,t_1) \sinh \frac{K}{L}(c-\xi)\,d\xi.$$

Integration by parts using

$$p_1(l,t_1) = 0, \tag{B.C.10}$$

yields

$$J(c,l) = -\frac{L}{K} p_1(c,t_1) + \frac{L}{K} \int_l^c \cosh \frac{K}{L}(c-\xi) \frac{d}{d\xi} p_1(\xi > l,t_1)\,d\xi.$$

Reference to Figures 2 and 3 shows that at the end of Phase I, not only is $p_1(l,t_1) = 0$, but $p_1(x > l,t_1)$ can be approximated very closely by a straight line, i.e., $p_1(x > l,t_1)$ varies almost linearly in the region $l < x < c$. The diffusion current at time $t = t_1$ must be constant throughout $l < x < c$ because of the straight-line variation in $p_1(x > l,t_1)$, i.e.,

$$I_{pd}(\xi > l,t_1) = -D_p \frac{d}{d\xi} p_1(\xi > l,t_1) = \text{constant}. \tag{18}$$

The subscript d in I_{pd} refers to diffusion. The drift current in the region $l < x < c$ during Phase II is zero because the field vanishes in that region due to the null potential difference between gate and n-i junction. Substituting for $(d/d\xi)p_1(\xi > l,t_1)$ in terms of $I_{pd}(\xi > l,t_1)$ in terms of $I_{pd}(\xi > l,t_1)$ into Equation (18) and treating $I_{pd}(\xi > l,t_1)$ as a constant in the region $l < \xi < c$ yields

$$J(c,l) = -\frac{L}{K} p_1(c,t_1) - \frac{L}{KD_p} I_{pd}(\xi > l,t_1) \int_l^c \cosh \frac{K}{L}(c-\xi)\,d\xi$$

$$= -\frac{L}{K} p_1(c,t_1) - \frac{L^2}{K^2 D_p} I_{pd}(\xi > l,t_1) \sinh \frac{K}{L}(c-l), \tag{19}$$

and

$$J(x,l) = -\frac{L}{K}\, p_1(x,t_1) - \frac{L^2}{K^2 D_p}\, I_{pd}(x > l, t_1) \sinh \frac{K}{L}(x-l). \tag{20}$$

Substituting Equations (19) and (20) into Equation (17) yields

$$P_2(x > l, s) = \frac{\sinh \frac{K}{L}(c+x)}{s \sinh K\lambda} - \frac{\tau\, p_1(c, t_1) \sinh \frac{K}{L}(c+x)}{K^2 \sinh K\lambda}$$

$$- \frac{\tau L\, I_{pd}(x > l, t_1) \sinh \frac{K}{L}(c-l) \sinh \frac{K}{L}(c+x)}{K^3 D_p \sinh K\lambda}$$

$$- \frac{\sinh \frac{K}{L}(c+l) \sinh \frac{K}{L}(c-x)}{s \sinh K\lambda \sinh \frac{K}{L}(c-x)}$$

$$+ \frac{\tau\, p_1(c, t_1) \sinh \frac{K}{L}(c+l) \sinh \frac{K}{L}(c-x)}{K^2 \sinh K\lambda \sinh \frac{K}{L}(c-l)} + \frac{\tau}{K^2} p_1(x, t_1)$$

$$+ \frac{\tau L\, I_{pd}(x > l, t_1) \sinh \frac{K}{L}(c-l) \sinh \frac{K}{L}(c+l) \sinh \frac{K}{L}(c-x)}{D_p K^3 \sinh K\lambda \sinh \frac{K}{L}(c-l)}$$

$$+ \frac{\tau L\, I_{pd}(x > l, t_1) \sinh \frac{K}{L}(x-l)}{D_p K^3}$$

Applying the Inversion Theorem[16] to $P_2(x > l, s)$ yields the Phase-II hole density $p_2(x > l, t')$. Since the field is zero in $l < x < c$ the current that flows in that region is

$$I_2\ (x > l, t') = -\ (D_p - D_n)\ \frac{d}{dx}\ p_2(x > l, t'),\ \text{since}\ n = p.$$

Note that $I_{pd}\ (x > l, t_1)$ is treated as constant in $l < x < c$ at the end of Phase I.

Since $p_1\ (x > l, t_1)$ varies linearly,

$$\frac{d}{dx}\ p_1\ (x > l, t_1) = \frac{p_1\ (c, t_1) - p_1\ (l, t_1)}{c - l} = \frac{p_1\ (c, t_1)}{c - l},$$

since $p_1\ (l, t_1) = 0$.

Thus $$I_{pd}\ (x > l, t_1) = -\frac{D_p\ p_1\ (c, t_1)}{c - l},$$

and hence $$I_d\ (x > l, t_1) = -\frac{(D_p - D_n)\ p_1\ (c, t_1)}{c - l}. \qquad (21)$$

$$I_2\ (x > l, t') = -\frac{(D_p - D_n)\ \cosh\left(\frac{x - l}{L}\right)}{\sinh\left(\frac{c - l}{L}\right)} + \frac{3\ (D_p - D_n)}{c - l}\ p_1\ (c, t_1)\ \exp\left\{-\frac{t'}{\tau}\right\}$$

$$+ \frac{3l\ (D_p - D_n)}{c\ (c - l)}\ p_1\ (c, t_1)\ \exp\left\{-\frac{t'}{\tau}\right\} + \frac{D_p - D_n}{c}\ p_1\ (c, t_1) \sum_{n=1}^{\infty} (-1)^n$$

$$\cos\left[\frac{n\pi}{2}\left(1 + \frac{x}{c}\right)\right] \exp\left\{-\left(1 + \frac{n^2\pi^2}{\lambda^2}\right)\frac{t'}{\tau}\right\} + \frac{D_p - D_n}{c} \sum_{n=1}^{\infty}\left(\cos\left[\frac{n\pi}{2}\left(1 - \frac{x}{c}\right)\right]\right.$$

$$\left. - (-1)^n \cos\left[\frac{n\pi}{2}\left(1 + \frac{x}{c}\right)\right]\right)\left(\frac{n\pi}{\lambda}\right)^2 \left(1 + \frac{n^2\pi^2}{\lambda^2}\right)^{-1} \exp\left\{-\left[\frac{n^2\pi^2}{\lambda^2}\right]\frac{t'}{\tau}\right\}$$

$$-\frac{4\ (D_p - D_n)}{(c - l)}\ p_1\ (c, t_1) \sum_{n=0}^{\infty} \frac{1}{(2n + 1)\pi} \exp\left\{-\left[1 + \frac{(2n + 1)^2\pi^2}{\lambda^2}\right]\frac{t'}{\tau}\right\}$$

$$\cos\left[\frac{(2n + 1)\ \pi l}{2c}\right] \sin\left[\frac{(2n + 1)\ \pi x}{2c}\right] + \frac{4\ (D_p - D_n)}{c - l}\ p_1\ (c, t_1) \sum_{n=1}^{\infty} \frac{1}{2\pi n}$$

$$\exp\left\{-\left[1+\left(\frac{2\pi n}{\lambda}\right)^2\right]\frac{t'}{\tau}\right\}\sin\left[\frac{n\pi l}{c}\right]\cos\left[\frac{n\pi x}{c}\right]-\frac{2\,(D_p-D_n)}{c-l}\sum_{n=1}^{\infty}\left(\frac{n\pi l}{c-l}\right)^2$$

$$\cos\left[\frac{n\pi\,(c-x)}{c-l}\right]\left[1+\left(\frac{n\pi l}{c-l}\right)^2\right]^{-1}\exp\left\{-\left[1+\left(\frac{n\pi L}{c-l}\right)^2\right]\frac{t'}{\tau}\right\}$$

$$\frac{2\,(D_p-D_n)}{c-l}\,p_1\,(c,t_1)\sum_{n=1}^{\infty}\exp\left\{-\left[1+\left(\frac{n\pi L}{c-l}\right)^2\right]\frac{t'}{\tau}\right\}\cos\left[\frac{n\pi\,(c-x)}{c-l}\right]$$

$$\frac{D_p-D_n}{c}\,p_1\,(c,t_1)\sum_{n=1}^{\infty}\exp\left\{-\left[1+\left(\frac{n\pi}{\lambda}\right)^2\right]\frac{t'}{\tau}\right\}\cos\left[\frac{n\pi}{2}\left(1-\frac{x}{c}\right)\right] \tag{22}$$

