RCA REVIEW

a technical journal

Published quarterly by

RCA LABORATORIES

in cooperation with all subsidiaries and divisions of

RADIO CORPORATION OF AMERICA

VOLUME XXVIII DECEMBER 1967 NUMBER 4

CONTENTS

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

RCA REVIEW

BOARD OF EDITORS

THE ORIGIN OF PHOTOELECTRONS IN THE SILVER–OXYGEN–CESIUM PHOTOCATHODE

By

A. H. Sommer*

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

Summary—The purpose of this paper is to attempt a consistent interpretation of many Ag-O-Cs characteristics on the basis of earlier as well as of some new experimental results. The conclusions can be summarized as follows. The Ag-O-Cs photocathode consists essentially of elementary silver and cesium oxide (Cs_2O). The response below 3000 A is due to electrons excited from the valence band of Cs_2O. The photoemission above 3900 A is due to a volume effect in elementary silver. The electrons are emitted from Ag into the surrounding Cs_2O and from there into vacuum. The relatively high quantum efficiency of Ag-O-Cs is attributed to the low reflection of the thin cathode films and to the great escape depth of electrons at the low photon energies required for emission from Ag into Cs_2O.

Introduction

Among the photoemissive materials with useful sensitivity to visible light the silver–oxygen–cesium (Ag-O-Cs) photocathode is of special interest because it is the only cathode with appreciable response in the near-infrared region. The material differs from other cathodes having high quantum efficiency in the visible region by its much greater complexity. Thus, while the emission process from cathode materials such as the alkali antimonides is reasonably well established in terms of photoelectrons released from the valence band of a homogeneous semiconductor (see, for instance Ref. (1)), the emission mechanism of Ag-O-Cs is still very inadequately understood.

The purpose of this paper is to show that on the basis of earlier as well as of some new experimental results a consistent, though still incomplete and not necessarily correct, interpretation of many Ag-O-Cs characteristics can be attempted. Before embarking on this interpretation it is necessary to review briefly what is presently known about the Ag-O-Cs material.

---

* This work was done while the author was on a temporary assignment at Laboratories RCA, Ltd., Zurich.

[1] A. H. Sommer and W. E. Spicer, "Photoelectric Emission," in *Photoelectronic Materials and Devices*, ed. S. Larach, D. Van Nostrand Co., Inc., Princeton, N. J., 1965.

Chemical Composition of Ag-O-Cs

The Ag-O-Cs photocathode is usually formed by reacting a thin film of silver oxide with Cs vapor. It is generally believed that this reaction is described by

$$Ag_2O + 2\,Cs \rightarrow 2Ag + Cs_2O$$

This would indicate that the cathode consists essentially of two constitutents, elementary silver and the cesium oxide, Cs_2O. In the past, it has sometimes been suggested that the presence of Ag_2O is also essential, but successful attempts [2,3] to prepare cathodes from Ag and Cs_2O without the intermediate formation of Ag_2O have disproved this suggestion. There exist, however, several other unsolved problems concerning the chemical composition of Ag-O-Cs. While chemical analysis has shown a Cs:O ratio [4,5] of 2:1 and an Ag:Cs ratio [6] of 1:1, both in agreement with the above reaction scheme, and while x-ray analysis [7] has confirmed the presence of Ag and Cs_2O, none of the analytical methods used are accurate enough to rule out the possibility that one or more of the other known cesium oxides and/or elementary Cs may be present in Ag-O-Cs in very small quantities and may even be essential.

Additional indications of the complexity of the chemical reactions involved are the following observations. First, additional deposition of Ag onto the finished cathode reduces the threshold wavelength of photoemission and the thermionic emmission, while increasing the quantum yield at shorter wavelengths.[2] Second, excessive layer thickness reduces photosensitivity. Third, excessive Cs reduces photosensitivity. The last effect may be due to the formation of a stable suboxide; the first two are discussed later.

To sum up, the two major constituents of Ag-O-Cs are undoubtedly Ag and Cs_2O. The presence of other Cs oxides and of elementary Cs is speculative and not detectable by any available physical or chemical method.

---

[2] S. Asao, "Behavior of Foreign Metal Particles in Composite Photocathodes," *Proc. Phys. Math. Soc. Japan*, Vol. 22, p. 448, 1940.

[3] P. G. Borzyak, V. F. Bibik, and G. S. Kramerenko, "Distinctive Characteristics of the Photoeffect in Cesium-Oxide-Silver Photocathodes," *Bull. Acad. Sci. U.S.S.R., Phys. Ser.*, Vol. 20, p. 939, 1956.

[4] N. R. Campbell, "The Photoelectric Emission of Thin Films," *Philos. Mag. & Jour. Sci.*, Vol. 12, p. 173, July 1931.

[5] C. H. Prescott, Jr., and M. J. Kelly, "The Cesium-Oxygen-Silver Photoelectric Cell," *Bell. Syst. Tech. Jour.*, Vol. 11, p. 334, July 1932.

[6] A. H. Sommer, unpublished work.

[7] W. H. McCarroll, unpublished work.

Structure of Ag-O-Cs

There seems to be general agreement that the Ag-O-Cs material is inhomogeneous in structure, at least insofar as the surface is concerned. The evidence comes from studies involving the use of optical microscope,[2,5] electron miscroscope,[9] emission electron microscope,[10] work-function measurements,[11] and emission-velocity measurements.[12] The electron microscope studies point to patches of dimensions in the 100 to 200 Å range, whereas the electron-emission experiments indicate the existence of small areas of very low work function. The former may be individual particles of elementary silver, but it is not known whether the low-work-function areas are identical with these particles.

Photoemission from Ag-O-Cs

The photoemissive properties of Ag-O-Cs vary considerably from sample to sample, but the spectral response curve has some features that are fairly reproducible. A typical curve is shown in Figure 1; its most important characteristics are as follows. Beginning at the long-wavelength threshold in the infrared, the quantum yield rises to a maximum below 1% near 8000 Å and then stays more or less constant throughout the visible spectrum. A peak occurs near 3500 Å, followed by a minimum near 3200 Å. Below 3200 Å the yield rises rapidly to values in the range of 0.1 to 0.2 electron per photon.

Two experiments must be mentioned here because of their relevance in connection with the infrared response of Ag-O-Cs. First, it was observed that the silver content of the cathode is very critical in that below a certain minimum amount no infrared response is obtained, while the response increases rapidly to its peak value when this critical amount is exceeded. Second, all attempts to replace silver by other metals[2,6] have produced cathodes with very low infrared response.

---

[8] N. A. Soboleva, A. S. Shefov, and V. N. Tolmasova, "Threshold Spectral Sensitivity Region of Ag-O-Cs Photocathodes and Its Relation with the Structure of the Photosensitive Layer," *Bull. Acad. Sci. U.S.S.R., Phys. Ser.*, Vol. 26, p. 1393, 1962.

[9] A. I. Frimer and A. M. Gerasimova, "Electron-Miscroscopic Investigation of the Structure of Photoelectric Cathodes," *Sov. Phys. (Tech. Phys.)*, Vol. 1, p. 705, 1956.

[10] L. N. Bykhoskaia and Iu. M. Kushnir, "Electron-Optical Investigation of Compound Photocathodes," *Jour. Tech. Phys.*, Vol. 25, p. 2477, 1955.

[11] A. Lallemand and M. Duchesne, "Some Properties of Complex Photoelectric Layers," *S. Angew. Math. Phys.*, Vol. 1, p. 195, 1950.

[12] N. A. Soboleva, "The Causes of the Anomalous Behaviour of Caesium-Oxide Photocathodes in a Spherical Condenser System," *Radio Eng. & Electronics*, Vol. 4, Pt. 2, No. 11, p. 204, 1959.

Optical Measurements on Ag-O-Cs

Spectral response and light-absorption curves of semiconductors such as the alkali antimonides tend to be similar in shape, except for a displacement on the photon energy scale corresponding to the electron affinity (see, for instance, Ref. (1)). Qualitatively, this is plausible

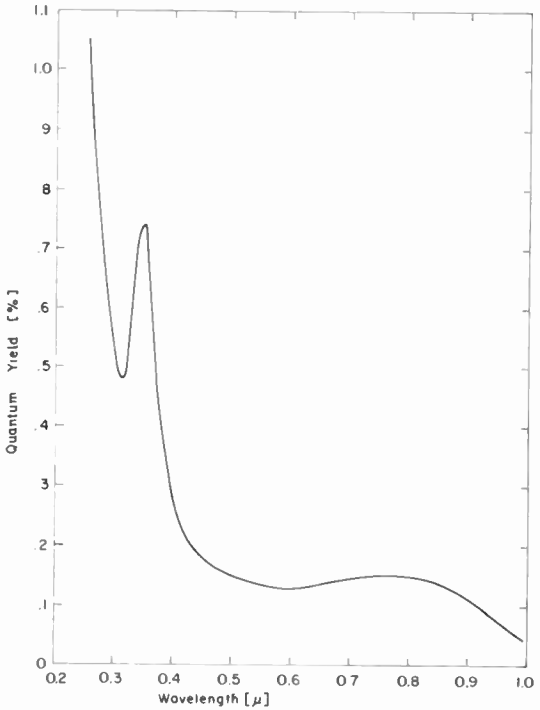


Fig. 1—Photoemission of Ag-O-Cs cathode.

because the photoemission will increase with increased light absorption. No equivalent correlation between photoemission and light absorption can be expected for Ag-O-Cs because the presence of elementary silver causes a more or less constant absorption throughout the visible and near-infrared region of the spectrum. An absorption curve, derived from Asao's [2] transmission and reflection data, is shown in Figure 2. It is apparent that, except for the minimum at 3200 Å, the photoemission and absorption curves have little in common.

Optical and Photoemissive Characteristics of Ag and Cs_2O

On the assumption that Ag-O-Cs consists basically of two com-

ponents, Ag and Cs_2O, Borzyak et al,[3] partly in continuation of Asao's earlier work,[2] measured light absorption and photoemission of these two materials separately.

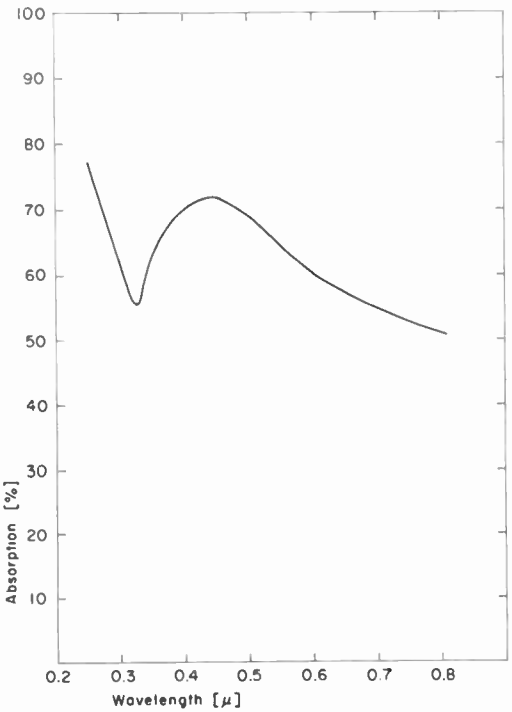


Fig. 2—Light absorption of Ag-O-Cs cathode.

a. Absorption measurements

Ag: Optical measurements on thin Ag films are notoriously irreproducible (see, for instance, Ref. (13)) because the silver tends to aggregate differently, depending on substrate temperature, rate of evaporation, etc. A typical absorption curve, based on Asao's [2] data is shown in Figure 3. It shows the same minimum at 3200 Å as Ag-O-Cs (Figure 2) but the maximum has moved to considerably longer wavelengths. Of course there is no a priori reason why the Ag particles in Ag-O-Cs should behave optically exactly the same way as Ag particles in a film of pure silver.

[13] R. S. Sennett and G. D. Scott, "The Structure of Evaporated Metal Films and Their Optical Properties," *Jour. Opt. Soc. Amer.*, Vol. 40, p. 203, April 1950.

Cs_2O: Borzyak's absorption curve for Cs_2O is shown in Figure 4. From the slope of the curve Borzyak estimates a band-gap energy $E_G \approx 2$ eV for Cs_2O. One comment must be made in connection with the absolute values shown in Figure 4. The absorption of 80% at

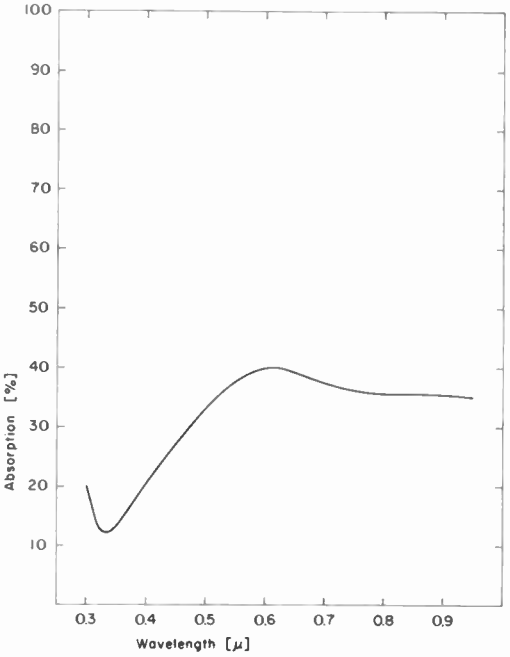


Fig. 3—Light absorption of thin Ag film (about 100 Å thick).

4000 Å is misleading because comparison with calibrated curves by Asao[2] indicates that the measured Cs_2O film was at least ten times thicker than that used in an Ag-O-Cs cathode. Hence the absorption at 4000 Å in the Ag-O-Cs cathode would be only a few percent. This conclusion is important for the later discussion.

b. Photoemission measurements

Ag: Borzyak et al[3] studied the photoemission of thin evaporated Ag films in the near UV. To reduce the work function and thus obtain response at longer wavelengths, they exposed the silver films to Cs vapor, but they made sure that this Cs treatment did not alter the optical properties of the silver. An example of the spectral response curve of such a film is shown in Figure 5. This curve is meaningful only with respect to the wavelengths at which minimum and maximum occur,

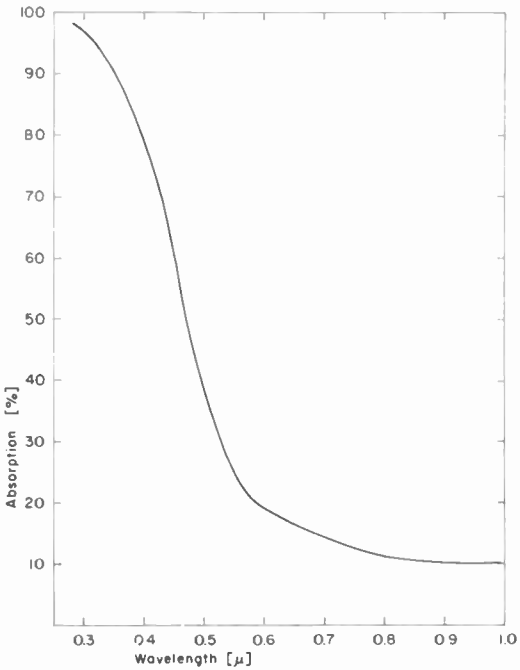


Fig. 4—Light absorption of Cs_2O.

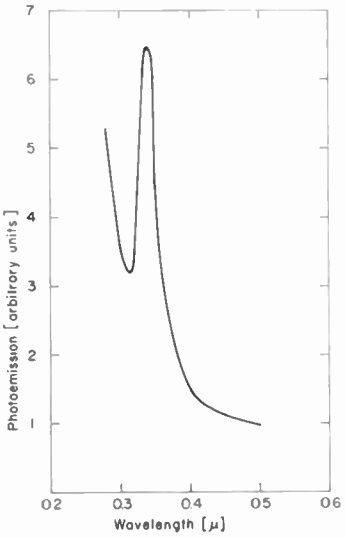


Fig. 5—Photoemission of silver with Cs surface film.

because the absolute values vary rapidly with the thickness and structure of the film. Besides, the figure does not indicate whether the ordinate represents relative or absolute values.

Cs_2O: Borzyak et al also measured the photoemission of Cs_2O and obtained the spectral response curve shown in Figure 6. This curve shows a threshold of photoemission near 3 eV. Since the difference between photoemission threshold and band-gap energy E_G is the elec-

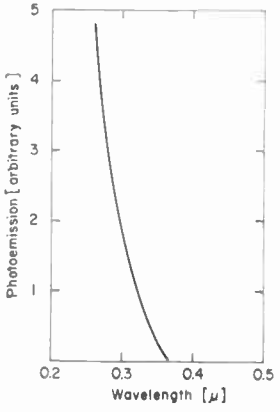


Fig. 6—Photoemission of Cs_2O.

tron affinity E_A, Borzyak obtained for Cs_2O the value $E_A \approx 1.0$ eV. (It is important to point out that the material on which the measurements were made may have contained a small stoichiometric excess of elementary Cs, since it was produced by baking Cs peroxide in Cs vapor. This excess is not likely to affect the optical properties, but it may have a considerable effect on the photoemission, especially near the threshold. Unfortunately no method for preparing Cs_2O exists in which the absence of excess Cs is assured.)

Interpretation of Ag-O-Cs Characteristics

During the first 25 years after the development of the Ag-O-Cs cathode, most workers believed that the infrared response was a surface effect and was in some way associated with free Cs adsorbed on cesium oxide. In this scheme the role of Ag was obscure, particularly after earlier suggestions that the silver was required for electrical conductivity and/or because of its optical (reflecting) properties had been disproved. An alternative theory that Ag acts as an impurity in a semiconductor (cesium oxide) host material also seems untenable because

the term impurity is inappropriate for a constituent that represents 50% or more of the bulk material. Moreover, the spectral response curve in the long-wavelength region differs in important aspects from that to be expected and obtained with impurity photoemitters.

A new, and more promising, approach was chosen by Borzyak et al[3] who suggested that of the two constitutents Cs_2O and Ag, the former is responsible for the UV sensitivity of the cathode, while the silver is essential for infrared response. The evidence for this combination of two photoeffects and a possible mechanism for the long-wavelength emission is discussed below.

a. The Role of Cs_2O

From the optical and photoelectric characteristics of Cs_2O one can draw the following conclusions concerning the effect of Cs_2O in the Ag-O-Cs cathode.

1. The photoemission threshold of 3 eV eliminates Cs_2O as a source of photoelectrons above 4000 Å, i.e., in the visible and infrared region of the spectrum.

2. The steep rise in quantum yield below 3000 Å (see Figure 1) is due to emission from the valence band of Cs_2O. Yields of 0.1 to 0.2 electron per photon at photon energies more than 1 eV above the threshold are consistent with the yields obtained with other semiconductors such as the alkali antimonides.

3. As was pointed out previously, the amount of Cs_2O contained in Ag-O-Cs cathodes is so small that only a small fraction of the incident radiation in the 3000 to 4000 Å region is absorbed. Moreover, the quantum yield for *absorbed* radiation is likely to be low in this region because it is so close to the photoemissive threshold of Cs_2O. As a result the quantum yield for *incident* radiation must be extremely low. This conclusion is relevant for the later discussion.

4. Asao [2], as well as Borzyak et al,[3] observed that the quantum yield below 3000 Å is reduced when silver is added to Cs_2O. Qualitatively this effect is easy to explain. While Cs_2O is a high-yield photoemitter in this spectral region, silver, in common with other metals, is likely to have a low quantum yield. Therefore the total yield must drop if part of the incident radiation is absorbed by silver.

To sum up the experimental results, Cs_2O, possibly containing a small stoichiometric excess of Cs, is not only photoelectrically insensitive above 4000 Å, but in the Ag-O-Cs cathode it does not even appreciably contribute to photoemission between 3000 and 4000 Å. On the other hand, it is responsible for the high quantum yield below 3000 Å.

b. The Role of Ag below 4000 Å

As was just shown, at wavelengths below 3000 Å the presence of silver tends, if anything, to reduce the quantum yield of Ag-O-Cs. In the 3000 to 4000 Å region the following experiments are relevant.

Comparison of Figures 1 and 5 with Figure 3 shows clearly the correlation between the minimum in light absorption of thin silver films near 3200 Å and the minimum in photoemission of both silver and Ag-O-Cs. On the other hand, the maximum photoemission of both silver and Ag-O-Cs occurs at considerably shorter wavelengths than the maximum absorpton of silver. The photoemission peak of Ag-O-Cs near 3500 Å seems to be quite reproducible, and was also found by earlier workers [14] and in more recent studies (see below).

Whatever the cause of the peak at 3500 Å, the important point is that the peak is the same for Ag-O-Cs and for silver alone. This is a strong indication that the photoemission of Ag-O-Cs in the 3000 to 4000 Å range is associated with Ag and not with Cs_2O, in agreement with the conclusion drawn from the measurements on Cs_2O that the yield of the latter must be very low in this spectral region. It must also be emphasized that the sharp rise in photoemission from 4000 to 3500 Å can not be attributed to emission from Cs_2O setting in because the emission should then continue to increase below 3500 Å instead of showing the observed sharp drop.

Additional evidence for the role of Ag in the 3000 to 4000 Å range was derived from the following experiment. The photoemission of a semitransparent Ag-O-Cs cathode of above optimum thickness, deposited on a quartz substrate, was measured with light incident from both directions. The absorption of the cathode was determined from transmission measurements. (No correction was made for reflection, but the changes with wavelength of the latter are small compared with those of transmission.) Figure 7 shows the results. The light-absorption curve resembles that of Figure 2 in that it shows a minimum at 3200 Å and a maximum below 5000 Å. If the light is incident from the vacuum side the photoemission shows the familiar minimum at 3200 Å and the maximum at 3500 Å (see Figure 1). However, with illumination from the substrate side the curve changes its character. A peak now occurs at the absorption minimum, and the maximum at 3500 Å shows only as a slight hump. In addition, the quantum

[14] T. F. Young and W. C. Pierce, "The Wave-Length-Sensitivity Curve of a Cesium Oxide Photocell; A New Light-Sensitive Instrument for the Ultraviolet," *Jour. Opt. Soc. Amer.*, Vol. 21, p. 497, Aug. 1931.

yield is lower throughout the spectrum. This behavior can be interpreted as follows.*

For illumination from the substrate side, the thickness of a semitransparent photocathode is very critical. If the cathode is too thin, too much of the incident light is lost by transmission; if it is too thick, photoelectrons are lost because they are released at a distance from the

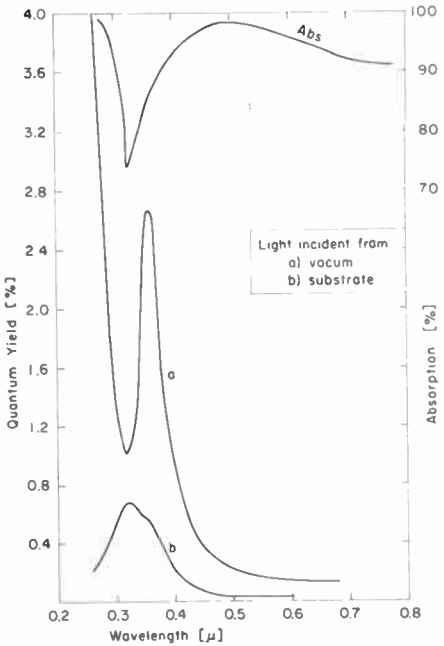


Fig. 7—Light absorption and photoemission of semitransparent Ag-O-Cs cathode of excessive thickness.

vacuum interface greater than the escape depth. In the present experiment the cathode was too thick, and therefore the radiation incident from the substrate side can penetrate to a region within the escape depth only at wavelengths where the absorption is low. As a result, the peak emission occurs at the absorption minimum.

The opposite holds, of course, for light incident from the vacuum side. Here only light in the wavelength range of lowest absorption is lost for the photoemission process because it is partly absorbed beyond the escape depth. Thus the structure at 3200 Å of both photoemission

* This interpretation was suggested by B. F. Williams of this laboratory.

curves in Figure 7 can be correlated with the light absorption of silver —a clear indication that the silver determines the photoemission in this spectral region.

Indirect evidence for the role of Ag in the 3000 to 4000 Å range was also obtained from experiments in which the silver in the Ag-O-Cs cathode was replaced by other metals. Cathodes corresponding to the formulae Pd-O-Cs, Ni-O-Cs, and Cu-O-Cs showed no structure in either absorption or photoemission curve in the relevant 3000 to 4000 Å region. While these materials have much lower photosensitivity than the Ag-O-Cs cathode (see also below), structure in the spectral response curve, particularly at the minimum at 3200 Å, would still be expected if Cs_2O, rather than Ag, were even partially responsible for this structure in the Ag-O-Cs curve. To sum up, it seems established that the sensitivity of the Ag-O-Cs cathode between 3000 and 4000 Å is essentially due to light absorption, and hence creation of photoelectrons, in elementary silver.

c. The Role of Ag above 4000 Å

From the foregoing discussion it is clear that the long-wavelength response must be associated with the silver, because at these wavelengths Cs_2O shows neither photoemission nor light absorption. Since silver alone is also insensitive to visible light because of its high work function, the most general conclusion would be as follows. Absorption of long-wavelength light and the conversion into photoelectrons take place in the silver. These electrons are able to escape with little energy loss into the surrounding Cs_2O, and the Cs_2O, which may contain a small stoichiometric excess of Cs, produces a very low surface barrier that enables the electrons to escape into vacuum.

A more specific interpretation of the photoemissive characteristics of Ag-O-Cs above 4000 Å must take into account the following experimental facts:

1. A minium amount of Ag is required for infrared response to occur.

2. Excessive cathode thickness is detrimental even if the light is incident from the vacuum interface.

3. Replacement of Ag by other metals produces shorter threshold wavelength.

4. Additional silver evaporation produces higher quantum yield in the near infrared, but shorter threshold wavelength and lower thermionic emission.

5. The quantum yield between approximately 4000 and 9000 Å does

not increase with photon energy, but remains practically constant in the 0.001 to 0.005 electron per photon range.

The following qualitative explanations for the experimental facts are suggested.

1. There is strong evidence (see previous discussion of the structure of Ag-O-Cs) that the silver in the Ag-O-Cs cathode is present in the form of individual small particles. If one makes the assumptions that these particles are essential for infrared response (for instance by forming a favorable Ag-Cs_2O configuration) and that a minimum surface density of Ag is required for the formation of the particles, one can understand why a minimum amount of Ag is needed for infrared response.

2. The detrimental effect of excessive cathode thickness may indicate that there exists an optimum particle size for optimum Ag-Cs_2O configuration.

3. The replacement of Ag by other metals may be ineffective for a number of reasons. For instance, other metals may not form particles or they may not form particles of the required size. Furthermore, the energy barrier between metal and Cs_2O may be higher than for silver thus reducing the threshold wavelength of the exciting light. This point is discussed in more detail below.

4. The effect of superficial silver evaporation is still largely unexplained. It seems certain that the effect is not simply an adjustment of the Ag to Cs ratio in the bulk material to an optimum value because the effect is not diminished by the incorporation of excess Ag at an earlier stage in the process. The reduction of threshold wavelength and of thermionic emission indicates that the silver deposition increases the work function, possibly by destroying the small patches of very low work function discussed previously. However, this still leaves the beneficial effect of the Ag deposition unexplained, i.e., the increased quantum yield at shorter wavelengths.

5. The long-wavelength response of Ag-O-Cs has often been attributed to impurity effects (see, for instance, Ref. (15)). As mentioned earlier, it is difficult to accept the silver in Ag-O-Cs in the role of an impurity center because it is a major constitutent of the material. Moreover, while impurity and surface effects undoubtedly determine the photoemissive characteristics near the threshold, the nearly constant and relatively high yield of Ag-O-Cs in the 4000 to 9000 Å region is at variance with the typical features of impurity photoemission.

---

[15] M. Berndt and P. Görlich, "Photoemission of Semiconductors," *Phys. Stat. Solid.*, Vol. 3, p. 963, 1963.

Impurity photoemission has been observed and studied in various materials, for instance in the alkali antimonides [16] and alkali tellurides,[17] and is characterized (1) by a steep rise of quantum yield with photon energy and (2) by very low absolute values of the yield, usually below 10^{-4} electron per photon. By contrast, the photoemission of Ag-O-Cs is almost constant over a photon energy range of about 1.6 eV and the quantum yield is close to 10^{-2} electron/photon.

d. Mechanism of Photoemission From Ag-O-Cs

The discussion of the Ag-O-Cs characteristics in the preceding sections points to the conclusion that the long-wavelength response is due to a volume effect in the silver particles of the cathode. At first it may seem surprising that the quantum yield from Ag in Ag-O-Cs should be close to 10^{-2} electron per photon, while the yield of bulk metals is usually below 10^{-4} electron per incident photon because of the high optical reflection and because electron–electron scattering causes a short escape depth. However, the photoemission from silver into Cs_2O in the Ag-O-Cs cathode differs from emission from silver into vacuum in the following important points.

First, the quantum yield of silver in the bulk form, in terms of electrons per incident photon, is low because only a small fraction of the incident light is absorbed. In the visible region, for instance, as much as 97% of the incident light is lost by reflection. By contrast, thin silver films have much lower reflection and correspondingly higher absorption (see Figure 3) and the absorption of Ag-O-Cs cathode films may be as high as 50% (see Figure 2). Thus the increased light absorption can account for the quantum yield of Ag-O-Cs being ten to twenty times larger than that of bulk silver.

A second reason for increased quantum yield is associated with the difference between the photon energies used for measurements on metallic silver and on Ag-O-Cs cathodes, respectively. Photoemission from silver into vacuum requires photon energies well above the work function of the metal, i.e., in the range above 4.5 eV. From theoretical and experimental work (for a summary see Ref. (18)) it is known that the mean free path of "hot" electrons decreases rapidly with increas-

[16] W. E. Spicer, "Photoemissive, Photoconductive, and Optical Absorption Studies of Alkali-Antimony Compounds," *Phys. Rev.*, Vol. 112, p. 114, Oct. 1, 1958.

[17] E. Taft and L. Apker, "Photoemission from Cesium and Rubidium Tellurides," *Jour. Opt. Soc. Amer.*, Vol. 43, p. 81, Feb. 1953.

[18] H. Thomas, "Excitation and Range of Electrons in Metals and their Determination by Thin Film Methods," in *Basic Problems in Thin Film Physics*, ed. by R. Niedermayer and H. Mayer, Vandenhoeck and Ruprecht, Göttingen, 1966.

ing electron energy because of the increasing probability of energy loss by electron–electron scattering. As a result, the escape depth of hot electrons decreases with increasing electron energy. To give an example, Sze et al [19] found for gold an escape depth of less than 70 Å at photon energies above the work function of gold (~ 5 eV), but escape depths of more than 1000 Å for electron energies below 1 eV. Corresponding high energy values for silver are not known, but Crowell et al [20] reported escape depths of 440 Å for electrons in the 1 eV range.

Applied to the photoemission from metals into vacuum, small escape depth means low quantum yield, even if the yield is expressed in electrons per absorbed rather than incident photon. Since the work function sets a minimum value for the photon energy that can produce photoemission, the energy of all the electrons is so high that the escape depth is short and the quantum yield correspondingly low. The long-wavelength response of the Ag-O-Cs cathode shows that photoelectrons are emitted from Ag into Cs_2O at much lower energy than is required for emission into vacuum. (To explain the small energy required for electrons to escape from a high-work-funtion metal into a low-work-function semiconductor, one probably must assume a tunneling process.) At these low energies, down to 0.5 eV, the escape depth is likely to be greater than 100 Å, so that it does not seriously limit the quantum yield.

An escape depth of the order of several hundred angstroms is also indicated by the observation that semitransparent Ag-O-Cs cathodes in this thickness range are as sensitive to visible light incident from the substrate side as to light incident on the vacuum interface. To prove that electrons with such a large escape depth actually originate in the silver the following experiment was performed. The work function of an evaporated silver film several hundred angstroms thick was reduced by depositing a surface film of Cs. In this way adequate sensitivity to visible light (low photon energies) was obtained. Photoemission measurements with light incident from substrate and vacuum interface, respectively, showed approximately equal sensitivity, similar to the result obtained with Ag-O-Cs (see above). This represents convincing evidence for a volume effect with long escape depth in metallic silver, provided the electron energy is sufficiently low.

In the preceding discussion an attempt was made to explain why the quantum yield of Ag in the Ag-O-Cs cathode is several orders of

[19] S. M. Sze, J. L. Moll, and T. Sugano, "Range-Energy Relation of Hot Electrons in Gold," *Solid State Electronics*, Vol. 7, p. 509, 1964.

[20] C. R. Crowell, W. G. Spitzer, L. E. Howarth, and E. E. LaBate, "Attenuation Length Measurements of Hot Electrons in Metal Films," *Phys. Rev.*, Vol. 127, p. 2006, Sept. 15, 1962.

magnitude higher than that of solid silver. If, as suggested, increased light absorption and escape depth favor high quantum yield, the question must now be reversed, i.e., one must ask why the quantum yield is still low compared with that of materials such as the alkali antimonides. A possible explanation [21] is that excitation of photoelectrons to energy levels above the vacuum level is more probable if the exciting light is absorbed in an interband rather than an intraband (absorption by free electron) transition. Since interband transitions in Ag are possible only at photon energies [22] above 3.5 eV, the absorption in the Ag of the Ag-O-Cs cathode at wavelengths above 3500 Å must be associated with intraband transitions and thus result in reduced escape probability of the photoelectrons and thereby in reduced quantum yield.

The foregoing discussion shows that arguments, though mostly of a hypothetical nature, can be adduced to explain the absolute value of the Ag-O-Cs quantum yield in the visible spectrum. However, the problem remains of explaining the almost constant quantum yield between approximately 9000 and 4000 Å, i.e., over a range of more than 1.5 eV. Since emission is always limited by the loss of those electrons that reach the vacuum interface at an angle at which the velocity normal to the surface is insufficient for escape, a common feature of spectral response curves is the rise of quantum yield with photon energy. A highly speculative interpretation of the absence of this rise in the Ag-O-Cs curve between 9000 and 4000 Å is that in this range the expected increase of yield with photon energy is balanced by the effect of reduced escape depth at higher electron energy.

Conclusions

On the basis of the available information, the following mechanism is proposed for the photoemission from the Ag-O-Cs cathode. The material consists essentially of the two constitutents silver and cesium oxide (Cs_2O). The silver is present in the form of individual particles embedded in Cs_2O. The response below 3000 Å is due to electrons excited from the valence band of Cs_2O and reaches quantum yields typical for similar materials such as the alkali antimonides. Most of the emission in the 3000 to 4000 Å region as well as the emission in the longer wavelength region of relatively high quantum yield, i.e., between 4000 and 9000 Å, is due to a volume effect in elementary silver. The photoelectrons produced in the silver are emitted into Cs_2O and from there into vacuum. Since photons of less than 1.5 eV energy cause electron

[21] H. Thomas, IBM Laboratories, Zurich, Private Communication.
[22] C. N. Berglund and W. E. Spicer, "Photoemission Studies of Copper and Silver: Experiment," *Phys. Rev.*, Vol. 136A, p. A1044, 16 Nov. 1964.

emission into vacuum, the energy barrier for both these processes must be very low. The response between 9000 Å and the threshold wavelength is largely determined by impurity and surface effects.

The quantum yield for emission from the silver in the Ag-O-Cs cathode is several orders of magnitude higher than that from bulk metals, mainly for two reasons. First, the reflection of the thin Ag-O-Cs cathode films is much lower and the absorption of incident light is therefore correspondingly higher. Second, the photon energies, and consequently the photoelectron energies, in the spectral region above 4000 Å are much smaller than those involved in the photoemission from bulk metals into vacuum. Since the escape depth of electrons increases rapidly with decreasing energy, photoelectrons can escape from greater depths in Ag-O-Cs than in bulk metals.

If the above interpretation is correct, one is led to the interesting conclusion that the Cs_2O component, possibly in combination with traces of other Cs oxides and of elementary Cs, performs three functions. First, it acts as photoemitter below 3000 Å. Second, it permits photoelectrons to escape from metallic silver at very low energies. Third, it produces a very low surface barrier and thus allows electrons of very low energy to escape into vacuum.

ACKNOWLEDGMENTS

I wish to express my appreciation to R. Widmer and A. Ziltener of Laboratories RCA Ltd., Zurich, for their help in the experimental work.

$GaAs_{1-x}P_x$ INJECTION LASERS*

BY

J. I. PANKOVE, H. NELSON, J. J. TIETJEN, I. J. HEGYI, AND H. P. MARUSKA

RCA Laboratories
Princeton, N. J.

Summary—The vapor-transport method of epitaxial growth, which yields highly homogeneous crystals, was used to produce alloys of $GaAs_{1-x}P_x$ containing a p-n junction. Injection lasers were made with x ranging up to 0.405. Lasing threshold densities comparable to those of pure GaAs can be obtained at 78°K up to $x = 0.25$. At room temperature, the threshold current density begins rising with $x = 0$. The emission efficiency in the incoherent mode decreases with increasing temperature but drops by only a factor of 10 up to $x = 0.33$. Nonradiative processes dominate the temperature dependence of the emission efficiency and the laser performance for $x < 0.34$. At higher values of x, the direct–indirect intervalley scattering seems to overwhelm the laser performance. Lasers have been obtained that at room temperature have yielded peak powers of 13 watts at 7200 Å. It has also been possible to operate visible lasers cw at 78°K emitting at 7259 Å.

PROCESSING

The fabrication of lasers capable of emitting in the range of 9000 to 6350 Å was made possible by the vapor growth of single-crystal $GaAs_{1-x}P_x$ alloys having a high degree of chemical homogeneity and crystalline perfection. This technique,[1] which involves the decomposition of a mixture of arsine and phosphine, permits the addition of impurities during crystal growth without any further processing. The alloy is grown on a GaAs substrate. To minimize strains due to lattice mismatch, the composition of the growing material is graded over 10 μ from pure GaAs to the desired $GaAs_{1-x}P_x$ alloy. The wafer containing the p-n junction is subsequently processed by a technique very similar to that used for making GaAs lasers: the wafer is lapped to the final thickness ($\sim 100\ \mu$); ohmic electrodes are applied to opposite surfaces; the crystal is cleaved along (110) planes to form Fabry–Perot cavities; and the cleaved bars are sawed into parallelepipeds that are finally mounted on a suitable header.

* The research reported in this paper was sponsored in part by the National Aeronautics and Space Administration, Manned Vehicle Space Center, Houston, Texas, under Contract Number NAS 9-6195.

[1] J. J. Tietjen and J. A. Amick, "The Preparation and Properties of Vapor-Deposited Epitaxial $GaAs_{1-x}P_x$ Using Arsine and Phosphine," *Jour. Electrochem. Soc.*, Vol. 113, p. 724, July 1966.

Threshold Current Density

Diodes were made with composition ranging from $x = 0$ to $x = 0.405$. The composition was determined from the lattice constant by x-ray back-reflection measurements. Figure 1 shows how the photon energy of the laser varies with the alloy composition. Although there is considerable scatter in the data, it is evident that the emission spectrum shifts with composition at the rate of about 12 meV/percent GaP as is

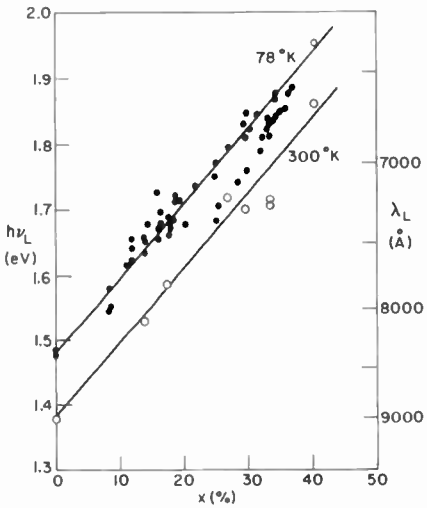


Fig. 1—Dependence of the photon energy of $GaAs_{1-x}P_x$ lassers on alloy composition at 78 and 300°K.

expected from the composition of the energy gap.[2] The scatter could be attributed to variations in doping level or to fluctuations in composition. The emission at $x = 0.405$ represents the highest photon energy reported for an injection laser. Its wavelength is 6350 Å at 78°K. Most of the lasers could be operated at room temperature using 25-nsec pulses. At room temperature, the photon energies are lower by 0.10 eV than the energies emitted at 78°K, as expected from the temperature dependence of the energy gap.

The lowest threshold current densities achieved to date are shown in Figure 2. Since for a given alloy composition the current density varied from wafer to wafer by one order of magniture, the data on

[2] H. Ehrenreich, "Band Structure and Electron Transport of GaAs," *Phys. Rev.*, Vol. 120, p. 1951, Dec. 15, 1960.

all diodes studied shows considerable scatter. However, only the lower envelope of this distribution is shown in Figure 2 in order to illustrate the best than can be obtained. At 78°K, low threshold density can be maintained over the composition $0 < x < 0.25$, above which it rises rapidly. The material is heavily doped and the junction is fairly abrupt to favor operation at room temperature.[3] Hence, still lower threshold

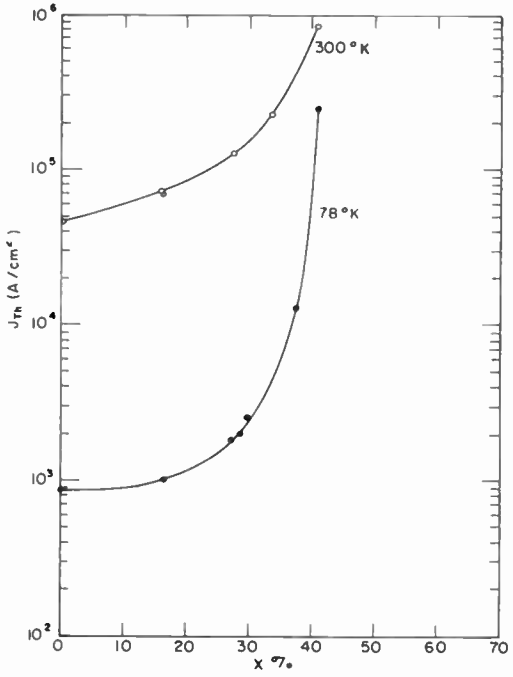


Fig. 2—Lowest limit of the threshold current density of $GaAs_{1-x}P_x$ lasers as a function of alloy composition at 78 and 300°K.

current densities could be obtained at 78°K by preparing more diffused junctions.