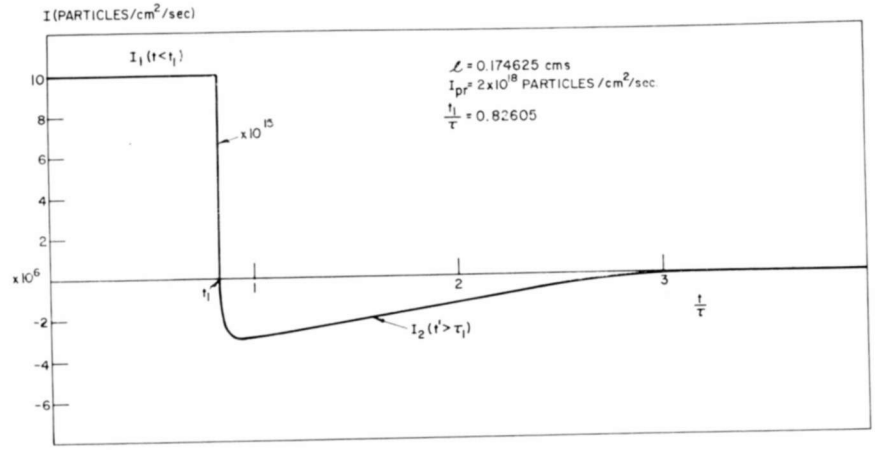


Fig. 5—Plot of Phase I and II currents. Phase-I current is constant for $t < t_1$. Phase-II current is given by Equation (21). Phase-I hole densities are shown in Figure 2.

Figures 5 and 6 are two graphs of plots of Phase-II current obtained by programming Equation (21) on a computer. The Phase II particle density was plotted for $x = (c + l)/2$, i.e., half-way between gate and n-i junction. When $l = (11/16)\ c = 0.174625$ cm and $I_{pr} = 2 \times 10^{18}$ particles/cm²/sec, Figure 5 shows that the Phase-II current at $x = (c + l)/2$, $I_2\ ((c + l)/2, t')$ is 15 orders of magnitude less than Phase-I current. Even though it takes several lifetimes for the Phase-II current to drop to zero, its amplitude is so much smaller than

Phase-I current that the switching time is effectively the duration of Phase I, t_1.

Figure 6 shows the same result as Figure 5. Since the duration of Phase I is clearly the switching time of the device, switching times (duration of Phase-I current) of $2.02 \times 10^{-3}\ \tau$ are distinctly possible.

Acknowledgments

The author is indebted to Harwick Johnson for suggesting the problem and for many interesting discussions. Tom Stiller is to be thanked for programming the calculations.

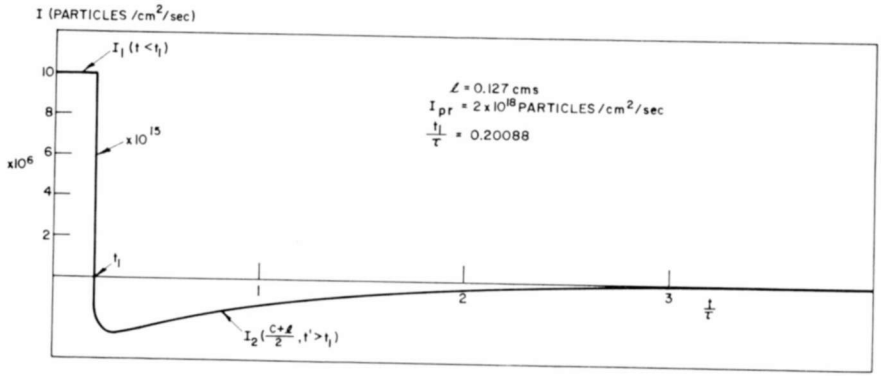


Fig. 6—Plot of Phase I and II currents. Phase-I current is constant for $t < t_1$. Phase-II current is given by Equation (21). Phase-I hole densities are shown in Figure 3.

Appendix A

It has been shown in References (1) and (14) that

$$\beta E = \frac{bI}{D_n\,(b+1)\,p} - \left[\frac{b-1}{b+1}\right]\frac{1}{p}\frac{dp}{dx}. \tag{23}$$

In Reference (1) it was shown that

$\left[\dfrac{b-1}{b+1}\right]\dfrac{1}{p}\dfrac{dp}{dx}$ is approximately three orders of magnitude less than

$\dfrac{bI}{D_n\,(b+1)\,p}$.

Thus, E depends strongly on the reciprocal of p, and the product βpE is constant to a very high degree, i.e., to within three significant figures. Thus, for our present purposes, we can neglect the last terms when evaluating E.

When reverse current is flowing out of the gate, in the region $x < l$,

$$\beta E\ (x < l, t) = \frac{b\ I\ (x < l, t)}{D_n\ (b + 1)\ p\ (x < l, t)}, \tag{24}$$

while in the region $x > l$,

$$\beta E\ (x > l, t) = \frac{b\ [I\ (x < l, t) - I_r]}{D_n\ (b + 1)\ p\ (x > l, t)}. \tag{25}$$

If I_r is within an order of magnitude of $I\ (x < l, t)$; then

$$\beta E\ (x > l, t) = \frac{b\ I\ (x < l, t)}{D_n\ (b + 1)\ p\ (x > l, t)} \left[1 - \frac{I_r}{I\ (x < l, t)} \right]$$

$$\simeq \frac{b\ I\ (x < l, t)}{D_n\ (b + 1)\ p\ (x > l, t)},$$

since $I_r/I\ (x < l, t) \ll 1$; this leads to

$$\beta p\ (x < l, t)\ E\ (x < l, t) = \beta p\ (x > l, t)\ E\ (x > l, t)$$

$$= \frac{b\ I\ (x < l, t)}{D_n\ (b + 1)}. \tag{26}$$

During Phase I, $I(x < l, t)$ is constant; therefore, the product βpE is constant.