At room temperature, the current density rises by a factor of two over the composition range $0 < x < 0.20$. Yet, the threshold current densities are sufficiently low that it is possible to operate reliably injection lasers that emit visible radiation at room temperature. The shortest wavelength to date at room temperature is 6750 Å, from an alloy containing 40.5% GaP. The threshold density for this material is 8.5×10^5 A/cm^2.

[3] G. C. Dousmanis, H. Nelson, and D. L. Staebler, "Temperature Dependence of Threshold Current in GaAs Lasers," *Appl. Phys. Letters*, Vol. 5, p. 174, 1 Nov. 1964.

Emission Efficiency

The performance of lasers depends on many factors. Among these, emission efficiency is perhaps the most important parameter. This factor can be singled out by measuring the radiant power output below threshold. Much attention was given to this study in the hope of finding those factors that influence efficiency. The power output was measured with a calibrated integrating sphere and photomultiplier.

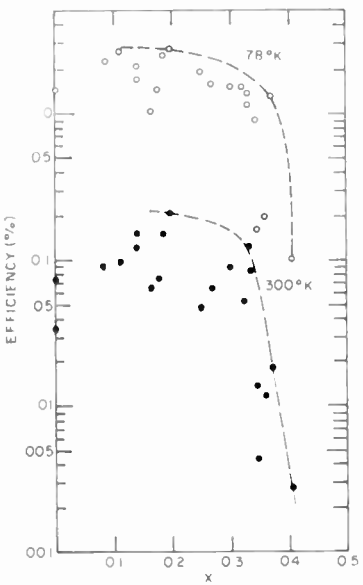


Fig. 3—Quantum efficiency of $GaAs_{1-x}P_x$ diodes below laser threshold. The dashed curves outline the upper limit obtained thus far.

A plot of radiant power output versus current through the diode below threshold permitted the determination of the external quantum efficiency. Measurements were made at 78° and 300°K. The results are shown in Figure 3. The large scatter in the data reflects the nonuniformity of parameters such as doping level and impurity gradient in the junctions. The maximum efficiency obtained at 78°K was 2.75% and changed little with GaP concentration until $x = 0.37$, when it dropped abruptly by an order of magnitude. These values agree quantitatively with those of Pilkuhn and Rupprecht.[4] The maximum efficiency at 300°K was 0.21% and decreased by a factor of about 50 above $x = 0.84$.

[4] M. Pilkuhn and H. Rupprecht, "Electroluminescence and Lasing Action in $GaAs_xP_{1-x}$," *Jour. Appl. Phys.*, Vol. 36, p. 684, 1965.

The concentration dependence of the emission efficiency at low temperature could be accounted for qualitatively on the basis of the interaction between direct and indirect valleys in sharing the electron population. Accordingly, the emission efficiency could be represented by the proportion of electrons in the direct valley of the conduction band. The relative population of the direct valley was calculated assuming parabolic bands.[5] Figure 4 shows the results of this calculation

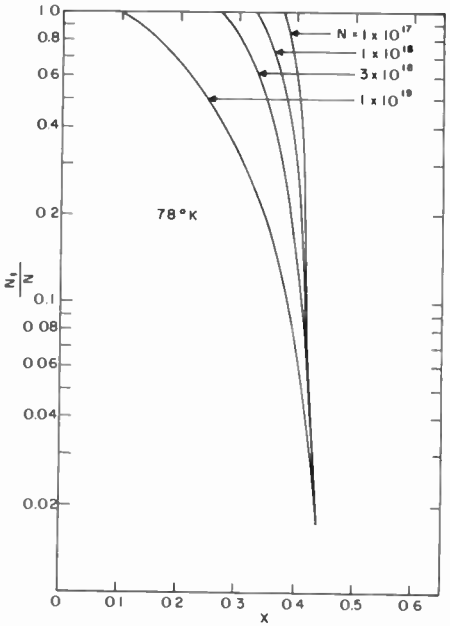


Fig. 4—Ratio of the number of electrons N_1 in the direct valley to the total number of electrons, N, in $GaAs_{1-x}P_x$ at 78°K.

and illustrates the fair agreement with our low-temperature data. It is worth noting that the doping level affects the relative population of the direct valley (the closer the Fermi level is to the indirect valleys, the more serious is the loss of carriers to the indirect valleys). Although the analysis does not apply rigorously to the case of a junction, where the electron concentration decreases as one penetrates deeper into the p-type layer, it indicates the maximum drop in efficiency to be expected.

The temperature dependence of the emission efficiency is believed

[5] H. P. Maruska and J. I. Pankove, *Solid State Electronics* (to be published).

to be dominated by nonradiative recombination processes. Hence, a small temperature dependence of efficiency should be found in those diodes that are most efficient at room temperature. Indeed, our data confirms this correlation. Therefore, the possibility of efficient room-temperature injection lasers should be associated with a small temperature dependence of efficiency η_{78}/η_{300} (a quantity that is easier to determine than the absolute efficiency). The ratio of efficiencies η_{78}/η_{300}

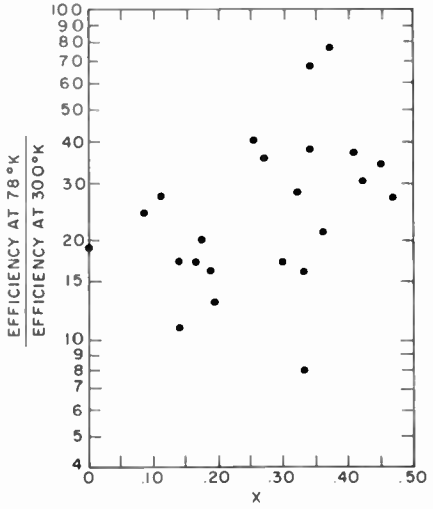


Fig. 5—Ratio of sub-threshold quantum efficiencies at 78 and 300°K obtained with $GaAs_{1-x}P_x$ diodes of various compositions.

is shown in Figure 5 plotted as a function of alloy composition. This figure shows that small temperature dependences are possible even with $x = 0.334$.

The hypothesis that the temperature dependence of the emission efficiency is due to increasing internal absorption with increasing temperature[6] is not consistent with the fact that there is no correlation between the temperature dependence of the efficiency and the size of the diode.

There seems to be an optimum doping of the n-type region for high emission efficiency below threshold. Figure 6 shows a plot of efficiency versus carrier concentration for diodes having $x < 0.33$. The results of

[6] T. Gonda, M. F. Lamorte, P. Nyul, and H. Junker, "Effect of Higher Absorption in Non-Lasing GaAs Diodes at 300°K," *Jour. Quantum Electronics*, Vol. QE2, p. 74, April 1966.

Herzog [7] and of Klein et al [8] for GaAs are also plotted for comparison. These results indicate that the efficiency is maximum in the vicinity of $N = 4 \times 10^{17}$ cm^{-3}. The scatter of the data may be due to compensation by uncontrollable acceptors in the n-type region or by the gradation of acceptors in the junction or by inhomogeneities in the material.

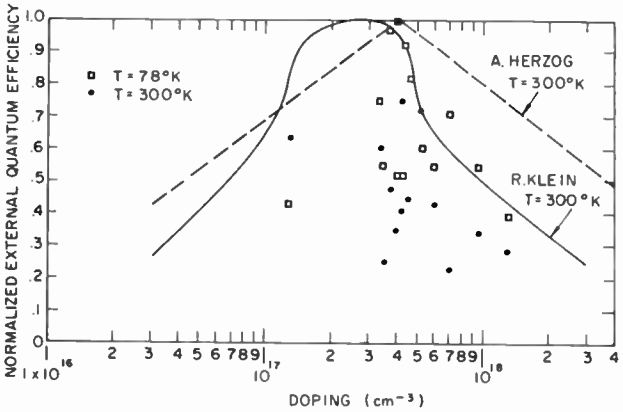


Fig. 6—Sub-threshold quantum efficiency as a function of n-type doping for $x < 0.33$.

Problems

Our experience indicates that both practical and fundamental limitations hinder the achievement of good injection lasers at GaP concentrations greater than 34%.

One fundamental limitation is the intervalley scattering, which robs carriers from the efficiently emitting direct valley. This loss of carriers to the vast set of indirect states rises as the temperature or doping (or both) increase. Of course, optimal (rather high) doping is wanted to reduce joulean losses.

Another fundamental limitation is the competition of nonradiative recombination. A better understanding of this process or processes is needed to make further progress. As stated above, it is nonradiative recombination rather than self-absorption that seems responsible for decreased efficiency when the temperature is raised.

The dominant limitation at present, however, is the practical diffi-

[7] A. H. Herzog, "Quantum Efficiency of GaAs Electroluminescent Diodes," Solid State Electronics, Vol. 9, No. 7, p. 721, July 1966.
[8] R. Klein, H. Kressel, L. Murray, and W. Agosto (to be published).

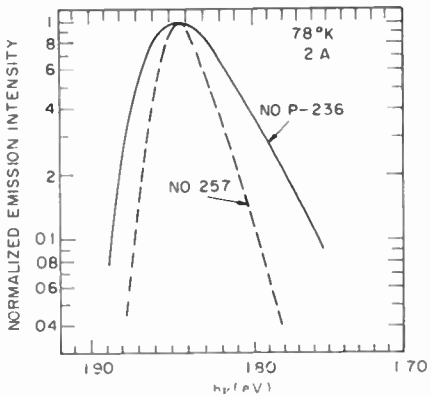


Fig. 7—Normalized emission spectra of two diodes.

culty of getting a sufficiently homogeneous alloy composition. This problem is best illustrated in Figures 7 and 8. Figure 7 shows the electroluminescent emission spectra of two different diodes. The better diode emits a 30-meV-wide peak, whereas the poorer diode's line width is 60 meV. The breadth of the emission peak and a combination of other characteristics, such as large spectral shift with current but high

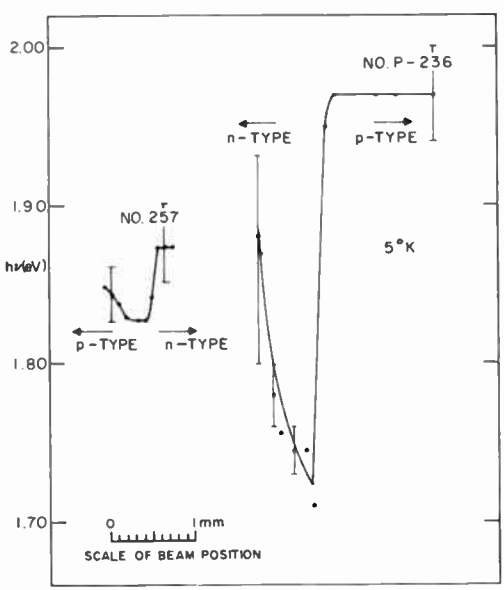


Fig. 8—Photon energy of cathodoluminescence as a function of position near the p-n junction of two alloy diodes. The dots mark the peak energy, the vertical bars represent the spectral width at half maximum.

reverse breakdown voltage, suggest that the poor diode was made of grossly inhomogeneous material. This hypothesis was verified by probing the material (which has been angle lapped through the junction) using a focused 20-kV electron beam. Figure 8 shows the variation of the photon energy during cathodoluminescence as the electron beam is scanned across the p-n junction. The measurement illustrates dramatically the large fluctuation in composition in the neighborhood of the p-n junction (resulting in a 250-meV fluctuation of photon energy). For

Table I—Performance of Selected $GaAs_{1-x}P_x$ Injection Lasers

	x	λ (Å)	P one side (watts)	Q.E. (%)	T (°K)
97	0	9000	6.5	7.5	300
P—221	.16	7860	12	15	300
107	.17	7780	14	13	300
P—230	.23	7540	10	6	300
P—164	.29	7200	13	15	300
134	.20	7259	cw operation		78

comparison, the same measurement was made on the material from which the better diode had been made. There, the fluctuation of cathodoluminescent photon energy is only 50 meV.

Hence, although a few problems remain to be solved (material inhomogeneity and intervalley population sharing at high x) substantial progress has been made in injection lasers emitting visible radiation at room temperature.

ACKNOWLEDGMENTS

The authors are grateful to J. E. Berkeyheiser, Mrs. M. Falk, H. Gossenberger, D. Marinelli, A. Matzelle, A. H. Maxwell, and N. Pastal for their assistance in this study.

METHODS OF DEPOSITION AND PHYSICAL PROPERTIES OF POLYCRYSTALLINE II-VI FILMS*

BY

F. V. SHALLCROSS

RCA Laboratories, Princeton, N. J.

Summary—Techniques for making polycrystalline films of II-VI compounds on amorphous and polycrystalline substrates are reviewed. The films can be prepared by vacuum deposition using single or multiple evaporation sources, by sputtering, or by various nonvacuum techniques. Many of the physical properties of the layers depend critically on the deposition parameters and on the type of post-deposition processing used. Structural and electrical properties are particularly sensitive to fabrication techniques. Properties of semiconducting CdS and CdSe films for use in thin-film active devices are emphasized, although films suitable for other applications are also considered.

INTRODUCTION

THE IIB-VIB compounds have long been of interest as photoconductors and phosphors.[1] More recently they have become important as semiconductors in field-effect devices[2] and piezoelectric materials for use in ultrasonic transducers.[3] They have also been employed in photovoltaic cells,[4] electroluminescent devices,[5] and Hall generators.[6] Thin films of these materials have been particularly important because of their ease of fabrication into precisely controlled geometries, their compatibility with other processing required to make complex devices, and the similarity between many properties of the polycrystalline thin films and of bulk single-crystal material.

This paper reviews a number of methods commonly used to make these films, and describes some of the physical properties of the films, with particular emphasis on those that are sensitive to details of the fabrication procedure. Because the author's research has been primarily concerned with semiconducting films of CdS and CdSe suitable for use in thin-film active devices,[7] films of that type are discussed in most detail.

Table I shows energy-gap data for the 12 compounds of interest.[8-11] There is a general correlation between energy gap and position of the

* Paper originally presented orally at Electrochemical Society Meeting, Dallas, Texas, May 1967.

Table I—IIB-VIB Compounds with Energy Gaps in eV

II B \ VI B	O	S	Se	Te
Zn	ZnO 3.2	ZnS 3.6	ZnSe 2.7	ZnTe 2.3
Cd	CdO 2.4	CdS 2.4	CdSe 1.7	CdTe 1.4
Hg	HgO ~2.3	HgS 2.0	HgSe ~0	HgTe ~0

components in the periodic table, except for some discrepancies in the case of oxides. The majority of the compounds have energy gaps with values corresponding to radiation in or very close to the visible part of the spectrum, making them important for opto-electronic applications. Nearly all have energy gaps large enough to make them suitable as wide-band-gap semiconductors or insulators.

Table II gives crystallographic data for II-VI compounds, listing phases that are stable or metastable under normal-room-ambient conditions.[8,9,12] Most of the compounds crystallize in either the wurtzite or the zinc-blende forms, which are relatively similar structurally. A few of the compounds, particularly oxides, exhibit other forms. In addition, a number of the compounds, such as ZnS and CdTe, form more complex structures built up from alternating layers of wurtzite and zinc-blende-type material by a stacking-fault mechanism. The struc-

Table II—II-VI Compounds: Crystallography

Wurtzite (Hexagonal)	ZnO, ZnS, ZnSe*, ZnTe* CdS, CdSe, CdTe*
Zinc Blende (Cubic)	ZnS, ZnSe, ZnTe CdS*, CdSe, CdTe HgS, HgSe, HgTe
NaCl (Cubic)	CdO
Cinnabar (Hexagonal)	HgO, HgS
Orthorhombic	HgO

*Less Common

tural similarities shown here are consistent with the fact that many solid solutions exist with properties intermediate between those of the parent compounds.

Deposition Methods

The choice of deposition method depends on the film properties required and on the type of other device fabrication and processing techniques employed. For convenience, the deposition methods can be divided into vacuum and nonvacuum techniques. The vacuum methods generally yield good thickness control and uniformity, and are compatible with precise masking techniques; they are capable of producing high-purity material. The nonvacuum methods are often carried out more nearly at thermal equilibrium, and therefore tend to have a lower defect density. They are particularly useful when controlled doping is important. They are also frequently simpler and cheaper than vacuum methods.

In the vacuum evaporation of most II-VI materials, the compound completely dissociates, as established by Goldfinger.[13] Somorgai has shown that the dissociation step is rate limiting, but is influenced by charge-transfer processes at the solid surface, and is thus sensitive to factors such as doping and illumination.[14,15] The material therefore evaporates in approximately stoichiometric proportions, and the condensation and re-evaporation processes at the substrate determine the film composition. As described by Günther,[16] approximately stoichiometric films can be expected provided the substrate temperature is sufficiently high to prevent permanent condensation of unreacted material, and a sufficiently high flux can be maintained. A single source of the type shown in Figure 1 can produce reasonably stoichiometric material provided a high substrate temperature is maintained. This source, consisting of a molybdenum wire basket coated with alumina and filled with a quartz wool plug to minimize spattering, has been used by the author to make moderately conducting films of CdS or CdSe suitable for thin-film transistors. As a small source, it is particularly useful in conjunction with fine-pattern deposition, and it operates at relatively low currents.

Quite a few workers have described other sources suitable for evaporation of II-VI compounds. For example, a modified Knudsen cell described by Escoffery[17] uses a rather large quartz tube containing the evaporant, which is heated by a tungsten coil. This source can be operated at a well-controlled, relatively low, source temperature. Indirect sources, such as described by Galla,[18] make use of a baffle above

particles.[25,26] In this technique, as shown applied to CdS, a fine powder is produced that is made into a suspension with an appropriate liquid. $CdCl_2$, used as a flux, and appropriate activators, such as copper, are added, and the layer is applied to the substrate by silk-screening, printing, or spraying. The layer is then fired at about 600°C to cause the grains to be sintered together and to drive off excess halide. This method is particularly important in making photoconductive layers, and is compatible with other thick-film techniques. The sintering minimizes the problems associated with interparticle contacts that are common to many deposition methods.

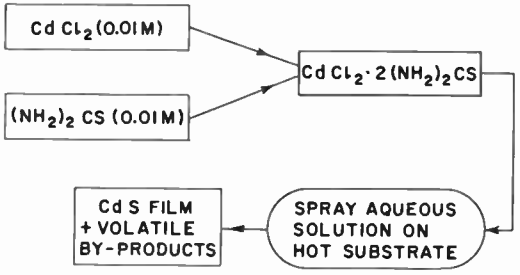


Fig. 4—Chemical-spray process.

In the chemical spray process,[27] an aqueous solution of a metal salt (such as cadmium chloride) is mixed with an aqueous solution of a sulfo- or seleno-organic compound, such as thiourea, to form a soluble complex, as outlined in Figure 4. This solution is sprayed onto a substrate held at 300° to 500°C, where it decomposes to produce the required film. The method can be applied to a large number of sulfides and selenides, and is readily adaptable to the preparation of doped films. Figure 5 shows a chemical-spray arrangement used by the author to make CdS films. The propellant (compressed air) and the appropriate spray solution are brought together in an atomizing spray nozzle through filters. The spray is directed onto a substrate heated by a hot plate. Suitable shields are used to surround the spray and substrate. Films prepared by this method are quite similar to those made by vacuum deposition.

Other chemical methods are also possible, including vapor transport,[28] in which a volatile intermediate permits transfer of II-VI compounds across a thermal gradient at relatively low temperature, and various vapor-deposition procedures.[1]

Physical Properties

Many of the physical properties of II-VI films are quite sensitive to the method of deposition used and to the details of the fabrication and processing technique. The microstructure, including crystallographic phases, crystallite size and orientation, and the surface features such as roughness and uniformity are dependent on many factors such as substrate temperature, deposition angle, and ambient pressure. The stoichiometry, impurity content, and degree of film stress vary appre-

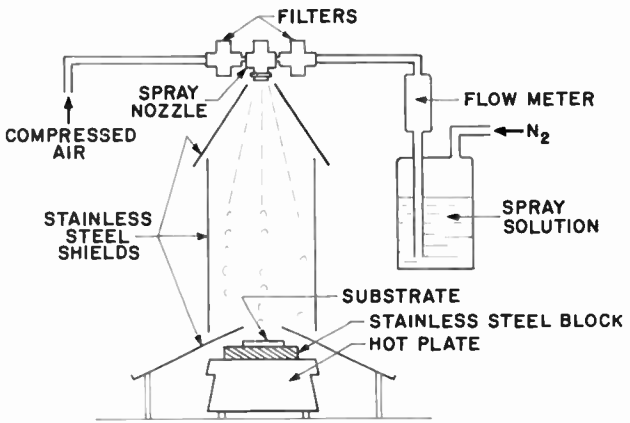


Fig. 5—Chemical-spray equipment.

ciably for different fabrication methods. These variations are particularly important in influencing the electrical and opto-electronic properties of the films.

The nature of the substrate can have a marked effect on film properties. In most of the deposition methods, film formation begins with the development of stable nuclei at certain preferred sites on the substrate. These sites are likely to be associated with imperfections or impurities on the surface. Many of the II-VI films can be grown epitaxially as single-crystal layers on appropriate substrates, such as mica or sodium chloride.[29,30] Even when the substrates are not single crystals, there are marked differences in density of nuclei and crystallography between films on polycrystalline and on amorphous substrates.[20,31] Thermal effects are important; differential expansion between the substrate and the film affects the film stress and is a factor in possible stress failure. The thermal conductivity of the substrate influences the film surface temperature during growth and also affects the performance of

the films in devices where power dissipation is appreciable. Chemical and electrical interactions between film and substrate are also significant; for example, high-temperature processing can cause diffusion of impurities from the substrate to the film. At lower temperatures adsorbed species are important.

In vacuum-deposited films of CdS and CdSe, the density of nuclei depends on the substrate temperature. Figure 6 shows electron micro-

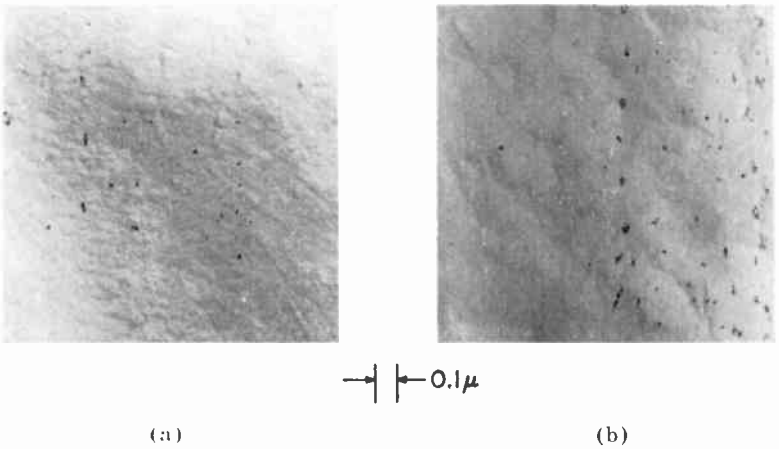


Fig. 6—Electron micrographs of surface replicas of CdS films deposited at 56Å/sec: (a) 35°C substrate, (b) 250°C substrate.

graphs of 7000 Å CdS layers. Films deposited at low temperature have a high density of initial nuclei and a resultant small grain size, together with poor orientation and a high defect density. When the substrate temperature is high, there is usually an initiation period before any nucleation occurs, and the subsequent density of nuclei is small. The resulting crystallites are considerably larger. Surface mobility and stoichiometry also increase with temperature, leading to improved crystallinity.

The surface crystallite size and degree of orientation increase as the films grow thicker. The electron micrographs in Figure 7 show the effect of thickness on CdS films. Some relatively large crystals appear in the thicker layer. Growth processes that minimize grain-boundary energy and a nucleation that is primarily limited to the substrate interface can lead to this effect. Although there is usually some preferred orientation in the initial nuclei, with the c-axis of the hex-

agonal phase perpendicular to the substrate, there is apparently a differential growth rate along the c-direction leading to an increase in the amount of preferred orientation with thickness.

As described above, many of the II-VI compounds can exist in both the zinc blende and the wurtzite structures. For several compounds (ZnS, ZnSe, ZnTe, CdTe) the effects of variations of deposition conditions on phase composition have been studied in some detail by Spînu-

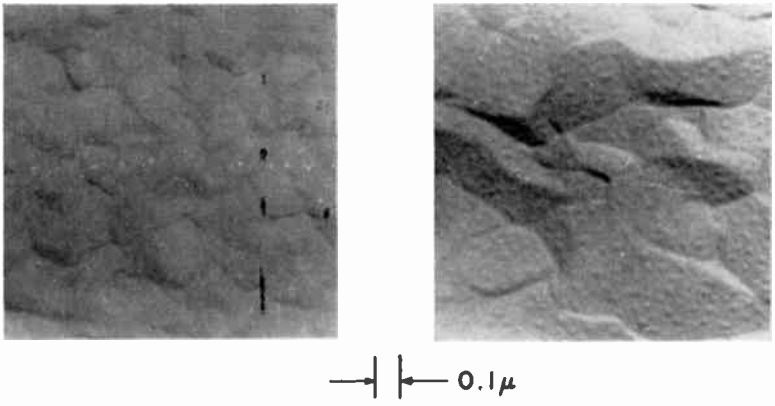


Fig. 7—Electron micrographs of surface replicas of CdS films deposited on 180°C substrate at 75Å/sec: (a) 0.5μ thick, (b) 3.2μ thick.

lescu-Carnaru and others.[32-34] In the case of a number of compounds that more commonly occur in the cubic phase, hexagonal material could be obtained by using an excess of the group II element, such as Zn or Cd. Excess of the group VI element favored the cubic phase. The optimum substrate temperature to obtain hexagonal material was around 300°C; lower temperatures yielded cubic material. In some cases substrate temperatures over 500°C also gave cubic films. The ambient pressure during deposition was also important. A residual atmosphere of about 10^{-2} torr of argon favored the formation of hexagonal films. The general conditions outlined here appeared to be similar for a number of the II-VI films.

Impurities, deviations from stoichiometry, and other crystal defects are particularly important in determining electrical properties of the films.[1,7,8] For many II-VI films the conductivity is largely determined by deviations from stoichiometry. Although both interstitials and va-

cancies apparently exist in the films, it is likely that the electrically active species are anion or cation vacancies. Chlorine is an important donor in II-VI films, particularly when it is used in the fabrication process as in chemically sprayed or sintered layers. Copper is a common acceptor introduced during deposition or by post treatment.

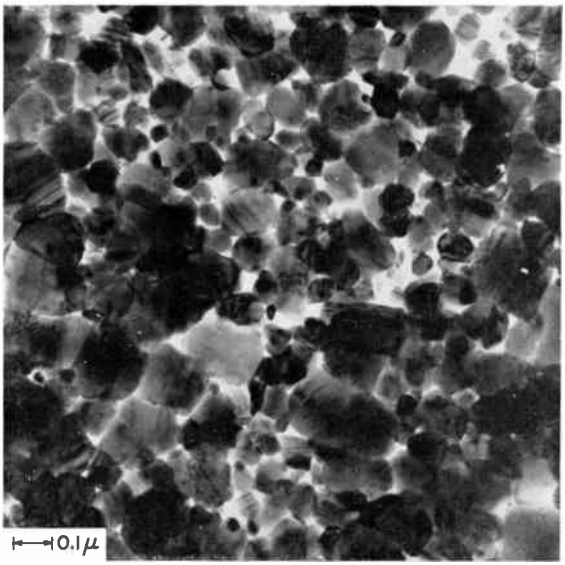


Fig. 8—Electron micrograph of 1300-A CdS film.

Many of the II-VI oxides, sulfides, and selenides tend to be obtainable as n-type materials only. Several factors are involved, including the tendency for acceptor levels to be comparatively deep lying, the low hole mobility, and the limited solubility of acceptor-type impurities.

Surface species are important in II-VI materials; this is particularly true for chemisorbed oxygen, which acts as a deep-lying donor, with marked effects on the observed conductivity, mobility, and stability of the films.[35] A large variety of trapping states have been observed in various experiments on CdS films. In many cases they appear to be associated with crystal defects such as interstitial sulfur, and are quite sensitive to annealing and other post-deposition processes.

Figure 8 is a transmission micrograph of a thin CdS film that emphasizes the importance of defects, grain boundaries, and intercrystalline contacts in the films. Understanding the electrical properties of material such as this is obviously more difficult than for single-

crystal material. For example, the observed Hall mobility for this type of film often has a complicated behavior.

The observed Hall mobility for many of the films increases with temperature, as illustrated in Figure 9. It is generally appreciably lower than single-crystal mobility, although it can be varied over a wide range by changes in crystallinity, impurity content, and adsorbed

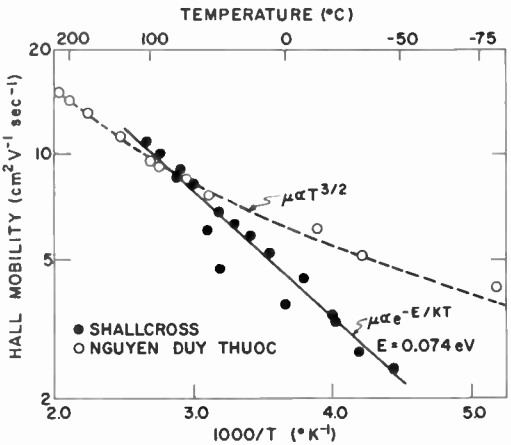


Fig. 9—Hall mobility data for CdS films.

surface species. The exact form of the temperature dependence apparently can vary. Usually there is an exponential dependence on $-1/T$, as indicated in the black dots;[7] this effect can be attributed to intercrystalline barriers or possibly other types of imperfections within crystallites. In some cases a $T^{3/2}$ dependence is reported, as shown in circles;[36] this result could arise from scattering by charged impurities. Experimentally it is rather hard to distinguish between an exponential dependence (particularly with several activation energies) and the $T^{3/2}$ relationship.

The properties of II-VI films can be modified by various post-deposition treatments. In practice, they frequently are important steps in film processing. They can be used to produce major changes in film stoichiometry by out-diffusion of excess material in a suitable ambient or by incorporation of material from an atmosphere containing one of the components. Desorption of surface species such as chemisorbed oxygen and change in impurity concentration by diffusion are also feasible. Annealing at temperatures of several hundred degrees C can lead to crystal growth, changes in film stress, and a reduction in defect density. As studied by Gilles and Van Cakenberghe, Vecht, and others,

massive recrystallization can occur in II-VI films by treatment that incorporates an appropriate flux.[37,38]

Table III illustrates the effects of post-deposition processing, based on previous work by Dresner and Shallcross on vacuum-deposited CdS films.[7,39,40] At low substrate temperature, the stoichiometry is poor and resistivity is low. A high-temperature air bake produces out-diffusion of cadmium together with some oxygen incorporation, leading to

Table III—Processing Effects on CdS Flms

Substrate Temperature (°C)	Processing	Resistivity (ohm-cm)	Crystal Size (μ)
35	None	0.1	~0.05
35	500°C Air Bake	~10^4	>0.05
100	None	10^2-10^3	—
100	305°C S Bake	~10^6	—
170	None	10^4-10^8	0.1-1
170	500°C Bake in CdS:Cu,Cl	~10^3	~10
170	Recrystallized Ag Flux	10^5-10^8	10^3-10^4

a marked increase in resistivity. Some increase in crystallite size occurs as well. For films deposited at 100°C, stoichiometry is still not very good; it can be increased by heating of the layer in sulfur vapor, as illustrated here.

A major change in doping and crystallite size and orientation can be accomplished by high-temperature processing of the films in contact with suitable materials such as cadmium sulfide powder doped with copper and chlorine, or with an adjacent layer of silver. This last process can lead to films that are essentially single-crystal layers. Doping and recrystallization processes of this type have pronounced effects on properties such as Hall mobility and photoconductivity.

Field-effect devices made with II-VI films are quite sensitive to many of the electrical properties of the layers.[41] An empirical approach to control of deposition and processing for semiconductor films used in thin-film transistors is often convenient. For example, by measuring

the source–drain resistance during CdSe deposition by vacuum evaporation, it is possible to adjust deposition rate and film thickness to obtain the desired conductivity in the film.[42] The effects of a post-deposition air bake on the completed CdSe units can also be observed by monitoring electrical properties during processing. This type of monitoring permits one to control the baking time and temperature to obtain optimum performance. When processing is well controlled, field-effect devices made with II-VI films have been particularly useful in large arrays such as thin-film scan generators and solid-state image sensors.[42,43]

In summary, it should be emphasized that most of the II-VI compounds have relatively similar physical properties when allowance is made for gradual changes associated with position of the components in the periodic table. There is a wide variety of film-deposition methods available, the optimum method depending on the specific applications intended. Many of the physical properties of the films are quite sensitive to deposition technique, and a high degree of control over these methods is desirable.

ACKNOWLEDGMENT

The author is indebted to P. K. Weimer and S. G. Ellis for valuable discussions.

BIBLIOGRAPHY

1. See, for example, R. H. Bube, *Photoconductivity of Solids*, John Wiley and Sons, Inc., New York, 1960.
2. P. K. Weimer, "The TFT—A New Thin-Film Transistor," *Proc. IRE*, Vol. 50, p. 1462, June 1962.
3. J. deKlerk and E. F. Kelly, "Coherent Phonon Generation in the Gigacycle Range via Insulating Cadmium Sulfide Films," *Appl. Phys. Letters*, Vol. 5, p. 2, July 1964.
4. D. C. Reynolds, G. Leies, L. L. Antes, and R. E. Marburger, "Photovoltaic Effect in Cadmium Sulfide," *Phys. Rev.*, Vol. 96, p. 533, Oct. 1964.
5. H. K. Henisch, *Electroluminescence*, Pergamon Press, New York, 1962.
6. O. D. Elpat'evskaia, "The Formation Mechanism of Thin Films of Mercury Selenide and Mercury Telluride," *Zhur. Tekh. Fiz.*, Vol. 28, p. 2669, 1958. Transl.: *Sov. Phys. Tech. Phys.*, Vol. 3, p. 2439, Dec. 1958.

7. F. V. Shallcross, "Effects of Fabrication Parameters on Structural and Electronic Properties of Thin CdS and CdSe Films," *Trans. Metall. Soc. AIME*, Vol. 236, p. 309, March 1966.
8. *Physics and Chemistry of II-VI Compounds*, M. Aven and J. S. Prener, Eds., North-Holland Publishing Company, Amsterdam, 1967.
9. N. A. Goryunova, *The Chemistry of Diamond-Like Semiconductors*, translation published by Chapman and Hall, London, 1965.
10. E. Mollwo and R. Stumpp, "The Reflectivity Spectrum of Cadmium Oxide Single Crystals in the Intrinsic Absorption Edge Region," *Z. für Physik*, Vol. 184, p. 286, April 1965.
11. G. Déchêne, "A New Type of Photoelectric Cell," *Compt. Rend.*, Vol. 208, p. 95, Jan. 1939.
12. R. W. G. Wyckoff, *Crystal Structures*, 2nd Ed., Vol. 1, Interscience, New York, 1963.
13. P. Goldfinger and M. Jeunehomme, "Mass Spectrometric and Knudsen-Cell Vaporization Studies of Group 2B-6B Compounds," *Trans. Farad. Soc.*, Vol. 59, p. 2851, Dec. 1963.
14. G. A. Somorgai and J. E. Lester, "Charge-Transfer-Controlled Vaporization of Cadmium Sulfide Single Crystals. I. Effect of Light on the Evaporation Rate of the (0001) Face," *J. Chem. Phys.*, Vol. 43, p. 1450, Sept. 1965.
15. G. A. Somorgai and H. B. Lyon, "Charge-Transfer-Controlled Vaporization of Cadmium Sulfide Single Crystals. II. Effect of Copper Doping on the Evaporation Rate of the (0001) Face," *J. Chem. Phys.*, Vol. 43, p. 1456, Sept. 1965.
16. K. Günther, "Interfacial and Condensation Processes Occurring with Multicomponent Vapours," in *The Use of Thin Films in Physical Investigations*, Ed. by J. C. Anderson, p. 213, Academic Press, London, 1966.
17. C. A. Escoffery, "Improved Knudsen-Cell Vapor Source for Vacuum Depositions," *Rev. Sci. Instr.*, Vol. 35, p. 913, July 1964.
18. R. T. Galla, "New Cadmium Sulfide Evaporation Source," *Rev. Sci. Instr.*, Vol. 36, p. 403, March 1965.
19. J. deKlerk and E. F. Kelly, "Vapor-Deposited Thin-Film Piezoelectric Transducers," *Rev. Sci. Instr.*, Vol. 36, p. 506, April 1965.
20. N. F. Foster, "Structure of CdS Evaporated Films in Relation to Their Use as Ultrasonic Transducers," *Jour. Appl. Phys.*, Vol. 38, p. 149, Jan. 1967.

21. R. M. Malbon, D. J. Walsh, and D. K. Winslow, "Zinc-Oxide Film Microwave Acoustic Transducers," *Appl. Phys. Letters*, Vol. 10, p. 9, Jan. 1967.
22. See, for example, *Symposium on the Deposition of Thin Films by Sputtering*, Consolidated Vacuum Corporation, Rochester, N. Y., 1966.
23. N. Schwartz, "Reactive Sputtering" in *Transactions of the Tenth National Vacuum Symposium* (1963), p. 325, MacMillan, New York, 1963.
24. T. K. Lakshmanan and J. M. Mitchell, "Some Properties of Sputtered Sulphide Films," in *Transactions of the Tenth National Vacuum Symposium* (1963), p. 335, MacMillan, New York, 1963.
25. S. M. Thomsen and R. H. Bube, "High-Sensitivity Photoconductor Layers," *Rev. Sci. Instr.*, Vol. 26, p. 664, July 1955.
26. Y. T. Sihvonen, S. G. Parker, and D. R. Boyd, "Printable Insulated-Gate Field-Effect Transistors," *Jour. Electrochem. Soc.*, Vol. 114, p. 96, Jan. 1967.
27. R. R. Chamberlin and J. S. Skarman, "Chemical Spray Deposition Process for Inorganic Films," *Jour. Electrochem. Soc.*, Vol. 113, p. 86, Jan. 1966.
28. I. J. Hegyi, "Preparation of CdS Films on Glass Substrates by Close-Spaced Chemical Transport," in *Extended Abstracts, Electronics Division, Electrochem. Soc. Meeting*, Washington, D. C., Oct. 1964; Vol. 13, No. 2, "Semiconductors," p. 72, Electrochemical Society, New York, 1963.
29. S. A. Semiletov, "Electron-Diffraction Study of the Structures of Thin Layers of Cadmium Sulfide, Selenide, and Telluride," *Kristallografiya*, Vol. 1, p. 306, May-June 1956. Transl.: *Sov. Phys. Cryst.*, Vol. 1, p. 236, May-June 1956.
30. A. Brunnschweiler, "Epitaxial Deposition of Cadmium Selenide," *Nature*, Vol. 209, p. 493, Jan. 1966.
31. R. R. Chamberlin, "Effects of Substrate on Films of Chemical Spray Deposited CdS," *Amer. Ceram. Soc. Bul.*, Vol. 45, p. 698, Aug. 1966.
32. I. Spînulescu-Carnaru, "The Crystalline Structure of ZnTe Thin Films," *Phys. Stat. Solidi*, Vol. 18, p. 769, 1966.
33. I. Spînulescu-Carnaru, "Growth of Hexagonal Crystallites in CdTe Thin Films," *Phys. Stat. Solidi*, Vol. 15, p. 761, 1966.
34. K. V. Shalimova, A. F. Andrushko, and I. Dima, "Polymorphism of Zinc Selenide," *Kristallografiya*, Vol. 10, p. 497, July-

Aug. 1965. Transl.: *Sov. Phys. Cryst.*, Vol. 10, p. 414, Jan.-Feb. 1966.

35. P. Mark, "Photo-Induced Chemisorption on Insulating CdS Crystals," *J. Phys. Chem. Solids*, Vol. 25, p. 911, Sept. 1964.
36. Nguyen Duy Thuoc, "Magnetogalvanic and Electrical Properties of CdS Layers Deposited in Different Vacua," *Compt. Rend.*, Vol. 263B, p. 826, Oct. 1966.
37. J. M. Gilles and J. Van Cakenberghe, "Photoconductivity and Crystal Size in Evaporated Layers of Cadmium Sulphide," *Nature*, Vol. 182, p. 862, Sept. 1958.
38. A. Vecht, "Recrystallization of Evaporated Films of ZnSe," *Nature*, Vol. 201, p. 486, Feb. 1964.
39. J. Dresner and F. V. Shallcross, "Rectification and Space-Charge-Limited Currents in CdS Films," *Solid-State Electronics*, Vol. 5, p. 205, July-Aug. 1962.
40. J. Dresner and F. V. Shallcross, "Crystallinity and Electronic Properties of Evaporated CdS Films," *Jour. Appl. Phys.*, Vol. 34, p. 2390, Aug. 1963.
41. P. K. Weimer, "The Insulated-Gate Thin-Film Transistor," in *Physics of Thin Films*, Ed. by G. Hass and R. E. Thun, Vol. 2, p. 147, Academic Press, New York, 1964.
42. P. K. Weimer, G. Sadasiv, L. Meray-Horvath, and W. S. Homa, "A 180-Stage Integrated Thin-Film Scan Generator," *Proc. IEEE*, Vol. 54, p. 354, March 1966.
43. P. K. Weimer, G. Sadasiv, J. E. Meyer, Jr., L. Meray-Horvath, and W. S. Pike, "A Self-Scanned Solid-State Image Sensor," *Proc. IEEE*, Vol. 55, p. 1591, Sept. 1967.

COPLANAR-CONTACT GUNN-EFFECT DEVICES

By

J. F. Dienst, R. Dean*, R. Enstrom, and A. Kokkas

RCA Laboratories
Princeton, New Jersey

Summary—The fabrication and operation of Gunn-effect devices with coplanar contacts is described. For the generation of low-frequency (~1 GHz) high-power oscillations, this geometry has some advantages over the conventional sandwich structure. However, high-power operation has not been attempted as yet with these devices. The best device produced a microwave output of 385 mW at 680 MHz with an efficiency of over 2%. The coplanar devices were operated in microstrip circuits and direct microscopic observation of the device during operation was possible.