This is the key factor in deriving Boundary Conditions (B.C. 2). During Phase I, the currents at the gate must satisfy

$$I_p\ (x < l, t) \Big|_{x=l} = I_p\ (x > l, t) \Big|_{x=l} + I_{pr}\ S\ (t), \tag{27}$$

where I_{pr} is the hole component of I_r, the number of particles/sec/cm^2 contained in the reverse current. $S(t)$ is the Heaviside unit step function defined by

$$S(t) = 0 \text{ for } t < 0; \qquad S(t) = 1 \text{ for } t > 0.$$

From Equation (1), at $x = l$,

$$-D_p \frac{d}{dx} p(x<l,t) + \beta D_p p(x<l,t)\ E\ (x<l,t)$$

$$= -D_p \frac{d}{dx} p(x>l,t) + \beta D_p p(x>l,t)\ E\ (x>l,t) + I_{pr} S(t)\,;$$

but at $x = l$,

$$\beta D_p p(x<l,t)\ E\ (x<l,t) = \beta D_p p(x>l,t)\ E\ (x>l,t) = \frac{bI(x<l,t)}{D_n(b+1)}.$$

Hence the two terms involving pE cancel, leaving

$$\frac{d}{dx} p(x<l,t)\bigg|_{x=l} = \frac{d}{dx} p(x>l,t)\bigg|_{x=l} - \frac{I_{pr}}{D_p} S(t).$$

If $p(x,t)$ now denotes the hole density divided by p_N, then

$$\frac{d}{dx} p(x<l,t)\bigg|_{x=l} = \frac{d}{dx} p(x>l,t)\bigg|_{x=l} - \frac{I_{pr}}{D_p p_N} S(t).$$

Note that the last term is now divided by p_N.

Appendix B

Applying Boundary Condition (B.C.7) to Equation (11) allows A_1 to be evaluated at $x = -c$.

$$A_1 \sinh K\lambda + \frac{1}{s} p(-c,0) = \frac{\sqrt{1-\gamma}}{Rs} \left[\frac{I}{I_{ps}} \right]^{1/2} \tag{28}$$

Boundary Condition (B.C.3) states that until $p(l,t) = 0$

$$p(-c,t) = \frac{\sqrt{1-\gamma}}{R} \left[\frac{I}{I_{ps}} \right]^{1/2} = p(-c,0),$$

i.e., Boundary Condition (B.C.3) holds for $t \geq 0$ until $p(l,t)$ vanishes.

Substituting for $p(-c,0)$ into Equation (28) gives $A_1 = 0$. Apply-

ing Boundary Condition (B.C.8) to Equation (12), observing that Boundary Condition (B.C.4) holds for $t \geq 0$ until $p(l,t)$ vanishes, yields

$$B_2 = \frac{\sqrt{1-\xi}}{s \sinh K\lambda}\left[\left(\frac{I-I_r}{I_{ps}}\right)^{1/2} - \left(\frac{I}{I_{ps}}\right)^{1/2}\right] \tag{24}$$

Applying Boundary Conditions (B.C.5) and (B.C.6) to Equations (11) and (12) results in the following equations:

$$A_1 \sinh \frac{K}{L}(c-l) + A_2 \sinh \frac{K}{L}(c+l) = B_1 \sinh \frac{K}{L}(c-l) + B_2 \sinh \frac{K}{L}(c+l) \tag{30}$$

$$-A_1 \frac{K}{L}\cosh \frac{K}{L}(c-l) + A_2 \frac{K}{L}\cosh \frac{K}{L}(c+l) = -B_1 \frac{K}{L}\cosh \frac{K}{L}(c-l) + B_2 \frac{K}{L}\cosh \frac{K}{L}(c+l) - \frac{I_{pr}}{sD_p p_N}. \tag{31}$$

Substituting $A_1 = 0$ into these equations along with the expression for B_2 given by Equation (29) leaves

$$A_2 \sinh \frac{K}{L}(c+l) - B_1 \sinh \frac{K}{L}(c-l) = \frac{\sqrt{1-\xi}}{s \sinh K\lambda}\left[\frac{I}{I_{ps}}\right]^{1/2}\left[\left(1-\frac{I_r}{I}\right)^{1/2} - 1\right]\sinh \frac{K}{L}(c+l) \tag{32}$$

$$A_2 \cosh \frac{K}{L}(c+l) + B_1 \cosh \frac{K}{L}(c-l) = \frac{\sqrt{1-\xi}\cosh \frac{K}{L}(c+l)}{s \sinh K\lambda}\left[\frac{I}{I_{ps}}\right]^{1/2}\left[\left(1-\frac{I_r}{I}\right)^{1/2} - 1\right] - \frac{I_{pr}L}{sKD_p p_N}. \tag{33}$$

Solving Equations (32) and (33) for A_2 and B_1 gives the values cited earlier.

APPLICATION OF CONTENT-ADDRESSED MEMORY FOR DYNAMIC STORAGE ALLOCATION

By

Yaohan Chu*

Summary—*This paper describes the application of a small-capacity content-addressed memory for dynamic storage allocation. The high-speed random-access memory consists of a hierarchy of ordinary words and macrowords, the latter being allocation units. A map of many chains of macrowords is stored in the content-addressed memory. A technique is presented in the paper that can allow automatic storage allocation and release of macrowords dynamically. The allocated program sequences can physically be stored in noncontiguous areas of the high-speed memory, and yet each sequence is protected from the others. A unique feature of the technique is that the original floating code remains unchanged during allocation and reallocation because of hardware-implemented address interpretation. The technique described in the paper may considerably reduce the task of "executive control".*

Introduction

When a sequence of instructions (which can be a program or a program segment) is to be executed by a processor, the instructions, together with data, must first be loaded into the HS (high-speed random-access) memory. By storage allocation, we mean the assignment of an absolute address of the HS memory to each word of the sequence and often the corresponding change of the operand address of the instruction of the sequence. If this assignment is made immediately before execution, or if the allocation can be extended during execution as the need arises, this is known as dynamic storage allocation of the HS memory. With dynamic storage allocation, the actual location of the sequence in the HS memory is not known until execution time.

There are a number of problems in allocating HS memory locations to the words of the sequence. One must know which HS memory locations are available for allocation. Are the available locations large enough? If so, how can they conveniently be allocated and still be protected from one program sequence to another, since these available locations are usually scattered rather than contiguous? After the HS memory addresses are assigned to the sequence, it is necessary to

* Formerly RCA Data Systems Center, Bethesda, Md., now associated with Control Data Corporation.

change the operand addresses. It may also be necessary to increase the number of allocated HS memory locations during the execution. There is the problem of storage release, the problem of sharing a portion of a sequence by other sequences, and the problem of reallocation of one or more sequences that have been dumped to make room for storing a longer sequence. This paper proposes a hardware approach to the solution of dynamic storage allocation of the HS memory.

When a problem-oriented language program is compiled, the object program is usually first in a floating code that is then converted into a fixed code by the assembly program. The proposed technique does not require conversion from the floating code to the fixed code. In this paper, it is assumed that there is an Executive Program which makes allocation decisions and selects a program sequence for execution by the processor. It is also assumed that the program sequences to be selected by the Executive Program are in floating codes with a single origin (i.e., each program sequence is ordered with addresses relative to an origin).

ORGANIZATION

The machine organization for dynamic storage allocation that is pertinent to this paper is shown in Figure 1. There is an HS memory and a CA (high-speed content-addressed) memory. The registers of interest are: HS memory address register, HS memory buffer register, next-instruction register, current-sequence-number register, available sequence counter, chain number counter, input register, mask register, and output register. The last three registers are associated with the CA memory. The use of these counters and registers is described.

High-Speed Random-Access Memory

The HS memory is conventional except that it is composed of macrowords that are allocation units. The characteristics of a representative HS memory are shown in Table I. The memory has 256 macrowords; each macroword consists of 128 48-bit words located contiguously. The memory address consists of two parts, an 8-bit macroword address for the 256 macrowords and a 7-bit local address for the 128 words in a macroword.

Content-Addressed Memory

The CA memory is one where bit comparison logic is part of the hardware at each memory element for simultaneous comparisons of all memory elements. The access to the memory is not by an address, but

Table I — Characteristics of the HS Memory

Characteristic	*Description*
storage medium	magnetic core
word length	48 bits
macroword length	128 contiguous words
capacity	256 macrowords or 32,768 words
macroword address	8 bits
local address	7 bits
total address	15 bits
cycle time	2 microseconds

by an interrogation word. The interrogation word is maskable by the combined use of the input register and the mask register as shown in Figure 1. Thus, if some known content is a portion of the interrogation word, the word or words that have the same known content can be read out of the memory.

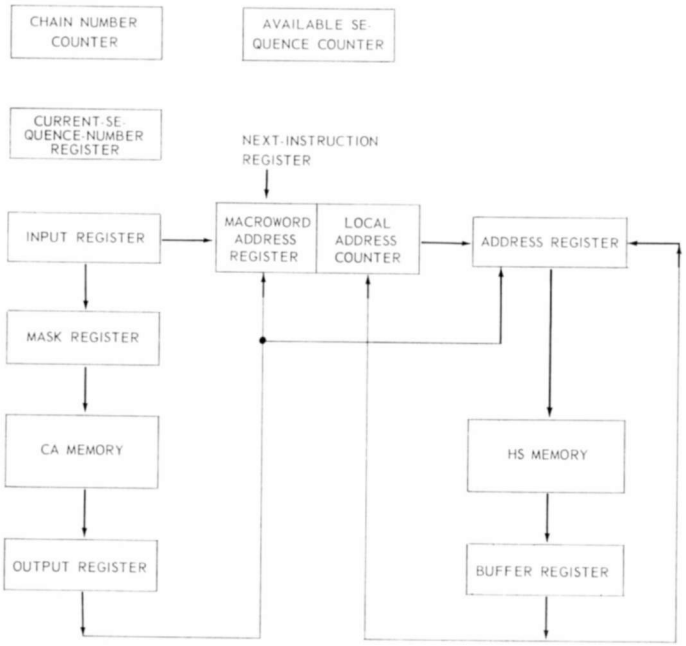


Fig. 1—Machine organization for dynamic storage allocation.

Table II — Characteristics of the CA Memory

Characteristic	*Description*
storage medium	magnetic thin film
word length	32 bits
capacity	512 words
associativity	fully associative and maskable
matching	single match only
operations	(a) read on equality (b) write a word (c) erase a word
cycle time	½ microsecond for any of the above operations

The characteristics of a representative CA memory are shown in Table II. The memory consists of 512 32-bit words. The contents of the entire CA memory can be stored in 2 macrowords of the HS memory. The 32 bits of the word consists of 5 fields; one or more fields can be used to interrogate the CA memory. The stored contents of each word of the memory are unique; therefore, multiple matching will not occur during interrogation. Operations required of the memory are: read on equality, write a word, erase a word. The capability required of the CA memory is relatively simple, except for the need of a very short cycle time.

The format of the CA memory word is shown in Figure 2. There are three 8-bit fields for macroword addresses A, B, and C; a 5-bit field for sequence number; and a 3-bit field for status indication. Macroword address A indicates the location of a macroword of the HS memory, and macroword address B the location of the next macroword. In other words, by means of 256 pairs of macroword addresses A and B, the 256 macrowords of the HS memory can be chained into one sequence of any desired manner. By means of the sequence number, the 256 macrowords can be chained into as many as 32 sequences. Macroword address C is used to indicate the order of the macroword

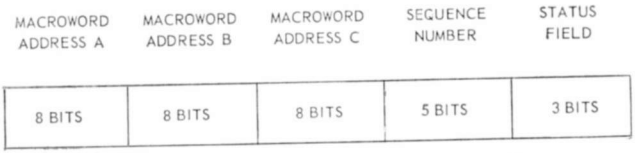


Fig. 2—Format of CA memory word.

Table III — Examples of Status Field Designation

Status Field	*Designation*
000	allocated
001	released
010	reserved
011	permanently allocated
100	available for allocation
101	shared
110	sequence number register

in a sequence, as will be further described. The status field is used to indicate the status of the corresponding macroword; examples of status field designation are shown in Table III.

An example showing the contents of the CA memory is shown in Figure 3, where it is assumed that there are only 12 macrowords in the HS memory. The four sequences of HS memory macrowords shown therein (the upper portion of Figure 3) are sequences 1-5-9,

MACROWORD ADDRESS A	MACROWORD ADDRESS B	MACROWORD ADDRESS C	SEQUENCE NUMBER	STATUS FIELD
0	*	3	3	011
1	5	1	1	000
2	10	1	4	100
3	11	1	3	011
4	8	2	2	001
5	9	2	1	000
6	4	1	2	001
7	*	3	4	100
8	*	3	2	001
9	*	3	1	000
10	7	2	4	100
11	0	2	3	011
1	w		1	110
6	x		2	110
3	y		3	110
2	z		4	110

Fig. 3—An example showing the contents of a CA memory (assume there are 12 macrowords).

6-4-8, 3-11-0, and 2-10-7 for sequence numbers 1, 2, 3, and 4 respectively. The asterisks in the macroword address B fields indicate the end of each sequence. As seen from the status designations in Table III, sequence number 4 contains the chained macrowords available for allocation (referred to as available sequence); sequence number 1 is an allocated sequence; sequence number 2 has been released and is ready to become a part of available sequence; sequence number 3 has been permanently allocated (for such use as the locations of the Executive Program). The available sequence may not use the macroword address C field.

For the example in Figure 3, four additional HS memory words (the lower portion of Figure 3) are used to store the first macroword addresses of these four sequences. The first macroword addresses are stored in the macroword address A fields. The macroword address B and C fields are used in combination to store return-after-interrupt addresses, as will be further explained. The sequence number and status fields remain the same as before. It is noted that the stored words in Figure 3 are not necessarily in the order shown therein and, in fact, the actual locations may not be known.

Storage Allocation

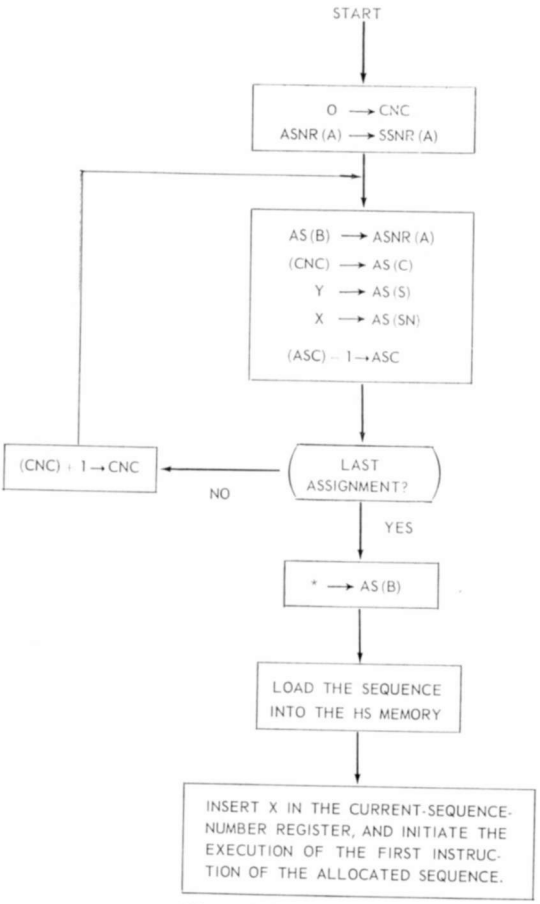
As mentioned, the sequence whose HS memory locations are to be assigned for execution is sequentially ordered with addresses relative to an origin. When the Executive Program selects a sequence, it compares the size of the sequence to be selected with the size of the available sequence; the latter is available from the sequence counter (see Figure 1). If there are enough locations, the Executive Program assigns a sequence number to the selected sequence and orders the allocation. If there are not enough locations, the Executive Program may have to select another sequence or initiate a dumping to provide more room before allocation is initiated. In any case, all sequences should have been segmented within an allowable maximum size.