Introduction

Since J. B. Gunn's [1] discovery of high-frequency current instabilities in n-GaAs, much additional research has gone into understanding and exploiting this phenomenon. The bulk of the work reported to date deals with devices of the "sandwich" type geometry, i.e., plane-parallel ohmic contacts on opposite sides of the GaAs crystal (Figure 1(a)). However, the sandwich configuration may not be the best geometry from the standpoint of such important factors as operating frequency, heat sinking, and mass-production techniques. An evolution in the geometry of these devices may be expected similar to that which has taken place in transistor technology where we now have, for example, planar transistors with intricate junction geometries. In this paper we discuss some experimental results for a device configuration in which the ohmic contacts are placed on the same face of the GaAs crystal. We call such configurations "coplanar" devices (Figure 1(b)).

The coplanar device is attractive for the following reasons. Firstly, in trying to generate low-frequency (~1 GHz) high-power oscillations using the transit-time mode in a sandwich device, the problem of providing a heat sink for the active region becomes severe. The active

* Now at Princeton University, Princeton, N. J.

[1] J. B. Gunn, "Instabilities of Current in III-V Semiconductors," *IBM Journal of Research and Development*, Vol. 8, p. 141, April 1964.

region must be about 100 μm long at this frequency, resulting in a heat flow path length of about 50 μm, assuming effective heat sinking at both contacts. For the coplanar geometry, the frequency is determined by the gap between the coplanar contacts, which is $\sim$ 100 μm for one GHz; the dimension perpendicular to the carrier flow can be made much smaller, say 10 μm. In addition to heat sinking at the contacts,

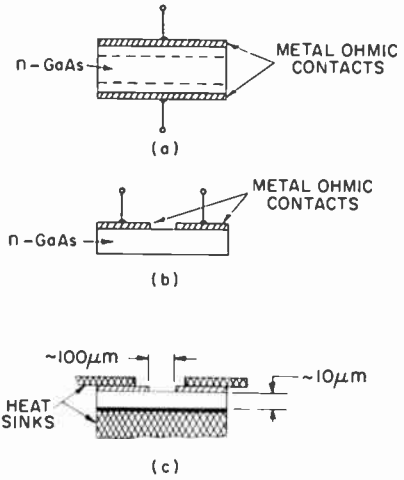


Fig. 1—(a) Sandwich geometry, (b) coplanar-contact, and (c) Gunn-effect devices.

an electrically isolated heat sink can also be applied to the opposite side of the device, thereby substantially reducing the heat flow path (Figure 1(c)). Secondly, the coplanar geometry allows all contacting to be done simultaneously on one side of the material, an obvious advantage for mass production. Thirdly, devices with coplanar access electrodes fit naturally into integrated circuit microstrip applications.

In this paper we give the results of some experimental work with coplanar-contact devices prepared for epitaxially grown GaAs. The following section describes the problems attendant upon the preparation of the starting material, fabrication of the devices, and, especially, the importance of carefully prepared surfaces for application of the contacts. The third section is devoted to a description of the experimental results and a discussion of the circuits used in the experiments.

Device Fabrication

All of the devices and results described herein were obtained from

epitaxially deposited layers of GaAs. However, coplanar devices have also been fabricated from bulk-grown GaAs and that work [2] was the foundation for the work reported here.

The system used to vapor deposit GaAs from AsH_3, Ga, HCl, and H_2 has been described in detail.[3] N-layers in thicknesses of from 1 to 100 μm have been prepared in n+-n-n+ structures with a modified version of this system. Measurements of electron carrier concentration and mobility in the n-layers have yielded concentrations as low as 10^{15} cm^{-3} and mobilities as high as 7200 cm^2/volt-sec at 300°K and 60,000 cm^2/volt-sec at 77°K.

For the coplanar devices it was necessary to grow thin, smooth n-layers of comparable purity over relatively large areas on semi-insulating rather than the usual high-conductivity GaAs substrates. The substrates were Cr-doped, < 100 > oriented GaAs with a resistivity of 4.7×10^8 ohm-cm and mechanically polished to a mirror finish. Layers 25 μm thick deposited on these substrates were very smooth and had electron carrier concentrations of approximately 2×10^{15} cm^{-3} as determined by point-contact-probe breakdown voltages. Successful devices require that the contacting surface be very smooth and clean. Microscopic scratches at the edge of the gap allow incipient spikes of contact metal to form during the metal evaporation and alloying processes, and these generate local high-field regions during operation. As a result, one or more of these spikes elongate and eventually short-circuit the gap. Such filament formation is shown in Figure 2.

The active layer had a resistivity of approximately 0.5 ohm-cm, and an evaporated layer of tin with a nickel overlay was used to make ohmic contact. The wafer was then alloyed in a hydrogen atmosphere at 400°C and sawed into individual units as shown in Figure 3.

Electrical Behavior

Terminal Characteristics

As is well known, the successful operation of Gunn-effect devices requires contacts that are ohmic. The usual criteria for ohmic contacts are: (a) that the low-field resistance ($E \ll E_T$) be constant, (b) that the low-field resistance be independent of the direction of current flow and (c) that the experimental low-field resistance of the device agree well with the resistance as calculated from the known

[2] C. T. Wu, RCA Laboratories, private communication.

[3] J. J. Tietjen and J. A. Amick, "The Preparation and Properties of Vapor-Deposited Epitaxial $GaAs_{(1-x)}P_x$ Using Arsine and Phosphine," *Jour. Electrochem. Soc.*, Vol. 113, p. 724, 1966.

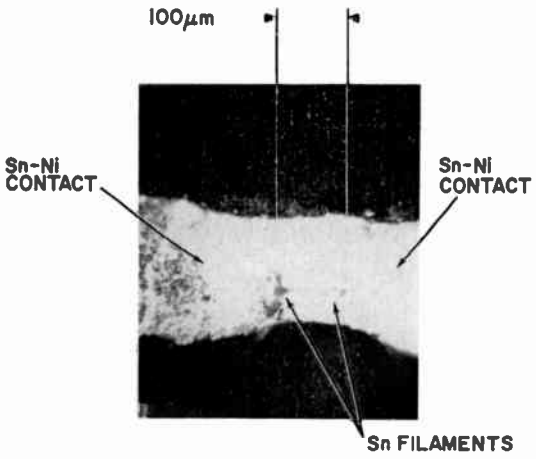


Fig. 2—Photomicrograph of a coplanar device clearly showing the filaments of tin across the gap.

resistivity of the material and the geometry of the device. The first two are determined by simply measuring the current–voltage characteristic for a given device, and are independent of the geometry. The last requirement can be checked rather easily for the sandwich geometry, where the current density is uniform. However, for the coplanar devices, the current density is not uniform, so that the calculation of the resistance of a device is more complicated.

Two-dimensional analogs of the coplanar devices were constructed from resistance paper to study the effects of altering the length of the contacts and the thickness of the active layer while keeping the gap length constant. These studies showed that the length, L, of the contacts does not substantially affect either the resistance of the device or the field distribution so long as L is greater than both the gap length, l, and the thickness, t, of the active layer. Figure 4 shows an

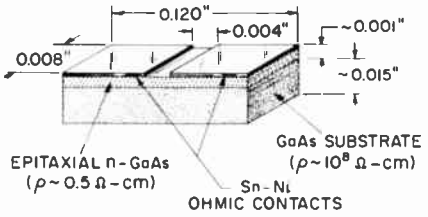


Fig. 3—Structural details of a typical coplanar Gunn-effect device.

experimental curve illustrating the variation of normalized device resistance (R/r), as a function of layer thickness (r is the resistivity of the resistance paper in ohms per square and R is the total resistance measured between the analog contacts). With this curve it is possible to calculate the active layer resistivity ρ from the dimensions of the sample and the total low-field resistance, R_o,

$$\rho = bR_o(R/r)^{-1},$$

where b is the width of the sample. On the basis of this equation and

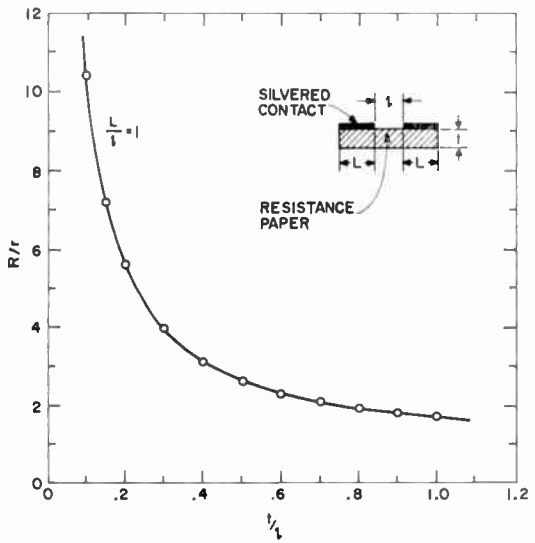


Fig. 4—Normalized resistance obtained from a resistance-paper model.

Figure 4, resistivities of 0.46 and 0.52 ohm-cm were calculated for two particular devices of widths 0.008 inch and 0.198 inch, respectively. These results agree well with a resistivity of 0.54 ohm-cm calculated from the carrier concentration obtained by breakdown voltages and an assumed mobility of 6000 cm^2/volt-sec and also attests to the uniformity of the epitaxially deposited n-layer.

Figure 5 shows a typical pulsed I-V characteristic for a coplanar device. It is from this characteristic that R_o was obtained.

Coherent Oscillations

Bias-voltage pulses of 80-100 nanoseconds duration were applied to the sample, which was mounted in a simple microstrip circuit. The

sample was placed in series with the hot side of the microstrip line and a small current-monitoring resistor. Contact to the coplanar electrodes of the sample was made by small probes soldered to the stripline. The voltage developed across the current-monitoring resistor was observed on a sampling oscilloscope.

A coherent current instability was observed when bias voltages exceeded a threshold of about 57 volts. A typical oscilloscope trace is shown in Figure 6(a); however, this is not a faithful reproduction of the current wave shape because of the reactance associated with the

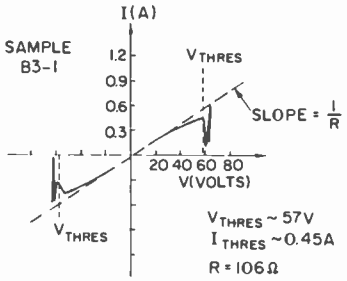


Fig. 5—Typical *I–V* characteristic of a coplanar device.

monitor resistor and its leads. To obtain the current wave shape, the voltage wave form of Figure 6(a) was Fourier-analyzed into d-c through third-harmonic components, and each of these was divided by the appropriate computed reactance of the monitor resistor. These current components were then added in the proper phase relationship and the resulting current wave form of Figure 6(b) was obtained. This current wave form is in good qualitative agreement with the published results of other workers,[4-6] and indicates that we are seeing domain formation and translation in the active layer.

The average threshold bias voltage for the onset of noisy current instabilities was about 57 volts. Within a small range of voltages (≈ 5 volts) above this threshold the oscillations were coherent. Since it is difficult to calculate the electric field in the active layer, we again made use of the resistance-paper analog. Figure 7 shows a field plot for a bias voltage of 60 volts, which is within the range for coherent

[4] A. G. Foyt and A. L. McWhorter, "The Gunn Effect in Polar Semiconductors," *IEEE Trans. ED*, Vol. ED-13, p. 79, Jan. 1966.

[5] J. S. Heeks, "Some Properties of the Moving High-Field Domain in in Gunn Effect Devices," *IEEE Trans. ED*, Vol. ED-13, p. 68, Jan. 1966.

[6] D. E. McCumber and A. G. Chynoweth, "Theory of Negative-Conductance Amplification and of Gunn Instabilities in Two-Valley Semiconductors," *IEEE Trans. ED*, Vol. ED-13, p. 4, Jan. 1966.

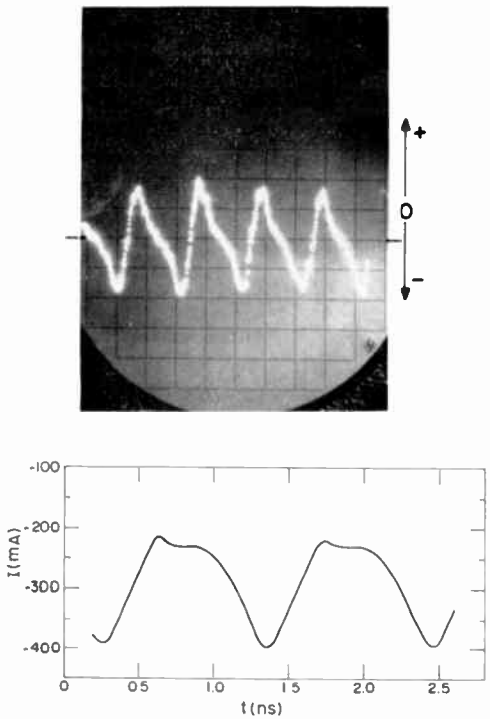


Fig. 6—(Top) Voltage developed across current-monitoring resistor (0.5 nanosecond/division horizontal scale, and 0.3 ampere/division vertical scale), and (bottom) current through the coplanar device producing the voltage waveform in (a).

oscillations. The electric field over most of the active layer is rather uniform and appreciably higher than the accepted threshold field of about 3.3 kilovolts/cm for the Gunn effect in GaAs. However, the field at the edge of the active layer opposite the contacts and to either side of the uniform-field region is approximately the threshold field (see Figure 7). Apparently for bias voltages less than ≈ 57 volts the

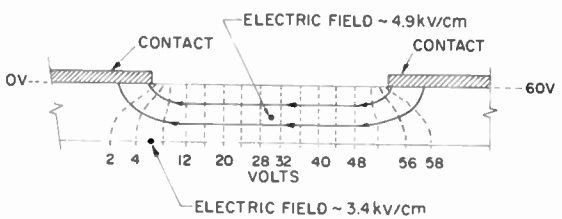


Fig. 7—Potential distribution in a resistance-paper analog of a coplanar device.

electric field at the edges of the uniform-field region falls below the threshold field even though the field in the vicinity of the source electrode is higher than the threshold field. This parallel low-field path tends to collapse any space-charge accumulation layer that tries to form near the source electrode. This condition leads to either no oscillations at all or noisy oscillations, depending upon the size of the low-field shunt path.

Bias voltages in excess of about 10-15% of the threshold voltage caused the coherent oscillations to break up into noisy oscillations, and again this is consistent with previously reported [7] results.

Using the microstrip measuring circuit, we were also able to observe the device under a microscope during operation. In general, the anode contact was the first to deteriorate at high bias voltages, and the spikes seen in Figure 2 seem to emanate from the anode. Prior to any noticeable change in either of the electrodes, microscopic white particles could be seen erupting from the crystal at random locations within the gap. Sporadic current fluctuations accompanied these eruptions. After a time the entire gap surface was covered with a film of this fine white powder that could easily be brushed away. The powder was not analyzed but is probably an oxide of gallium.

Circuit Effects

The fundamental transit-time frequency for all of the samples tested was about 600-700 MHz, which is quite a bit lower than the 1000 MHz that was expected on the basis of a gap length of 100 μm. The lower transit-time frequency in low-reactance circuits suggests that the domains follow a curved path that is longer than the gap length.

To improve the power output and investigate the effects of the circuit surrounding the device, a variety of microstrip circuits were designed and used. In all cases the output voltage was taken as that voltage developed across the 50-ohm impedance of the sampling oscilloscope. The microstrip-circuit technique proved very useful from several aspects. First, it was easy to change line impedance and construct a variety of tuning elements. Second, since the sample was not mounted in a cartridge, parasitic package reactances were avoided. Third, since the sample was not soldered into the circuit, the same sample could be used in the various circuits as they evolved.

The best performance was obtained with the circuit shown in the photograph in Figure 8. The tuners and shorting stub were set so

[7] D. G. Dow, et al., "High-Peak-Power Gallium Arsenide Oscillators," *IEEE Trans. ED*, Vol. ED-13, p. 105, Jan. 1966.

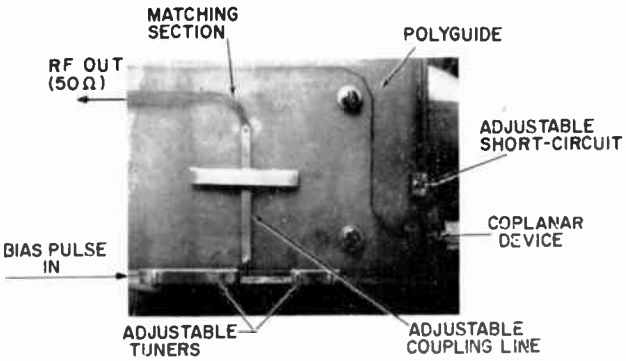


Fig. 8—Microstrip circuit constructed of Polyguide.

that, at the fundamental frequency, the device was presented an inductive reactance approximately equal to the low-field d-c resistance. The circuit, when viewed from the device, should be open-circuited at the second harmonic, and the $\lambda/4$ tuner nearer the device was adjusted accordingly. These conditions are in agreement with the findings of Carroll.[8] A device inserted in the circuit adjusted according to these criteria produced a useful output of 385 mW peak pulsed power at 680 MHz with an efficiency of better than 2%. This was an 0.008-inch-wide sample, and the circuit required only a minimum of readjustment from the predicted values to achieve this result. It is estimated that at least a 3-db improvement in power output can be achieved by using

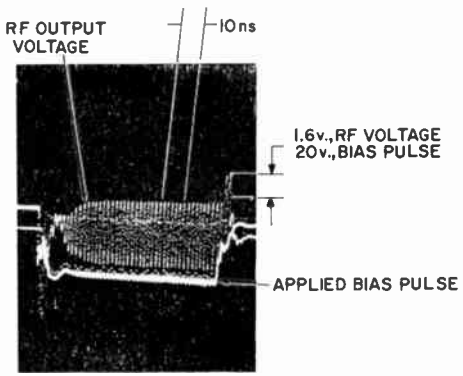


Fig. 9—RF output from the circuit shown in Figure 8.

[8] J. E. Carroll, "Resonant Circuit Operation of Gunn Diodes: a Self Pumped Parametric Oscillator," *Electronics Letters*, Vol. 2, p. 215, June 1966.

a lower-loss dielectric for the microstrip and fixed, rather than adjustable, tuning elements. Figure 9 illustrates the output-voltage wave form as viewed with the sampling oscilloscope.

Summary

The operation of Gunn-effect devices having coplanar contacts has been successfully demonstrated. The units were used in microstrip circuits, and coherent peak pulsed power outputs as high as 385 mW at 680 MHz were obtained from 0.008-inch-wide devices, with conversion efficiencies of over 2%.

During the course of these preliminary experiments, we were able to directly observe the devices during operation and identify the anode as the contact first subject to deterioration under excessive bias conditions. We believe that this is probably also true of the conventional sandwich geometry devices and underlines the importance of providing an adequate heat sink at this electrode for high-power operation. With improved epitaxial material and contacts, i.e., contacting material capable of withstanding higher operating temperatures, it should be possible to realize significant increases in power output from these devices.

EFFECT OF NOISY SYNCHRONIZATION SIGNALS ON SPLIT-PHASE CODED SYSTEM PERFORMANCE*

BY

A. B. GLENN AND G. LIEBERMAN

RCA Defense Electronic Products
Camden, N. J.

Summary—The performance of a noncoherent frequency-shift-keyed system utilizing split-phase waveforms was investigated under conditions of noisy synchronizing signals. The analysis showed the probability of degradation of signal-to-noise ratio or bit-error probability due to time jitter at the switching and sampling operations.

WHEN both the frequency and the phase of a received digital signal are unknown, a receiver with post-detection matched filters represent the optimum. This condition becomes significant when the frequency uncertainty of the carrier is greater than the bandwidth required by the information.[1] In actual communication systems it is impossible to maintain exact synchronization between the sampling times and the end of the data pulses. Since the bit probability of error depends upon the ratio of signal energy to noise-power spectral density at the instant of sampling, it is evident that lack of synchronization will deteriorate the system performance. The degradation of the system performance will increase as the lack of synchronization or time jitter on the sampling pulses increases. This paper analyzes system degradation as a function of the time jitter on the sampling pulses for the split-phase coded waveform. This type of waveform is analyzed because of its desirable bandwidth requirements and transition density characteristics.

SYNCHRONIZER DESCRIPTION

Figure 1 shows the detector and synchronizer for the split-phase waveform of a noncoherent frequency-shift-keyed receiver operating

* This work was done under NASA Contract NAS2-3772 for AMES Research Center.

[1] A. B. Glenn, "Analysis of Noncoherent FSK Systems with Large Ratios of Frequency Uncertainties to Information Rates," *RCA Review*, Vol. XVII, p. 272, June 1966.

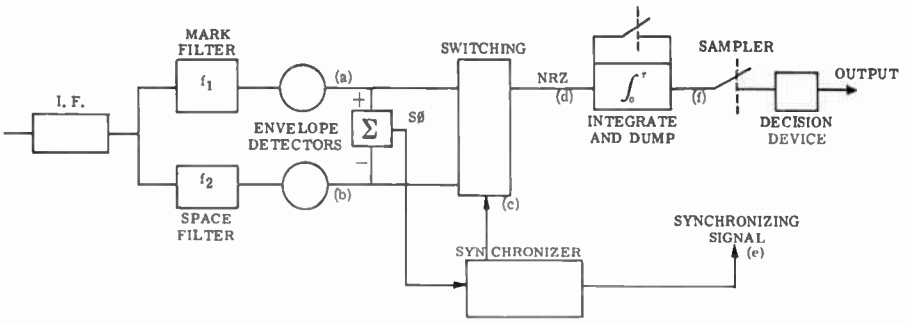


Fig. 1—Noncoherent FSK receiver.

under conditions where the exact frequency and phase of the received signal are unknown. Under these conditions, the optimum receiver is one with post-detection matched filters.

The structure of the signal for a mark consists of frequency f_1 transmitted during the first half of the signal duration, T, and frequency f_2 transmitted during the second half. As shown in Figure 2, the envelope detector at the output of the band-pass filter tuned to f_1, shown by waveform (a), consists of a positive pulse (or mark) during the first half, and zero during the second half of the signal duration.

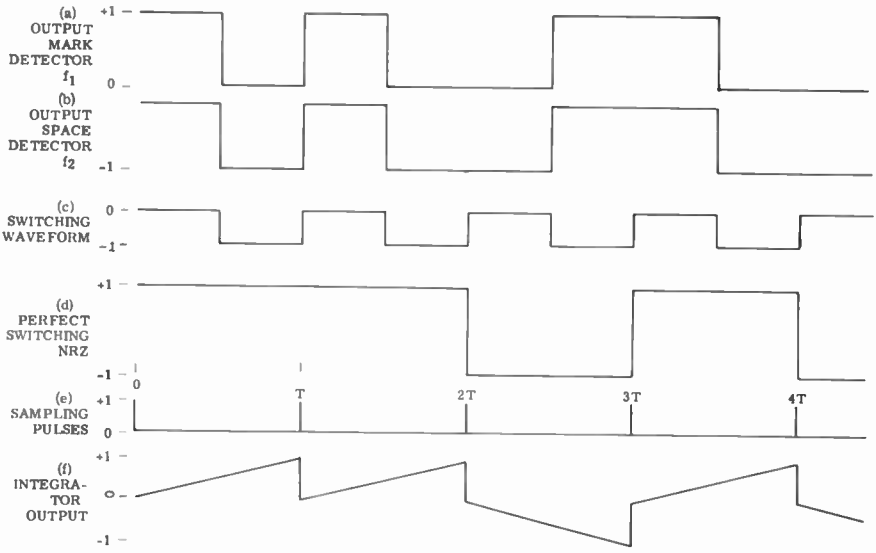


Fig. 2—Effect of zero timing errors for split-phase coded waveform.

The other envelope detector output, shown by waveform (b), is reversed in polarity; i.e., it has a zero output during the first half and a negative pulse (or space) output during the second half. Similarly, the structure of the signal for a space consists of frequency f_2 transmitted during the first half of the information bit and frequency f_1 transmitted during the second half.

The synchronizing signal is obtained at the output of the summer. After this signal is suitably processed, the output is used for the switching and the sample and dump operations. Let us first consider zero timing errors for these operations, as shown in Figure 2. The switching waveform is shown by (c). The sync pulses at the reversing switch input alternately produce a polarity reversal of the respective inputs from the two envelope detectors. If a mark is transmitted, the positive pulse from the f_1 envelope-detector output keeps its polarity, whereas the second negative pulse from the polarity reversed f_2 envelope-detector output is reversed in polarity. The resulting output shown by waveform (d) consists of a positive pulse of duration T. This assumes that the sync pulses occur at integral multiples of $T/2$ corresponding to the f_1, f_2 switching times.

This analysis shows that a split-phase waveform enters the switching circuit and a non-return-to-zero (NRZ) waveform leaves at point (d) in Figure 1. The NRZ waveform is then applied to an integrate and dump circuit. Sampling pulses, shown by waveform (e), from the synchronizer will sample and dump the integrator output. The resulting waveform is indicated by (f). The above discussion illustrates the results for perfect switching and sampling of a split-phase coded waveform.

The two basic modes for synchronization are acquisition and tracking.[2] If the mission requires fast acquisition, the bandwidth of the phase-locked loop (PLL) should be large. The limit is determined by the S/N ratio and the resulting phase jitter, which determines the probability of unlock.[3] For example, the time for a 1% probability of unlock during the acquisition mode should be at least ten times the acquisition time. The acquisition time should be about one-hundredth the total communication time. The switching, sample, and dump operations will depend upon the phase jitter in the synchronizing signal. In other words, the probability of making a decision at the end of each information bit depends upon the S/N ratio. But, as will be discussed, S/N will depend upon the phase jitter. Since the bandwidth of the

[2] F. M. Gardner, *Phaselock Techniques*, John Wiley & Sons, Inc., New York, 1966.

[3] R. W. Sanneman and J. R. Rowbotham, "Unlock Characteristics of the Optimum Type II Phase-Locked Loop," *Trans. IEEE*, Vol. ANE-11, p. 15, March 1964.

phase jitter in the synchronizing signal depends upon the S/N in the PLL bandwidth, this bandwidth during the tracking mode should be as small as possible. Since the PLL requirements during the fast acquisition and tracking modes are directly opposite, the two-mode PLL will be optimum for these mission requirements.

Synchronization Extraction from a Split-Phase Binary Signal

In order to extract synchronization information from a split-phase message, the split-phase format must be converted to a return-to-zero (RZ) form. This must be done in order to obtain a discrete component at the keying frequency.[4,5] This is accomplished by nonlinear processing, such as delay, inversion, addition, and squaring. Bennett[5] showed that the power at the discrete keying frequency is 4 db below the total RZ power. A second-order, two-mode phase-locked loop[2] will be used to filter this discrete component from the output of this band-pass filter. If the frequency difference between the received signal and the voltage-controlled oscillator (VCO) is less than the PLL bandwidth, the PLL will acquire in the lock-in frequency mode.* In this mode, the acquisition time is inversely proportional to the natural radian frequency (ω_n) of the PLL. After acquisition, or lock-in, the loop will switch to the tracking mode where the PLL bandwidth B_L is less than one-tenth of the acquisition PLL bandwidth. This tracking bandwidth is also about one hundredth the information bandwidth. When the sine of the phase difference between the VCO output and the incoming signal plus noise can be approximated by the phase difference, the phase-locked loop performance can be described by a linear model. With this model, the VCO output phase equals the phase angle of the loop-filter output for an input equal to the loop-input signal plus noise. The VCO phase jitter equals the difference between this signal-plus-noise phase angle, and the phase angle in the absence of noise.

For loop-filter-output signal-to-noise ratios greater than 10 db, the VCO phase jitter is approximately normally distributed with RMS value[2]

$$\sigma \doteq \left[2\left(\frac{S}{N}\right) \right]^{-1/2} \text{ radians}. \qquad (1)$$

* Lock-in time as defined as the first zero-time crossing of the transient phase error.

[4] S. J. Martin, "A Comparison of the Synchronization Efficiency Offered by Several PCM Code Formats in a Noisy Channel," *Proc. Nat. Telemetry Conf.*, p. 342, 1966.

[5] W. R. Bennett and J. R. Davey, *Data Transmission*, Chap. 19, p. 315, McGraw-Hill Book Co., New York, 1965.

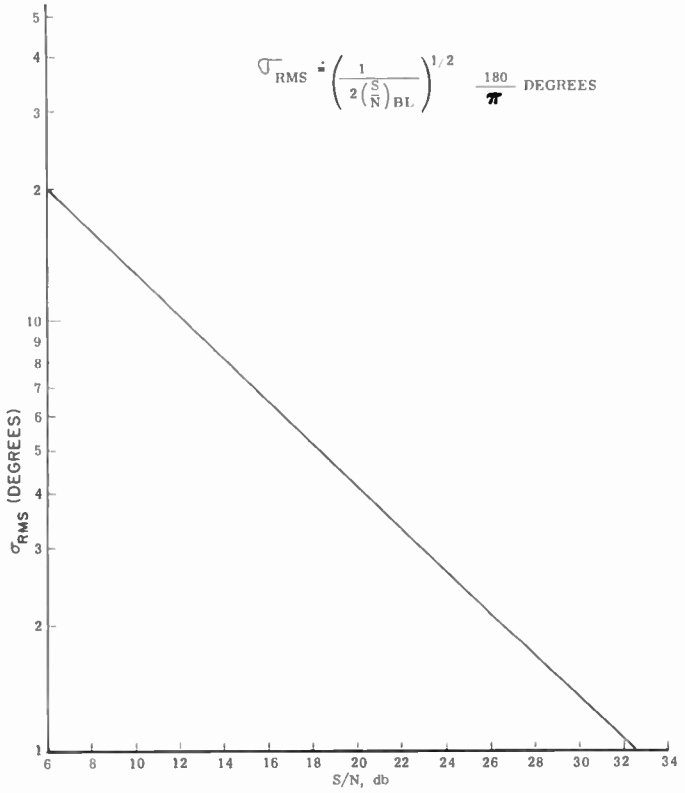


Fig. 3—Phase jitter versus signal-to-noise ratio.

This relationship is plotted in Figure 3.

Assuming that the total communication time is about 20 seconds, the acquisition time should be less than 0.2 second. The limitation on this acquisition time is that the time for a 1% probability of unlock should be greater than 20 seconds. Figure 4, which is taken from Reference (4) and extended (dotted curve), shows the normalized mean time to unlock τ versus the phase jitter for a second-order PLL.

Decrease in S/N Ratio Due to Phase Jitter in a Receiver Employing Extracted Timing

The extracted reference signal plus noise is

$$r(t) = A \cos\left(\frac{\pi t}{T}\right) + n(t) , \qquad (2)$$

where T is the information bit duration. This produces a timing signal with successive axis crossings T seconds apart. If this function is processed in a phase-locked loop, the VCO output signal will be

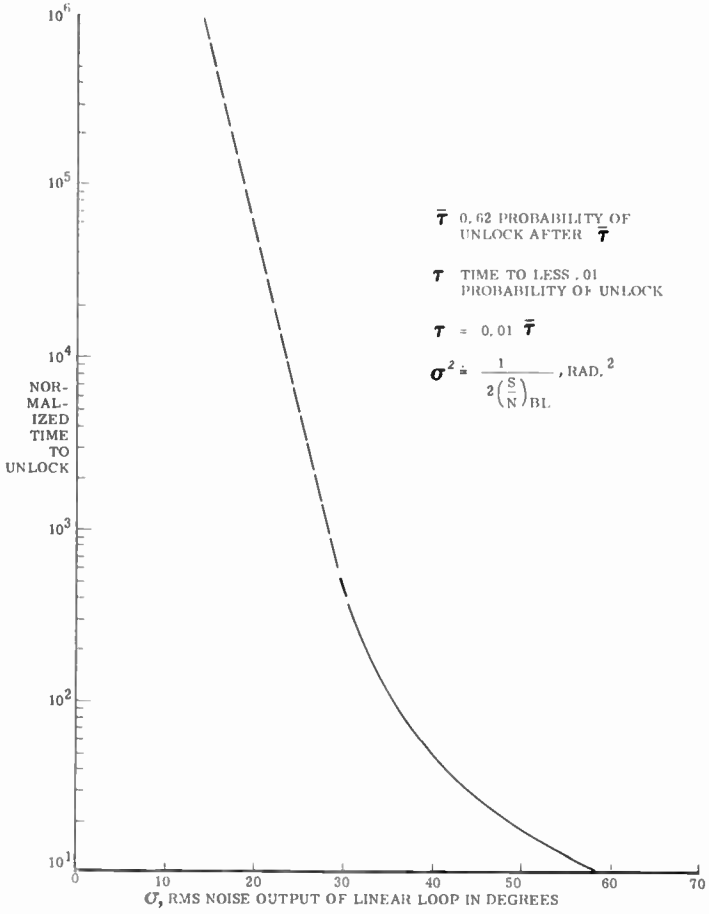


Fig. 4—Normalized mean time to unlock versus phase jitter (Reference (3)).

$$v_o(t) = \cos\left[\frac{\pi t}{T} + \Delta\theta(t)\right], \tag{3}$$

where $\Delta\theta$ represents phase jitter at the output of the loop filter.

This phase jitter has to be related back to the timing jitter in the sampling process. This timing jitter can be related back to an equivalent loss in signal-to-noise ratio. This loss in S/N ratio can then be

directly applied to the error probability versus S/N characteristics for any system. The degradation in error probability can then be easily determined.

The instantaneous output signal-to-noise-power ratio at the video matched-filter output is of the form:

$$\rho(t) = \begin{cases} \left(\frac{t}{T}\right)^2 \frac{E}{N_o}; & 0 \leq t \leq T \\ \left(2 - \frac{t}{T}\right)^2 \frac{E}{N_o}; & T \leq t \leq 2T \end{cases} \tag{4}$$

where E/N_o is the maximum signal-energy to white-noise-power spectral density at this low-pass-filter output. The equivalent time jitter for phase jitter $\Delta\theta$ is

$$\Delta t = \frac{\Delta\theta T}{\pi}; \qquad |\Delta\theta| < \pi. \tag{5}$$

Equation (5) assumes that the reference signal in Equation (2) is time delayed by Δt, and that this delay can either be positive or negative. This time-delayed reference signal can be written

$$\cos\left(\frac{\pi(t + \Delta t)}{T}\right) = \cos\left(\frac{\pi t}{T} + \Delta\theta\right), \tag{6}$$

from which the definition of $\Delta\theta$ in Equation (5) becomes evident. From Equation (4), the maximum degradation in output signal-to-noise ratio occurs when one samples at time $t = 0$ or $t = 2T$. Let t_s denote the actual sampling time. One can write

$$t_s = T + \Delta t \tag{7}$$

From Equation (4), there will be no degradation when $\Delta t = 0$. Positive Δt corresponds to sampling late. Negative Δt corresponds to sampling early. The worst cases occur when $\Delta t = \pm T$. Assuming $|\Delta t| \leq T$, it follows that $|\Delta\theta| \leq \pi$, as stated in Equation (5). In practical systems, Δt will be considerably less than T.

Figure 5 shows the switching effects of successive positive (late)

timing errors. Waveform (a) represents the f_1 envelope detector output. Waveform (b) represents the negative f_2 envelope detector output. Waveform (c) shows a set of time-displaced switching pulses of duration $T/2$ due to a sequence of positive timing errors. Waveform (d) shows the desired switch output waveform with no timing errors. The last waveform (e) shows the consequences of imperfect switching.

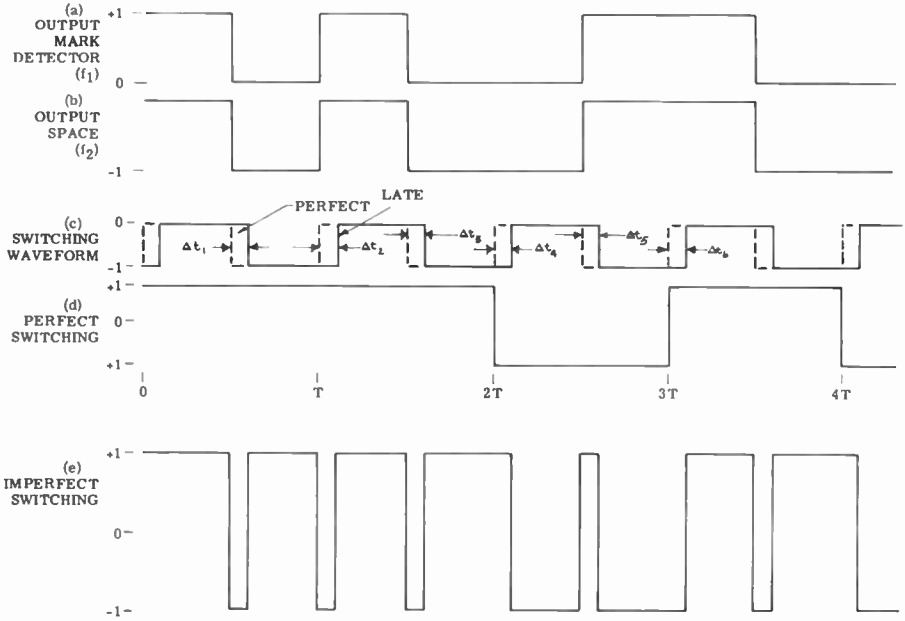


Fig. 5—Effect of late timing errors in the switching operation for split phase.

To illustrate how this last waveform was obtained, let us consider the first information bit. Due to the positive timing error Δt_1, the negative f_2 envelope is switched late, thereby causing a notch of width Δt_1 in waveform (e). The switching operation consists of multiplying the negative switching waveform (c) and the sum of waveforms (a) and (b). No operation takes place during the first part of the information bit until this switching process begins. The switching waveform is initiated by the same timing source as the sampling waveform used for the integrate and dump operations. When the switching waveform occurs, it reverses the polarity of either waveform (a) or (b). Assuming $B_l T \ll 1$, the switching pulse will be turned off late by an amount Δt_{k+1}, which is approximately equal to Δt_k. Δt_k is the time that the switching pulse was late.

The net switching effects are as follows. When there are two successive marks or spaces, there will be two notches per information pulse, one near the center and one at the end. When there are alternate mark and space, there will be one notch near the center of the information bit. The results are essentially the same for negative or early timing errors.

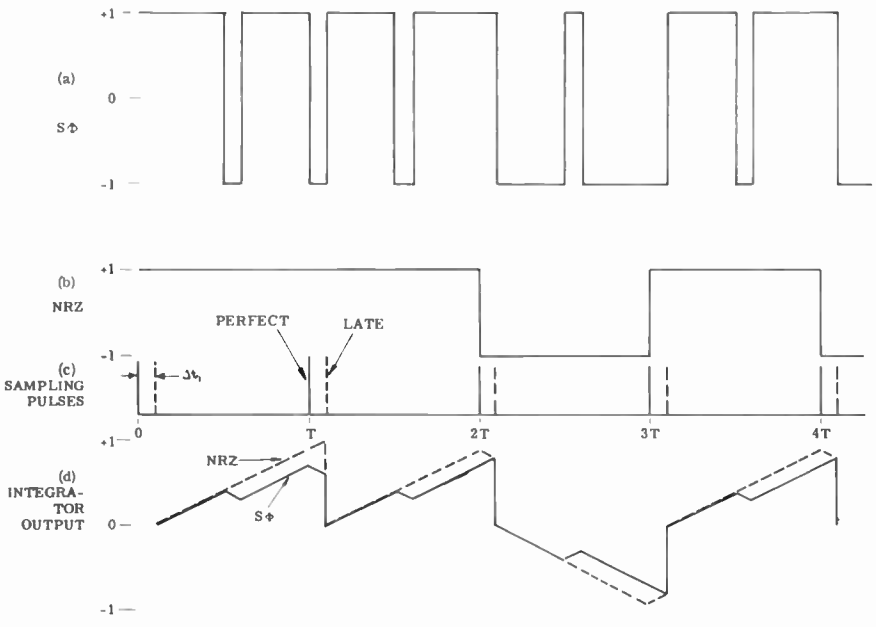


Fig. 6—Effect of late timing errors on NRZ and split-phase coded waveforms.

Figure 6 shows the combined effects of the switching and sample and dump errors due to imperfect late timing. Waveform (a) shows the switching output for positive (late) timing errors. Waveform (b) shows the ideal waveform for perfect switching. Waveform (c) shows the sampling pulses. The waveforms in (d) show the outputs of the integrate and dump outputs for waveforms (a) and (b).

In the absence of switching errors or for an NRZ waveform, the integrator output is shown by the dashed curve. The solid waveform shows the integrator output for the split-phase coded waveform, which was subject to combined switching and sample and dump timing errors. Let us assume that the integrator starts to build up late by Δt_1. During this build-up process, a notch occurs due to switching, causing the integrator to reverse the direction of integration. At the

end of the notch, the integrator starts to integrate positively. The switching notch at the end of the information causes the integrator to again reverse direction until it is finally dumped. The amplitude for this first integrator output is $[1 - 4\,(\Delta t/T)]$.

The four errors occur as follows. One error due to late integration; two errors due to the integrator reversal in the middle of the information bit; and one error due to the integrator reversal at the end of the information bit.

On the second pulse, the integrator starts late again. There is a reversal of the integrator in the middle due to a switching error.

Table I—Amplitude Factors for Combined Switching and Sample and Dump Errors (Split-Phase Waveform); $B_L T \ll 1$.

t_k \ $t_k + 1$	Mark	Space
Mark	$1 - \dfrac{4\lvert\Delta t\rvert}{T}$	$1 - \dfrac{2\lvert\Delta t\rvert}{T}$
Space	$1 - \dfrac{2\lvert\Delta t\rvert}{T}$	$1 - \dfrac{4\lvert\Delta t\rvert}{T}$

The switching error at the end of this second information bit produces an elongation rather than a notch. Since the sampling pulse is late, the integrator continues its positive build-up until it is dumped. This additional build-up cancels one of the timing errors. Therefore, the resulting amplitude factor is $[1 - 2(\Delta t/T)]$. It follows that when there is a polarity reversal at the end of the information, there are four timing errors. This is true for either late or early timing errors. Table I summarizes these results.

From Table I, we obtain an expression for the signal-to-noise ratio corresponding to any sampling pulse for equal mark and space probabilities:

$$\rho = \begin{cases} \left(1 - \dfrac{4\lvert\Delta\theta\rvert}{\pi}\right)^2 \dfrac{E}{N_o} : \text{with probability } \tfrac{1}{2}, \\ \left(1 - \dfrac{2\lvert\Delta\theta\rvert}{\pi}\right)^2 \dfrac{E}{N_o} : \text{with probability } \tfrac{1}{2}, \end{cases} \tag{8}$$

Using the normal distribution approximation to $\Delta\theta$, it follows that $\Delta\theta$ has probability density

$$\begin{aligned} p(y) &= \frac{2}{\sigma}\phi\left(\frac{y}{\sigma}\right); \qquad y \geqq 0 \\ &= 0\ ; \qquad \text{otherwise} \end{aligned} \tag{9}$$

where σ is the standard deviation of $\Delta\theta$ in Equation (1), and

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\}. \tag{10}$$

Since $\Delta\theta$ is a random variable, ρ is characterized by its probability distribution function

$$\begin{aligned} F\rho(z) &= P_r\{\rho \leq z\} \\ &= \frac{1}{2} P_r\{\rho \leq z \,|\, A\} + \frac{1}{2} P_r\{\rho \leq z \,|\, B\}\ , \end{aligned} \tag{11}$$

where A refers to successive marks or spaces, and B refers to successive mark-space reversals. From Equations (8) and (9),

$$\begin{aligned} P_r\{\rho \leq z \,|\, A\} &= P_r\left\{\left(1 - \frac{4|\Delta\theta|}{\pi}\right)^2 \frac{E}{N_o} \leq z\right\} \\ &= P_r\left\{|\Delta\theta| \geqq \frac{\pi}{4}\left[1 - \sqrt{\frac{N_o z}{E}}\right]\right\} \\ &= \int_{\frac{\pi}{4}\sqrt{1-\frac{N_o z}{E}}}^{\infty} \frac{2}{\sigma}\phi\left(\frac{y}{\sigma}\right) dy \\ &= 2\left[1 - \Phi\left\{\frac{\pi}{4\sigma}\sqrt{1 - \frac{N_o z}{E}}\right\}\right]. \end{aligned} \tag{12}$$

Similarly, $$P_r\{\rho \leq z \mid B\} = 2\left[1 - \Phi\left\{\frac{\pi}{2\sigma}\sqrt{1 - \frac{N_o z}{E}}\right\}\right]. \qquad (13)$$

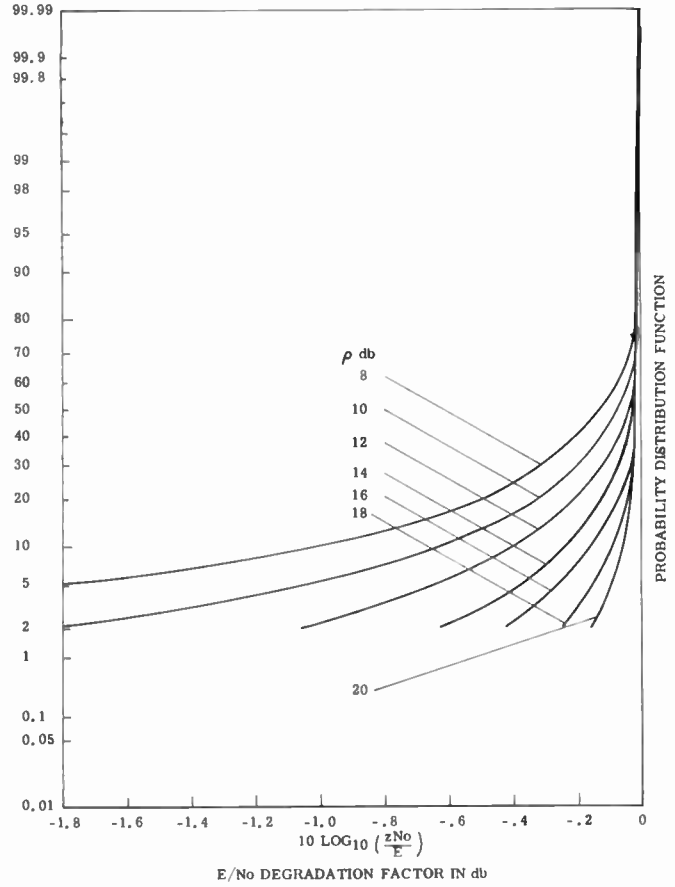


Fig. 7—Distribution function for output signal-to-noise ratio of 8 to 20 db.

Applying Equations (12) and (13) to Equation (11),

$$F_\rho(z) = 2 - \Phi\left\{\frac{\pi}{4\sigma}\sqrt{1 - \frac{N_o z}{E}}\right\} - \Phi\left\{\frac{\pi}{2\sigma}\sqrt{1 - \frac{N_o z}{E}}\right\}, \qquad (14)$$

where σ is the standard deviation of $\Delta\theta$ defined by Equation (1). Figure 3 contains a plot of σ versus ρ_o in decibels.

The distribution function given by Equation (14) is plotted in Figure 7. Since the information bit error probability varies as a negative exponential of E/N_o, the error probability is very sensitive to changes in E/N_o. These curves show, for example, that for a signal-to-noise ratio ρ out of the PLL of 20 db, the S/N ratio degradation will

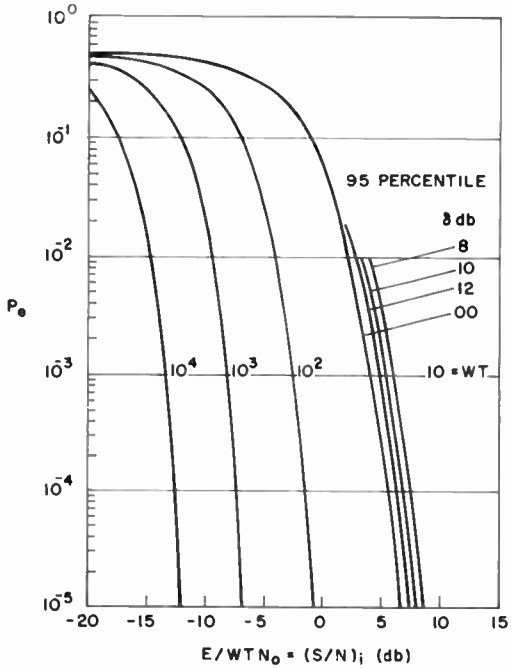


Fig. 8—Bit error probability curves for a wide-band frequency-shift-keyed system for $WT = 10$ to $WT = 10^4$.

be greater than 0.1 db for 5% of the time. Decreasing S/N to 10 db results in an increase in degradation of greater than 1.8 db for 5% of the time.

Figure 8, which is obtained from Reference (1) shows the bit error probability curves for a wide-band frequency-shift keyed system for $WT = 10$ to 10^4, where W is the bandwidth of the predetection filter. The signal-to-noise ratio at the input to the envelope detector is $E/(N_oWT)$, where E is the energy of an information bit and N_o is the noise-power spectral density. These curves were derived for perfect sampling. The curves for $WT = 10$ show the result of imperfect switching and sampling for split-phase coded waveforms. These curves are the 95 percentile points for PLL signal-to-noise ratios (ρ) of 8, 10,

and 12 db. This data is obtained directly from the curves in previous figure.

Conclusions

Since both the time for synchronization and time jitter on the synchronizing signal are important, a two-mode synchronizer utilizing the split-phase coded format was investigated. The analysis showed the probability of degradation of signal-to-noise ratio due to time jitter at the switching and sampling operations.

Acknowledgment

During the course of this investigation the authors were fortunate in receiving contributions from H. Rouland, S. Roth, and H. Zelin of the RCA Astro-Electronics Division.

PRACTICAL ASPECTS OF INJECTION LASER COMMUNICATION SYSTEMS*

By

W. J. Hannan,† J. Bordogna,** and D. Karlsons†

Summary—This paper describes several pertinent design factors for laser line-of-sight communication systems utilizing GaAs laser diodes operating at room temperature in a pulsed mode. Analysis, experimentation, and construction of several such systems have revealed that (1) reliable voice communication can be achieved effectively over ranges of at least five miles with a one-watt pulsed injection laser; (2) if background radiation is high (e.g., daylight operation), there is a receiver field of view beyond which a photodiode receiver minimizes required transmitter power; and (3) for a photodiode receiver that is optimized for detecting the pulsed radiation from a room-temperature-operated GaAs injection laser, use of an optical filter to suppress background radiation will not (in most cases of practical interest) improve overall performance.

Introduction

Several varieties of GaAs room-temperature IR laser communication systems have been developed that provide fair-weather voice communication over a line-of-sight range of more than five miles at sea level. In most of these systems, the equipment design parameters have been optimized to achieve small size, light weight, and low power consumption consistent with desired performance and requirements for portable systems.

Typical components of these systems are the hand-held transmitter and receiver shown in Figures 1 and 2. The transmitter shown in Figure 1 is composed of three modules, as evidenced by the sectional construction shown in the photograph. The upper section is the optics module and contains four GaAs laser diodes,‡ four collimating lenses, a sighting telescope, a microphone, and circuitry to drive the laser diodes. The center section, or electronics module, contains the modulation circuits. The power module (the lower and smallest section) con-

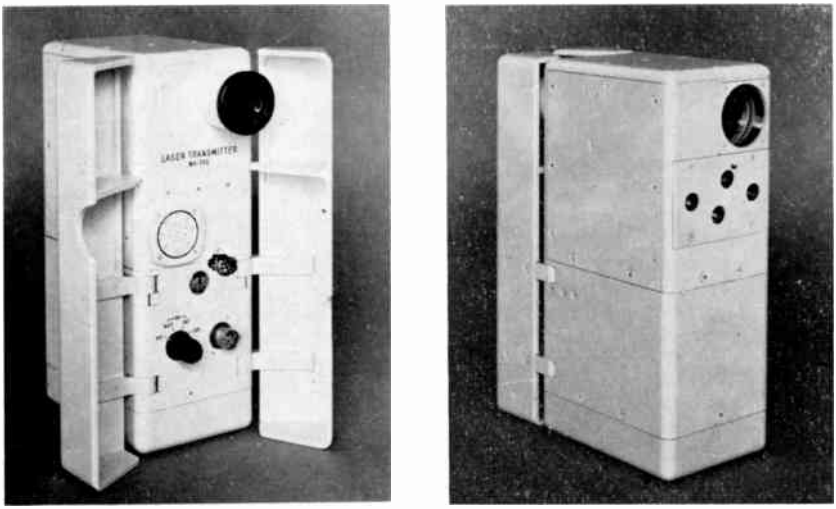
* Research reported in this paper was sponsored by the U.S. Army Electronics Command, Fort Monmouth, N.J., under contract No. DA28-043-AMC-01284(E).

† RCA Laboratories, Princeton, N.J.

** Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pennsylvania.

‡ Arrays of several diodes are used frequently to provide greater transmitter power than is available at the present state of the art with one diode.

tains rechargeable NiCd batteries and a dc-to-dc converter to provide the necessary system operating voltages. The receiver, shown in Figure 2, consists of a hand-held receiver and sighting telescope and separate power module and earphones.



(a) (b)

Fig. 1—Typical hand-held laser voice transmitter: (a) view showing microphone face and handles, and (b) view showing light-emitting face with four lenses that collimate light beams from four GaAs diodes.

Comparison of Laser and Nonlaser Diodes

A comparison of currently available laser and nonlaser diodes reveals that nonlaser diodes can provide higher average output power, but laser diodes can provide higher peak output power. Another significant difference is that nonlaser diodes can operate continuously, whereas laser diodes must be operated in a pulsed mode (at room temperature). Hence, the basic question is which device will provide a greater communication range. One could also consider a pulsed nonlaser diode, but it can be shown that the reduction of average output power (caused by the higher I^2R losses when the device is operated in the pulsed mode) offsets any advantages gained.

Theoretical analysis leads to the conclusion that the laser diode is the better choice for the communication system considered here. In particular, it can be shown that a nonlaser diode would have to emit at least 15 times the average power of a laser diode to provide equivalent

performance. At the present state of the art, laser diodes can emit an average power of about 4 milliwatts. Therefore, a nonlaser diode must emit at least 60 milliwatts to be competitive. Since currently available nonlaser diodes can emit only about 30 milliwatts (at room temperature), the laser diode will provide better performance.

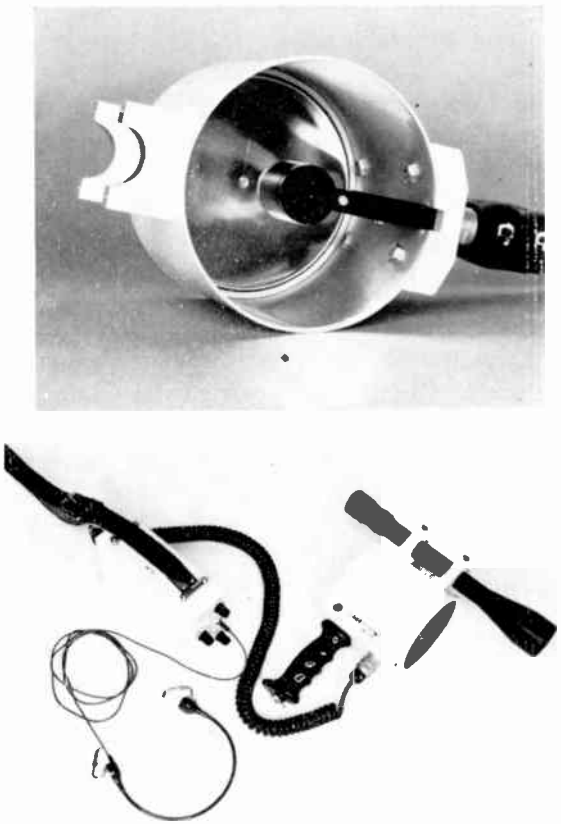


Fig. 2—Typical hand-held receiver: (bottom) hand-held receiver, power pack, and earphones, and (top) close-up view of 4-inch receiving mirror with low-noise photodiode preamplifier at focal point.

When one considers the size of the transmitter optics needed for laser and nonlaser diodes, it is evident that the laser diode has an effective power advantage much greater than 15:1, especially if a very narrow transmitter beam width is needed. Since the radiation from a laser diode is emitted in a relatively narrow cone, it can be collected efficiently with a small lens. For example, consider the situation where the transmitter beam width must be less than 10 milliradians and where

the diameter of the transmitter lens is restricted to 1 inch for maximum portability. Since the active diameter of a typical noncoherent source is about 0.05 inch, the focal length of the transmitter lens must be at least 5 inches. Assuming all the power emitted from a nonlaser diode is radiated in a 60° cone, then only 1/25th of the emitted radiation is collected by the 1-inch lens. On the other hand, all the light emitted from a laser diode can be collected and focused into a 3-milliradian beam by a 16-mm microscope objective lens with a diameter of three-eighths inch. In fact, an array of four laser diodes can be assembled within a 1-inch diameter, increasing the power output 4:1 over that of a single laser transmitter. Thus, for this case, the laser diode obtains an additional power advantage of as much as 100:1.

Optimum Duty Factor of an Injection-Laser Diode

Since currently available room-temperature diodes must be operated in a pulsed mode, the question of optimum duty factor is an important one, especially in relation to its effect on radiated output power. The answer to this question lies in the experimental relationship between peak output power and peak drive current. This relationship is shown in Figure 3, where three regions of diode operation are clearly evident—the nonlasing region below threshold current, the lasing region between the threshold knee (I_{th}) and the saturation knee (I_{sat}), and the lasing region beyond the saturation knee.

Clearly, the optimum operating point lies somewhere in the region above the lasing threshold. The location of this point can be determined by mathematical analysis. It can be shown that when the diode drive current is less than I_{sat} (and greater than I_{th}), laser output power increases approximately as the seventh power of drive current; whereas above I_{sat}, laser output power increases in direct proportion to the drive current. In either region, diode power dissipation increases as the square of the drive current. Therefore, maximum efficiency is achieved at the highest value of drive current in the region between I_{th} and I_{sat}, i.e., at I_{sat}. At this value of drive current, duty factor is limited by the maximum average power dissipation $\overline{P}_{d_{max}}$ that the diode can tolerate. With this limit, optimum duty factor q_{opt} can be shown to be

$$q_{opt} = \frac{\overline{P}_{d_{max}}}{(I_{sat})^2 r}, \tag{1}$$

where r is the diode resistance.

The above expression neglects diode heating effects that occur during the pulse. If the laser diode is driven by long pulses (e.g., greater than 100 nanoseconds), then diode temperature rise during the pulse may lower output power; if the pulse is long enough, the diode may stop lasing. Therefore, the value of q_{opt} given above is valid only for

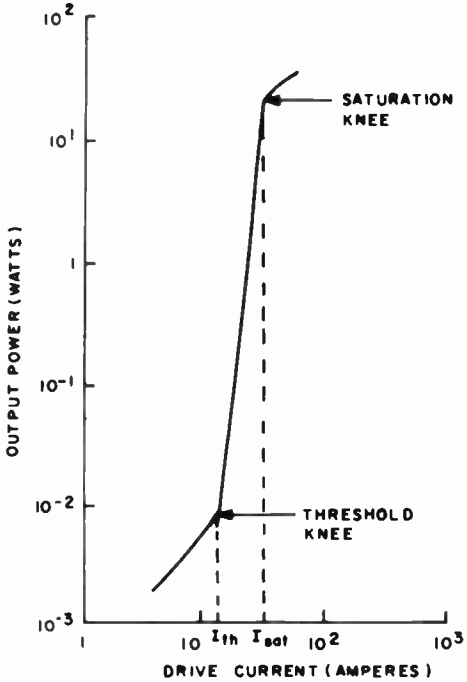


Fig. 3—Light output power versus diode current.

the case where the pulse duration is of the order of 100 nanoseconds or less. Typical parameter values of state-of-the-art diodes are $\overline{P}_{d_{max}} = 0.2$ watt, $I_{sat} = 40$ amperes and $r = 0.2$ ohm. Substitution of these values into Equation (1) gives $q_{opt} \approx 6 \times 10^{-4}$. The duty factor is

$$q = \tau f , \tag{2}$$

where τ is the pulse duration and f is the pulse repetition rate. Since the pulse repetition rate required for 3-kHz voice transmission is $f = 8$ kHz, the optimum pulse duration is 75 nanoseconds.

Receiver Threshold Detection Level

Communication range is limited by the threshold detection level

of the received signal. In particular, the threshold detection level of a pulse-modulation receiver marks a received signal level below which output signal-to-noise ratio (SNR) deteriorates rapidly. As shown in Figure 4, noise is suppressed in a pulse-modulation receiver by the use of limiting and squelch circuits, which, in effect, slice the midsection from the received pulses, thus suppressing noise. It is evident that noise peaks that exceed one-half the peak signal amplitude will generate extraneous pulses. Similarly, if the addition of noise and signal drops the resultant pulse amplitude below its nominal half-amplitude

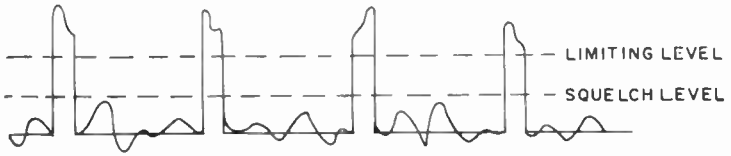


Fig. 4—Signal at output of pulse modulation receiver.

value, then a signal pulse will be suppressed. Hence, the threshold detection level corresponds to the level at which the peak signal amplitude is twice the peak noise amplitude. The signal-to-noise power ratio that corresponds to this condition can be determined as follows.

The SNR of the receiver will be a maximum[1] when the receiver bandwidth is $B = 0.4/\tau$, where τ is the pulse duration of the laser transmitter. For this condition, the ratio of peak-to-rms signal amplitude is approximately $\sqrt{2}$. It is well known that the noise prevalent in an optical receiver has a gaussian amplitude distribution, the ratio of peak-to-rms noise amplitudes being about 4:1.* Accordingly, the ratio of signal power to noise power at threshold is

$$\left(\frac{S}{N}\right)_{th} = \frac{(\text{peak signal current}/\sqrt{2})^2}{(\text{peak noise current}/4)^2} = \left[\frac{(2)(4)}{\sqrt{2}}\right]^2 = 32. \quad (3)$$

From this result, it follows that the maximum communication range is that range at which the signal-to-noise ratio has dropped to 32, or 15 db.

[1] C. Cherry, *Pulses and Transients in Communication Circuits*, p. 219, Chapman and Hall, Ltd., London, 1949.

* Over a sufficiently long period of time (i.e., a period much greater than the reciprocal of the receiver bandwidth), the peak amplitude of gaussian noise will exceed four times the rms amplitude approximately 0.006% of the time.

Comparison of Photodiode and Multiplier Phototube Receivers

The salient feature of the multiplier phototube is the high, essentially noise-free, internal gain of its electron multiplier. For most applications, the gain of this device is high enough to make amplifier noise negligible compared to the amplified shot noise from the photocathode. On the other hand, a photodiode has no internal gain, and amplifier noise may be significant. For applications where received background radiation is high, the shot noise from a photodiode may be comparable to (or even higher than) the noise introduced by the amplifier. For these applications, the photodiode will provide better performance than a multiplier phototube, because the responsivity of a photodiode is about two orders of magnitude higher than that of the photocathode of a multiplier phototube.

The expression for received background noise power is

$$P_b = \frac{M\alpha_r^2 A_r \xi T_a T_o B_{\text{opt}}}{4} \tag{4}$$

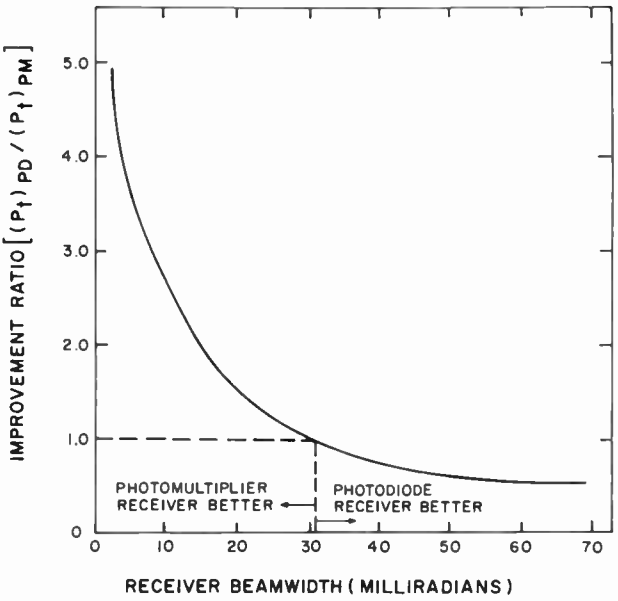
where M is the solar irradiance, α_r the receiver beam width, A_r the area of receiver optics, ξ the background reflection coefficient, T_a the transmission of atmosphere, T_o the transmission of optics, and B_{opt} is the pass band of optical filter.

As can be seen, the received background power is proportional to the square of receiver angular beam width. It follows that increasing the receiver beam width increases the required transmitter power because background power is increased. The effect of beam width on required transmitter power is different in the case of the photodiode receiver than it is for the multiplier phototube receiver. To determine this difference, the ratio of required transmitter power for the photodiode receiver to required transmitter power for the multiplier phototube receiver [i.e., $(P_t)_{PD}/(P_t)_{PM}$] was determined and plotted in Figure 5. This plot is based on typical system specifications and the equation for signal-to-noise ratio at the output of an optical receiver:

$$\frac{S}{N} = \frac{\rho^2 P_s^2 R_l G^2}{2eB(\rho P_s + \rho P_b + I_d) R_l G^2 + 2FkTB} \tag{5}$$

where ρ = responsivity of photodetector (amperes/watt),
P_s = received signal power (watts),
P_b = received background power (watts),
R_l = load resistance (ohms),

e = charge on an electron (1.6×10^{-19} coulomb),
B = bandwidth (Hz),
F = noise factor of preamplifier,
k = Boltzmann's constant (1.38×10^{-23} joule/°K),
T = temperature (°K),
I_d = detector dark current (amperes),
G = internal gain of photodetector.



$A_r = 8 \times 10^{-3}$ meter^2
$B_{opt} = 100°$ A
$\tau = 75$ nanoseconds
$M = 0.1$ watt/meter^2/A
$T = 300$°K
$T_a = 0.2$
$\alpha_t = 3$ milliradians
$\xi = 0.5$
$R_l = 1000$ ohms
$I_d = 2 \times 10^{-7}$ ampere (photodiode)
1×10^{-10} ampere (phototube)
$\rho = 0.3$ ampere/watt (photodiode)
3×10^{-3} ampere/watt (phototube)
$S/N = 32$
$F = 2$
$\lambda = 9000$ A

Fig. 5—Power improvement ratio versus receiver beam width.

The curve in Figure 5 was obtained by substituting the given parameters into Equation (4) and solving for background power P_b as a function of receiver beam width α_r. This value of P_b was substituted into Equation (5) along with the given parameters, and received signal power [both $(P_s)_{PD}$ and $(P_s)_{PM}$] at the 15-db threshold signal-

to-noise ratio was determined. Since required transmitter power is directly proportional to received signal power, the ratio of $(P_s)_{PD}/(P_s)_{PM}$ yielded $(P_t)_{PD}/(P_t)_{PM}$ directly.

The curve in Figure 5 indicates that when the receiver beam width is wider than 31 milliradians, the photodiode receiver provides better performance, since required transmitter power for the photodiode receiver is less than that for the multiplier phototube receiver in this region.

Of course, the 31-millirad crossover point is valid only for the particular set of parameters selected for the analysis. If a different set of parameters were used the crossover point could change radically. For example, if the curve shown in Figure 5 were plotted for night operation, when received background power is very low, the result would show no crossover point, indicating that the multiplier phototube would provide better performance regardless of receiver beam width. The important fact revealed by this analysis is that (1) there are certain conditions under which a photodiode will provide better performance than a multiplier phototube and (2) these conditions are likely to exist in injection-laser communication systems.

Optical Filtering

Of the two terms present in the denominator of the noise expression of Equation (5), the first term expresses the effect of shot noise on the signal-to-noise ratio and the second gives the effect of noise introduced by the load resistor and the preamplifier following the optical detector. Since the optical filter serves to reduce the background power P_b, it is apparent that this filter affects the magnitude of only the first term in the denominator.

It can be shown that a photodiode receiver designed to detect 75-nanosecond laser pulses is amplifier-noise limited. Accordingly, the second term in the denominator of Equation (5) predominates in such a receiver, and, therefore, the expression for SNR reduces to

$$\frac{S}{N} = \frac{\rho^2 P_s^2 R_l G^2}{2FkTB}. \tag{6}$$

Equation (6) indicates that the SNR of the photodiode considered here is virtually independent of background radiation. Thus, it is evident that the addition of an optical filter to suppress background radiation will have little effect on SNR. In fact, addition of an optical filter may actually deteriorate performance because of the transmission loss due to the filter itself.

Photodiode Bias in a Laser Receiver

In a photodiode receiver, the signal-to-noise ratio is dependent on photodiode bias, since shot noise varies as a function of bias voltage. This bias can be critical or noncritical, depending on the value of the detector load resistor used. Compare, for example, a noncoherent (non-laser) voice communication system that utilizes a narrow receiver bandwidth of about 6 kHz and a coherent (laser) system that employs a 5 MHz bandwidth. In either system, the maximum permissible load resistance is given by

$$R_l = \frac{1}{2\pi CB}, \tag{7}$$

where B is the bandwidth and C is the total capacitance shunting the load resistor. Since C is essentially the same in either system, it is evident that the detector load resistor is orders of magnitude smaller in the laser system. As a result, the performance of the laser system is virtually independent of shot noise generated within the photodiode (i.e., photodiode bias remains essentially constant despite large changes in received background power). Thus, the laser system is affected little by changes in background radiation and, therefore, complex automatic bias controls are not needed.

Required Transmitter Power Versus Range

Based on the state-of-the-art and the desire for minimum size and maximum portability, a typical set of laser communication system parameters can be listed as follows:

Diameter of receiver optics	4 inches
Responsivity of photodiode	0.3 ampere/watt
Transmission wavelength	9020 Å
Transmission of optics	0.7
Information bandwidth	3 kHz
Noise factor of preamplifier	2
Pulse repetition rate	8000 Hz
Transmitter beam width	3 milliradians
Receiver beam width	23 milliradians
Pulse duration	75 nanoseconds
Detector bandwidth	5 MHz
Detector load resistance	1000 ohms
Threshold signal-to-noise ratio	15 db
Solar irradiance at 9020 Å	0.1 watt/meter^2/Å
Atmospheric attenuation coefficient	0.2 km^{-1}

Based on the above system parameters, Equation (6), and the expression for required transmitter power,

$$P_t = \frac{P_s \alpha_t^2 R^2}{A_r T_o T_a} ; \tag{8}$$

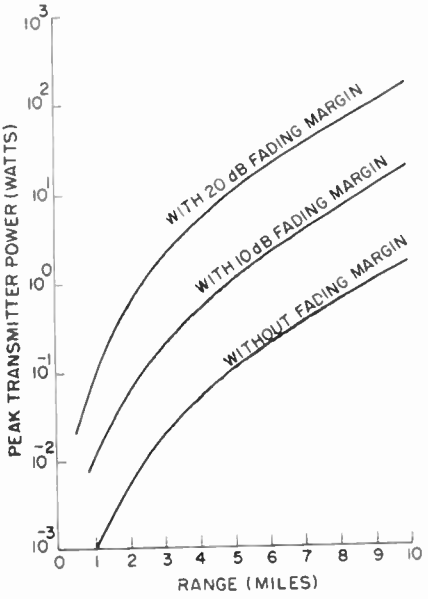


Fig. 6—Required transmitter power versus range.

the required transmitter power-versus-range is plotted in Figure 6 for several values of fading margin. Test results with such a system have revealed that the received signal level frequently fluctuates 10 db due to atmospheric perturbations over a 5-mile path. Occasionally, the signal level fluctuates as much as 20 db. Hence, as shown in Figure 6, reliable communication over a 5-mile range calls for a peak transmitter power of more than 1 watt.

ACKNOWLEDGMENTS

Contributions to the development of the equipment described here were made by C. W. Reno, T. E. Penn, G. D. Clubine, and R. Tarzaiski of RCA Defense Electronic Products. The laser diodes were developed at RCA Laboratories under the direction of H. R. Lewis.

AN IMPROVED CASSEGRAIN MONOPULSE FEED SYSTEM

By

C. E. Profera, Jr. and L. H. Yorinks

RCA Missile and Surface Radar Division
Moorestown, New Jersey

Summary—The efficiency and, hence, the gain of a Cassegrain antenna may be increased beyond that which has been previously attainable by illuminating the antenna subreflector with a wave-guide horn excited with a combination of the TE_{10} and LSE_{12} wave-guide modes. The wave-guide horn is positioned so that the subreflector is in the near field of the horn aperture. The radiation characteristics of such a feed are discussed and the expected antenna performance is deduced. The feed efficiency, that is the product of spillover efficiency and antenna aperture taper efficiency, is greater than 80% throughout a band of 10 to 15%. Monopulse operation is obtained by exciting the feed with additional higher-order wave-guide modes.

Introduction

THE inherent low-noise-temperature characteristics of Cassegrain antennas are well known. Incorporation of a high-efficiency feed, one that provides a low spillover-loss characteristic with no compromise of antenna-aperture efficiency, results in a low-noise high-efficiency reflector antenna system suitable for radar, satellite communications, or radio astronomy applications.

Improved performance of the Cassegrain antenna is provided by improving feed radiation characteristics. A successful technique provides for improved feed performance by exciting a single wave-guide horn with a combination of modes. A monopulse tracking capability is conveniently provided in the single horn by generation of additional modes. The resulting single-aperture multimode monopulse feeds have been shown to provide increased antenna efficiency relative to four- or five-horn monopulse feeds.

The feed system described here is a monopulse system that provides a significant efficiency improvement over the multimode monopulse feed mentioned above. It combines multimode and near-field radiation characteristics and produces patterns that are essentially side-lobe free over 10 to 15% bandwidths, resulting in significant reduction of feed spillover loss in the Cassegrain geometry. The feed spillover-loss reduction results in improvement of both sum and differ-

ence modes of operation so that increased sum and difference mode antenna gains result.

Description of Feed System

The improved Cassegrain monopulse feed combines the beam-collimating properties of near-field feeds with the beam-shaping and monopulse capabilities of controlled higher-order-mode radiation. It takes advantage of the fact that very near the aperture (within a maximum limit determined by aperture dimensions and wavelength), the radiation pattern is essentially a reproduction of the aperture electric-field distribution.

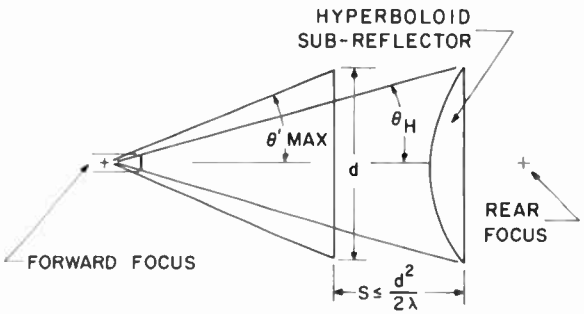


Fig. 1—Near-field Cassegrain feed system.

For horn feeds, the near-field pattern closely resembles the electric-field mode distribution in the horn, which is a spherical wave centered at the feed vertex. For a square aperture with side dimension d, the limit of the near-field region is determined approximately by the relation

$$S \leq \frac{d^2}{2\lambda} \tag{1}$$

where S is the distance from the center of the feed aperture to a point of observation in the near field and λ is the free-space wavelength. In the resulting spherical-wave near-field Cassegrain feed system, which is illustrated in Figure 1, the feed vertex is coincident with the forward focal point of a hyperboloid subreflector.

Verification of the near-field radiation properties of simple pyramidal horns was obtained by pattern computation. A program was written to accommodate any radiating pyramidal horn mode or combination of modes. Beam-width characteristics of computed dominant

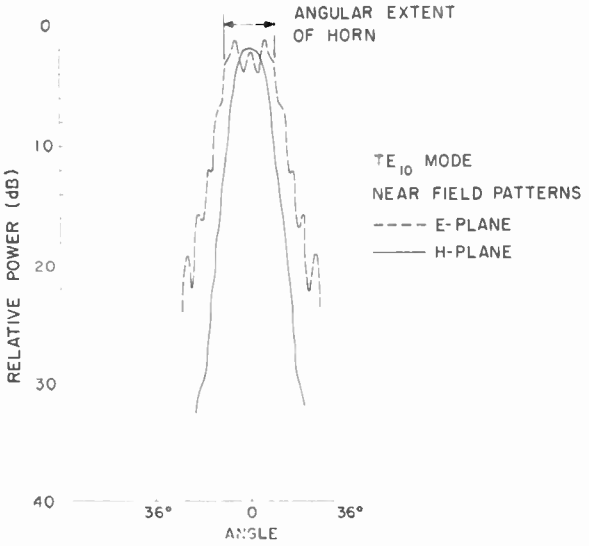


Fig. 2—E- and H-plane near-field patterns.

(TE_{10}) mode near-field radiation were compared to experimentally obtained pattern data for specific horn geometries. Excellent agreement was obtained between analytical and actual results. Typical TE_{10} mode *E* and *H*-plane near-field patterns are shown in Figure 2. The *E*-plane pattern resembles the *E*-plane aperture mode distribution, which is uniform over the angular extent of the horn, $2\theta'_{MAX}$. The *H*-plane pattern is tapered and side-lobe free, resembling the TE_{10} mode *H*-plane cosine distribution.

Use of the near-field technique for improved feed efficiency in a Cassegrain geometry requires that the feed pattern be rotationally symmetric and virtually side-lobe free. These characteristics are provided by tapering the TE_{10} mode *E*-plane aperture distribution by generation of a higher-order mode. The required higher-order mode is the LSE_{12} (hybrid $TE_{12} + TM_{12}$) mode. Figure 3 illustrates the TE_{10}

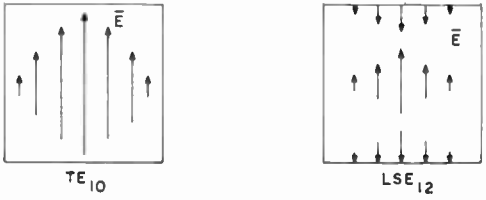


Fig. 3—Sum monopulse modes.

and LSE_{12} mode electric-field distributions. The spherical-mode field distribution resulting from a combination of the TE_{10} and LSE_{12} modes for the horn geometry shown in Figure 4 and for the field polarized in the θ' direction is

$$E_{10+12}(\theta',\phi') = \cos\frac{\pi\phi'}{2\phi'_{MAX.}} + B\left[\cos\frac{\pi\phi'}{2\phi'_{MAX.}}\right]\left[\cos\frac{\pi\theta'}{\theta'_{MAX.}}\right]. \quad (2)$$

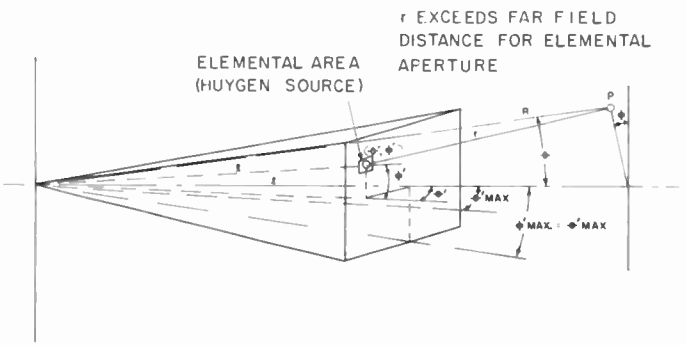


Fig. 4—Near-field horn geometry for feed-pattern computation.

Here,

$$B = \frac{LSE_{12}\text{ mode amplitude}}{TE_{10}\text{ mode amplitude}}\, e^{j\psi},$$

and ψ is the phase difference between TE_{10} and LSE_{12} modes at the feed aperture. A typical tapered E-plane feed aperture distribution for $B = 0.6$ is shown in Figure 5.

The near-field patterns were derived from the field distribution at the aperture of a conventional pyramidal horn. In order to arrive at the specific aperture distribution, the horn was assumed to be an infinitely long quasi-pyramidal waveguide [1] propagating the desired mode. The field distribution at those points in the infinite guide corresponding to the aperture of the finite guide was taken as the aperture distribution. The aperture was then divided into elemental areas, or Huygen's sources, of dimensions such that the near-field pattern

[1] F. E. Borgnis and C. H. Papas, "Electromagnetic Waveguides and Resonators," *Handbuch Der Physik*, Vol. XVI, 1958.

observation points were well in the far-field of each elemental source. The pattern was then obtained from the following expression

$$\mathbf{G}_P(\theta,\phi) = \sum_{-\phi'_{MAX}}^{\phi'_{MAX}} \sum_{-\theta'_{MAX}}^{\theta'_{MAX}} \frac{\mathbf{E}_M(\theta',\phi')}{|\bar{r}|} e^{j\beta|\bar{r}|} \left[\frac{1+\cos\theta''}{2} \right], \quad (3)$$

where $\mathbf{G}_p$ is the complex vector field at a given observation point "p" in the near field of the horn aperture; $\mathbf{E}_M(\theta',\phi')$ is the vector field at

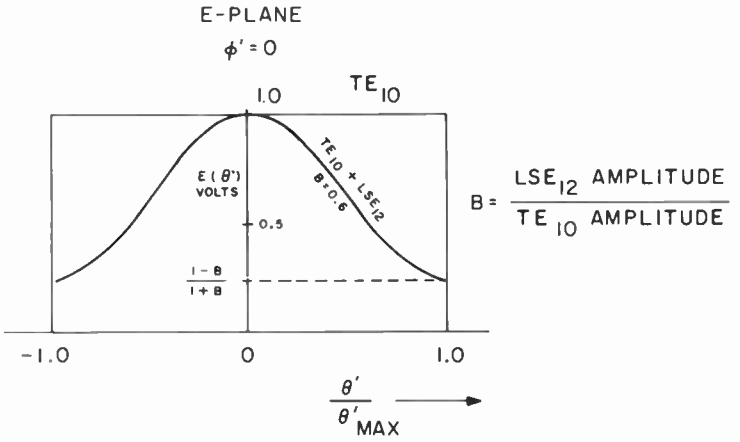


Fig. 5—Normalized E-plane distributions.

the aperture point (θ',ϕ') (electric field of Huygen source) for the desired mode; $\bar{r}$ is the vector from the aperture point (θ',ϕ') to the observation point p; θ'' is the angle between the normal to the Huygen source and a ray from the source to p, and $\beta = 2\pi/\lambda$.

The effect of the LSE_{12} mode on the E-plane near-field radiation pattern is illustrated by comparison of computed TE_{10} and $TE_{10} + LSE_{12}$ near-field patterns in Figure 6. Computed E, H, and 45° plane near-field patterns of a horn excited with the $TE_{10} + LSE_{12}$ modes are shown in Figure 7. The computed phase variation relative to a spherical wave centered at the feed vertex is also shown in Figure 7 for the various planes. For optimum Cassegrain subreflector illuminations, the phase variation is a maximum of 20°. This phase variation has only a negligible effect on antenna performance.

Incorporation of the spherical-wave near-field feed in a Cassegrain geometry requires that the feed and subreflector diameters be as large as practical and that their separation be as small as practical.

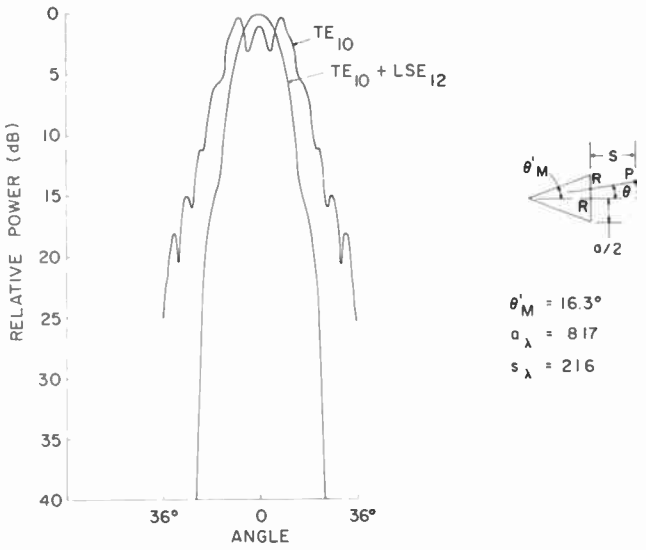


Fig. 6—E-plane near-field patterns.