The allocation is shown in the flow chart of Figure 4. The flow chart is as follows:

Step 1: This step initiates the allocations. There are two operations:

(a) set the contents of the chain number counter (CNC) to 0;

(b) insert macroword address A of the available sequence number register (ASNR) to the macroword address A field of the selected sequence number register (SSNR) so as to store the start address of the sequence whose locations are being assigned.



List of Symbols

( )	contents of	ASNR(A)	macroword address A of available sequence number register
→	transferred to		
A	macroword address A field	AS	the first CA memory word of the available sequence. It is located by the macroword address A of the available sequence number register.
B	macroword address B field		
C	macroword address C field		
S	status field		
SN	sequence number field	AS(B)	macroword address B of the first CA memory word of the available sequence.
ASNR	available sequence number register		
SSNR	selected sequence number register	X	the assigned sequence number
ASC	available sequence counter	Y	the status of the assigned macroword
CNC	chain number counter		

Fig. 4—Flow chart showing the storage allocation.

Step 2: This step assigns the macroword at the top of the available sequence to the 128 words at the top of the selected sequence. There are five operations:

(a) move the macroword address B of AS to macroword address A field of ASNR to establish the new beginning of the available sequence;

(b) insert the contents of the chain number counter to the macroword address C of AS;

(c) insert sequence number X to the sequence number field of AS;

(d) insert status Y to the status field of AS;

(e) step the available sequence counter (ASC) one count down.

Step 3: This step forms the loop by asking whether the assignment just made is the last assignment of the sequence;

(a) if it is not, step the chain number counter one count up, and then repeat step 2,

(b) if it is, insert an asterisk to macroword address B field of AS to signify the end of the allocated sequence.

Step 4: The Executive Program now orders the loading of the sequence to the allocated locations of the HS memory. When the loading is completed, the Executive Program inserts the sequence number in the current-sequence-number register and initiates execution by the processor of the instruction in the first local address of the first macroword of the allocated sequence.

As an example, a sequence of 600 words of instructions and data is to be allocated. Assume that sequence number 3 is assigned to the sequence, and the available sequence having a sequence number 9 is 56-34-103-6-89-201-153-55. The contents of the CA memory before allocation is illustrated in Figure 5, where 000 represents the allocated status. The first 128 words of the sequence are assigned to macroword address 56; the required operations for this assignment are also indicated in Figure 5. The second 128 words are assigned to macroword address 34, and similarly a third, fourth, and fifth 128 words are assigned, although 40 words in the fifth macroword are left unused. After the allocation of the sequence is completed, the contents of the CA memory are shown in Table IV.

Allocation Time

Allocation of each macroword requires about two CA memory cycles. By modifying the algorithm in Figure 4, allocation can be

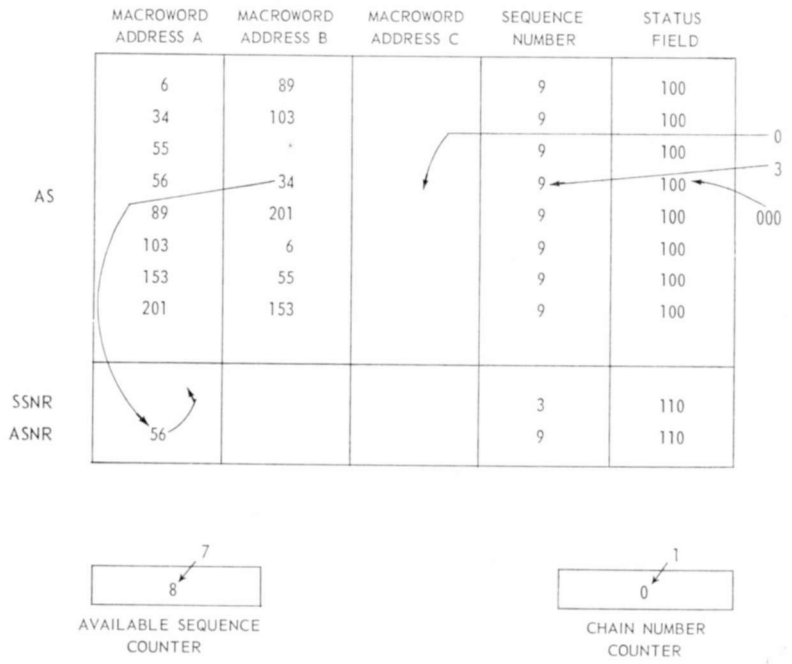


Fig. 5—Contents of the CA memory before allocation (arrows indicate operations required in assignment of first 128 words of sequence to macroword address 56).

reduced to about one CA memory cycle per macroword. For 256 macrowords, the allocation time is about 64 microseconds. For an average program sequence of 64 macrowords (or 8,192 HS memory words),

Table IV — Contents of the CA Memory After Allocation

	Macroword Address A	Macroword Address B	Macroword Address C	Sequence Number	Status Field
	56	34	0	3	000
allocated	34	103	1	3	000
sequence	103	6	2	3	000
	6	89	3	3	000
	89	*	4	3	000
available	201	153		9	100
sequence	153	55		9	100
	55	*		9	100
SSNR	56			3	110
ASNR	201			9	110

the allocation time is about 16 microseconds. Although this simple estimate of allocation time is not precise it indicates the order of magnitude of the allocation time.

Allocation During Execution

During the execution of a sequence, additional locations of the HS memory may be needed. The sequence can request the Executive Program to make additional allocation, and the Executive Program chains additional macrowords to the sequence.

Instruction Sequencing

Instruction sequencing is controlled by the next-instruction register (15 bits); the lower 7-bit of the register is a local-address counter for sequencing the local address, and the upper 8-bit is merely a register for storing the macroword address. The execution of a sequence begins at the first local address of the first macroword address. Within a macroword, the local address is advanced by the local-address counter. When the local-address counter recycles from the contents of 1111111 to 0000000, the macroword address should be changed.

The new macroword address is obtained from the CA memory. Two types of interrogation are required for interrogating the CA memory: macroword-address-A interrogation and macroword-address-C interrogation. The former is for instruction sequencing here; the latter for operand address interpretation. In the macroword-address-A interrogation, the interrogation fields are the macroword address A and the sequence number. The macroword address A is taken from the macroword address portion of the next-instruction register, and the sequence number from the current-sequence-number register. The output of this interrogation is the macroword address B of the interrogated CA memory word; this macroword address is transferred to the macroword address portion of the next-instruction register and then the instruction sequencing continues.

Memory Protection

Each macroword of the HS memory has a corresponding CA memory word, and each CA memory word has a sequence number field. Since each HS memory word is labeled by a sequence number, one sequence can not access the macrowords of other sequences and all 32 sequences are thus protected. Note that the macrowords of each protected sequence need not lie in one contiguous area of the HS memory.