Therefore, the Cassegrain geometry must be designed so that feed and subreflector aperture shadowing are essentially equal. In addition, it is generally desirable to make the feed length as short as

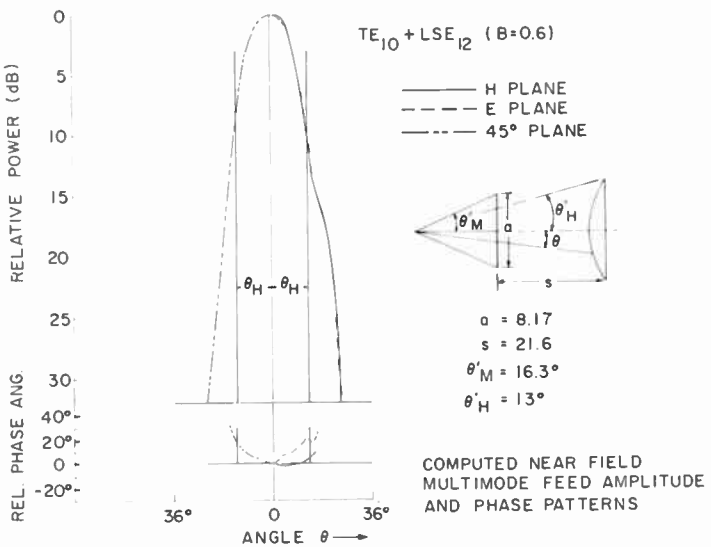


Fig. 7—E, H, and 45° plane near-field patterns.

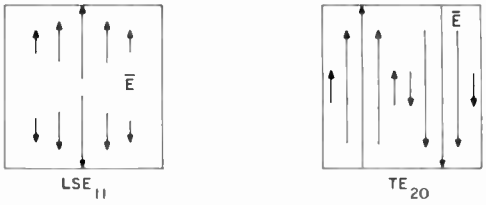


Fig. 8—Difference modes.

possible to provide space within the feed support cone for microwave network packaging. Optimum subreflector illumination occurs for $\theta_{11} \approx 0.8\,\theta'_{MAX}$, so that the hyperboloid included angle at the feed vertex is less than the horn included angle.

Monopulse capability is provided by independent generation of two additional modes, namely the LSE_{11} ($TE_{11} + TM_{11}$) and TE_{20} modes, which are illustrated in Figure 8. Computed near-field E (LSE_{11}) and H (TE_{20}) plane difference patterns are shown in Figure 9. A significant improvement in antenna difference-mode performance will result using the spherical-wave near-field multimode feed. The feed difference-pattern envelope falls partially within the sum-pattern envelope and is virtually side-lobe free, resulting in a difference-mode spillover loss that is lower than that exhibited by more conventional monopulse feeds.

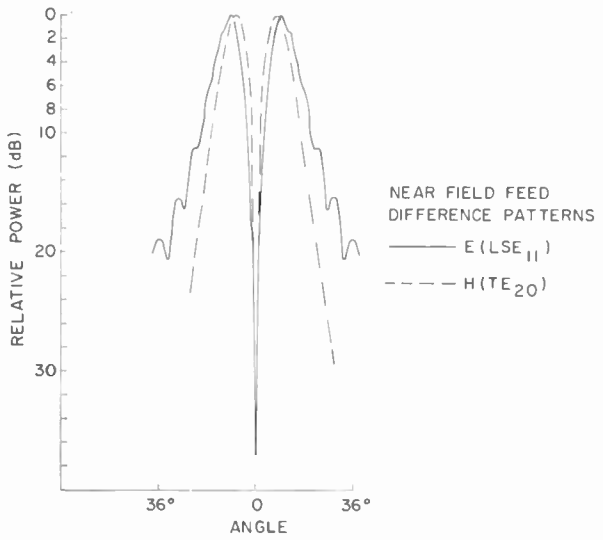


Fig. 9—E- and H-plane near-field difference patterns.

Feed Performance

A C-band spherical-wave near-field feed system designed for a 29-foot-diameter Cassegrain antenna is shown in Figure 10. The feed is a pyramidal horn with 33° included angle and 18-inch-square aperture. In an optimum configuration, the feed subtends a 28° total angle at the subreflector resulting in an equivalent paraboloid system

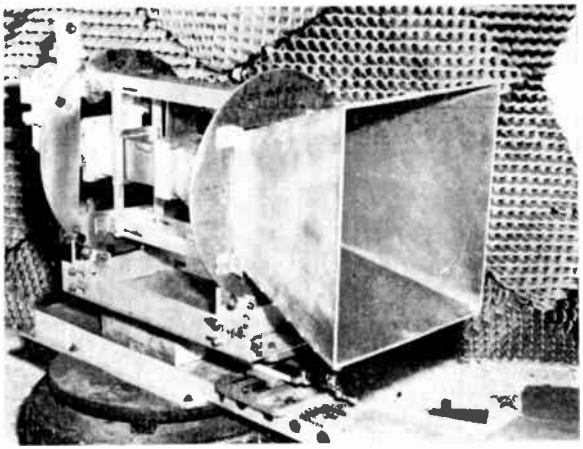


Fig. 10—C-Band spherical-wave near-field feed.

with focal length to diameter ratio of 2.04. The feed is designed to provide an LSE_{12} to TE_{10} mode amplitude ratio B of 0.6. Analytical investigation has indicated this to be a near-optimum value. A conventional C-band monopulse combiner is utilized to excite the feed in the TE_{10}, LSE_{11}, and TE_{20} modes. LSE_{12} mode generation is provided by a mode-generating step discontinuity[2] in the feed. For proper utilization of the feed pattern-shaping capabilities of the LSE_{12} mode the phase difference between the two modes at the feed aperture must be an integral (n) multiple of 360° for the design frequency. For optimum bandwidth, n should equal unity near the center frequency.

The antenna performance for a particular feed system is found by determining the total expected antenna efficiency. For reflector antennas, the antenna efficiency η_T is defined as

[2] C. E. Profera, Jr., L. H. Yorinks, "A High Efficiency Dual Frequency Multimode Monopulse Antenna Feed System," Supp. to *IEEE Trans. on Aerospace and Electronic Systems*, Vol. AES-2 No. 6, Nov. 1966.

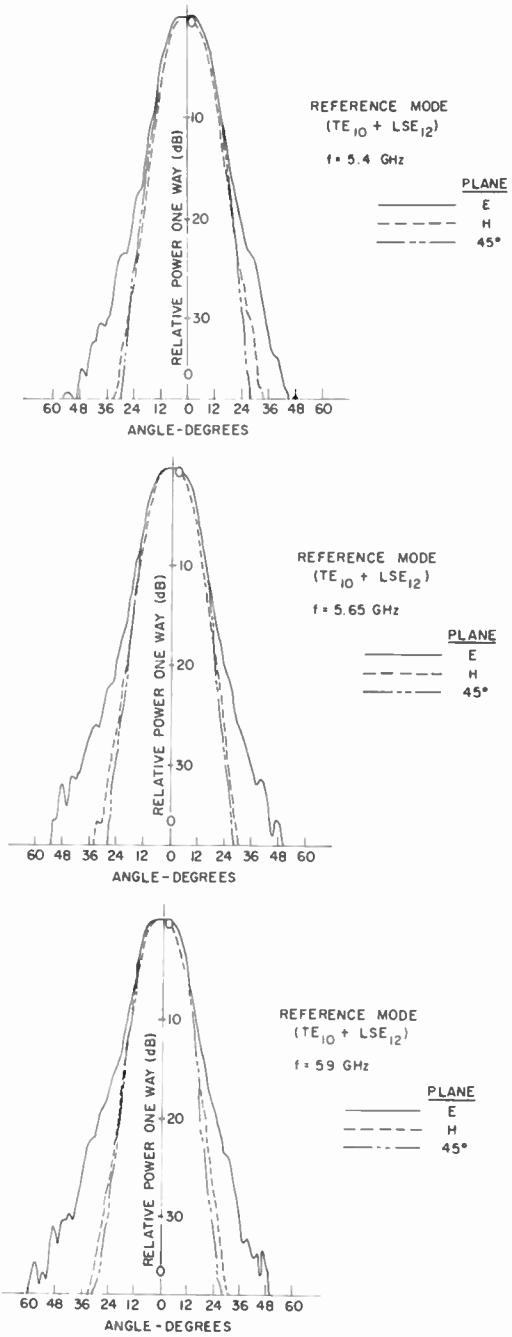


Fig. 11—Spherical-wave near-field feed reference pattern.

$$\eta_T = \eta_a \eta_{sp} \eta_{b1} \eta_\phi \eta_p \eta_o \,, \tag{4}$$

where

η_a = antenna aperture taper efficiency,

η_{sp} = spillover efficiency,

η_{b1} = aperture blockage efficiency,

η_ϕ = aperture phase efficiency,

η_p = polarization efficiency,

η_o = other efficiencies (diffraction losses of subreflector, main reflector, etc.).

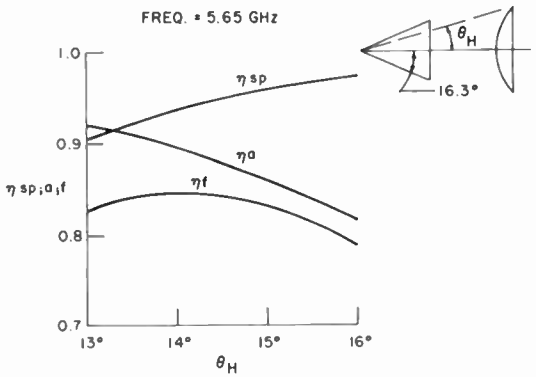


Fig. 12—Spillover, aperture, and feed efficiency versus θ_H.

For a common-reflector configuration, a feed system comparison can be made by considering only two of the above parameters, η_a and η_{sp}. Feed efficiency is defined as

$$\eta_F = \eta_a \eta_{sp} \,. \tag{5}$$

Monopulse sum patterns of the spherical-wave near-field feed measured at 5.4, 5.65, and 5.9 GHz are shown in Figures 11(a) through (c). To provide a complete set of data for an efficiency evaluation, patterns were recorded in the $E(0°)$, $H(90°)$, diagonal (45°), and intermediate (22.5°; 67.5°) planes of the feed aperture at 0.1 GHz intervals over the 5.4 to 5.9 GHz frequency band. The observed patterns indicate near beam width equality in all planes over the full 10% frequency band. Feed-pattern side lobes are virtually nonexistent, indicating extremely high spillover efficiency. The spillover, aperture, and feed efficiencies at 5.6 GHz are plotted in Figure 12 versus the angle θ_H subtended at the feed vertex by the subreflector edge. Feed efficiency

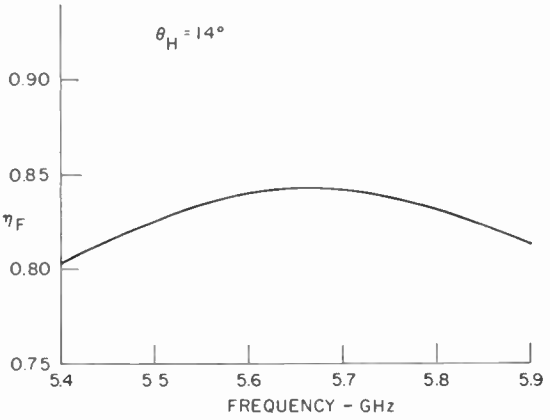


Fig. 13—Feed efficiency versus frequency.

versus frequency at $\theta_H = 14°$ is plotted in Figure 13. The feed efficiency varies between 81 and 84% over the full 10% band, indicating excellent wide-band performance.

A similar analysis was made for the conventional multimode feed and the dominant-mode feed (center horn of 5-horn feed). Optimized feed efficiencies are compared in Figure 14 for the spherical-wave near-field multimode, conventional multimode, and five-horn feeds. Over the operating frequency band, the spherical-wave near-field multimode feed shows an improvement of from 0.1 to 0.6 db relative to the conventional multimode feed and from 1.5 to 2 db relative to the 5-horn feed.

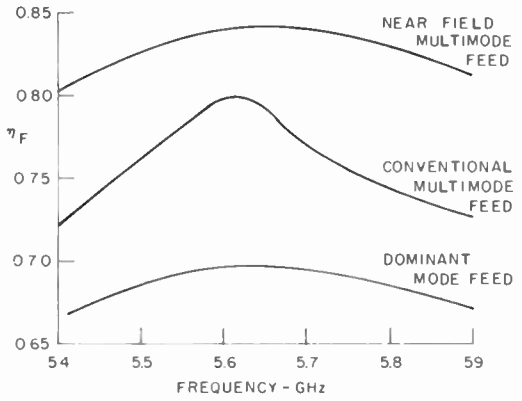


Fig. 14—Optimized feed efficiency versus frequency.

Phase-front data was obtained from the experimental C-band feed by measuring the relative phase of the sum modes on a spherical surface centered at the approximate feed vertex. Phase-front data at 5.4, 5.64, and 5.9 GHz are shown in Figure 15. The feed phase efficiency determined from the above data was essentially unity at the center frequency and 0.98 at the band edges, indicating that the phase-front deviation from a pure spherical wave has only a negligible effect on antenna gain.

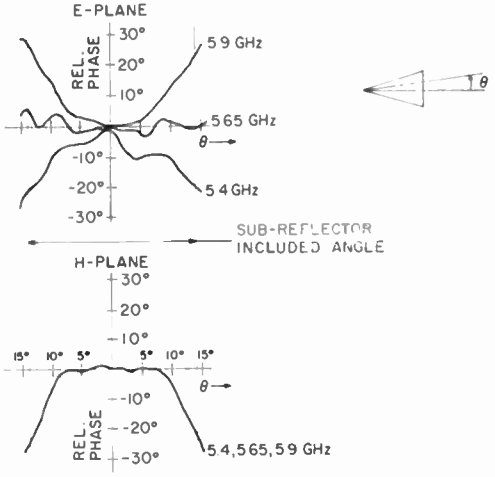


Fig. 15—Phase front data.

E- and *H*-plane principal-plane difference-mode patterns of the spherical-wave near-field feed are shown in Figure 16. Patterns were recorded in the 0°, 22.5°, 45°, 67.5°, and 90° planes to obtain the typical *E*- and *H*-plane difference-mode contours shown in Figure 17 and to provide sufficient information for determination of difference-mode spillover efficiency. The difference-mode contours are invarient with frequency over the 5.4 to 5.9 GHz band.

The difference-mode aperture efficiency is obtained by comparing the difference-mode contour on the antenna aperture to known difference-mode functions of the form

$$f_D(\rho,\phi) = J_1(k\rho)\cos\phi \tag{6}$$

where $J_1(k\rho)$ is the Bessel function of first kind of order one and ρ is the normalized fractional radius for a circular aperture geometry.

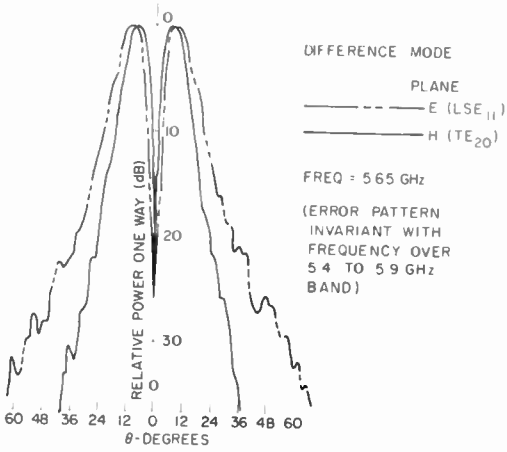


Fig. 16—E- and H-plane error patterns.

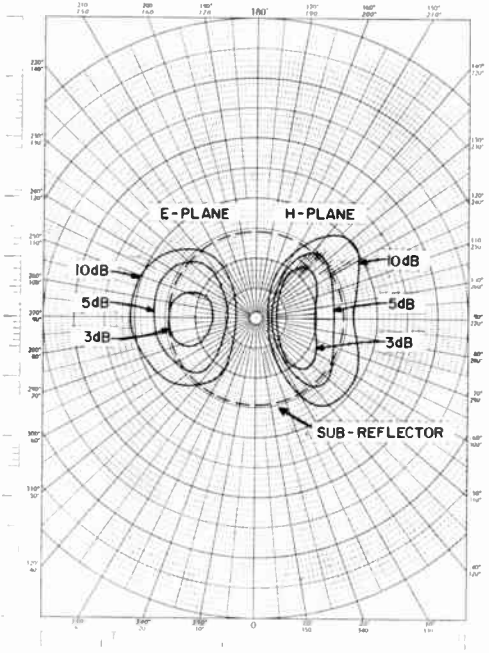


Fig. 17—E- and H-plane difference-mode contour patterns.

Both difference patterns and aperture efficiency have been computed for distributions of the above form. The near-field feed difference-mode contours are approximated by Equation (6), with $k = 2.8$ in the E plane and 3.0 in the H plane, resulting in aperture efficiencies of 0.54 and 0.53, respectively. Feed spillover loss was determined from the measured patterns and found to be 28% and 38% in the H and E planes, respectively.

The difference-mode feed efficiency η_{FD} is

H-plane

$$\eta_{FD} = (.53)(.72) = .38$$

E-plane

$$\eta_{FD} = (.54)(.62) = .34$$

The resulting difference-mode antenna gain will be approximately 3.4 db and 3.9 db below the sum-mode gain in the H and E planes, respectively.

Conclusions

The spherical-wave near-field feed has extremely desirable radiation characteristics. Further improvement by the employment of multimode beam-shaping techniques yields a feed with an efficiency not previously attainable. The resulting feed patterns are rotationally symmetric and virtually side-lobe free over bandwidths of 10% to 15%. The efficiency of such feeds has been found to be greater than 80% over a 10% band.

Use of the spherical-wave near-field feed as a multimode single-aperture monopulse feed yields monopulse difference-mode feed efficiencies of 38% and 34% in the H and E planes, respectively.

THE ACOUSTOELECTRIC EFFECTS AND THE ENERGY LOSSES BY HOT ELECTRONS—PART III

SOME ASPECTS OF THE LARGE-SIGNAL ACOUSTOELECTRIC EFFECTS

By

A. Rose

RCA Laboratories

Editor's Note: This is the third part of a several part paper. Part I appeared in the March 1966 issue (Vol. XXVII, No. 1, p. 98) and Part II in the December 1966 issue (Vol. XXVII, No. 4, p. 600).

Summary—The small-signal theories of acoustoelectric gain $l < \lambda$ and $l > \lambda$ are cast in a form that allows a ready graphical comparison of their approach to the large-signal regime. For $l < \lambda$ the onset of net acoustoelectric gain is also the onset of large-signal saturation of the drift velocity at the velocity of sound. For $l > \lambda$ the onset of net acoustoelectric gain is separated from the onset of large-signal saturation at the velocity of sound by an intermediate range, where the flux is sufficient to cause a departure from Ohm's law but not sufficient to completely bunch the carriers. It is shown that the large-signal regime must inherently terminate at high fields in a current instability. The classical and quantum treatments for $l > \lambda$ are shown to be in substantial agreement and to be equivalent to a Landau-damping type of argument. The threshold fields are derived for the onset of net acoustoelectric gain and for the high-field current instability. Also a criterion is obtained for the existence of net acoustoelectric gain. The incubation time is expressed in terms of the applied field and the time constant for acoustic losses.

Introduction

THE small-signal theory for acoustoelectric amplification of sound waves for piezoelectric coupling and for the mean free path l of an electron $< \lambda$ was derived by White [1] and Hutson [2] and confirmed experimentally in their own work [3] and that of many subsequent investigators. On the other hand, the attempts to understand

[1] D. L. White, "Amplification of Ultrasonic Waves in Piezoelectric Semiconductors," *Jour. Appl. Phys.*, Vol. 33, p. 2547, Aug. 1962.

[2] A. R. Hutson and D. L. White, "Elastic Wave Propagation in Piezoelectric Semiconductors," *Jour. Appl. Phys.*, Vol. 33, p. 40, Jan. 1962.

[3] A. R. Hutson, J. H. McFee, and D. L. White, "Ultrasonic Amplification in CdS," *Phys. Rev. Letters*, Vol. 7, p. 237, Sept. 15, 1961.

the large-signal behavior reported by Smith [4] in CdS and by Bray and coworkers [5] in several III-V compounds have raised a number of questions concerning the large-signal aspects of the acoustoelectric effect and the role of quantum effects that have not been resolved. A group of papers [6] presented at the Kyoto Conference on Semiconductors characterize both the state of the art and its major ambiguities.

In the present paper, the large- and small-signal analyses of the acoustoelectric effects for both $l < \lambda$ and $l > \lambda$ are presented in such form as to allow certain conclusions to be drawn with some confidence and at the same time to emphasize the area where a good theoretical understanding is still lacking.

We begin with the small-signal theories of White [1] for $l < \lambda$ and of Spector,[7] Conwell,[8] and Yamashita [9] for $l > \lambda$. Spector's classical analysis and the quantum treatments by Conwell and by Yamashita are shown to be in essential agreement (meaning that the effect is essentially classical). It is shown further that the same result can be obtained from a Landau-damping argument.

The acoustic gain for large flux amplitudes, sufficient to bunch all of the carriers, is written out of hand and is the same for $l < \lambda$ and $l > \lambda$.

The small-signal theories are then cast in a form that shows by inspection a relatively simple transition to the large-signal form for $l < \lambda$ and a more complex transition for $l > \lambda$. It is in this transition region for $l > \lambda$ where theory is mostly lacking.

Several significant parameters are derived—the minimum coupling constant for the existence of acoustoelectric amplification, the threshold fields for departure from Ohm's law and for saturation of the drift velocity, the incubation time, and the maximum field beyond which the current should depart abruptly from its saturated value towards its ohmic value.

---

[4] R. W. Smith, "Current Saturation in Piezoelectric Semiconductors," *Phys. Rev. Letters*, Vol. 9, p. 87, Aug. 1, 1962.

[5] P. O. Sliva and R. Bray, "Oscillatory Current Behavior in GaSb and Its Relation to Spontaneous Generation and Amplification of Ultrasonic Flux," *Phys. Rev. Letters*, Vol. 14, p. 372, 8 March 1965; J. B. Ross and R. Bray, "Acoustic Buildup Versus Intrinsic Breakdown in p-InSb," *Bull. Amer. Phys. Soc.*, Vol. 11, p. 173, Feb. 1966; R. Bray, C. S. Kumer, J. B. Ross, and P. O. Sliva, "Acoustoelectric Domain Effect in III-V Semiconductors," *Pro. Int. Conf. on Semicon.* (Kyoto 1966), p. 483.

[6] Proc. Int. Conf. on Semicon. (Kyoto 1966), Section XI, pp. 443-501.

[7] H. N. Spector, "Amplification of Acoustic Waves through Interaction with Conduction Electrons," *Phys. Rev.*, Vol. 127, p. 1084, Aug. 15, 1962.

[8] E. Conwell, "Effect of Phonon Drift on Low Temperature Conduction in Semiconductors," *Physics Letters*, Vol. 13, p. 285, 15 Dec. 1964.

[9] J. Yamashita and K. Nakamura, "Instability of the Electron-Phonon System in Strong Electric Fields," *Prog. Theor. Phys.*, Vol. 33, p. 1022, June 1965.

The last effect, the high-field instability, was treated by Beale [10] under a special set of conditions and was ascribed by him to a heating of the electrons. In the present paper, the high-field instability is obtained as a general result of large-signal behavior and does not require the heating of the electrons. Our result is in agreement with that obtained by Ridley [11] by an independant argument and communicated to the author while the present analysis was in progress.

Small-Signal Theory ($l < \lambda$)

The following expression [12] for the small-signal theory is a simple generalization of the White [1] theory to include the various types of coupling:

$$\frac{dE}{dt} = \beta E_\omega \frac{\omega_r \omega_c}{(\omega_c + \omega_D)^2 + \omega_r^2} . \tag{1}$$

In Equation (1)

E = total energy density of sound wave = $K\mathcal{E}_W^2/(8\pi\beta)$

K = relative dielectric constant

$\mathcal{E}_W$ = peak electric field of sound wave

β = ratio of electrical energy to total energy of sound wave

$\omega_r = \omega(v_d - v_s)/v_s$

$\omega_c = 4\pi ne\mu/K$

$\omega_D = kT\mu/(e\lambda^2)$

v_s = phase velocity of sound

v_d = drift velocity of electrons

$\lambda = \lambda/2\pi$

---

[10] J. R. A. Beale, "Acoustoelectric Effects with Hypersonic Waves of Large Amplitude," *Phys. Rev.*, Vol. 135, p. A1761, 14 Sept. 1964.

[11] B. K. Ridley (Private Communication).

[12] A. Rose, "The Acoustoelectric Effects and the Energy Losses by Hot Electrons, Part I," *RCA Review*, Vol. 27, p. 98, March 1966. All of the entries in Table I of this reference should satisfy the form of Equation (1). The diffusion term ω_D was erroneously omitted from the last two entries: "Intervally Deformation Potential" and "Metals". The reason given for the omission was that the energy relaxation process occurred in phase space rather than in physical space so that the density of relaxed carriers remained uniform in physical space (see, e.g., Figure 4 of this reference). The argument is erroneous because the density of carriers in physical space must be evaluated for each valley separately. Under these conditions the density in physical space is nonuniform, and diffusion plays its usual role.

Equation (1) is rewritten in the form

$$\frac{dE}{dt} = \frac{nev_s}{2}(\mathcal{E} - \mathcal{E}_0)\frac{\mathcal{E}_W'^2}{(\mathcal{E}_s' + \mathcal{E}_D')^2 + (\mathcal{E}' - \mathcal{E}_0')^2} \tag{2}$$

by making use of the relations

$$\omega_c = \frac{4\pi ne\mu}{K} = \frac{4\pi ne\lambda}{K}\frac{\mu}{\lambda} \equiv \mathcal{E}_s'\frac{\mu}{\lambda}$$

$$\omega_D = \frac{kT\mu}{e\lambda^2} = \frac{kT}{e\lambda}\frac{\mu}{\lambda} \equiv \mathcal{E}_D'\frac{\mu}{\lambda} \tag{3}$$

$$E = \frac{K\mathcal{E}_W'^2}{\beta 8\pi}$$

$$\frac{v_d - v_s}{v_s} = \frac{\mathcal{E} - \mathcal{E}_0}{\mathcal{E}_0} \qquad \text{[This is valid only in the small-signal approximation]}$$

where $\mathcal{E}' \equiv$ applied field and $\mathcal{E}_0' \equiv v_s/\mu$.

Note that the condition for the small-signal theory is

$$\mathcal{E}_W'^2 \ll (\mathcal{E}_s' + \mathcal{E}_D')^2 + (\mathcal{E}' - \mathcal{E}_0')^2 \tag{4}$$

The small-signal theory was derived on the assumption that only a small fraction of the carriers is bunched in the troughs of the sound wave. This means that $\mathcal{E}_W \ll \mathcal{E}_s'$ where $\mathcal{E}_s'$ is the field required to bunch all of the carriers. It also means that the effective diffusion field is $|\Delta n/n|\,|kT/(e\lambda)| \ll \mathcal{E}_D'$ where $\Delta n/n$ is the fraction of carriers bunched Finally, since the small-signal theory assumes a constant drift velocity, the perturbation of this velocity by $\mathcal{E}_W'$ must be small compared with the relative velocity $(\mathcal{E}' - \mathcal{E}_0')\mu$.

Large-Signal Theory ($l < \lambda$)

In the limit of a very-large-amplitude coherent sound wave, the carriers are completely bunched in the troughs of the sound wave and are locked to the velocity of sound. This means that $\mathcal{E}_W' > \mathcal{E}_s'$ in order to bunch all of the carriers and that $\mathcal{E}_W' > |kT/(e\lambda)|$ and $\mathcal{E}_W > \mathcal{E} - \mathcal{E}_0$ in order to prevent carriers from escaping from the troughs either by

diffusion or by the excess applied field. Under these conditions the rate of doing work on the wave (per unit volume) can be written out of hand as

$$\frac{dE}{dt} = ne(\mathcal{E}' - \mathcal{E}'_n)v_s . \tag{5}$$

Equation (5) states that when the carriers are locked in the troughs of the sound wave the force $ne\mathcal{E}'_n$ causes the carriers to drift at the velocity of sound while the excess force $ne\ (\mathcal{E}' - \mathcal{E}'_n)$ is transmitted to the sound wave by pushing the carriers against its flank.* The excess force does work on the sound wave at the rate $ne(\mathcal{E}' - \mathcal{E}'_n)v_s$. This is the model originally invoked by Smith [4] to account for the saturation of drift velocity at or near the velocity of sound. It is in contrast with the model used by Bray [5] and by Yamashita [9] in which the drift velocity approaches the velocity of sound asymptotically at high sound-flux densities but cannot in principle be equal to the sound velocity. Since the gain in Equation (5) is proportional to the electric field it contains the necessary feedback mechanism for the formation of domains, i.e., a local increase of electric field due to excess flux tends to exaggerate the rate of build up of flux and field in the place where the excess occurred.

Recent observation by Many and Balberg [13] on the time rate of decay of current towards its saturated value strongly favor the model expressed by Equation (5). As they point out, the "lock-in" model of Equation (5) leads to a much more rapid decay of current, once the sound flux has reached a magnitude sufficient to bunch most of the carriers, than does the model of Bray and Yamashita.

Note that Equation (5) is based on a single coherent sound wave. In practice, a range of frequencies in the neighborhood of the frequency for maximum gain is amplified to form a noisy sound flux. Thus, while the average energy of the sound flux may be more than sufficient to bunch all of the carriers, the instantaneous energy will at times fall below this value and permit the drift velocity to exceed the velocity of sound. Qualitatively, the effect of the noisy flux is to relax the lock-in model in the direction of the Bray model. Quantitatively, however, the lock-in model leads to a much more rapid approach to the

* In the case of "domains", $\mathcal{E}'$ is the field in the domain as measured by potential probes.

[13] A. Many and I. Balberg, "On the Mechanism of Acoustic Gain Under Large Signal Conditions," *Physics Letters*, Vol. 24A, p. 705, 19 June 1967; I. Balberg and A. Many, "Current Decay Characteristics Associated with Acoustic Amplification in CdS Crystals," *Physics Letters*, Vol. 24A, p. 707, 19 June 1967.

drift velocity of sound for average fluxes exceeding that necessary to bunch all of the carriers than does the Bray model. For the lock-in model the approach should vary as

$$\frac{v_d - v_s}{v_s} \propto \exp\left\{-\frac{E}{E_o}\right\};$$

while for the Bray model the variation is

$$\frac{v_d - v_s}{v_o} \propto \frac{E_n}{E},$$

where E_n is the energy density of flux needed to completely bunch the carriers.

Moore [14] has shown that the combination of a noisy flux and the lock-in model gives good agreement with his observation on current–voltage curves in CdS.

Combined Large- and Small-Signal Theory ($l < \Lambda$)

The small- and large-signal expressions, Equations (2) and (5), are plotted in Figure 1 as a function of $\mathcal{E}'_W{}^2$, which is proportional to the energy density of sound flux (see Equation (1)). The nominal demarcation between large and small signals is $\mathcal{E}'_W{}^2 = (\mathcal{E}'_s + \mathcal{E}_D)^2 + (\mathcal{E} - \mathcal{E}_n)^2$. At this point, the two theories fail to meet by a factor of two. The actual transition is certain to be smooth rather than abrupt. We have chosen, for convenience of discussion, to use a simple extrapolation of a small-signal theory (shown dashed) to represent the transition. If the transition is curved above or below the small-signal line (see dotted lines), the error involved in our choice is likely less than a factor of two.

Combined Gain and Loss Curves ($l < \Lambda$)

To compute the steady-state flux we need to compare the rates of gain and loss of acoustic energy. The latter is assumed to be characterized by a time constant τ_{ph} and is plotted in Figure 2 as a linear function of $\mathcal{E}'_W{}^2$. Note that τ_{ph} refers to the loss of acoustic flux by processes *other* than the interaction with the free carriers that are

[14] A. R. Moore, "Acoustoelectric Current Saturation in CdS as a Fluctuation Process," *Jour. Appl. Phys.*, Vol. 38, p. 2327, April 1967.

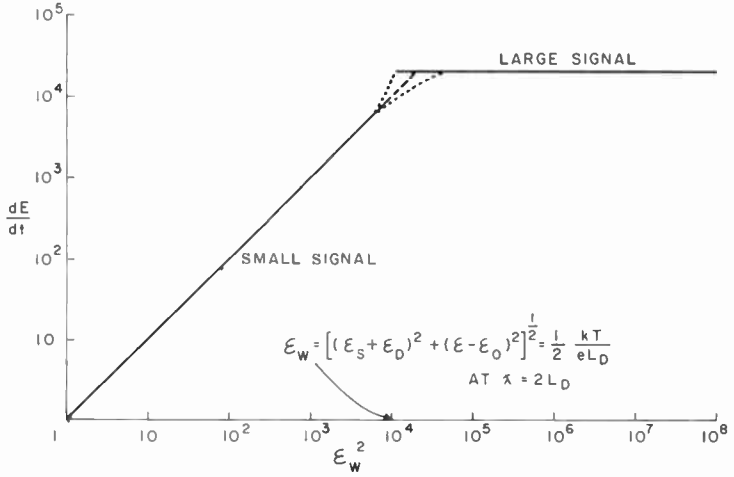


Fig. 1—Rate of energy gain by acoustic flux as a function of energy density of flux ($l < \lambda$). (Note: $\mathcal{E}_w^2 = \beta \times$ energy density of flux.)

responsible for amplifying the flux. Two representative gain curves are plotted in Figure 2, the higher curve being for a higher value of $\mathcal{E} - \mathcal{E}_0$.

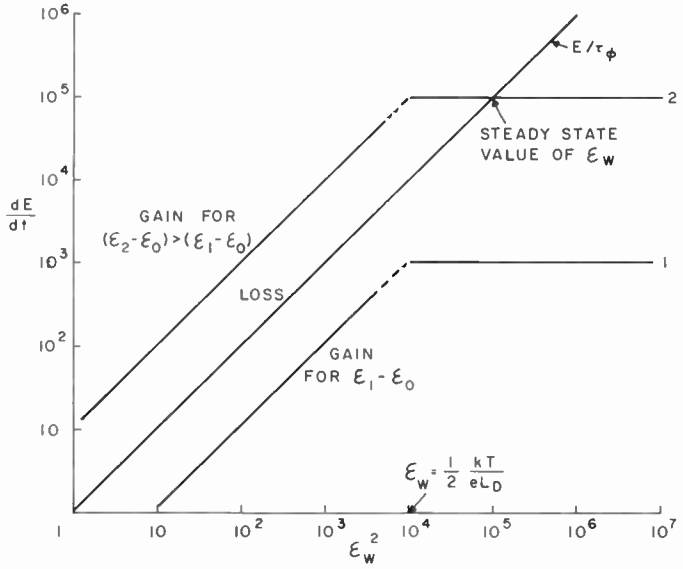


Fig. 2—Combined rates of energy gain and loss by acoustic flux as a function of energy density of flux ($l < \lambda$).

Since gain curve 1 in Figure 2 lies always below the loss curve, there will be no amplification of acoustic flux. Gain curve 2 lies above the loss curve and will accordingly lead to an amplification of flux such that the steady-state value is determined by its point of intersection with the loss curve. By inspection, therefore, the onset of net acoustic amplification will occur when the gain and loss curves first coincide, that is, when

$$\left.\frac{dE}{dt}\right|_{\text{gain}} = \frac{E}{\tau_\phi} \tag{6}$$

where $\left.\frac{dE}{dt}\right|_{\text{gain}}$ is the small-signal form of the gain curve. *But this onset of acoustic amplification is also, by inspection of Figure 2, essentially the onset of the large-signal theory and of a drift velocity saturated at the velocity of sound.* The choice of the upper dotted transition curve in Figure 1 would lead to the same abrupt transition of the drift velocity to the velocity of sound at a somewhat lower value of $\mathcal{C}' - c'_n$. The choice of the lower dotted transition curve would lead to a value of $\mathcal{C} - c'_n$ for saturation somewhat higher than the value of $c' - c'_n$ for the onset of acoustic amplification. In both cases, less than a factor of 2 is involved.

We note that curve 2 in Figure 2 is simply a constant factor higher than curve 1. This will be true according to Equation (2) if $(c' - c'_n)^2 < (c'_x + c'_p)^2$ and hence is negligible in the denominator. If, on the other hand, $(c' - c'_n)^2 > (c'_x + c'_p)^2$, an increase in $c' - c'_n$ will, according to Equation (2), *lower* curve 1 rather than *raise* it. Accordingly, the gain curve could not be made to intersect the loss curve and there would be no condition for net acoustic gain. By this argument, the criterion for the *existence* of net acoustic gain is

$$\mathcal{C} - c'_n \leq c'_x + c'_p . \tag{7}$$

where $c' - c'_n$ has its threshold value determined by Equation (6).

Since $c' - c'_n < c'_x + c'_p$, we can maximize the gain given by Equation (2) by minimizing the value of $c'_x + c'_p$. From their definitions (see Equation (3)), the minimum occurs at $\lambda = L_p$ = Debye length. The wavelength for optimum amplification is shifted by about a factor of two to the long-wavelength side of L_p because $\tau_\phi \propto \lambda$. (Actually a range of wavelengths near $2L_p$ will be amplified.) Hence the denom-

inator of Equation (2) becomes $(1/4)\,\mathcal{E}'_D{}^2 = 1/4\left(\frac{kT}{eL_D}\right)^2$, and Equation (2), rewritten for the above conditions, is

$$\frac{dE}{dt} = 2\,nev_s\,(\mathcal{E} - \mathcal{E}'_o)\,\frac{\mathcal{E}'_W{}^2}{\mathcal{E}'_D{}^2} \tag{8}$$

In the large-signal range, where the loss curve intersects the gain curve (e.g., curve 2 of Figure 2), the point of intersection is given by the large-signal theory:

$$\left.\frac{dE}{dt}\right|_{\text{gain}} = \frac{E}{\tau_\phi}$$

or

$$nev_s\,(\mathcal{E} - \mathcal{E}'_o) = \frac{K\mathcal{E}'_W{}^2}{8\,\pi\beta\,\tau_\phi}\,. \tag{9}$$

The large-signal behavior, that is $v_d = v_s$, will persist so long as $\mathcal{E}' - \mathcal{E}'_o < \mathcal{E}'_W$. Since, however, $\mathcal{E}'_W$ increases only as $(\mathcal{E}' - \mathcal{E}'_o)^{1/2}$, there will be an applied field $\mathcal{E}'$, above which $\mathcal{E}' - \mathcal{E}'_o > \mathcal{E}'_W$ and the current rises abruptly to its ohmic value.

Calculation of Threshold Parameters

Equation (8) is inserted into Equation (6) to obtain the threshold field for the onset of amplified acoustic flux and, at the same field, the onset of large signal behavior. The result is

onset of net acoustic gain

$$\frac{\mathcal{E} - \mathcal{E}'_o}{\mathcal{E}_o} = \frac{1}{4\beta\tau_\phi\omega^*} \tag{10}$$

Where $\omega^* = v_s{}^2 e/(kT\mu)$ and is the same as the ω_D of the Hutson–White terminology. Equation (9) with $\mathcal{E}'_W = \mathcal{E}' - \mathcal{E}'_o$ is used to compute the high-field instability:

onset of high-field instability

$$\frac{\mathcal{E}' - \mathcal{E}_o}{\mathcal{E}'_o} = 2\beta\tau_\phi\omega_c\,. \tag{11}$$

Figure 3 shows schematically the expected current–voltage curves* for several values of coupling constant β.

The condition for the possibility of net acoustic gain is obtained by inserting $c' - c'_0$ from Equation (10) in Equation (7) to obtain:

condition for possibility of net acoustic gain

$$2\beta\tau_\phi\omega_{Deb} > 1 \tag{12}$$

where $\omega_{Deb} \equiv v_s/L_D$.

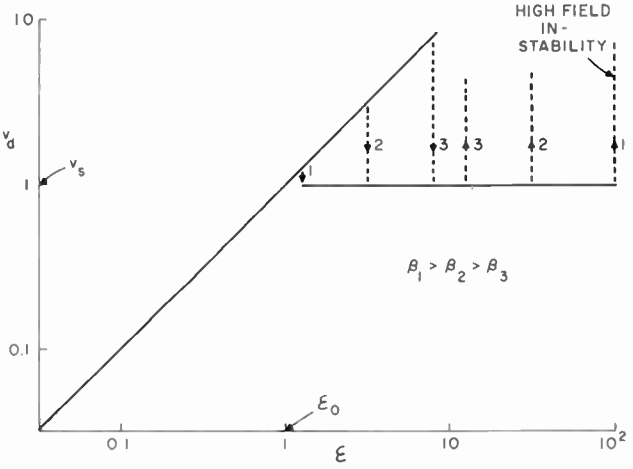


Fig. 3—Drift velocity as a function of applied field for several values of coupling constant ($l < \lambda$).

We note further by inspection of Figure 2 that at the onset of net acoustic gain, the time constant for the gain curve is equal to the time constant for the loss curve. Further, the time constant for gain varies as $(c' - c'_0)^{-1}$. The time constant for *net* gain then is

$$\tau_g = \frac{c'^* - c'_0}{(c' - c'_0) - (c'^* - c'_0)}\,\tau_\phi = \frac{c'^* - c'_0}{\mathcal{E} - c'^*}\,\tau_\phi . \tag{13}$$

where c'^* is the onset field determined by Equation (10).

The total time required to build up acoustic flux from some small

* We have shown the current as saturated at the drift velocity of sound in the large-signal range—a result consistent with a single value of ω. A. R. Moore [14] has shown that for a range of ω the flux is noisy and thereby leads to a positive slope of v_d versus $\mathcal{E}$.

initial value, taken to be either the thermal flux or some shock excited flux, to the magnitude required to bunch a significant fraction of the carriers is called the incubation time. By Equation 13, the incubation time is

$$t_I = \frac{\mathcal{E}^* - \mathcal{E}_o}{\mathcal{E} - \mathcal{E}^*} \tau_\phi \ln G , \tag{14}$$

where G is the ratio of final to initial flux. The gain G needed to amplify a narrow cone of the thermal flux up to energy densities sufficient to show large-signal effects is in the order of 10^{10}. Whether the actual gain is as low as 10^7 or as high as 10^{13} affects the incubation time by less than a factor of two. Hence, using 10^{10} as a nominal value for G, the incubation time is

$$t_I \doteq 20 \frac{\mathcal{E}^* - \mathcal{E}_o}{\mathcal{E} - \mathcal{E}^*} \tau_\phi . \tag{15}$$

The minimum length of crystal to accommodate the incubation process will then be

$$\begin{aligned} L_I &\doteq v_s t_I \\ &= 20 \frac{\mathcal{E}^* - \mathcal{E}_o}{\mathcal{E} - \mathcal{E}^*} \tau_\phi v_s \text{ cm} . \end{aligned} \tag{16}$$

That is, for a finite length of crystal, the applied field must exceed $\mathcal{E}^*$ by an amount sufficient to satisfy Equation (16). This will tend to blur the onset of a saturated drift velocity. The onset is shown in Figure 3 as an abrupt transition for an infinite crystal.

Comparison with Experiment

Table I lists the values of the parameters used to compute the threshold fields of Equations (10) and (11). A carrier density of $10^{16}/\text{cm}^3$ was selected as representative of a number of experiments reported for CdS [15] and GaAs.[16]

---

[15] A. R. Moore and R. W. Smith, "Effect of Traps on Acoustoelectric Current Saturation in CdS," *Phys. Rev.*, Vol. 138, p. A1250, 17 May 1965.

[16] P. Leroux-Hugon, Conference on Phonons, Grenoble, April 1966.

Onset of Acoustoelectric Effect

$$\frac{\mathcal{E} - \mathcal{E}_0}{\mathcal{E}_0} = \begin{cases} 0.05 \text{ for CdS } (300^\circ\text{K}) \\ 0.5 \text{ for GaAs } (77^\circ\text{K}) \end{cases}$$

Onset of High-Field Instability

$$\frac{\mathcal{E} - \mathcal{E}_0}{\mathcal{E}_0} = \begin{cases} 100 \text{ for CdS } (300^\circ\text{K}) \\ 6{,}000 \text{ for GaAs } (77^\circ\text{K}) \end{cases}$$

Table I

	CdS (300°K)	GaAs (77°K)
β	0.05 (Ref. (2))	2.4×10^{-3} (Ref. (2))
ω^*	2.9×10^{10}/sec (Ref. (2))	10^9/sec (Ref. (2))
v_s	4.3×10^5 cm/sec	3.3×10^5 cm/sec
n	10^{16}/cm^3	10^{16}/cm^3
L_D	5×10^{-6} cm	2.5×10^{-6} cm
v_s/L_D	9×10^{10}/sec	1.3×10^{11}/sec
τ_ϕ	3×10^{-9} sec (Ref. (17))	2×10^{-7} sec (Ref. (18))
μ	200 cm^2/volt sec	5,000 cm^2/volt sec
ω_c	2.7×10^{11}/sec	6×10^{12}/sec

The results for the onset of the acoustoelectric effect reflect the observation [15] that the drift current in CdS departs from its ohmic value essentially as soon as the applied field exceeds $\mathcal{E}_0$. On the other hand, this departure for GaAs, as reported by Leroux-Hugon [16] for long samples, occurs at values of $(\mathcal{E} - \mathcal{E}_0)/\mathcal{E}_0$ above unity and consistent with his carrier densities of $\approx 5 \times 10^{16}$/cm^3.

The results for the onset of the high-field instability show that for a density of carriers, 10^{16}/cm^3, the threshold field is over 10^5 volts/cm, or well beyond the usual range of applied fields. Such fields may be attained within the localized domains of CdS and thereby limit the maximum field in the domain and set the minimum width of domain. The product of domain width and domain field should equal the voltage applied to the sample in excess of that needed to attain the velocity of sound.

[17] J. E. May, Jr., "Electronic Signal Amplification in the UHF Range with the Ultrasonic Traveling-Wave Amplifier," *Proc. IEEE*, Vol. 53, p. 1465, Oct. 1965.

If we adopt Pomerantz [18] suggestion that $\tau_\phi \propto \lambda \propto L_D$, all of the above threshold fields vary as the square root of the free-carrier density. Moreover, the same relation inserted in Equation (12) yields the following condition for the possibility of net acoustoelectric gain:

$$\begin{aligned} 4 \times 10^2 \beta &> 1 \text{ for CdS at } 300^\circ\text{K} \\ 4 \times 10^4 \beta &> 1 \text{ for GaAs at } 77^\circ\text{K} \end{aligned} \tag{17}$$

From Table I, it can be seen that the values of β readily satisfy Equation (17). It is interesting that this result is independent of the carrier density. It is implied that at low carrier densities the specimen length must be increased so that it is always sufficient to achieve the full steady-state amplification. For CdS, for example, the specimen length would need to exceed one centimeter at a carrier density of $10^{12}/\text{cm}^3$ and a value of $(c'^* - c'_n)/(c' - c'^*) = 1$. At higher fields the crystal length, by Equation (16), could be shorter.

The incubation times reported by Smith [4] for CdS are less than 0.1 microsecond. Those reported by Bray for GaAs are in order of a few microseconds. In brief, the incubation times for GaAs are over ten times larger than those for CdS and in qualitative agreement with the corresponding ratios of τ_ϕ, the time constant for acoustic loss, shown in Table I.

Small-Signal Theory ($l > \lambda$)

Spector has derived an expression for the small-signal acoustoelectric effect for $l > \lambda$ using a classical Boltzmann transport argument together with a shifted Maxwellian distribution. In the terminology of Equation (1), Spector's [7] result is

$$\frac{dE}{dt} = \beta E \omega \frac{l\, \omega_r \omega_c}{\lambda\, (\omega_c + \omega_D)^2}. \tag{18}$$

Equation (18) differs from Equation (1) in the absence of the ω_r term in the denominator and in the presence of the additional factor l/λ. Conwell [8] has obtained essentially the same result using perturbation theory for acoustic phonons coupled by a deformation potential. While Conwell's result, as published, contains Planck's constant h explicitly, other factors, also containing Planck's constant implicitly, combine to eliminate h from the final expression for induced emission of phonons.

[18] M. Pomerantz, "Temperature Dependence of Microwave Phonon Attenuation," *Phys. Rev.*, Vol. 139, p. A501, 1965.

Conwell's result does not include the ω_r term in the denominator. This is a screening term that takes into account the reduction or relaxation of the electric field of the coherent sound wave by the conductivity of the free carriers. The screening term was introduced phenomenologically by Yamashita [9] in his perturbation treatment of the acoustoelectric effect. Yamashita's result, after elimination of the various h factors by self cancellation, agrees with Spector's. The fact that the two quantum treatments agree with the classical treatment indicates that the problem is essentially a classical one.

Note in Equation (18) that, while the mean free path l appears explicitly, it is cancelled by the fact that both ω_r and ω_f are proportional to l. We have chosen this form for ease of comparison with Equations (1) and (2).

Both the quantum and classical treatments assume a Maxwellian distribution for the free carriers shifted to their mean drift velocity, $v_d = \mathcal{E}'\mu$. This means that electron–electron collisions must be frequent enough to maintain the Maxwellian distribution centered at $v_d = \mathcal{E}'\mu$—a condition that may be met in the range of small signals but becomes impossible as the flux density approaches that necessary to completely bunch the carriers and saturate the current at the velocity of sound. To emphasize this point we outline another method of obtaining Spector's result. This method is essentially a Landau-damping type of argument.

In Figure 4, a shifted Maxwellian distribution centered on v_d is shown schematically. In the neighborhood of v_s, Figure 4 shows a slice of the distribution of width Δv on each side of v_s. This is the slice of the distribution that is energetically trapped in the trough of the sound wave. Hence,

$$\tfrac{1}{2}\, m\,(v_s + \Delta v)^2 - \tfrac{1}{2} m v_s^2 \leq \mathcal{E}'_W e\lambda\,, \tag{19}$$

or, for $\Delta v > v_s$,

$$(\Delta v)^2 \leq \frac{\mathcal{E}'_W e\lambda}{m}\,. \tag{20}$$

The slice to the right of v_s is trapped on the forward or up flank of the wave and does work on the wave, while the slice on the left of v_s is trapped on the back or down flank of the wave and abstracts energy from it. Hence, the net work done on the wave per unit volume is

$$\frac{dE}{dt} = (n_f - n_b)\, e\mathcal{E}'_W v_s\,. \tag{21}$$

where n_f is the total number of carriers in the right-hand slice and n_b is that in the left-hand slice. From the geometry of Figure 4 and the Maxwellian distribution, the quantity $n_f - n_b$ is*

$$n_f - n_b \doteq \frac{3}{2} n_o \frac{(v_d - v_s)(\Delta v)^2}{v_t^3} \qquad \text{for } \Delta v \ll v_t \text{ and } v_d < v_t \,, \quad (22)$$

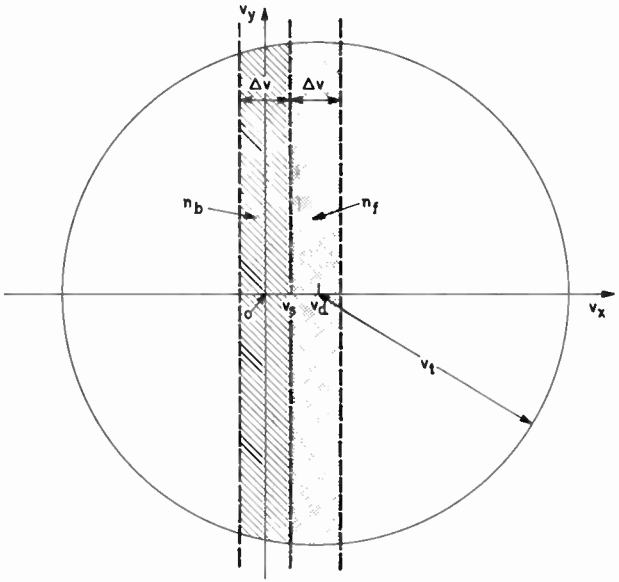


Fig. 4—Schematic diagram for calculation of acoustic gain by a Landau-damping argument.

where n_o is the total density of carriers and v_t is the thermal velocity. Equations (20) and (22) are inserted into Equation (21) to give

$$\frac{dE}{dt} = \frac{3}{2} \mathcal{E}_W^2 n_o \frac{e^2 v_s^2}{m v_t^3 \omega} (v_d - v_s) \,. \quad (23)$$

Equation (23) reduces to Equation (18) without the ω_c term in the denominator of Equation (18), i.e., without screening. Screening is introduced by recognizing that the field $\mathcal{E}_W$ of the acoustic wave is par-

* For $v_d > v_t$ the number of carriers energetically trapped in accordance with Equation (19) rapidly vanishes as does the acoustoelectric effect.

tially relaxed by the electron density n_0 so that

$$\mathcal{E}_W|_{\text{screened}} = \mathcal{E}'_W \frac{1}{1 + \left(\frac{\lambda}{L_D}\right)^2} \tag{24}$$

where L_D is a Debye length $(KkT/4\pi n_0 e^2)^{1/2}$. Insertion of the screening factor of Equation (24) into Equation (23) yields Equation (18).

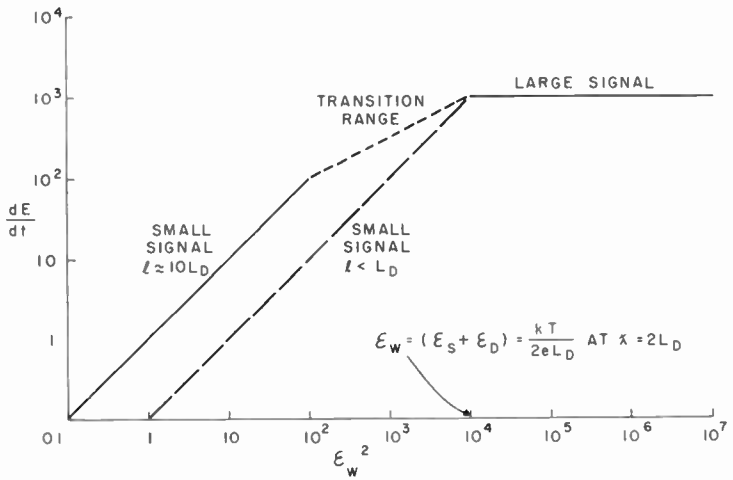


Fig. 5—Rate of energy gain by acoustic flux as a function of energy density of flux ($l > \lambda$).

The point to be emphasized here is that electron–electron collisions due to the distribution of electrons outside the Δv slices must maintain the Maxwellian distribution, centered on $v_d = \mathcal{E}'\mu$, within these slices. When the slices are a small fraction of the total, this is a reasonable condition to be satisfied. However, as the large-signal limit is approached, the slices by definition tend to encompass all of the electrons, and the maintenance of the Maxwellian distribution *centered on* v_d becomes impossible. This is the transition region shown dashed in Figure 5 and it merits further theoretical study. In the course of this transition the center of the Maxwellian distribution shifts from $v_d = \mathcal{E}'\mu$ to a value close to v_s. Also, in the course of this transition and as the large-signal limit is approached, it is likely that the random-phase approximation, on which the quantum arguments are based, breaks down.

Combination Large- and Small-Signal Theory ($l > \lambda$)

We re-write Equation (18) to make it directly comparable with Equation (8) using the same arguments that led to Equation (8):

$$\frac{dE}{dt} = 2\beta n e v_s (\mathcal{E}' - \mathcal{E}'_w) \frac{l}{\lambda} \frac{\mathcal{E}'^2_w}{\mathcal{E}'^2_D}. \tag{25}$$

The difference between Equation (25) and Equation (8) is the factor

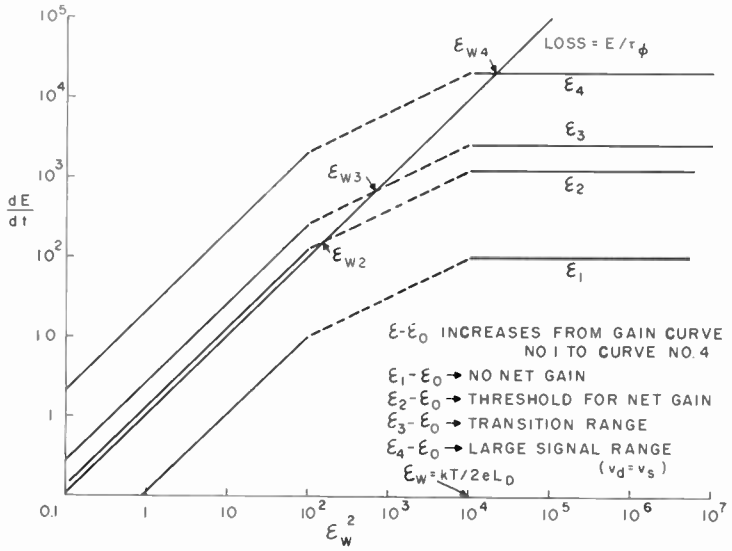


Fig. 6—Combined rates of energy gain and loss by acoustic flux as a function of energy density of flux ($l > \lambda$).

l/λ, which by assumption is greater than unity. Figure 5 shows a plot of Equation (25) for $l/\lambda = 10$ and shows its approach to the large-signal region as a dashed segment of an as yet uncertain form. The major point to be made here is that the small-signal theory for $l > \lambda$ introduces a large intermediate or transition region that is substantially absent for $l < \lambda$. The factor l/λ causes the linear extrapolation of the small-signal theory to miss joining on smoothly to the large-signal theory and hence leads to the dashed transition segment.

Figure 6 shows the same graphical operation for the $l > \lambda$ curve as was shown in Figure 2 for the $l < \lambda$ curve. As the applied field is increased, the gain curve intersects the loss curve earlier than in Figure 2 by the factor l/λ. At the first intersection, however, the flux

is considerably less than that needed to saturate the drift velocity. There is, therefore, a range of fields for which the equilibrium flux increases with increasing field toward the large-signal saturation limit. This transition region can account for the slow departure from ohmic behavior observed in GaAs as compared with the abrupt departure observed in CdS.

Figure 6 shows that, at sufficiently high fields, the drift current should be saturated at the velocity of sound. At fields intermediate

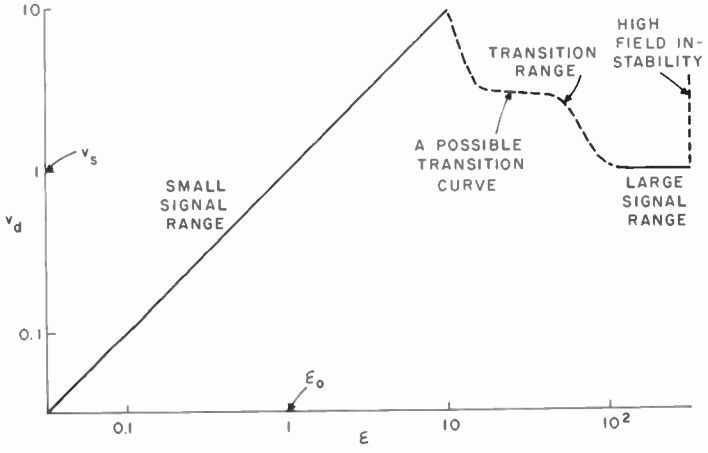


Fig. 7—Drift velocity as a function of electric field ($l > \lambda$).

between the onset of net acoustoelectric gain and final saturation at the velocity of sound, it is possible that the model used by Bray may be a good approximation to account for saturation at velocities well above the velocity of sound. This behavior is sketched in Figure 7. A more reliable estimate of the current–voltage curve in the intermediate region will depend on a better understanding of the departures from the v_d-drifted Maxwellian distribution encountered as the large-signal limit is approached.

Conceptually, this means understanding the transition from a small-signal formulation of the gain,

$$\text{Gain} \propto (v_d - v_s) \times E_{\text{flux}} . \tag{26}$$

to the large-signal formulation of Equation (5),

$$\text{Gain} \propto \mathcal{E} - \mathcal{E}_0 . \tag{27}$$

The small-signal formulation is proportional to the energy density of sound flux; the large-signal formulation is independent of the flux. This means that, in the small-signal formulation, the electrons sample the full strength of the electric field of the flux. In the large-signal formulation, the electrons are bunched near the bottoms of the troughs of the sound waves and do not sample or are not "aware" of the full amplitude of the electric fields of the sound waves.

Drift velocities saturated close to the velocity of sound favor Equation (27). Drift velocities, as observed by Bray for the III-V compounds, that are saturated well above the velocity of sound, favor Equation (26) and the corresponding analysis carried out by Bray.

Characteristic Parameters for $l > \lambda$

The onset of net acoustoelectric gain, as pointed out above, is given by Equation (10) with the value of β enhanced by the factor l/λ or $l/(2L_D)$. The onset of final saturation at the velocity of sound should be essentially the same as for $l < \lambda$, since the large-signal theory is the same for both. The criterion for the existence of net acoustoelectric gain should again follow Equations (12) and (17) for $l < \lambda$ with β replaced by $\beta l/(2L_D)$. The incubation times are given as before by Equation (14), where $\mathcal{E}^*$ is determined by Equation (10) with β replaced by $\beta l/(2L_D)$.

Acknowledgments

The writer is indebted to Professor Andrew Schultz, Jr., Dean of the College of Engineering and to Professor Herbert J. Carlin, Director of the School of Electrical Engineering, for the invitation to spend a semester in residence at Cornell University. He is grateful to Professors L. F. Eastman and G. C. Dalman for their hospitality during this stay. The writer has had the advantage of numerous discussions of the subject of this paper with Professor R. Bray of Purdue University, with Professor Brian K. Ridley of the University of Essex, during a joint residence at Cornell, and with Drs. L. Friedman and A. R. Moore of RCA Laboratories.

ADAPTIVE DETECTION WITH REGULATED ERROR PROBABILITIES*

By

HAROLD M. FINN

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

Summary—A class of automatic detection modes has been developed for operation in a target environment of statistically non-stationary extended clutter. These modes yield both a false-alarm probability invarient with changes in clutter level and a lower bound of the detection probability over a specified range of clutter-level changes. Spatially sampled maximum-likelihood estimates made of the clutter levels at each of the radar resolution cells on a first sequential step are used to control aspects of the radar policy affecting discrimination against the clutter on subsequent sequential steps. In addition to effecting uniform performance, the adaptive-detection procedures offer the potential of a significant reduction in sensor-system cost over the case of a radar system operating with a non-adaptive mode in the same environment. A complete performance analysis of such an adaptive mode for the Swerling Case #2 fluctuating target model is presented.

The false-alarm-rate regulation mechanism is a threshold control procedure—the detection threshold at a resolution cell is made proportional to an updated maximum-likelihood estimate of the output variance due to the clutter phenomena at the resolution cell. The mechanism for controlling a lower bound of the detection probability, in this analysis, involves making the number of transmissions yielding statistically independent clutter samples on the second sequential step a function of the maximum of the clutter-level estimates made on the first sequential step. Other adaptive discriminants that may be employed in such a mode are also discussed.

INTRODUCTION

THE WORK described is associated with the maintenance of uniform performance and a minimized cost of a surveillance radar system designed to automatically detect a point target embedded in a statistically non-stationary extended-clutter target environment. Where the noise involved is statistically stationary, the detection procedure normally employed is to compare the linearly detected

* This work was partially supported under Air Force Contract No. AF30 (602)-4145.

matched filter output with a fixed threshold. The extreme sensitivity of the false-alarm probability of such a detection mode to small changes in the noise level is indicated in Figure 1(a). As can be seen, a 3-db increase in the total noise-power density (clutter plus thermal noise) causes an increase of 10,000 times in the single-cell false-alarm prob-

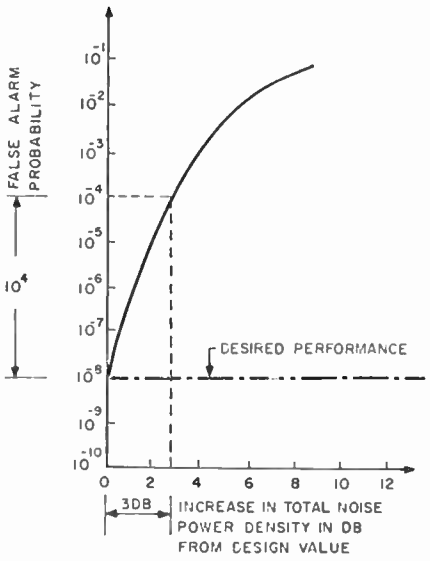


Fig. 1(a)—Change in false-alarm probability with noise level (fixed-threshold system).

ability (P_{FA}). This represents a change of four orders of magnitude from the design value ($P_{FA} = 10^{-8}$ to $P_{FA} = 10^{-4}$). The dual nature of the problem is revealed when one considers a system where, in some manner, the detection threshold can be reset to compensate for this 3-db increase in noise-power density, i.e., where the false-alarm probability is again set at 10^{-8}. The 3-db increase in noise level would then cause the detection probability to fall from 0.9 to 0.15. Because a non-adaptive detection mode is so sensitive to a statistically non-stationary clutter environment, the sensor system must, in most cases, be designed to handle a worst-case clutter level at each beam position. The system expense of a look at each beam position would be equally high, matching the case of a uniformly high clutter level in all beam positions of the surveillance volume. For those surveillance situations where typical clutter encountered is of a heterogeneous nature, i.e.,

where some beam positions have high clutter levels and others have light clutter or experience a clear environment, the non-adaptive mode would provide a serious mismatch condition. Such a hypothetical case is shown schematically in Figure 2, where the clutter level and non-

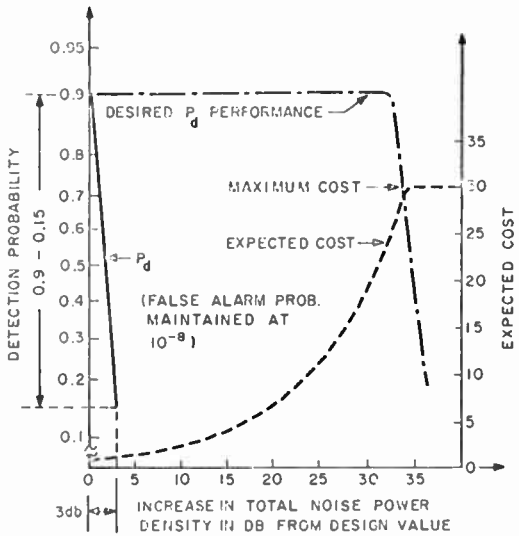


Fig. 1(b)—Change in detection probability with noise level for the case where the threshold is reset for $P_{FA} = 10^{-8}$.

adaptive system cost are plotted as functions of the position of the beam within the surveillance volume.

This paper describes an adaptive detection procedure that provides (1) a false-alarm probability that is invariant with changes in the clutter power density, (2) a specified lower bound of the detection probability for an hypothesized target type and range of clutter level changes, and (3) a cost that is roughly proportional to the clutter encountered at a beam position. Typical regulated forms of the detection and false-alarm probabilities for this adaptive mode are shown in Figures 1(a) and 1(b), respectively. The cost per beam position is related to the encountered clutter, and this type of cost function is schematically illustrated in Figures 2 and 1(b). The average cost for the adaptive mode for this hypothesized heterogeneous clutter environment is lower than the cost for the corresponding non-adaptive mode, indicating the possibility of a reduction in the overall system costs.

Basic Adaptive-Detection-Mode Design Philosophy

It is assumed that the surveillance beam can be programmed through the surveillance volume in discrete steps and that the dwell time at a beam position can change from one beam position to another. Consequently, a multiple-stage decision processor can be employed at a

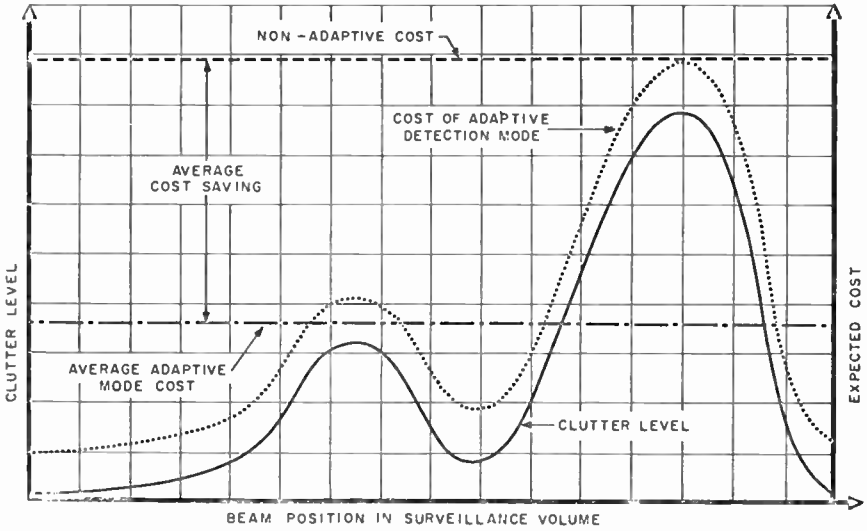


Fig. 2—Indication of potential system cost saving of adaptive detection mode in non-uniform clutter environment.

beam position to perform the two-alternative statistical hypothesis test relating to the presence of a target, the number of stages employed being a random variable. The adaptive-mode design philosophy involves embedding within a multiple-stage decision processor framework those adaptive mechanisms that most efficiently regulate the error probabilities when a non-uniform and statistically non-stationary clutter environment is encountered. These adaptive modes are developed here within the context of a decision processor having a maxmium of two sequential steps. (However, there are no limitations that would preclude the use of more than two steps.) The first sequential step at a beam position in these modes provides an estimate of the encountered clutter levels in the admissible target parameter space. This estimate of the target-scattering function of the extended clutter target phenomena is then used to determine the transmission, reception, and data-processing requirements on the second step. Thus, the

overall error probabilities, false alarm, and false dismissal are regulated over the specified range of clutter-level changes. The basic clutter-level estimation procedure employed in these adaptive modes[1] is based on a priori assumption of the shape of the continuous target scattering function of the clutter $\rho(\tau, f)\, d\tau df$. That is to say, it is assumed that

$$\rho(\tau,f)\, d\tau df = C\rho'(\tau,f)\, d\tau df\,, \tag{1}$$

where $\rho'(\tau,f)\, d\tau df$ is the assumed shape of the target-scattering function and C is the magnitude to be estimated. $\rho(\tau,f)\, d\tau df$ represents the average value of the energy returned from the clutter targets in the elementary region having a range of delay lying between τ and $\tau + d\tau$ and a doppler shift lying between f and $f + df$. For the uncorrelated-clutter models considered here (i.e., where clutter returns from disjoint portions of the parameter space are statistically uncorrelated), the ensemble average of the square of the linearly detected matched-filter output $E\{|MF(\tau_i,f_i)|^2\}$ may be represented as the two-dimensional convolution of Woodward's signal ambiguity[2] $|X(\tau,f)|^2$ and the target-scattering function $\rho(\tau,f)\, d\tau df$, plus a term N_n representing the additive thermal noise power density of the receiver. Thus,

$$E\{|MF(\tau_i,f_i)|^2\} = \int\int |X(\tau,f)|^2 \rho(\tau + \tau_i, f + f_i)\, d\tau df + N_n\,. \tag{2}$$

τ_i and f_i are the coordinates of the ith resolution-cell delay and doppler frequency; the coordinate system has been centered on the resolution cell under test.[3,4] For this model, the clutter return is the result of a spatially distributed random collection of point scatterers; the ensemble average of the clutter returns in non-overlapping portions of the range doppler-angle space is uncorrelated. For this clutter model the maximum likelihood estimate (MLE) of the variance of the output of the resolution cell under test, $\hat{E}\{|MF(\tau_n,f_n)|^2\}$, may be approximated as

---

[1] H. M. Finn, "Adaptive Detection in Clutter," *Proc. Nat. Electronics Conf.*, Vol. XXII, p. 562, 1966.

[2] P. M. Woodward, *Probability and Information Theory, with Applications to Radar*, McGraw-Hill Book Co., Inc., New York, 1953.

[3] E. C. Westerfield, R. H. Prager, and J. L. Stewart, "Processing Gains Against Reverberation (Clutter) Using Matched Filters," *IRE Trans. on Information Theory*, Vol. IT-6, p. 342, June 1960.

[4] R. Price and P. E. Green, Jr., "Signal Processing in Radar Astronomy-Communication via Fluctuating Multipath Media," Lincoln Laboratory Tech. Report 234, Oct. 1960.

$$\hat{E}\{|MF(\tau_o,f_o)|^2\} = \frac{1}{N}\sum_{1}^{N}\frac{X_i^2}{W_i} \tag{3}$$

for two important cases of interest in this paper. One is the case where the predominant component of the output variance is due to the clutter return rather than the additive thermal noise of the receiver; the second case is where the clutter level is assumed to be uniform over the breadth of the resolution cells involved in the estimation (see Appendix I and Reference (1)). The random variables $\{X_i = |MF(\tau_i,f_i)|\}$ are the linearly detected matched-filter outputs of the N resolution cells surrounding the cell under test and covered by the clutter cloud. The weights $\{W_i\}$ are defined by

$$W_i = \frac{\sigma_i^2}{\sigma_o^2}, \tag{4}$$

or, using Equations (1) and (2) for the clutter limiting environment, by

$$W_i = \frac{\gamma_i^2}{\gamma_o^2} \tag{5}$$

where

$$\gamma_i^2 = \int\int |X(\tau,f)|^2 \rho'(\tau+\tau_i, f+f_i)\, d\tau df\,. \tag{6}$$

The numbers $\{\gamma_i^2\}$, and consequently the $\{W_i\}$, are based on the known signal-ambiguity function $|X(\tau,f)|^2$ and the a priori assumed 'shape-factor' of the target-scattering function. For the case where the clutter level is assumed to be uniform over the resolution cells involved in the estimation procedure, the weights $\{W_i\}$ are each equal to unity. A functional diagram of this estimation procedure employed in a specific adaptive-detection mode, which is the subject of this paper, is shown in Figure 3.

The receiver system may be assumed to be composed of a number of contiguously spaced matched filters covering the admissible range-doppler parameter space of a target, where the dimension of each of the resolution cells in range is $\Delta\tau = 1/B$, and for the doppler dimension $\Delta f = 1/T$, where B and T are the bandwidth and duration of the transmitted signal, respectively.

If the decision process used to compare the output of each resolution cell with a threshold made proportional to the clutter-level estimate of that cell is the MLE procedure defined by Equation (3), the false-alarm probability would be a constant and would not vary with changes in the clutter level.[1] The detection efficiency of this automatic threshold-control procedure when applied to a fixed-sample-size test is also

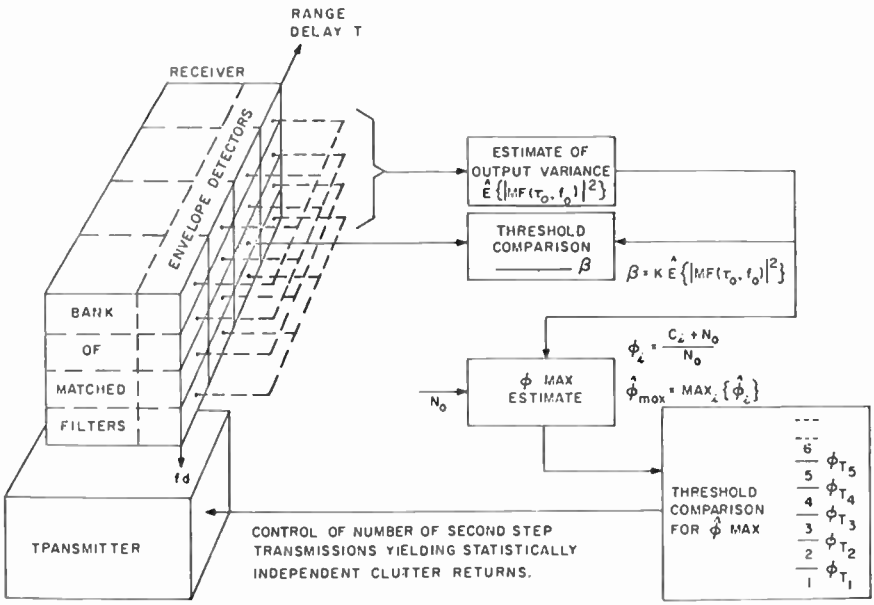


Fig. 3—Adaptive detection mode yielding regulated error probabilities.

noted to be a monotonically increasing function of the number of resolution cells employed in the threshold-control (or the clutter-level-estimation) procedure. The detection efficiency of this procedure is comparable to that of an ideal detection situation when 100 resolution cells are employed (the ideal detection situation is defined as one where the total clutter plus thermal-noise-power density is a known quantity, and a fixed threshold can be employed). This threshold-control procedure becomes the adaptive mechanism for regulating the false-alarm probability in the multiple-stage adaptive modes described here. For each resolution cell, the decision statistic of that cell, which is a function of the first and second sequential step outputs, is compared with a threshold made proportional to an up-dated estimate of the cell's clutter level.

This threshold control procedure does not, however, solve the other major problem associated with detection in a clutter environment, namely the sensitivity of the detection probability of a conventional detection mode to small changes in the clutter level. As stated previously, an adaptive mechanism must be embedded within the multiple-stage decision processor to achieve a control of the lower bound of the detection probability over a specified range of clutter levels. This mechanism involves making the degree of the second-step discrimination against the extended clutter return a monotonically increasing function of the estimate of the clutter level obtained on the first step. Roughly speaking, if large clutter-level estimates are made on the first step, a proportionately more expensive second-step look, involving a greater discrimination against the clutter, is employed for the second-step radar policy; a less expensive second step is employed if the estimated clutter levels on the first step are relatively low. In this manner, a degree of regulation of the detection probability is effected, and the effort expended by the sensor system at a beam position is a monotonically increasing function of the clutter encountered. The variable clutter discriminant appropriate for this adaptive detection mode depends, for any one application, on the clutter model, the sensor system constraints, and the cost function involved. This adaptive discrimination may involve a control of one or more of the following second-step transmission characteristics—the transmitted signal bandwidth, the signal duration, and the number of transmissions yielding statistically independent clutter returns.

Specific Class of Adaptive Detection Modes

The specific class of adaptive detection modes developed in this paper employs, as the adaptive discriminant, a control of the number of second-step transmissions (N_2) designed to yield statistically independent samples of the clutter return. The number of transmissions is based on the maximum of the clutter level estimates made on the first step. (N_1 transmissions are assumed to be employed on the first step, and this number is a deterministic quantity.) The decision statistic at each resolution cell employed in this mode is the noncoherent sum of the first- and second-step outputs, and this statistic is compared with a threshold that is proportional to an updated estimate of the variance of the ouput of the cell in question.

The method for accomplishing the statistical independence of the clutter returns of the ($J = N_1 + N_2$) transmissions may be accom-

plished by employing one or more of the following measures. (1) The interval between transmissions may be made long with respect to the correlation time of the clutter. (2) A sufficient carrier-frequency 'jump' may be employed on each transmission to effect the statistical independence of the sum of the field vectors representing the discrete point-scatter returns of the cloud at a resolution cell. (3) The waveform design may be changed for each transmission in such a manner that the intersection, or convolution, of the clutter-target-scattering function and the waveform-signal-ambiguity functions involve different and statistically independent contributions of the extended clutter-target phenomena. The particular measures actually employed to approximate this statistical independence would depend on the particular surveillance situation involved.

A functional diagram of this adaptive mode is presented in Figure 3. The number of second-step transmissions is made a monotonically increasing function of the estimate made at each resolution cell of the quantity ϕ_i. This quantity is defined as the ratio of the variance of the detected matched filter output of the ith resolution cell (which may be in a clutter region) to the variance of the receiver output when only thermal noise is present (the clear environment case);

$$\phi_i = \frac{E\{|MF(\tau_i, f_i)|^2\}}{E\{X_{TN}^2\}}. \tag{7}$$

X_{TN} is the sampled and detected filter output when only the additive thermal noise is present. This receiver thermal-noise variance, $E\{X_{TN}^2\} = N_o$, is assumed to be a known quantity or to be accurately measured. For example, this measurement can be made periodically by sampling the receiver output during the radar dead-time. This time interval of no radar backscattering is typically sufficient to make the error in the thermal-noise-level estimate negligibly small.

For the radar cases of interest where the power spectrum of the noise is uniform over the receiver bandwidth and the target scattering function of the clutter is relatively constant over the dimensions of the resolution cell, ϕ_i is approximately equal to the ratio of the thermal-plus-clutter noise-power density to the thermal noise-power density

$$\phi_i \approx \frac{C_i + N_o}{N_o}.$$

The estimates

$$\left\{\hat{\phi}_i = \frac{\hat{E}\{|MF(\tau_i, f_i)|^2\}}{N_o}\right\}$$

are obtained using the result of Equation (3) (see Appendix III and Reference (1))

$$\hat{\phi}_i = \frac{1}{N_o} \sum_{i=1}^{N} \frac{X_i^2}{W_i}. \tag{8}$$

The maximum of these estimates, $\hat{\phi}_{MAX} = \text{MAX}\ \{\hat{\phi}_i\}, i = 1, 2, \ldots n$, is then compared with a set of thresholds $\{\phi_{TJ}\}$ to determine the number of second-step transmissions N_2 that are needed to yield statistically independent clutter returns. J transmissions are employed (or $J - N_1$ second-step transmissions) whenever $\phi_{TJ-1} < \hat{\phi}_{MAX} < \phi_{TJ}$; when $\hat{\phi}_{MAX} > \phi_{TN}$, N transmissions are employed on the second step. This latter policy rule places a limit on the expenditure of the sensor system at a beam position, i.e., regardless of the encountered clutter, the number of transmissions at a beam position never exceeds N. Consequently, when the true value of the clutter level exceeds a value where more than N transmissions are required for the hypothesized average target cross section at maximum range, the detection probability will fall below the desired lower bound. Rather than placing a fixed maximum or "cap" of N transmissions at each beam position, the expenditure at prior beam positions could be memorized so that an adjustable cap could be employed.

The adaptive mode is also designed so that at least one second-step transmission is employed even in the clear environment. That is, $N_1 + 1$ total transmissions are employed whenever $0 \leq \hat{\phi}_{MAX} < \phi_{TN_1+1}$.

The general test procedure described, that of allowing the maximum of the clutter-level estimates obtained on a first sequential step at a beam position to control the second-step radar policy, follows a min–max type of strategy. An alternative Bayes procedure (not considered here) could involve basing the number of second-step transmissions on an average of these clutter-level estimates.

Decision Rule for Adaptive Mode

The observed random variables of the adaptive mode are the sampled matched-filtered and square-law detected outputs and may be rep-

resented as the set of cross-correlator outputs

$$\{Y_{iK} = |\int S_K(t_i,\tau_i,f_i)\, V_K(t)\, dt|^2\} \tag{9}$$

where Y_{iK} is the output of the ith resolution cell on the Kth transmission. $S_K(t_i,\tau_i,f_i)$ is a delayed (τ_i) and doppler shifted (f_i) replica of the Kth transmitted signal $S_K(t)$ for the ith resolution cell. $V_K(t)$ is the input after the Kth transmission. The cells are assumed to be spaced by the range and doppler resolution so that

$$\tau_{i+1} = \tau_i + \frac{1}{B}, \text{ and } f_{i+1} = f_i + \frac{1}{T}.$$

The decision statistics (the functions of the observed random variables that are employed in deciding on the presence or absence of the target) are formed in this adaptive mode after the second sequential step. They are simply the sum of the (J) sampled square-law detected outputs on both steps in a common resolution cell, N_1 transmissions being employed on the first step. For the ith cell, the decision statistic is

$$Z_i = \sum_{K=1}^{J} Y_{iK}\,. \tag{10}$$

The target-present hypothesis is accepted in the ith resolution cell if $Z_i \geqq T_J$, where the threshold T_J is made proportional to an estimate of the variance of the output of the ith cell. This estimate is based on the sampled outputs of the $(J - N_1)$ second-sequential-step transmissions of the cells surrounding the cell under test, which, it is assumed, are also covered by the extended clutter target phenomenon. That is,

$$T_J = K_J \hat{E}_{(J)}\{|MF(\tau_o,f_o)|^2\}$$

or

$$T_J = \frac{2K_J}{N_C(J - N_1)} \sum_{l=1}^{N_C} \sum_{K=N_1+1}^{J} \frac{Y_{lK}}{W_l},$$

where K_J is the threshold control constant used when $(J - N_1)$ trans-

missions are employed on the second sequential step. If no threshold crossings occur, that is

$$\bigcap_{i=1}^{M} [E:(Z_i < T_J)]$$

is true, (where M is the number of resolution cells covering the admissible target parameter space) then the no-target hypothesis is accepted at the beam position.

Analytical Formulation

This analytical procedure for setting the $\{\phi_{TJ}\}$ and $\{K_J\}$ thresholds and the performance evaluation of the adaptive mode are considered in this section. An evaluation of the adaptive detection mode involves a determination of the detection probability (P_D) in the resolution cell with the maximum clutter level (i.e., the one where $\phi_{MAX} = MAX\ \{\phi_i\}, i = 1, 2, \ldots n$), the single-cell false-alarm probability P_{FA}, and the expected number of transmissions $E(J) = N_1 + E(N_2)$. All these parameters are functions of the following:

(1) The number (N_C) of resolution cells employed in the control of both the receiver thresholds and the average number of transmissions.