Operand Address Interpretation

The instruction being executed by the processor has a 15-bit operand address. This address is the relative address of the original sequence and not the actual HS memory address, because the operand addresses of the original floating code have not been changed during the allocation. The operand address where the operand is actually stored in the HS memory must be found. The local address of the actual operand address, however, is the same as the original relative address which, now in the buffer register of the HS memory, is transferred to the local address field of the HS memory address register. The macroword address of the actual operand address is found by using the above-mentioned macroword-address-C interrogation. In this interrogation, the interrogation fields are the sequence number and the macroword address C. The sequence number is taken from the current-sequence-number register, and the macroword address C is taken from the macroword address field of the original operand address, now in the buffer register. The output of this interrogation is the macroword address B of the interrogated PS memory word; this macroword address is then transferred to the macroword address field of the address register of the HS memory.

Shared Macroword

By shared macroword, we mean that one macroword may be a part of two or more sequences. If a macroword is shared by m sequences, m CA memory words are required. This makes macroword address A alone nonunique for interrogation, and both macroword address A and sequence number are required in the macroword-address-A interrogation. The shared macroword provision requires more than 256 CA memory words for the 256 HS memory macrowords. Of the 512 CA memory words, 32 words are needed for the 32 sequence number registers, and the remaining 480 words are available for allocation use.

Branching

Branching may occur when a transfer instruction is encountered. In this case, the operand address is an instruction address. The operand address is again interpreted in the same manner, except the local address in the HS memory buffer register and the macroword address from the CA memory are now transferred to the next-instruction register, instead of the HS memory address register.

Indexing and Indirect Addressing

The operand address interpretation does not prevent the use of

indexing and indirect addressing. To include indexing, the contents of the index register are added to (or subtracted from) the original relative address before the operand address interpretation. For indirect addressing, an operand address is obtained from the operand address interpretation instead of an operand, and the operand address interpretation is repeated.

Program Switching

When a sequence is complete, an interrupt, which can be initiated by an instruction, occurs. The Executive Program takes over and selects the next sequence. If the next sequence is already in the HS memory, it is only necessary for the Executive Program to insert the sequence number to the current-sequence-number register and the start address to the next instruction register. If the next sequence is not in the HS memory, the Executive Program must perform the storage allocation and then load the sequence into the HS memory in the manner described previously.

Storage Release

When a sequence is completed and the Executive Program takes over, the latter may decide to release the locations of the sequence for other use. The release of such locations can be done by simply chaining them to the available sequence.

Program Return

During the execution of a sequence, an interrupt occurs when an instruction of the sequence requires data transfer to or from the mass memory (or other device). The Executive Program places the return address (both macroword and local addresses) in the combined macroword address B and C fields of the corresponding sequence number register as illustrated by symbols w, x, y and z in Figure 3. The Executive Program now selects another sequence to make use of the processor. When the data transfer is completed another interrupt occurs, and the Executive Program may initiate the processor to resume the original sequence by obtaining the return address from the appropriate sequence number register.

Dumping and Reallocation

When the Executive Program finds that the available sequence is not long enough for allocation, it can dump some sequences in the HS

memory to a mass memory. Reallocation of the dumped sequences does not present any problem since the values in the floating codes of the original references are not changed during the allocation, nor need the values be changed when the dumped sequences are again allocated.

Conclusions

The above-described technique for dynamic storage allocation can considerably reduce the task of executive control. Since address interpretation takes place at execution time by the hardware, no special effort is needed at assembly time for dynamic allocation and reallocation. The codes can be assembled relative to an origin. The assignment of a sequence number and chaining together of a series of macrowords into which consecutive macrowords of the sequence are loaded permits the floating code to be executed without being modified at the load time to adapt to the allocated HS memory locations. Furthermore, the sequence need not be specially segmented prior to loading, since the chain of macrowords into which the sequence are loaded are logically contiguous.

The storage allocation technique offers an important feature of memory protection for many program sequences, and each program sequence need not lie in a physically contiguous area of the HS memory. The status bits in the CA memory offer a convenient and logical place for record keeping and, if desired, can also be used for additional memory protection.

Since dynamic storage allocation function occupies a significant portion of the Executive Program, incorporation of hardware-oriented dynamic storage allocation can reduce the size of the Executive Program and the overhead time. It may also limit the need for HS memory capacity to a reasonable amount, not only as by shortening the Executive Program, but also by increasing use of the HS memory.

Acknowledgment

The author wishes to acknowledge stimulating discussions which he had with John Robert Goodroe, who provided valuable suggestions during the course of the work.

RCA Technical Papers†

Fourth Quarter, 1964

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Blue Gas Laser Using Hg^{2+}," H. J. Gerritsen and V. P. Goedertier, *Jour. Appl. Phys.* (Communications) (October) 1964
"Cryoelectric Memories," L. L. Burns, *Proc. IEEE* (October) 1964
"CW Operation of GaAs Injection Lasers," M. F. Lamorte, R. B. Liebert, and T. Gonda, *Proc. IEEE* (Correspondence) (October) 1964
"Design Considerations for Transistorized Television Deflection Circuits, Part I," C. F. Wheatley, *Solid/State/Design* (October) 1964
"Diplexed TV Antenna Radiates WHEC and WROC Signals," B. K. Kellom, *Broadcast News* (October) 1964
"High-Speed Adders and Comparators Using Transistors and Tunnel Diodes," J. J. Amodei, *Trans. IEEE PTGEC* (October) 1964
"Lattice Parameter and Density in Germanium-Silicon Alloys," J. P. Dismukes, L. Ekstrom, and R. J. Paff, *Jour. Phys. Chem.* (October) 1964
"Low-Temperature Elastic Constants of Indium Arsenide," D. Gerlich, *Jour. Appl. Phys.* (Communications) (October) 1964
"Measurement Accuracy," W. D. Moon, *Trans. IEEE PTGAS* (October) 1964
"Microwave Properties of Superconductors," J. I. Gittleman and B. Rosenblum, *Proc. IEEE* (October) 1964
"Optical and Electrical Properties of Single-Crystal GaP Vapor-Grown on GaAs Substrate," H. Flicker, B. Goldstein, and P-A Hoss, *Jour. Appl. Phys.* (October) 1964
"On the Optimization of Random-Access Discrete Address Communications," R. C. Sommer, *Proc. IEEE* (Correspondence) (October) 1964
"Plasma Effects in Solids," M. C. Steele and coauthor, *Proc. IEEE* (October) 1964
"Remote Control for Radio Transmitters," D. L. Wright, *Broadcast News* (October) 1964
"Resolving Multiple Responses in an Associative Memory," H. S. Miiller, *Trans. IEEE PTGEC* (Short Notes) (October).... 1964
"The Ryotron—A New Cryogenic Device," R. A. Gange, *Proc. IEEE* (October) 1964
"The Ryotron—A Variable Inductance Cryogenic Device," J. C. Miller, C. M. Wine, and L. S. Cosentino, *Proc. IEEE* (October) 1964
"Theory of One-Carrier, Space-Charge-Limited Currents Including Diffusion and Trapping," M. A. Lampert and F. Edelman, *Jour. Appl. Phys.* (October) 1964
"Thermal and Electrical Properties of Heavily Doped Ge-Si Alloys up to 1300°K," J. P. Dismukes, L. Ekstrom, E. F. Steigmeier, I. Kudman, and D. S. Beers, *Jour. Appl. Phys.* (October) ... 1964
"A 320 Mc Broadband Amplifier," L. C. Drew, H. F. King, and A. G. Atwood, *Solid/State/Design* (October) 1964
"Anomalous Resistivity of Nb_3Sn," D. W. Woodard and G. D. Cody, *Phys. Rev.* (5 October) 1964

† Report all corrections to *RCA Review*, RCA Laboratories, Princeton, New Jersey.