(2) The set of actually encountered ratios $\{\phi_i\}$, and in particular the maximum of this set ϕ_{MAX} (the target scattering function $\rho(\tau,f)\,df$ and the thermal-noise-power density N_o provide this information).

(3) The specified stochastic properties of the target fluctuations.

(4) The hypothesized ratio (for a single transmission) of the average energy of the signal received from the target to be detected to the thermal noise power density, $\bar{X}_o = \bar{E}/N_o$.

(5) The maximum number of transmissions permitted at a beam position (N_M).

The detection probability in the resolution cell with the maximum clutter level may be expressed as:

$$P_D = \sum_{J=N_1+1}^{N_M} \int_{T_J=0}^{\infty} \mathrm{P}\left\{Z_o = \sum_{i=1}^{J} Y_i > T_J \,|\, \bar{X}, J\right\} f(T_J, J; N_C, \{\phi_i\})\, dT_J \quad (12)$$

The single-cell false-alarm probability P_{FA} is symbolically represented in the same manner, but on the hypothesis that $\bar{X}$ equals zero.

In Equation (12), $P\{Z_o > T_J \,|\, \bar{X}, J\}$ represents the conditional proba-

bility of detection based on the following: (1) the sum of the first-step (N_1) and second-step (N_2) transmissions equals J, (2) the threshold for the noncoherent sum of the J square-law-detected outputs in a common resolution cell $\{Y_i\}$ is T_J, and (3) the single-transmission average signal energy to clutter-plus-thermal-noise-power density is $\overline{X}$ (and expressing $\overline{X}$ in terms of the encountered ϕ_{MAX} and the hypothesized $\overline{X}_o$, $\overline{X} = \overline{X}_o/\phi_{MAX}$).

The probability density function $f(T_J, J; N_C, \{\phi_i\})\,dT_J$ in Equation (12) represents the joint probability that J transmissions will be employed in the adaptive mode, and that the threshold lies between the value T_J and $T_J + dT_J$. This joint probability density of the discrete variable J and the continuous variable Z_o is a function of the number of resolution cells N_C employed in making the clutter-level estimates and the actually encountered clutter levels at each of these cells (the set $\{\phi_i\}$). Expressed in terms of conditional probabilities

$$f(T_J, J; N_C, \{\phi_i\})\,dT_J = f(T_J | J; N_C, \{\phi_i\})\,dZ_o P\{J\}\,. \tag{13}$$

The discrete probability $P\{J\}$ may be expressed

$$P\{J\} = \int_{\phi_{TJ-1}}^{\phi_{TJ}} f(\phi_{MAX}; N_C, \{\phi_i\})\,d\phi_{MAX}\,. \tag{14}$$

That is to say, the probability of employing J transmissions in this adaptive mode is the probability that the estimate ϕ_{MAX} lies between the two ϕ thresholds ϕ_{TJ-1} and ϕ_{TJ}.

An exact evaluation of the probability density function $f(\phi_{MAX}; N_C, \{\phi_i\})\,d\phi_{MAX}$ is difficult since the set of $\{\phi_i\}$ clutter-level estimates made for each of the N_C resolution cells and employed in obtaining ϕ_{MAX} are not statistically independent of one another. These $\{\phi_i\}$ estimates depend on overlapping sets of resolution cell outputs. The formulation difficulties are overcome if it is assumed that the resolution cell yielding the largest estimate of the set $\{\phi_i\}$ is actually the cell containing the largest clutter level, i.e., $\phi_o = \text{MAX}\{\phi_i\}$ where the subscript zero designates the cell that actually encounters the maximum clutter level. The density function of ϕ_o on this hypothesis, $f(\phi_o | \phi_o = \phi_{MAX})\,d\phi_o$ may now be employed in obtaining an approximate formulation of $P\{J\}$.

$$P\{J\} \approx P'\{J\} = \int_{\phi_{TJ-1}}^{\phi_{TJ}} f(\phi_o | \phi_o = \phi_{MAX})\,d\phi_o\,. \tag{15}$$

The probability density function of the estimator

$$\hat{\phi}_o = \frac{2}{N_C N_1 N_o} \sum_{i=1}^{N_C} \sum_{K=1}^{N_1} \frac{Y_{iK}}{W_i} \tag{16}$$

based on Equation (8), but with square-law-detector outputs $\{Y_{iK}\}$, is shown in Appendix III to be

$$f(\hat{\phi}_o | \phi_o = \phi_{MAX})\, d\hat{\phi}_o = \frac{\left(\frac{N_C N_1}{\underline{\phi_o}}\right)^{N_C N_1} \hat{\phi}^{N_C N_1 - 1} \exp\left\{-\frac{N_C N_1 \hat{\phi}_o}{\underline{\phi_o}}\right\} d\hat{\phi}_o}{(N_C N_1 - 1)!} . \tag{17}$$

Here $\underline{\phi_o}$ represents the 'true' value of the ratio of the clutter-plus-thermal-noise-power density to the thermal-noise-power density at the zeroth resolution cell. $P\{J\}$ may now be approximated as

$$P'\{J\} = F_G\left(\frac{N_C N_1 \phi_{TJ}}{\underline{\phi_o}}\right) - F_G\left(\frac{N_C N_1 \phi_{TJ-1}}{\underline{\phi_o}}\right) \tag{18}$$

where

$$F_G(X) = \frac{1}{(N_C N_1 - 1)!} \int_0^X \xi^{N_C N_1 - 1} e^{-\xi}\, d\xi . \tag{19}$$

The detection probability of Equation (12) may now be expressed as

$$P_D \approx \sum_{J=X_1+1}^{X_{MAX}} \int_{T_J=0}^{\infty} P\left\{Z = \sum_{i=1}^{J} Y_i > T_J | \bar{X}, J\right\} f(T_J | J ; N_o, \{\phi_i\})\, dT_J P'\{J\} . \tag{20}$$

It is convenient to consider the change of variables $\xi_i = Y_i/\sigma_o$ and $\beta_i = T_i/\sigma_o \cdot \sigma_o^2$ as the output variance for the clutter-plus-thermal-noise case. Then

$$P_D \approx \sum_{J=X_1+1}^{X_M} \int_{\beta_J=0}^{\infty} P\left\{Z = \sum_{i=1}^{J} \xi_i > \beta_J | \bar{X}, J\right\} f(\beta_J | J)\, d\beta_J P'\{J\} . \tag{21}$$

The conditional probability density function of the normalized threshold β_J on the hypothesis of J transmissions is developed in Appendix II, and is shown to be

$$f(\beta_J|J)d\beta_J = \frac{1}{(N_c(J-N_1)-1)!}\left(\frac{\beta_J}{C_J}\right)^{N_c(J-N_1)-1} e^{-\beta_J/C_J}\frac{d\beta_J}{C_J}, \quad (22)$$

where $C_J = 2K_J/(N_c(J-N_1))$ and the normalized threshold β_J is made proportional to the clutter-level estimate made with the resolution cell outputs of the N_c surrounding resolution cells obtained on the $J-N_1$ second-step transmission, or

$$\beta_J = \frac{2K_J}{\sigma_v^2 N_c(J-N)}\sum_{i=1}^{N_c}\sum_{K=N_1+1}^{J}\frac{Y_{iK}}{W_i}. \quad (23)$$

The development of the conditional probability of detection

$$P\left\{Z = \sum_{i=1}^{J}\xi_i > \beta_J|\bar{X},J\right\}$$

is now considered. In the interest of expressing the adaptive detection-mode performance for some specific cases of interest, this analysis considers only the case where the target cross section is assumed to undergo relatively rapid fluctuations relative to the time interval between transmissions. Thus, the target cross sections on each transmission are statistically independent of one another. This is the case, for example, for the Swerling Cases #2 and #4 targets.[5] Since the samples of the clutter return and the additive thermal noise are also assumed to be statistically independent between transmissions, the probability density function $f(Z|J,\bar{X})dZ$ of the normalized noncoherent sum of the square-law-detected outputs $(Z = \sum_{i=1}^{J}\xi_i)$ may be expressed as the convolution of the individual density functions $f_i(\xi_i)d\xi_i$ or

$$f(Z) = f_1(\xi_1)*f_2(\xi_2)*\ldots*f(\xi_J\xi_J). \quad (24)$$

[5] P. Swerling, "Probability of Detection for Fluctuating Targets," *IRE Trans. on Information Theory*, Vol. II-6, p. 269, April 1960.

where
$$f_i(\xi_i)\,d\xi_i = \int_{X=0}^{\infty} f_i'(\xi_i|X)\,f(X,\overline{X})\,dX\,d\xi_i\,. \tag{25}$$

$f_i'(\xi_i|X)\,d\xi_i$ is the conditional probability density function of the square-law-detected output on the ith transmission on the hypothesis that the input signal energy to the total noise power density is X, and

$$f_i'(\xi_i|X)\,d\xi_i = e^{-\xi_i - X} I_o(2\sqrt{X\xi_i})\,d\xi_i\,. \tag{26}$$

(See, for example, Helstrom.[6]) This function is recognized as the Rice distribution where $I_o(X)$ is the modified Bessel function of zero order. Performing the integration of Equation (25) for the swerling Case #2 target where $f(X,\overline{X})\,dX = (e^{-X/\overline{X}}/\overline{X})\,dX$:

$$f(\xi_i)\,d\xi_i = \frac{1}{1+\overline{X}}\, e^{-\xi_i/(1+\overline{X})}\,d\xi_i\,. \tag{27}$$

The probability density function $f(Z|J,\overline{X})\,dZ$ is now obtained by performing the convolution indicated by Equation (24),

$$f(Z|J,\overline{X})\,d\overline{X} = \alpha^J Z^{J-1} e^{-\alpha z} dZ\,, \tag{28}$$

where $\alpha = 1/(1+\overline{X})$. Now performing the integration

$$P\{Z > \beta_o|\overline{X},J\} = \int_{Z=\beta_J}^{\infty} f(Z|\overline{X},J)\,dZ,$$

we obtain

$$P\{Z > \beta_J|\overline{X},J\} = e^{-\alpha\beta_J} \sum_{K=0}^{J-1} \frac{(\alpha\beta_J)^K}{K!} \tag{29}$$

Employing the developed distribution, the probability of detection P_D of Equation (21) may now be expressed as

---

[6] C. Helstrom, *Statistical Theory of Signal Detection*, Pergamon Press, 1960.

$$P_D = \sum_{J=N_1+1}^{N_M} \int_{\beta_J=0}^{\infty} \left[e^{-\alpha\beta_J} \sum_{K=0}^{J-1} \frac{(\alpha\beta_J)^K}{K!} \right] \left[\frac{\left(\frac{\beta_J}{C}\right)^{N_C(J-N_1)-1}}{(N_C(J-N)-1)!} e^{-\beta_J/C} \frac{d\beta_J}{C} \right] P'\{J\}, \tag{30}$$

and finally

$$P_D = \sum_{J=N_1+1}^{N_M} \sum_{K=0}^{J-1} \frac{\left(\frac{\alpha\tau_J}{M}\right)^K (M+K-1)!}{\left(1+\frac{\alpha\tau_J}{M}\right)^{M+K} K!\,(M-1)!} P'\{J\}, \tag{31}$$

where $M = N_C(J - N_1)$ and $\tau_J = 2K_J$ and $P'\{J\}$ is defined by Equations (18) and (19).

Selection of $\{\tau_J\}$ Threshold Constants and the False-Alarm Probability

The single-cell false-alarm probability is obtained from Equation (31) by setting $\overline{X} = 0$, so that $\alpha = 1$. The result is

$$P_{FA} = \sum_{J=N_1+1}^{N_M} \sum_{K=0}^{J-1} \frac{\left(\frac{\tau_J}{M_J}\right)^K (M_J+K-1)!}{\left(1+\frac{\tau_J}{M_J}\right)^{M_J+K} K!\,(M_J-1)!} P'\{J\}. \tag{32}$$

By setting

$$P_{FA_\theta} = \sum_{K=0}^{J-1} \frac{\left(\frac{\tau_J}{M_J}\right)^K (M_J+K-1)!}{\left(1+\frac{\tau_J}{M_J}\right)^{M_J+K} K!\,(M_J-1)!}, \tag{33}$$

the threshold constants $\tau_J = 2K_J$ are determined for all $J = N_1 + 1, N_1 + 2, \ldots, N_M$. Then

$$P_{FA} = P_{FA_0} \sum_{J=N_1+1}^{N_M} P'\{J\} = P_{FA_0}, \text{ since } \sum_{J=N_1+1}^{N_M} P'\{J\} = 1 , \qquad (34)$$

and it is clear that the false-alarm probability of this adaptive mode is a constant and is invariant to changes in the clutter plus thermal noise power density.

Selection of $\{\phi_{TJ}\}$ Thresholds

The method employed in this adaptive mode for setting the $\{\phi_{TJ}\}$ thresholds, which determine the number of second-step transmissions that yield statistically independent clutter returns, is best illustrated by putting the expression of P_D (Equation (30)) in the form

$$P_D = \sum_{J=N_1+1}^{N_M} (P_D|J,\overline{X})\, P'\{\phi_{TJ-1} < \phi_o < \phi_{TJ}\} , \qquad (35)$$

where the conditional probability of detection $(P_D|J,\overline{X})$ for J transmissions and an average return-signal-energy to thermal-plus-clutter-noise-power density, $\overline{X}$, is expressed as

$$P_D|J,\overline{X} = \sum_{K=0}^{J-1} \frac{\left(\dfrac{\alpha\tau_J}{M_J}\right)^K (M_J + K - 1)!}{\left(1 + \dfrac{\alpha\tau_J}{M_J}\right)^{M_J+K} K!\,(M_J - 1)!} . \qquad (36)$$

The discrete probability of employing a total of J transmissions in the adaptive mode is

$$P'\{\phi_{TJ-1} < \phi_o < \phi_{TJ}\} = P'\{J\} , \qquad (37)$$

as defined by Equations (18) and (19). The aim in the threshold-determination procedure is to keep the detection probability P_D greater or equal to a specified lower bound P_{D_0} over a range of clutter-level changes for the hypothesized target. The threshold ϕ_{TJ} is set equal

to the clutter level yielding the desired detection probability,

$$P_{D_0} = P_D | J, \overline{X} \tag{38}$$

when J transmissions are employed and where the hypothesized target return-signal energy to thermal-noise-power density is $\overline{X}_0$. To determine the threshold ϕ_{TJ}, we solve Equation (38) for $\overline{X}_J$ and employ the relationship

$$\phi_{TJ} = \frac{\overline{X}_0}{\overline{X}_J} = \frac{C_J + N_0}{N_0}.$$

The $\overline{X}_0$ is determined in this analysis by setting $J = N_1 + 1$ in Equation (38) and solving for $\overline{X}$, and the first threshold ϕ_{TN_1} is set equal to one. By this procedure, the detection probability is regulated for a range of clutter levels, and the specified lower bound P_{D_0} is maintained under the conditions

$$P_D \geq P_{D_0} \qquad \text{and} \qquad 1 < \phi_0 \leq \phi_{N_M}. \tag{40}$$

That is to say, the detection probability bound is maintained for the range of ϕ_0 not exceeding the clutter level ϕ_{N_M}. To maintain the detection probability above the lower bound for clutter levels exceeding ϕ_{N_M}, the number of transmissions would have to be greater than the maximum number N_M allowed at a beam position.

Performance Evaluation

The expression for the false-alarm probability in this adaptive mode (Equations (32) through (34)) is noted to be independent of the encountered clutter level so that a constant probability of false alarm is maintained.

Typical sets of computations of the detection probability based on Equation (31) are presented in Figures 4 and 5 as a function of ϕ_0. The detection characteristics curves exceed the specified bound of 0.9 in Figure 4 and of 0.5 in Figure 5. As can be seen, a relatively sharp fall occurs in the detection probability when the encountered clutter level exceeds the value where the number of transmissions required to maintain the lower bound are greater than the maximum permitted at the beam position. For clutter levels below this value, however, a degree of regulation of the detection probability is achieved.

The expected number of transmissions of the adaptive modes are also plotted in these curves as a function of the clutter level. In addition, the number of transmissions that an ideal detector requires to achieve the specified P_{D_0} in a given clutter environment is also plotted.

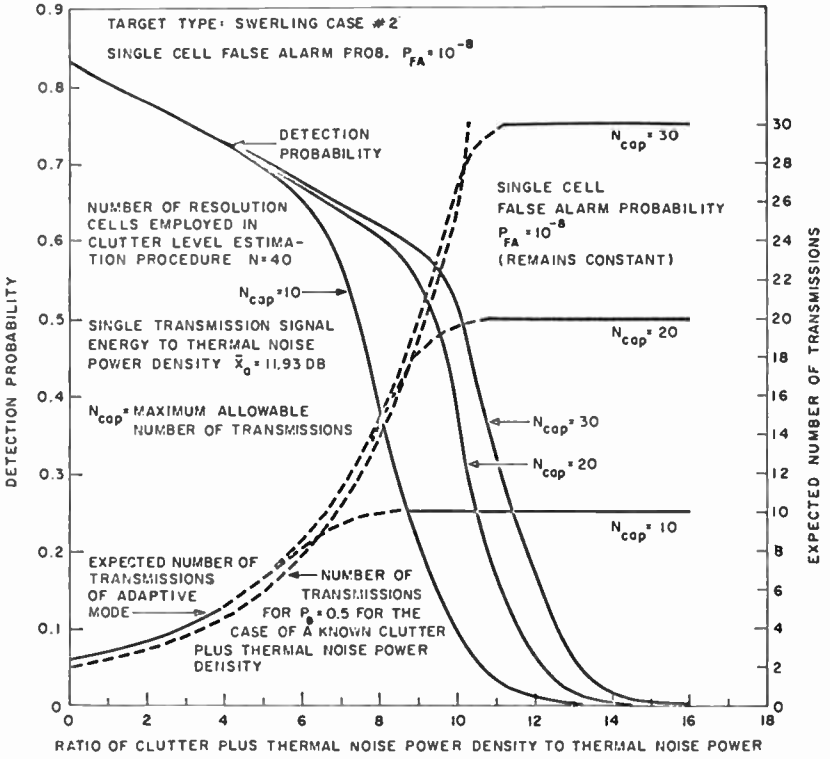


Fig. 4—Detection probability as a function of the ratio of clutter-plus-thermal-noise-power density to thermal-noise-power density (lower bound $P_D = 0.9$).

The ideal detection situation is defined here as one where the clutter level is a known value and the number of transmissions and detection thresholds can consequently be deterministic quantities. The 'loss' of the adaptive mode from the ideal detection situation for these examples is noted to be less than 1 db.

ACKNOWLEDGMENT

The computer implementations of the probability distributions describing the adaptive detection mode performance were developed by R. S. Johnson.

APPENDIX I—DEVELOPMENT OF SPATIALLY SAMPLED MAXIMUM LIKELIHOOD ESTIMATE OF RECEIVED CELL OUTPUT VARIANCE

The development of a maximum-likelihood estimate (MLE) of the ensemble average of the square of the matched filter output $E\{|MF(\tau_o,f_o)|^2\}$ based on using the set $\{X_i = |MF(\tau_i,f_i)|\}$ of the N

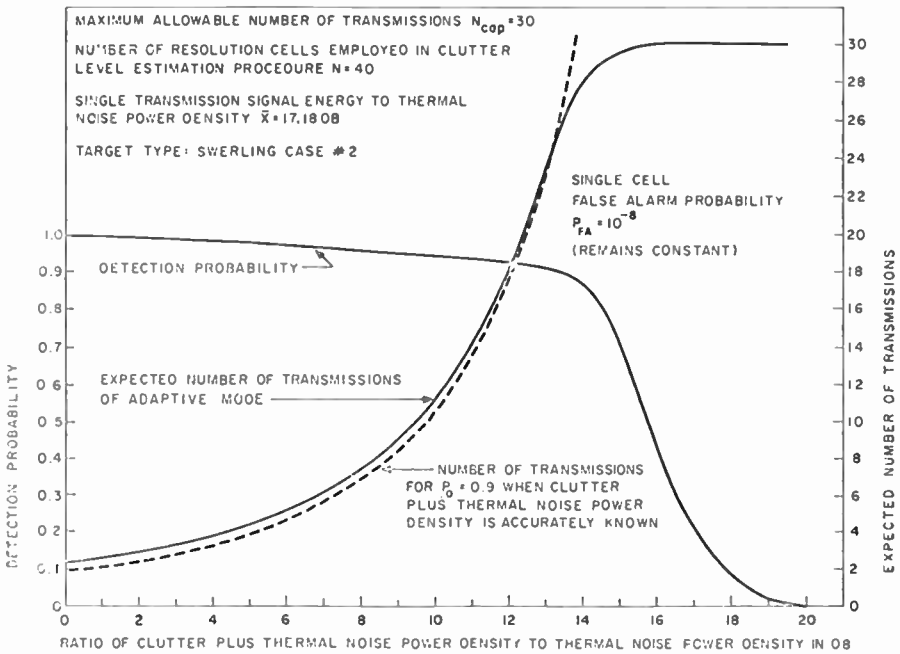


Fig. 5—Detection probability as a function of ratio of clutter-plus-thermal-noise-power density to thermal-noise-power density (lower bound $P_D = 0.5$).

outputs of the resolution cells surrounding the cell under test (centered at τ_o,f_o) begins with the expression of the output variance of the ith cell:

$$E\{|MF(\tau_i,f_i)|^2\} = \int\int |X(\tau,f)|^2 P(\tau+\tau_i, f+f_i)\,d\tau\,df + N_o, \quad (41)$$

τ_i and f_i are the coordinates of the ith resolution cell delay and doppler frequency, respectively, N_o is the noise power density due to the additive thermal noise of the receiver, and the component of variance due to the incoherent clutter cloud is expressed as the two-dimensional convolution of Woodward's ambiguity function $|X(\tau,f)|^2$ and the

target scattering function of the extended clutter phenomenon $\rho(\tau,f)d\tau df$.

$X(\tau,f)$ is the combined time and frequency auto correlation function of $U(t)$ where the transmitted signal may be represented as $AU(t)\exp\{2\pi i f_o t\}$ and where $U(t)$ is normalized so that $\int U(t)U^*(t)dt = 1$ and

$$X(\tau,f) = \int U(t)U^*(t+\tau)\exp\{-2\pi i f t\}dt. \tag{42}$$

A basic assumption involved in the development of the estimate $E\{|MF(\tau_o,f_o)|^2\}$ is that the target scattering function $\rho(\tau,f)d\tau df$ may be expressed as

$$\rho(\tau,f)d\tau df = C\rho'(\tau,f)d\tau df\,, \tag{43}$$

where the shape factor $\rho'(\tau,f)d\tau df$ of the clutter 'cloud' target scattering function is assumed to be known, and where C represents the unknown magnitude that must be estimated. The estimate $\hat{E}\{|MF(\tau_o,f_o)|^2\}$ may then be represented as:

$$\hat{E}\{|MF(\tau_o,f_o)|^2\} = \hat{C}\int\int |X(\tau,f)|^2\rho'(\tau+\tau_o,f+f_o)d\tau df + N_o\,, \tag{44}$$

since the additive thermal noise power density N_o is also assumed to be a known quantity. In effect, the $\hat{C}$ estimate of the magnitude of the target scattering function is what is desired. To develop $\hat{C}$, we first form $L(X_1,X_2,X_3,\ldots X_N;C)$, the likelihood of the receiving set $\{X_i\}$ when the magnitude of the target scattering function is C.

For the assumed clutter model, the likelihood may be expressed

$$L(X_1,X_2\ldots X_N;C) = \prod_{i=1}^{N}\frac{X_i}{\sigma_i^2}\exp\left\{-\frac{X_i^2}{2\sigma_i^2}\right\}dX_i \tag{45}$$

where

$$\sigma_i^2 = \frac{1}{2}(C\gamma_i^2 + N_o)\,. \tag{46}$$

and

$$\gamma_i^2 = \int\int |X(\tau,f)|^2 \rho'(\tau+\tau_i, f+f_i)\, d\tau df \,. \tag{47}$$

The likelihood of Equation (45) is an expression of the conditional probability of the $(X_1, X_2, \ldots X_N)$, the outputs of the resolution cell surrounding the cell under test and also covered by the clutter cloud, on the hypothesis that the target scattering function may be expressed as $C\rho'(\tau,f)d\tau df$. For the non-coherent clutter model assumed here, where a large number of randomly oriented point scatterers appear in each resolution cell, the outputs $\{X_i\}$ for the no-target case in Equation (45) are statistically independent of one another and Rayleigh distributed. The maximum likelihood estimator $(\hat{C})$ is the value of C that gives the maximum $L(X_1, X_2, \ldots X_N; \hat{C})$, and where we determine MAX L by solving $\partial L/\partial\tau = 0$ for $\hat{C}$.

Working with the monotonically related quantity $\ln L$:

$$\ln L = \sum_{i=1}^{N} \left(\ln X_i - \ln\sigma_i^2 - \frac{X_i^2}{2\sigma_i^2} \right) \tag{48}$$

and

$$\frac{\partial \ln L}{\partial \tau} = 0 = \sum_{i=1}^{N} \frac{\gamma_i^2}{\hat{C}\gamma_i^2 + N_o} - \sum_{i=1}^{N} \frac{X_i^2 \gamma_i^2}{(\hat{C}\gamma_i^2 + N_o)^2} \,. \tag{49}$$

A solution of this $2N$th degree polynominal for the maximum likelihood estimator C is feasible, but in general, it is not practical in a radar adaptive detection mode where real-time computations must be made. However, for some important cases of interest, relatively simple solutions are obtained. The first of these cases is the clutter-limiting case where the clutter-power density is much greater than the thermal-noise-power density (i.e., $C\gamma_i^2 \gg N_o$). In this case we may write

$$\hat{C} = \frac{1}{N} \sum_{i=1}^{N} \frac{X_i^2}{\gamma_i^2} \tag{50}$$

and

$$\hat{E}\{|MF(\tau_o, f_o)|^2\} = \frac{1}{N} \sum_{i=1}^{N} \frac{X_i^2}{W_i} \,, \tag{51}$$

where

$$W_i = \frac{\gamma_i^2}{\gamma_o^2}. \tag{52}$$

The set $\{\gamma_i^2\}$ are defined in Equation (47).

The second case of interest is the one where the target scattering function is relatively constant over the resolution cells involved in the estimation procedure so that $\gamma_i^2 = \gamma_J^2$ for all i and J. For this case, Equation (51) is again a good approximation where $W_i = 1$ for all i:

$$\hat{E}\{|MF(\tau_o, f_o)|^2\} = \frac{1}{N} \sum_{i=1}^{N} X_i^2. \tag{53}$$

Appendix II—Probability Density Function of the Detection Threshold β

The normalized detection threshold β_J (for the case where J transmissions are employed) is expressed as

$$\beta_J = \frac{2K_J}{\sigma_o^2 N_C (J - N_1)} \sum_{i=1}^{N_C} \sum_{K=N_1+1}^{J} \frac{Y_{iK}}{W_i},$$

(see Equation (23)).

The conditional probability density function of β_J on the hypothesis of J transmissions, $f(\beta_J|J)d\beta_J$ (Equation (22)), is developed in this Appendix.

β_J may be expressed as

$$\beta_J = C \sum_{i=1}^{N_C} \sum_{K=N_1+1}^{J} \xi_{iK} \tag{54}$$

where $C = 2K_J/(N_C(J - N_1))$ and $\xi_{iK} = Y_{iK}/\sigma_i^2$. Since $Y_{iK} = X_{iK}^2/2$, ξ_{iK} may be represented as $\xi_{iK} = X_{iK}^2/(2\sigma_i^2)$.

The probability density function of X_{iK}, the linearly detected output of the ith resolution cell on the Kth transmission, is assumed to be Rayleigh distributed, i.e.,

$$f(X_{iK})dX_{iK} = \frac{X_{iK}}{\sigma_i^2} \exp\left\{\frac{-X_{iK}^2}{2\sigma_i^2}\right\} dX_{iK} \tag{55}$$

The probability density function of ξ_{iK} becomes

$$f(\xi_{iK})d\xi_{iK} = \exp\{-\xi_{iK}\}\, d\xi_{iK} \tag{56}$$

The probability density function of

$$l = \sum_{i=1}^{N_C} \sum_{K=N_1+1}^{J} \xi_{iK} \tag{57}$$

is first developed. Since for the assumed clutter model, the ξ_{iK} are statistically independent, the Laplace transform

$$L(f(l)) = \left(\frac{1}{J+1}\right)^{N_C(J-N_1)}. \tag{58}$$

Taking the inverse transform, we obtain

$$f(l)dl = \frac{1}{(N_C(J-N_1)-1)!}\, l^{N_C(J-N_1)-1} e^{-l} dl \tag{59}$$

Introducing the change of variable $\beta_J = Cl$, Equation (22) is obtained;

$$f(\beta_J|J)d\beta_J = \frac{1}{(N_C(J-N_1)-1)!}\left(\frac{\beta_J}{C}\right)^{N_C(J-N_1)-1} \exp\left\{\frac{-\beta_J}{C}\right\}\frac{d\beta_J}{C}. \tag{60}$$

Appendix III—Development of the Probability Density Function $f(\phi_o|\phi_o = \mathrm{MAX}\,\phi_i)d\phi_o$

The estimate of the ratio of the variance of the output of the zeroth cell for the clutter plus thermal noise case to the variance of the cell output for thermal noise alone (ϕ_o) is expressed as

$$\phi_o = \frac{2}{N_C N_1 N_o} \sum_{i=1}^{N_C} \sum_{K=1}^{N_1} \frac{Y_{iK}}{W_i} \text{ (Equation (16))}.$$

It is an estimate made after the first sequential step employing the square-law-detected outputs Y_{iK}, the ith resolution cell, and the

Kth transmission, and where the outputs of the N_1 first sequential step transmissions of each of the N_C resolution cells surrounding the cell under test are employed. The zeroth indexed cell is assumed to contain the maximum clutter level, i.e., $\phi_o = \text{MAX}\ \{\phi_i\}$.

The term ϕ_o of Equation (16) may be expressed as

$$\phi_o = \frac{\underline{\phi_o}}{N_C N_1} \sum_{i=1}^{N_C} \sum_{K=1}^{N_1} \xi_{iK} . \tag{61}$$

where

$$\underline{\phi_o} = \frac{2\sigma_o^2}{N_o} = \frac{E\{|MF(\tau_o, f)|^2\}}{N_o} \text{ (Equation (7))}. \tag{62}$$

$\underline{\phi_o}$ is the 'true' value of the ratio of the output variances for the clutter plus thermal noise and thermal noise alone case, N_o is the thermal noise power density, and $\xi_{iK} = Z_{iK}/\sigma_i^2$.

The probability density function of

$$l = \sum_{i=1}^{N_C} \sum_{K=1}^{N_1} \xi_{iK}$$

is shown in a manner similar to that employed in Appendix II to be

$$f(l)\,dl = \frac{1}{(N_C N_1 - 1)!}\, l^{N_C N_1 - 1} e^{-l} dl \ ; \tag{63}$$

employing the change of variable

$$\phi_o = \frac{\underline{\phi_o}}{N_C N_1}\, l \tag{64}$$

Equation (17) results:

$$f(\phi_o | \phi_o = \phi_{MAX})\, d\phi_o = \frac{\left(\dfrac{N_C N_1}{\underline{\phi_o}}\right)^{N_C N_1}}{(N_C N_1 - 1)!}\, \phi_o^{N_C N_1 - 1} \exp\left\{ - \frac{N_C N_1 \phi_o}{\underline{\phi_o}} \right\} d\phi_o . \tag{65}$$

A SYSTEMS LOOK AT SATELLITE-BORNE HIGH-RESOLUTION RADAR

By

J. S. Greenberg

Summary—*In recent years, advances in radar-data-processing techniques have made possible high-resolution radars of the synthetic-aperture type that are capable of achieving angular resolution comparable to conventional systems employing much shorter wavelengths. The synthetic-aperture radar can, using coherent data-processing techniques, obtain angular resolution much finer than the angular dimension of the radiated beam itself. This resolution capability, along with the inherent all-weather and all-time-of-day capability of active UHF systems, makes radar an attractive candidate for performing surveillance of terrestrial targets from space. This paper briefly evaluates the performance capabilities of satellite-borne high-resolution radar when used for terrestrial surveillance. The evaluation is from a systems point of view. Performance capabilities and limitations are discussed and limiting constraints determined.*

Introduction

SATELLITE-BORNE reconnaissance systems that operate in the visible portion of the spectrum and rely upon reflected solar radiation are limited to surveillance of terrestrial targets during daylight hours. Their surveillance capabilities, because of the short wavelengths utilized, are further limited by cloud cover. Radar systems operating in the UHF portion of the spectrum could eliminate these undesirable restrictions. However, because of the relatively long wavelengths and moderate antenna size employed by conventional radar systems, it has not been possible to achieve the desired resolution capability.

In recent years, advances in radar-data-processing techniques have made possible high-resolution radars of the synthetic-aperture type that are capable of achieving angular resolution comparable to conventional systems employing much shorter wavelengths. Thus, the possibility exists of achieving a high-resolution capability without forsaking the desired all-weather (and all times of day) reconnaissance capability. The synthetic-aperture radar is capable of obtaining angular resolution that is much finer than the angular dimension of the radiated beam itself. The angular resolution is achieved by synthetically creating a large linear array. The prerequisites for a synthetic-aperture

radar system are predictable radar motion with respect to the targets to be observed, coherent radar techniques (i.e., the receiver makes use of a reference signal from which the transmitted signal was derived), and a data-processing system that can efficiently process all of the collected coherent data.

The following discussion briefly evaluates the performance capabilities of satellite-borne high-resolution radar when used for terrestrial surveillance. The approach is from a systems point of view. Performance capabilities and limitations are discussed and limiting constraints determined.

Radar System

The satellite-borne high-resolution radar system under consideration is of the side-looking or synthetic-aperture type.[1-5] The synthetic array radar consists basically of a coherent radar whose antenna (hereafter referred to as the real aperture) is in motion with respect to the terrestrial targets under surveillance. Thus the real aperture becomes a new element of the synthetic array with each succeeding transmission. Data is collected by the real aperture as it traverses a known flight path and is stored with both amplitude and phase preserved. At the end of the time required to traverse the synthetic aperture, the data is summed vectorially with the result that a narrow beam, which is substantially equivalent to that achievable with a large linear array, is achieved. Thus, by using synthetic array techniques, a single small antenna can have an effective beam width as narrow as that possible with a large linear array. Range resolution is obtained by transmitting wide-band signals.

The angular resolution attainable with the radar system is a function of the length of the synthetic aperture and is determined from a consideration of the received target signal history. Since relatively long apertures are synthesized, focused systems must be considered.*

[1] C. W. Sherwin, J. P. Ruina, and R. D. Rawcliffe, "Some Early Developments in Synthetic Aperture Radar Systems," *IRE Trans. Military Electronics*, Vol. MIL-6, April 1962.

[2] H. L. McCord, "The Equivalence Among Three Approaches to Deriving Synthetic Array Patterns," *IRE Trans. Military Electronics*, Vol. MIL-6, April 1962.

[3] L. J. Cutrona and G. O. Hall, "A Comparison of Techniques for Achieving Fine Azimuth Resolution," *IRE Trans. on Military Electronics*, Vol. MIL-6, April 1962.

[4] R. C. Heimiller, "Theory and Evaluation of Gain Patterns of Synthetic Arrays," *IRE Trans. on Military Electronics*, Vol. MIL-6, April 1962.

[5] W. G. Hoefer, "Optical Processing of Simulated IF Pulse-Doppler Signals," *IRE Trans. Military Electronics*, Vol. MIL-6, April 1962.

* Targets are in the near field of the synthetic aperture.

The system geometry is shown in Figure 1. Each target or point observed by the real aperture as it traverses the distance d_{as} exhibits a particular doppler history that is known a priori (because of the fixed system geometry). The target signal characteristics are illustrated in Figure 2 for three targets, two of which are at the same distance from the satellite ground track but displaced from each other along the

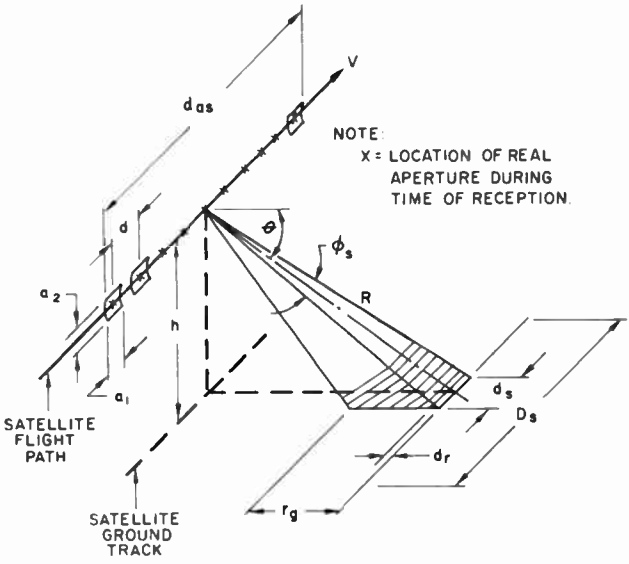


Fig. 1—System geometry.

ground track. Two of the three targets are displaced from each other in a direction perpendicular to the satellite ground track. It should be noted that targets at the same range from the ground track exhibit the same doppler histories, but delayed in time. It should also be noted that the slope of the doppler history curve is a function of the distance from the ground track. The angular resolution capability of the synthetic aperture can be determined from a consideration of the doppler history of two targets that are at the same distance from the satellite ground track but displaced in the direction of the ground track. Referring to Figure 2,

$$f_{d_1} = \frac{2v^2}{\lambda R} t , \tag{1}$$

$$f_{d_2} = \frac{2v^2}{\lambda R} t - \frac{2v^2}{\lambda R} (t_2 - t_1) . \tag{2}$$

To resolve the two targets, the bandwidth associated with each target (taken to be twice the reciprocal of the observation time, $2/T_{obs}$) must be less than the doppler frequency difference. Therefore,

$$|f_{d_1} - f_{d_2}| \geq \frac{2}{T_{obs}} , \tag{3}$$

and

$$\frac{2v^2}{\lambda R} (t_2 - t_1) \geq \frac{2}{T_{obs}} .$$

But

$v\,(t_2 - t_1)$ = separation of targets or resolution capability = d_s'

and

$$v\,T_{obs} = d_{as}$$

Therefore, the angular resolution capability of the system, ϕ_s, is seen to be

$$\phi_s|_{\text{null-to-null}} = \frac{d_s'}{R} = \frac{\lambda}{d_{as}} , \tag{4a}$$

or

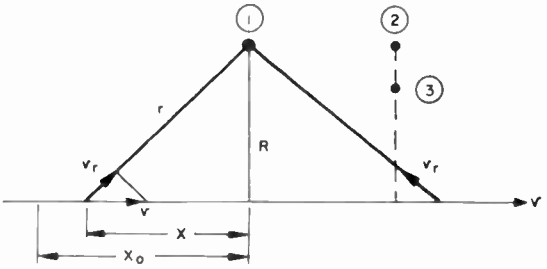
$$\phi_s|_{-4\text{db}} = \frac{d_s}{R} = \frac{\lambda}{2d_{as}} . \tag{4b}$$

Throughout the following pages, the —4-db beamwidth (Equation (4b)) is referred to as the resolution capability. d_s' corresponds to the null-to-null ground resolution and d_s to the —4-db ground resolution ($d_s' = 2d_s$).

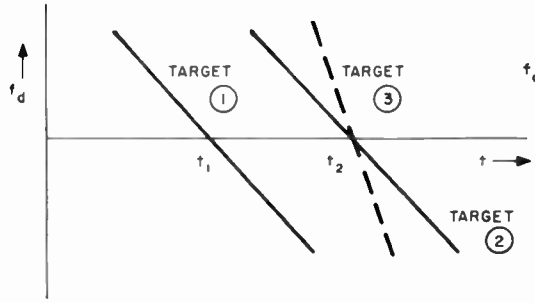
It can be seen that the two-way power gain of a synthetic array of length d_{as} is equivalent to the one-way power gain of a real antenna of length $2d_{as}$. The ground resolution, d_s, is

$$d_s = \frac{\lambda R}{2\mathrm{d}_{as}} = \frac{\lambda h}{2d_{as}\sin\theta} \tag{5}$$

where h is the satellite orbit altitude.

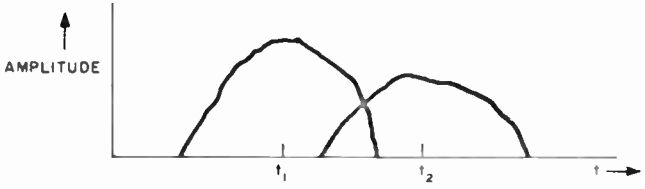


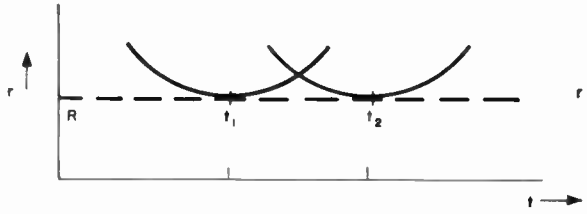
v = SATELLITE VELOCITY
R = MINIMUM SLANT RANGE
X_0 = REFERENCE POINT
f_d = DOPPLER FREQUENCY
$r(t)$ = SLANT RANGE



$$f_d = \frac{-2v^2}{\lambda R}\left(t - \frac{X_0}{v}\right)$$

(WHEN $X << R$)





$$r(t) = R + \frac{1}{2}\frac{v^2}{R}\left(\frac{X_0}{v} - t\right)^2$$

Fig. 2—Target signal characteristics.

The distance, d, that the real aperture moves during an interpulse period is equal to the satellite orbital velocity divided by the radar pulse repetition frequency (PRF). Since the orbital velocity is quite large, d can become appreciable and can be many times larger than a_1, the real antenna aperture dimension in the direction of the satellite velocity vector. This results in a discontinuous aperture that intro-

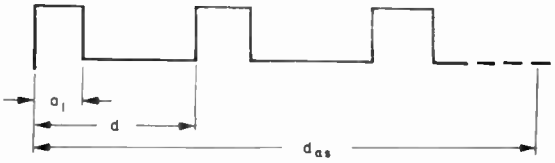


Fig. 3a—Antenna illumination function.