"Electrolytic Reduction of Trivalent Rare-Earth Ions in Alkaline-Earth Halides," F. K. Fong, *Jour. Chem. Phys.* (15 October) 1964
"Development of the RCA 25-Inch 90-Degree Rectangular Color Picture Tube," A. M. Morrell and A. E. Hardy, *Trans. IEEE PTGBTR* (November) 1964
"Fabrication of Ultrathin Pinhole-Free Insulation of Aluminum Oxide," R. B. Schilling, *Proc. IEEE* (Correspondence) (November) 1964
"Gas Lasers in Magnetic Fields," S. A. Ahmed, R. C. Kocher, and H. J. Gerritsen, *Proc. IEEE* (Correspondence) (November) 1964
"High-Efficiency Injection Laser at Room Temperature," H. Nelson, J. I. Pankove, H. Hawrylo, and G. C. Dousmanis, *Proc. IEEE* (Correspondence) (November) 1964
"The Impedance of an Electric Dipole in a Magneto-Ionic Medium," H. Staras, *Trans. IEEE PTGAP* (November) 1964
"Infrared Transmission Characteristics of Various Color Release Prints and Their Effects on Color Television Reproduction," H. N. Kozanowski, *Jour. S.M.P.T.E.* (November) 1964
"Investigation of an L-Band Tunnel-Diode Amplifier at Low Temperatures," V. Stachejko, *Proc. IEEE* (Correspondence) (November) 1964
"Metal-Semiconductor Barrier-Height Measurement by the Differential Capacitance Method—Degenerate One-Carrier System," A. M. Goodman and D. M. Perkins, *Jour. Appl. Phys.* (November) 1964
"New Perspectives for Consumer Electronics," J. Hillier, *Trans. IEEE PTGBTR* (November) 1964
"A New UHF/VHF Silicon Transistor for the Consumer Market," L. Plus, R. A. Santilli, and H. Thanos, *Trans. IEEE PTGBTR* (November) 1964
"Note on the Scattering of Waves by Rough Surfaces," R. Ruffine, *Trans. IEEE PTGAP* (Communications) (November) 1964
"Protection of Transistors in Class B Audio Output Stages," M. S. Fisher, *Trans. IEEE PTGBTR* (November) 1964
"Room-Temperature GaAs Laser Voice-Communication System," D. Karlsons, C. W. Reno, and W. J. Hannan, *Proc. IEEE* (Correspondence) (November) 1964
"A Self-Oscillating Vertical Circuit Using a Dual-Control Pentode Tube," W. M. Austin and J. A. Dean, *Trans. IEEE PTGBTR* (November) 1964
"A Study of the Need for Color Controls on Color TV Receivers in a Color TV System Operating Perfectly," C. J. Hirsch, *Trans. IEEE PTGBTR* (November) 1964
"The Systematic Design of the Butler Matrix," H. J. Moody, *Trans. IEEE PTGAP* (Communications) (November) 1964
"Two-Phase Mercury Plasma Tunnel," A. I. Carswell, *Rev. Sci. Instr.* (November) 1964
"The Wave Tilt of an Electromagnetic Wave Propagating in the Earth-Ionosphere Cavity," B. E. Keiser, *Trans. IEEE PTGAP* (Communications) (November) 1964
"Quantum-Mechanical Effects in Stimulated Optical Emission. II," R. C. Williams, *Phys. Rev.* (2 November) 1964
"Boosting D-C Voltage with Silicon Transistors," H. T. Breece, III, *Electronics* (November 16) 1964
"Scattering of Phonons by Electrons in Germanium-Silicon Alloys," E. F. Steigmeier and B. Abeles, *Phys. Rev.* (16 November) .. 1964
"Dynamics of Energy Transfer from $3d$ to $4f$ Electrons in $LaAlO_3$: $Cr^{3+}Nd^{3+}$," Z. J. Kiss, *Phys. Rev. Letters* (30 November) .. 1964
"Elastic Constants of Strontium Fluoride Between 4.2 and 300°K," D. Gerlich, *Phys. Rev.* (30 November) 1964
"Electron Spin Resonance of Concentrated Copper Phthalocyanine Crystals," J. M. Assour and S. E. Harrison, *Phys. Rev.* (30 November) 1964

"Metal-Oxide-Semiconductor Field-Effect Transistors," F. P. Heiman and S. R. Hofstein, *Electronics* (November 30) 1964

"Computer Solutions Aid Rectifier Filter Design," W. Austin and K. Angel, *Electronic Industries* (December) 1964

"Dielectric Behavior of Nonrigid Molecules. II. Intramolecular Interactions and Dielectric Relaxation," F. K. Fong, *RCA Review* (December) 1964

"Dislocation Etch Pits in GaAs," M. S. Abrahams, *Jour. Appl. Phys.* (Communications) (December) 1964

"The 'Dynagroove' System," H. F. Olson, *Audio* (December) 1964

"Effect of High Magnetic Field on Electron-Beam Noise," J. M. Hammer and C. P. Wen, *RCA Review* (December) 1964

"Generating Network Functions with an Infinite Potential Analog Plane," F. M. Brock and R. Binks, *RCA Review* (December) 1964

"Grown-Film Silicon Transistors on Sapphire," C. W. Mueller and P. H. Robinson, *Proc. IEEE* (December) 1964

"High Temperature Differential Thermal Analysis Apparatus," M. M. Hopkins, *Rev. Sci. Instr.* (December) 1964

"Integrated Circuits Incorporating Thin-Film Active and Passive Elements," P. K. Weimer, H. Borkan, G. Sadasiv, L. Meray-Horvath, and F. V. Shallcross, *Proc. IEEE* (December) 1964

"Microwave Electron Cyclotron Resonance Pumping of a Gas Laser," S. A. Ahmed and R. Kocher, *Proc. IEEE* (Correspondence) (December) 1964

"New Cartridge Tape Equipments," C. B. Meyer, *Broadcast News* (December) 1964

"New Look Transistorized Sync Generator," R. J. Smith, *Broadcast News* (December) 1964

"Nonlinear-Admittance Mixers," L. Becker and R. L. Ernst, *RCA Review* (December) 1964

"Optimum Synthesis of Inductorless Bandpass Filters Using a Combination of Identical RC Coupled Amplifiers," J. Pearl, *Trans. IEEE PTGCT* (December) 1964

"Parallel Operation of TV Transmitters at WNAC-TV," *Broadcast News* (December) 1964

"Planning Data and Equipment for 12,000 Mc Microwave Relay Service," J. B. Bullock, *Broadcast News* (December) 1964

"Switching Response of Complementary-Symmetry MOS Transistor Logic Circuits," J. R. Burns, *RCA Review* (December) 1964

"Theory and Application of the B-Chart," J. Breckman, *RCA Review* (December) 1964

"Thermoplastic Organic Photoconductive Recording Media—Electrophotographic Characteristics and Processing Techniques," E. C. Giaimo, *RCA Review* (December) 1964

"Understanding and Using the MOS FET," D. M. Griswold, *Electronics* (December 14) 1964

"Paramagnetic Resonance Absorption in an Optically Excited State in CaF_2:Dy^{2+}," E. S. Sabisky and C. H. Anderson, *Phys. Rev. Letters* (21 December) 1964

"Advancements in the Facsimile Art During 1963," W. H. Bliss, *IEEE International Convention Record*, Part 5, *Wire and Data Communication* 1964

"Code Division Multiplex System," A. B. Glenn, *IEEE International Convention Record*, Part 6, *Radio Communication; Recording and Audio; TV* 1964

"The Optimum Allocation of Mass in Dispersed Reflector Clusters," N. S. Potter, *IEEE International Convention Record*, Part 6, *Radio Communication; Recording and Audio; TV* 1964

"Report on IEEE PTGEWS/PTGE Seminar on Writing-Improvement Programs for Engineers," C. A. Meyer, *IEEE International Convention Record*, Part 10, *Education; Engineering Management; Engineering Writing and Speech; Information Retrieval* 1964

AUTHORS

ERWIN F. BELOHOUBEK received the degree of Diplom-Ingenieur in 1953 and the Ph.D. in Electrical Engineering in 1955 from the Technical University in Vienna, Austria. From 1953 to 1955 he worked as Research Assistant at the Institute for High-Frequency Techniques of the same university. He joined RCA Electronic Components and Devices, Harrison, New Jersey, in 1956 and transferred to Princeton in 1957, where he has been engaged in work on magnetrons, electrostatically and magnetically focused traveling-wave tubes, and more recently, crossed-field microwave delay tubes.

Dr. Belohoubek is a member of the Institute of Electrical and Electronics Engineers.

DONALD J. BLATTNER received the BS (EE) and MA (Physics) degrees from Columbia University. He taught physics and did solid-state research at Columbia. At RCA he has worked on microwave tubes, microwave light modulators and detectors, and solid-state microwave power sources. He is a member of Sigma Xi.

YAOHAN CHU received the B.S. degree from Chiao-Tung University (China) in 1942, the B.S. degree from M.I.T. in 1945, and the Sc.D. from M.I.T. in 1953. He was associated with the Scientific Laboratory of Ford Motor Company from 1953 to 1955, with Westinghouse Electric Corporation (Baltimore) from 1955 to 1960, and with Melpar, Inc. from 1960 to 1962. He joined RCA Data Systems Center, Bethesda, Md., in 1962. He became a consultant to the Control Data Corporation in 1964. His work has been in the areas of feedback control systems, radar systems, analog computers, and digital computers. Dr. Chu has been teaching courses on digital and analog computers since 1953, currently at the University of Maryland.

He is a member of Sigma Xi, the Association for Computing Machinery and the Institute of Electrical and Electronics Engineers.

THOMAS B. HOWARD received his B.S. degree from the U.S. Naval Academy in 1943 and his MSEE degree from the University of Pennsylvania in 1951. From 1943 and 1947 he served in the U.S. Navy. Following his release from active duty, he entered the field of electrical engineering and for the next two years was employed by the Ozone Processes Division of the Welsbach Corporation, Philadelphia, Pa. In 1951 he joined Sperry Gyroscope Co., Great Neck, N. Y. While there, he did research and development in radar and missiles and in the digital control of machine tools and, on special assignment at Lincoln Laboratory, MIT, worked on pulse doppler radars and FM delay line cancellation receivers. In 1955 he joined the Radio Corporation of America. His assignments have included study and development work in the fields of missile range instrumentation, radar pulse compression, wideband signal processing, detection and systems optimization. His most recent work has pertained to advanced technique development for interference reduction in radar.

He is a member of the Institute of Electrical and Electronics Engineers and of Sigma Xi.

HENRY C. JOHNSON attended Johns Hopkins University and has recently received the B.S. (M.E.) degree from Drexel Institute of Technology. He joined RCA Electronic Components and Devices, Harrison, New Jersey in 1956 working on pencil tube design. In 1959 he transferred to the Microwave Applied Research Laboratory in Princeton, N.J., where he has been actively engaged in the development of tunnel diodes, optical modulators and detectors. He has also worked on tunnel diode oscillators, sub-harmonic generators and is presently working on the design and development of microwave solid-state circuits.

GEORGE A. MORTON received the degree of B.S. in Electrical Engineering in 1926, the M.S. in 1928, and the Ph.D. in Physics in 1932, all from the Massachusetts Institute of Technology. Dr. Morton was a Research Associate and Instructor at MIT from 1927 to 1933. In 1933, he joined the research laboratory of RCA Manufacturing Co. as a research engineer and was transferred to RCA Laboratories in 1941 as a research section head. During World War II, Dr. Morton was a Section Member, National Defense Research, and a member of the AAF Advisory Board, NDRC, RDB Panel IR. He received the IEE Overseas Premium Award and Air Force, Navy, and Army Certificates for War Research. He became Associate Director of the Physical and Chemical Research Laboratory of RCA Laboratories in 1954, and was made Director of the Conversion Devices Laboratory of RCA Electronic Components and Devices in 1961.

Dr. Morton is a Fellow of the Institute of Electrical and Electronics Engineers and the American Physical Society and a member of Sigma Xi and the American Association for the Advancement of Science.

JOHN E. RUEDY received the A.B. degree from Western Reserve University in 1924 and the Ph.D. degree in Physics from Cornell University in 1933. From 1925 to 1927 he was an engineer with Bell Telephone Laboratories, from 1927 to 1934 held teaching and research assistantships at Cornell University, from 1934 to 1936 was with the Evaporated Metal Films Company. He joined the RCA Manufacturing Company in 1936, and later transferred to RCA Laboratories, Princeton, N. J. In 1961 he joined the Conversion Devices Laboratory of RCA Electronic Components and Devices. His work has been principally in the fields of photoelectron and secondary electron emission and of electron imaging devices.

Dr. Ruedy is a member of Sigma Xi, the American Physical Society, and is a Fellow of American Association for the Advancement of Science.

FRED STERZER received the B.S. degree in physics from the College of the City of New York in 1951, and the M.S. and Ph.D. degrees in physics from New York University in 1952 and 1955, respectively. From 1952 to 1953 he was employed by the Allied Control Corporation, New York, New York. During 1953 and 1954 he was an instructor in physics at the Newark College of Engineering, Newark, New Jersey, and a research assistant at New York University. He joined the RCA Electron Tube Division in Harrison, New Jersey, in October, 1954, and transferred to the Princeton, New Jersey branch in 1956, where he is now Manager of the Microwave Applied Research Group. His work has been in the field of microwave spectroscopy, microwave tubes, parametric amplifiers, tunnel-diode microwave amplifiers, frequency converters and oscillators, microwave computing circuits, and light modulators and demodulators.

Dr. Sterzer is a member of Phi Beta Kappa, Sigma Xi, the American Physical Society, and the Institute of Electrical and Electronics Engineers.

ROSCOE C. WILLIAMS graduated from the University of Manitoba in 1951 with an Honours Bachelor of Science in both Physics and Mathematics. He received the Master's degree in Theoretical Physics from McGill University in 1953, and the Ph.D. in Experimental Physics at the University of Manitoba in 1958. He joined Bell Telephone Laboratories in 1956, where he worked on magnetostrictive delay lines. At RCA Laboratories, which he joined in 1958, he engaged in cyclotron resonance studies in semiconductors and research in lasers and space-charge-limited emission. He was a Guest Lecturer in the Electrical Engineering Department of Princeton University in 1963. Dr. Williams is presently with the Physics Department of the University of British Columbia, Vancouver, B. C.

OAKLEY M. WOODWARD, JR. received the degree of Bachelor of Science in Electrical Engineering from the University of Oklahoma in 1938. He joined the Seismograph Service Corporation of Tulsa, Oklahoma, after graduation. During 1941, he was a research engineer at the RCA Manufacturing Company, Camden, N. J. In January 1942, he became a member of the technical staff at RCA Laboratories, Princeton, N. J. In 1961, he transferred to the RCA Antenna Skill Center, Missile & Surface Radar Division, Moorestown, N. J. Mr. Woodward is a Member of Sigma Xi, and a Senior Member of the Institute of Electrical and Electronics Engineers.

RCA REVIEW