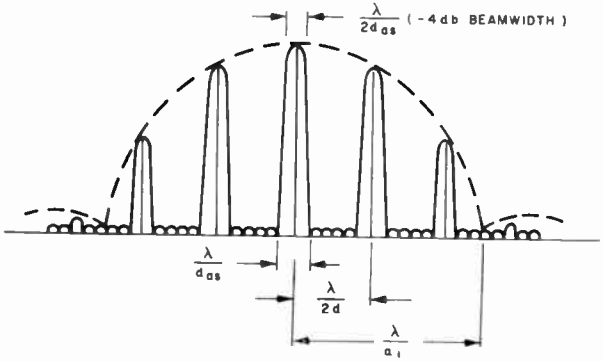


Fig. 3b—Intensity pattern.

duces angular ambiguities. This is illustrated as follows. Let the illumination function be as shown in Figure 3(a), where a_1 is the real aperture dimension along the flight path and d is the distance the real aperture moves during an interpulse period.* The intensity pattern resulting from this illumination pattern is shown in Figure 3(b). The factor of one half (in $\lambda/2d$) results from the two-way phase shifts of the synthetic antenna pattern. Therefore, to eliminate the angle ambiguities, $d - a_1$ must be decreased (by increasing the PRF or increasing a_1) to the point where the first major side lobe of the synthetic

* It is assumed that the real aperture transmits and receives in the same position and then jumps to a position a distance d away where $d = v/\text{PRF}$.

aperture falls outside the main lobe of the real aperture. In the absence of a true antenna pattern, it is assumed that the major lobes of interest of the real aperture extend an angular dimension of approximately two beam widths. Therefore

$$\frac{\lambda}{2d} \geq \frac{2\lambda}{a_1}$$

and

$$a_1 \geq 4d = 4\left(\frac{v}{\text{PRF}}\right). \tag{6}$$

Therefore, the PRF necessary to eliminate angular ambiguities is given by

$$\text{PRF} \geq \frac{4v}{a_1}. \tag{7}$$

The unambiguous ground range, r_g, resulting from the elimination of angular ambiguities is given by

$$r_g \leq \frac{a_1 c}{8v \cos\theta}, \tag{8}$$

where c is the velocity of light. Equation (8) is illustrated in Figure 4. It can be seen that the unambiguous ground range is relatively small when the real aperture is on the order of 10 feet. The ground range, r_g, that may be observed from the satellite is also a function of the real antenna pattern in the vertical dimension and the possibility of "second-time-around" echoes. Consideration of these factors may tend to reduce r_g from the value given in Equation (8).

The restriction (Equation (8)) imposed upon r_g may be eliminated if successive transmissions occur at different frequencies. For example, if two frequencies are employed (i.e., consecutive transmitted pulses employ different carrier frequencies), the available ground coverage becomes $2r_g$. The frequencies employed must be separated by at least several times the system i-f bandwidth. The burden for decoding must be borne by the receiving system (for example, multiple stable reference signals) and data processor. It should be mentioned here, and will be discussed again later, that the available ground coverage determines the size of the real aperture (dimension perpendicular to

the flight path)—larger available ground coverage means smaller apertures, which in turn means higher transmitter power.

The targets to be observed are illuminated for the length of time that it takes to synthesize or traverse a distance d_{as}. This means that the real aperture should have a beam width that is broad enough to

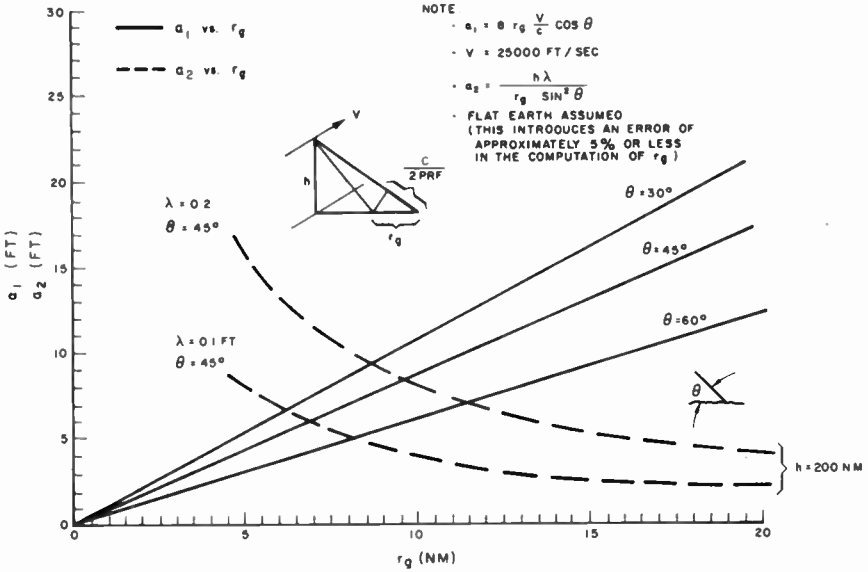


Fig. 4—Relationship of real aperture to maximum ground mapping interval (r_g)

constantly illuminate the target during the synthesis of d_{as}. Thus, the real antenna resolution capability (on the ground) should be equal to or greater than d_{as}. Therefore

$$\frac{R\lambda}{a_1} \geq d_{as}\,. \tag{9}$$

Since, from Equation (5),

$$d_{as} = \frac{\lambda R}{2d_s}\,,$$

there results

$$a_1 \leq 2d_s\,. \tag{10}$$

Equation (10) states that the resolution capability is at best one half of the real aperture dimension along the flight path.

Figure 5 illustrates the synthetic aperture requirements in terms of desired ground resolution and wavelength. The radar transmitter

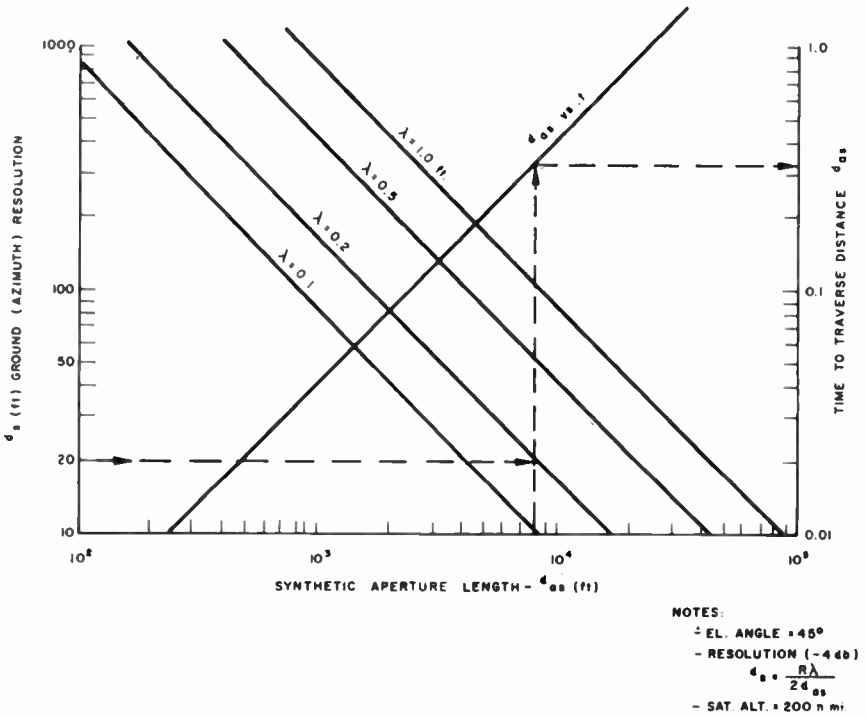


Fig. 5—Synthetic aperture in terms of desired resolution and wavelength.

power requirements may be determined as follows. The received signal power, P_r, is given by

$$P_r = \frac{P_t G_t}{4\pi R^2} \frac{\sigma}{4\pi R^2} \frac{A_r}{L}. \tag{11}$$

Since n pulses are to be integrated coherently,

$$\frac{P_t G_t}{4\pi R^2} \frac{\sigma}{4\pi R^2} \frac{A_r}{L} n = \left(\frac{S}{N}\right) KTB, \tag{11a}$$

and the average transmitter power is given by

$$P_{ave} = \frac{(4\pi)^2 R^4 \left(\frac{S}{N}\right) KTL}{G_t A_r \sigma T_{int}} . \tag{12}$$

In the above equations,

P_r = peak received signal power (watts)
P_t = peak transmitted power (watts)
P_{ave} = average transmitted power (watts)
σ = target echo area (square feet)
R = slant range (feet)
A_r = effective receiver aperture (square feet)
L = system loss factor
(S/N) = integrated signal to noise ratio
K = Boltzmann's constant
T = effective receiver temperature (°K)
G_t = transmitting antenna gain
T_{int} = integration time or the time for the satellite to achieve a synthetic aperture equal to d_{as} that will yield the desired ground resolution ($T_{int} = n/\mathrm{PRF}$)
n = number of pulses integrated
B = bandwidth.

Equating Equation (5) and $d_{as} = vT_{int}$,

$$T_{int} = \frac{h\lambda}{2vd_s \sin\theta} . \tag{13}$$

The radar transmitting antenna gain is given by

$$G_t = \frac{4\pi a_1 a_2}{\lambda^2} \delta , \tag{14}$$

where δ is the efficiency of the aperture. As stated previously a_1 is limited by the need to eliminate angular ambiguities and is given by Equations (8) and (10). The receiver aperture, A_r, is

$$A_r = a_1 a_2 \delta . \tag{15}$$

Substitution of Equations (13), (14) and (15) and the following two equations,

$$R = \frac{h}{\sin \theta} \tag{16}$$

$$a_2 = \frac{h\lambda}{r_g \sin^2 \theta} \tag{17}$$

into Equation (12) results in the following form of the radar range equation in terms of the important system parameters:

$$P_{ave} = \frac{8\pi h r_g^2 \sin \theta v d_s \left(\frac{S}{N}\right) KTL}{a_1^2 \lambda \delta^2 \sigma} . \tag{18}$$

Equation (18) is subject to the conditions that

$$2d_s \geq a_1 \geq 8r_g \frac{v}{c} \cos \theta .$$

It is important to note that the power requirements vary linearly with satellite altitude provided the system resolution capability does not vary with satellite altitude.. As discussed later, the side-looking radar resolution capability is dependent only upon the realizable value of d_{as}. The limit upon d_{as} is determined from a consideration of system errors including tropospheric and ionospheric perturbations, attitude stabilization, etc.

In order to obtain an estimate of the average transmitter power requirements, typical parameters are assumed; the average power versus target echo area is shown in Figure 6. The average power is given versus the signal-to-noise ratio, typical values of which are discussed in following paragraphs.

Equation (17), and hence Equation (18) (and the data* shown in Figure 6), imply that the transmit and receive beam widths are determined by the ground mapping interval r_g, as determined from PRF

* Occasionally, in the following pages, specific values are given for a number of the variables. These values are chosen to be typical or representative of possible values and are given only so that estimates may be made for values of the dependent variables.

requirements. This need only be true for the transmit beam width, since multiple receive beams may be employed resulting in larger receiver apertures and reduced transmitter power requirements. Therefore, if different frequencies are employed on consecutive transmissions and the transmitter gain reduced accordingly, power requirements are proportional to ζr_g, where ζ is the fractional increase in the basic ground mapping interval, r_g, which is established by PRF or sampling requirements. If multiple beam widths (on receive) are not

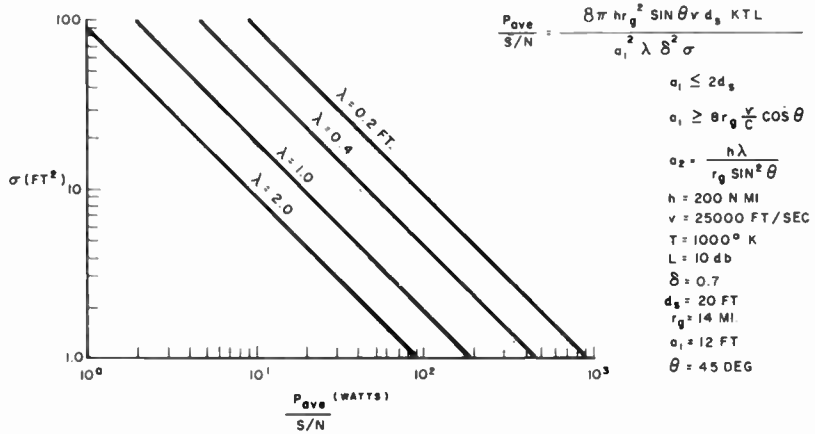


Fig. 6—Average transmitted power versus target cross section.

employed, power requirements are proportional to $(\zeta r_g)^2$, as shown in Figure 7. The average power requirements when multiple receive beams are employed can be further reduced if multiple receive beams are used to cover the basic r_g interval.

The transmitter peak power, P_t, requirements are given by

$$P_t = \frac{P_{ave}}{\tau(\mathrm{PRF})} = \left(\frac{a_1}{4v}\right)\left(\frac{c}{2d_r \cos\theta}\right) P_{ave} = \frac{5000\, a_1}{d_r \cos\theta} P_{ave}, \qquad (19)$$

where τ is the effective pulse width and d_r is the range resolution (perpendicular to the satellite ground track).

When typical parameters such as $a_1 = 12$ feet, $d_r = 20$ feet, and $\theta = 45°$ are substituted in Equation (19), $P_t = 4200\, P_{ave}$. Obviously pulse compression, possibly of the "CHIRP" type, is desirable. When the pulse compression ratio is 100, $P_t = 42\, P_{ave}$, which is much more realistic. When pulse compression techniques are used, r_g, the map-

pable ground range (or unambiguous ground range), is reduced by twice the compression ratio multiplied by the compressed pulse width, for the case illustrated, this amounts to a reduction in r_g of approximately 2000 feet, which may be considered negligible.

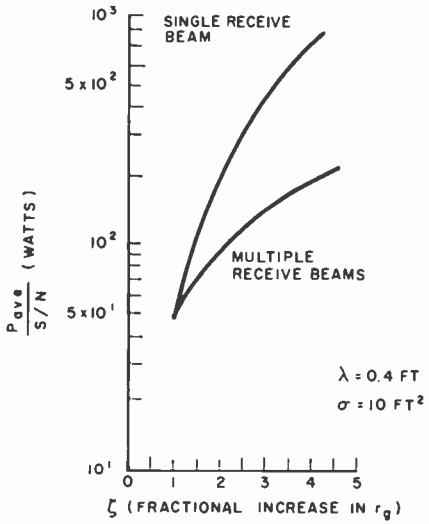


Fig. 7—Average power requirements in terms of mappable interval.

Signal-to-noise ratio has significance only when the clutter level is low. Such is not necessarily the case in a side-looking radar that observes the surface of the earth. Here, undesired terrain echoes may be significantly larger than the echoes from targets of interest. The general terrain return parameter γ for many types of terrain, incidence angle, and frequency, has been measured. γ is related to the radar cross section by*

$$\sigma = \gamma A_i , \tag{20}$$

where

σ = radar cross section

γ = radar cross section per unit area of terrain

A_i = incidence area, i.e., the area perpendicular to the line of sight through which passes all the energy contributing to the return at a given range and time.

* Some authors prefer to characterize the terrain by the parameter σ_o, equal to σ divided by the area of the terrain illuminated. The relation between σ_o and γ is given by $\sigma_o = \gamma \sin \theta$.

For the usual case of pulsed radars,

$$A_i = \frac{c\tau}{2} R\phi_s \tan\theta \,. \tag{21}$$

The terrain influence is thus isolated in the single factor γ. Some typical values[6] of γ are shown in Figures 8, 9, and 10. Values of γ vary

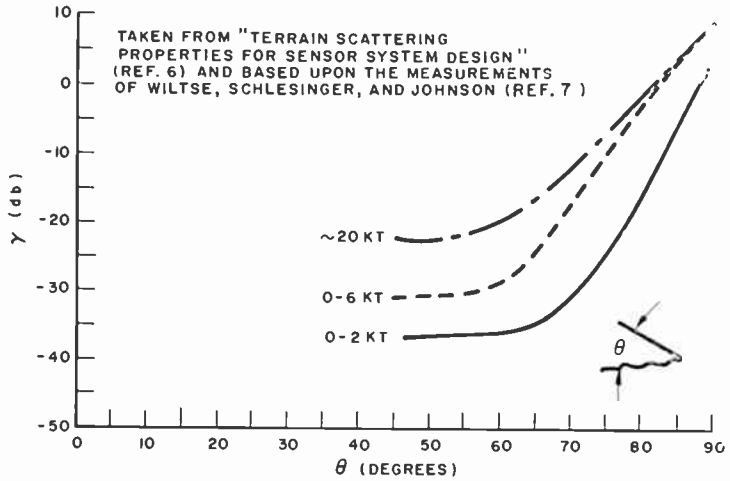


Fig. 8—Sea return at X-band.

from an upper limit of about 0.5 for very rough surfaces to less than 0.00005 for very smooth surfaces. The magnitude of γ generally increases with a corresponding increase in incidence angle, but as the surface becomes rough (in terms of wavelength), γ becomes independent of incidence angle. For a smooth surface such as a concrete road, the increase in γ from $\theta = 10°$ to $\theta = 80°$ may be as large as 25 db, while for a vegetation-covered surface there may be no noticeable change in γ. When the terrain has a heavy vegetation cover, little difference is noted in the value of γ for vertical and horizontal polarization. For smooth terrain there is a definite dependence of γ upon polarization, especially at grazing angles. The magnitude of γ from smooth terrain using horizontal polarization may be as much as 15 db

[6] R. L. Cosgriff, W. H. Peake, and R. C. Taylor, "Terrain Scattering Properties for Sensor System Design (Terrain Handbook II)," Engineering Experiment Station Bulletin 181, The Ohio State University, May 1960.

[7] J. C. Wiltse, S. P. Schlesinger, and C. M. Johnson, "Back-Scattering Characteristics of the Sea in the Region from 10 to 50 KMc," *Proc. IRE*, Vol. 45, Page 220, 1957.

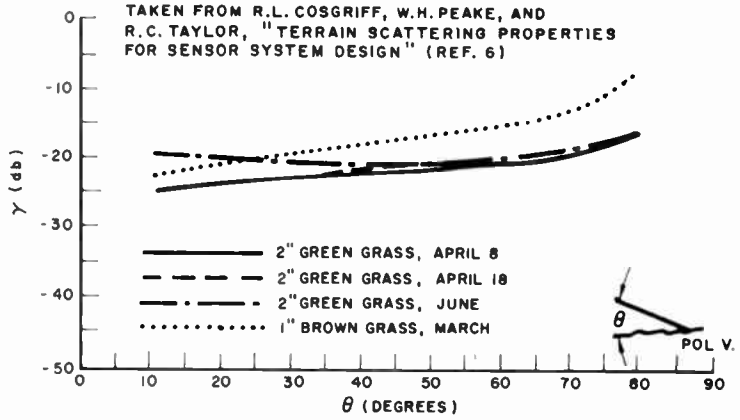


Fig. 9—Seasonal changes at grass at X-band.

lower than the return from the same surface using vertical polarization. Water on a relatively smooth surface such as a road tends to lower the return as much as 3 db.*

Signal-to-noise ratio has significance when considering the effect of receiver noise upon false-alarm rate and probability of detection. Signal-to-noise ratio is used to establish the power requirements in terms of acceptable receiver noise. The tolerable false alarm rate and probability of detection are functions of the number of resolution elements or decision elements subtended by the targets of interest. Since the side-looking radar observes the surface of the earth, the effect of spurious returns from targets that are of little or no interest must

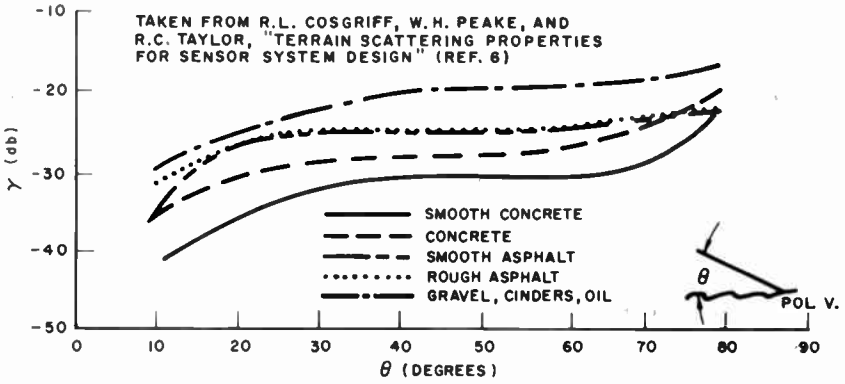


Fig. 10—Contrast of γ for various smooth surfaces at X-band.

* The above discussion of terrain echo area is based on Reference (6).

be determined—that is, the clutter level must be considered. The clutter background is a function of the class of targets of interest. For example, if it is desired to map the surface of the earth with all of its man-made as well as natural features, the set of targets is infinite and the set of clutter is null. If it is desired to recognize ships at sea, then the set of targets is small (i.e., the ships) and the set of the clutter is also small (i.e., the sea returns). In the former case, signal-to-noise ratio is important; in the latter, both signal-to-noise and signal-to-clutter (or contrast) ratios are important. The signal-to-noise ratio determines the minimum discernable signal and the signal-to-clutter ratio determines the minimum discernable difference that can be detected between target and background. In the latter example, the echo area statistics of sea return must be different from that of ships in order to detect ships no matter how high the signal-to-noise ratio.

From Figures 9 and 10 it can be seen that, typically, contrast ratios of from 5 to 10 db between smooth concrete roads and grass are to be expected (at X-band). Therefore, it may be concluded that smooth concrete roads that subtend many resolution elements, e.g., when passing through fields, will be easily discernable. On the other hand, a concrete slab that does not subtend many resolution elements may not be readily discernable. Thus, shape as well as echo area characteristics, are important.

A reasonable value of signal-to-noise ratio may be determined as follows. The number of resolution elements per range sweep (r_g/d_r) is on the order of 5000. If it is assumed that the false-alarm rate should be small per range sweep, then p_{fa}, the probability of a false alarm, should be on the order of 10^{-4} to 10^{-5}. It is usually desired to detect a given class of targets with a high degree of confidence. Therefore, it seems reasonable to assume that for targets having specified echo area characteristics, the probability of detection, p_d, should be on the order of 0.8 to 0.9, or perhaps greater. Therefore, S/N will assume values [7(a)] on the order of 10-12 db.

Data Processing*

A simplified block diagram of the overall system is shown in Figure 11, where the coherent radar operation is as described in the previous paragraphs. The data collected by the radar may be processed in the

[7(a)] W. Hall, "Prediction of Pulse Radar Performance," *Proc. IRE*, Vol. 44; Feb. 1956.

* If an optical system is used, a doppler signal offset from zero frequency is required. This produces image lobes in the synthetic antenna pattern, necessitating a somewhat higher PRF than given in Equation (6).

satellite and the final, or processed, information transmitted back to ground terminals, or the raw unprocessed data may be transmitted to ground terminals where the processing is then accomplished. Figure 11 illustrates the latter approach. Thus data from the radar is to be recorded and stored in the satellite and transmitted to the ground read-out terminals during the time that the satellite is observable by the ground terminals. The recorder must have sufficient bandwidth to record the raw radar data, but, as discussed later, the communica-

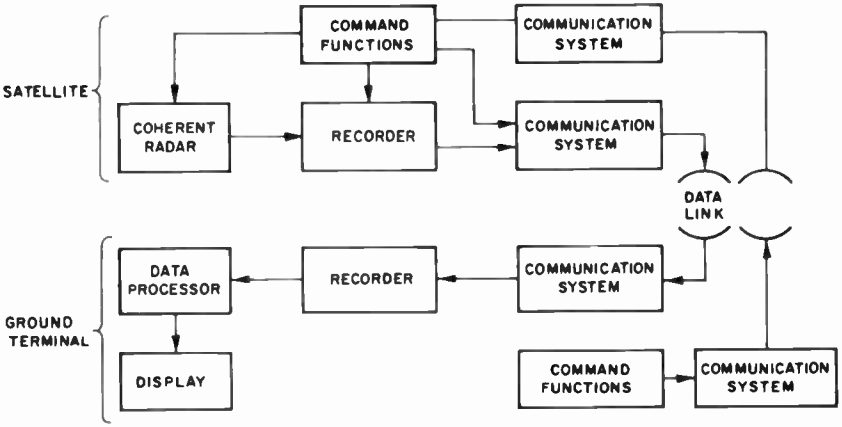


Fig. 11—System block diagram.

tion system data bandwidth need not be the same as the radar bandwidth. A communication link from ground to satellite is required so that the radar may be pointed in the proper direction at the proper time.

The processing of the data of a side-looking radar is usually accomplished with optical recording and processing systems.[5, 8-10] In a pulse-doppler radar it is desired to obtain the frequency spectrum of the return from a large number of range elements independently. A simplified diagram of an optical processing system that accomplishes multi-channel processing simultaneously is shown in Figure 12.* Co-

[8] L. J. Cutrona, E. N. Leith, C. J. Palermo, and L. J. Porcello, "Optical Data Processing and Filtering Systems," *IRE Trans. Information Theory*, June 1960.

[9] L. J. Cutrona, "Optical Computing Techniques," *IEEE Spectrum*, Oct. 1964.

[10] L. J. Cutrona, E. N. Leith, L. J. Porcello, and W. E. Vivian, "On the Application of Coherent Optical Processing Techniques to Synthetic-Aperture Radar," *Proc. IEEE*, Vol. 54, No. 8, Aug. 1966.

* The following discussion of the optical data processing is based upon Reference (5).

herent video is recorded transversely on a slowly moving film. Successive radar range sweeps lie side by side. Amplitude or density modulation is assumed. Returns from a single range element fall on a vertical line indicated by the row of dots. This assumption is valid if the range resolution element of the radar is several times larger than the variation in range to a given target as the synthetic aperture is traversed (see Figure 2 as an example, when $d_{as} = 2$ miles† and $h = 200$ miles, the maximum variation of range to a given target is 15 feet). The

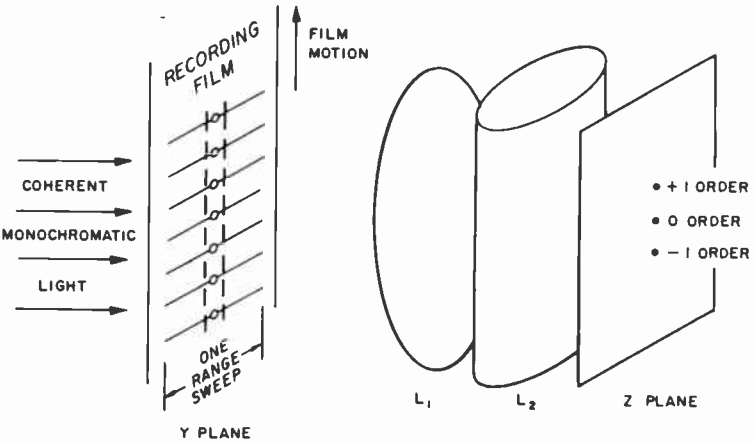


Fig. 12—Simplified pulse-doppler processing system.

film is illuminated by coherent, monochromatic light, and is located in the focal plane of the spherical lens L_1. The Fourier transform of the function recorded on the film then appears in the other focal plane of L_1. The cylindrical lens L_2 is required to focus the Y-plane upon the Z-plane in the horizontal direction so that the different range elements will be processed independently. The zero-order pattern,‡ due to the biasing (because negative amplitudes cannot be recorded) of the optical recording, is independent of doppler frequency, whereas the principal maxima of the first-order patterns move outward from the center as doppler frequency increases. Thus, the pertinent mapping information is in the +1 order terms, and the 0 and −1 order terms may be disregarded.

For targets in the near field, or Fresnel zone, of the synthetic aper-

† Throughout this paper, distances given in miles represent nautical miles.

‡ The zero- and first-order patterns correspond to the principal maxima of the Fourier transform of a biased sinusoid pulse.

ture, doppler frequency of the signal varies linearly with time (Figure 2). The Fresnel-integral transform of the linear FM signal also results in zero-order and positive and negative first-order terms. The useful information is contained in the positive first-order term; its focal length is a measure of the rate of change of doppler frequency, and its position on the Z-axis is a measure of the initial value of the doppler frequency.

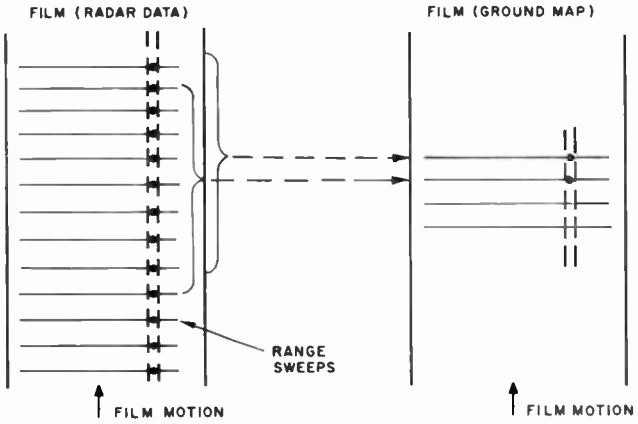


Fig. 13—Radar data to ground map transformation.

The optical data processor therefore collapses (coherently) all of the information obtained from n range sweeps to a single sweep on the final "map." Each range resolution element is treated independently, as illustrated in Figure 13. For a detailed discussion of optical processing see References (5), (8), (9) and (10).

Data Communications

The data collected by the side-looking radar may be processed in the satellite or at a ground terminal. In either case the data must be transferred from satellite to earth. It is desirable to accomplish this without physically recovering the satellite. The bandwidth required for a communication link to transfer the radar data from satellite to ground is a function of many things; for example, location of data processor (space or ground), radar beam width, read-out time, etc.

Consider first the communication system bandwidth when data processing is accomplished in the satellite, i.e., when processed data is to be communicated. The data consists of a map or format having dimen-

sions D_s along the ground track and D_r across the ground track. This information must be stored in the satellite until the satellite is in view of a ground read out terminal. When the resolution capability of the system is d_s along the ground track and d_r across the ground track, the total number of resolution elements, N, for which data is to be transmitted is given by

$$N = \frac{D_s D_r}{d_s d_r}. \tag{22}$$

The time, t, available to transmit the picture format is equal to the time per resolution element multiplied by the number of resolution elements per format. The data bandwidth is approximately equal to the reciprocal of the time per resolution element of the format. The format data may be used to modulate an r-f carrier. The bandwidth (B_{ds}) required by the video signal or data is

$$B_{ds} = \frac{1}{t} \frac{D_s D_r}{d_s d_r}. \tag{23}$$

D_r is equal to the cross track ground range mapped by the radar and is equal to ζr_g. When $\zeta = 1$, D_r is equal to r_g and is given by Equation (8). Substitution of Equation (8) into Equation (23) yields

$$B_{ds} \leq \frac{1}{8t} \frac{D_s a_1}{d_s d_r} \frac{c}{v \cos \theta}, \tag{24}$$

where

$$d_s \geq a_1/2 .$$

When $D_s = 100$ miles, $a_1 = 10$ feet, $d_s = 20$ feet, $d_r = 20$ feet and $\theta = 45°$, $B_{ds} \doteq 103/t$ MHz, where t is in seconds.

When the data processing is performed on the ground, the radar data must be stored in the satellite until it can be read out to a ground terminal. To obtain a single meaningful line on the ground map, the side-looking radar must traverse a distance d_{as} (Figure 13). Thereafter, there is a one-to-one correspondence between radar range sweeps and ground map traces. Therefore to obtain a ground map D_s long, the satellite must move a distance equal to approximately $D_s + d_{as}$. The communication system data bandwidth B_{dg}, when processing is performed on the ground is given by

$$B_{dg} = \frac{\text{time to obtain data}}{\text{time to transmit data}} \times \text{radar bandwidth}, \tag{25}$$

and

$$\begin{aligned} &\text{time to obtain data}^* = (D_s + d_{as})/v \\ &\text{time to transmit data} = t, \\ &\text{radar bandwidth} \doteq 1/\tau. \end{aligned} \tag{25A}$$

The radar pulse width τ is given by

$$\tau = \frac{2d_r \cos\theta}{c}. \tag{26}$$

Substitution of Equations (25A) and (26) into Equation (25) yields

$$B_{dg} = \frac{(D_s + d_{as})c}{2vtd_r \cos\theta}. \tag{27}$$

Substitution of Equation (5) for d_{as} into Equation (27) yields

$$B_{dg} = \frac{\left(D_s + \dfrac{h\lambda}{2d_s \sin\theta}\right)c}{2vtd_r \cos\theta}. \tag{28}$$

When $D_s \gg d_{as}$ Equation (28) reduces to

$$B_{dg} = \frac{D_s c}{2vtd_r \cos\theta}, \tag{28A}$$

and

$$\frac{B_{dg}}{B_{ds}} = \frac{\dfrac{D_s c}{2vtd_r \cos\theta}}{\dfrac{1}{8t}\,\dfrac{D_s a_1}{d_s d_r}\,\dfrac{c}{v\cos\theta}}.$$

* It is assumed that the total time between radar pulses is used to obtain data.

This reduces to

$$B_{dg} = 4 \frac{d_s}{a_1} B_{ds} . \tag{29}$$

When the numbers of the previous example are used, $B_{dg} = 8\, B_{ds} = 825/t$ MHz, where t is in seconds.

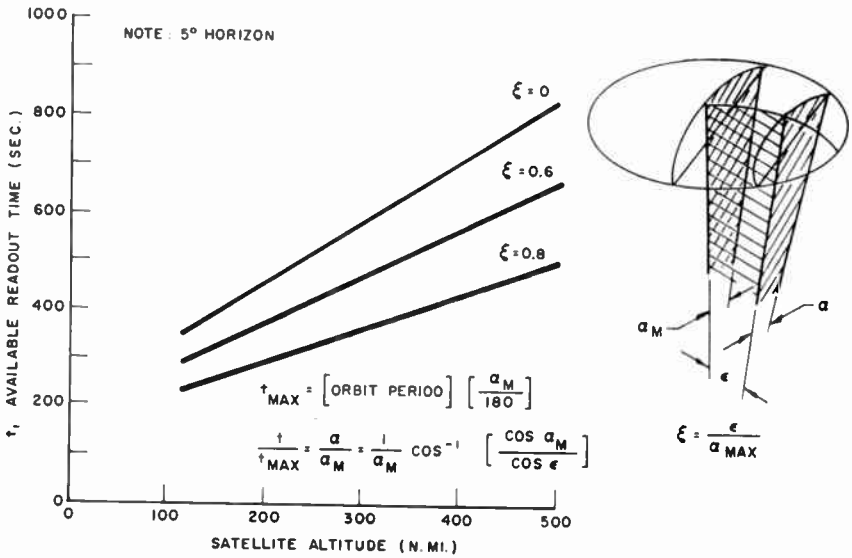


Fig. 14—Time available for data transmission.

The time, t, available for data transmission from satellite to ground is shown in Figure 14 as a function of satellite altitude and angular distance, ϵ, of closest approach of ground track to the read-out location. α_M is the maximum angle (as seen from the center of the earth) of visibility of the satellite from the ground read-out station. It can be seen that, typically, several hundred seconds are available for data transmission without any appreciable loss in coverage available to the read-out terminals.

The effect of transmission time and bandwidth upon the amount of data per satellite passage that can be read out through the ground read-out terminal coverage can best be illustrated by an example. Let the data bandwidth be 25 MHz (limited by recorder read-out) and let the time available for communication be 100 sec. Then, from Equation (28A),

$$\frac{D_s}{d_r} = 0.9 \times 10^5 .$$

When $d_r = 20$ feet, D_s is approximately 300 miles. The above applies to the case where ζ is unity and the mappable swath width is approximately 14 miles.

Power requirements for the communication link from satellite to ground may be estimated as follows:

$$P_t = \frac{(C/N)AKT'BL}{G_t G_r}, \tag{30}$$

where

P_t = transmitter power

C/N = carrier-to-noise ratio = 20 db

A = path loss = $(4\pi R/\lambda)^2 \leq 170$ db (when $h = 200$ miles and $\lambda = 1/5$ foot)

$K = 1.38 \times 10^{-23}$ joules/deg Kelvin = -228.6 db

T' = effective receiver temperature = 20 db

B = receiver noise bandwidth

L = system losses = 10 db

G_t = satellite transmitter antenna gain

G_r = ground terminal (20-foot dish) receiving antenna gain = 48 db.

Using the above typical parameters,

$$P_t(\text{db}) = B(\text{db}) - G_t(\text{db}) - 56.6$$

When the receiver noise bandwidth is about 50 MHz,

$$P_t(\text{db}) = 20.4 - G_t(\text{db}) .$$

Since the satellite will be stabilized and directional references will undoubtedly be available, directional antennas may be used to reduce the transmitter power requirements. If a 20-db antenna (approximately one foot in diameter having a beam width of about 14 degrees) is employed in the satellite, transmitted power is on the order of a few watts.

Coverage

In discussion of coverage capability of the side-looking radar satellite, two terms must be defined—the mappable ground coverage and

the field of view. The mappable ground coverage is defined as the area that is actually mapped (the swath) by the side-looking radar and that has a dimension perpendicular to the flight path of ζr_g, where r_g is determined by the antenna aperture and PRF and is given by Equation (8) and Figure 4. The field of view is defined as that area of the earth observable from the satellite and limited theoretically by line of sight considerations. There are, however, some practical considerations that limit the field of view—(1) the increase in range resolution element size as the depression, or elevation angle, θ, is increased and (2) the increase in the angle-resolution dimension and required slant-range capability as the elevation or depression angle decreases. The ground resolution along and perpendicular to the ground track, are, respectively,

$$d_s = \frac{\lambda h}{2d_{as} \sin \theta}, \tag{5}$$

and

$$d_r = \frac{c\tau}{2 \cos \theta}. \tag{31}$$

When $\theta = 60°$, the resolution *perpendicular* to the ground track has degraded to ½ its theoretical maximum; when $\theta = 30°$, the resolution *along* the ground track has degraded to ½ its theoretical maximum. Thus, if one-half theoretical maximum is used as a limiting value, the field of view from the satellite is limited to depression or elevation angles between approximately 30 and 60 degrees. Since it is possible to observe from either or both sides of the satellite, the field of view has a gap in it directly below the satellite. The mappable ground coverage can occur anywhere in the field of view, as illustrated in Figures 15 and 16 for satellites at 200 and 500 miles altitude, respectively. Slant range, ground range, and field of view capability as a function of altitude are shown in Figures 17 and 18.

From Figures 15 and 16, it can be implied that the mappable coverage (or mapped swath) can occur at any position, as desired, within the confines of the field of view. In order to fully utilize the data-gathering capability of the side-looking radar system, it must be designed with this fact in mind, i.e., the radar should be so designed that it can be pointed upon command from the ground in a given direction to observe a given target of interest. With this in mind, the capability of observing targets of interest as a function of time can be seen from Figures 15 and 16. It can be seen that a single 200-mile-altitude polar satellite has the capability of observing all targets of

interest at latitude greater than 50 degrees (except for a small gap) every 24 hours. This latitude coverage ($> 50°$) can be reduced by using 2 or more satellites in polar orbit. The number of satellites must be dictated by the time allowed or desirable between successive observa-

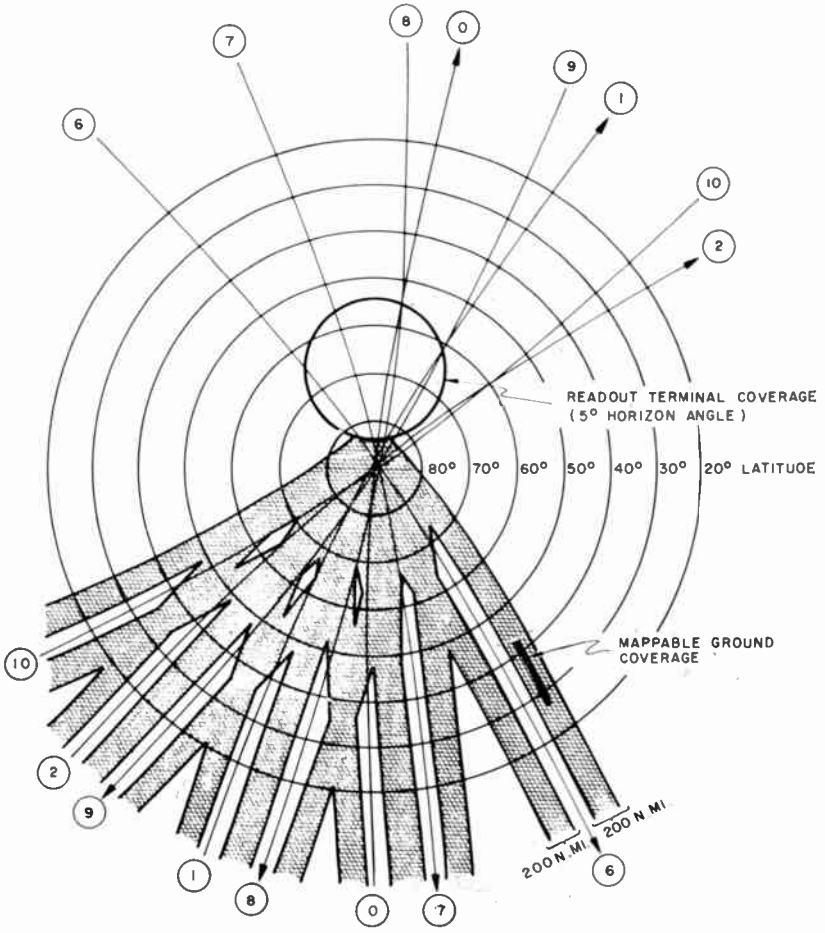


Fig. 15—Ground coverage of 200-mile altitude satellite (polar orbit) $30° < \theta < 60°$ period = 92 min.

tions of the same target and the target handling capability of each satellite. The value of going to higher altitudes can be seen by referring to Figure 16. From an altitude of 500 miles the total earth can be observed (field of view) from a single polar satellite once every 24 hours except for some small gaps at latitudes 30 degrees and 0 degrees.

JAMMING

The vulnerability of a satellite-borne side-looking radar system should be assessed in terms of electronic countermeasures (jamming), target camouflage, concealment or simulation, and space radiation effects. The following paragraphs discuss only electronic counter-

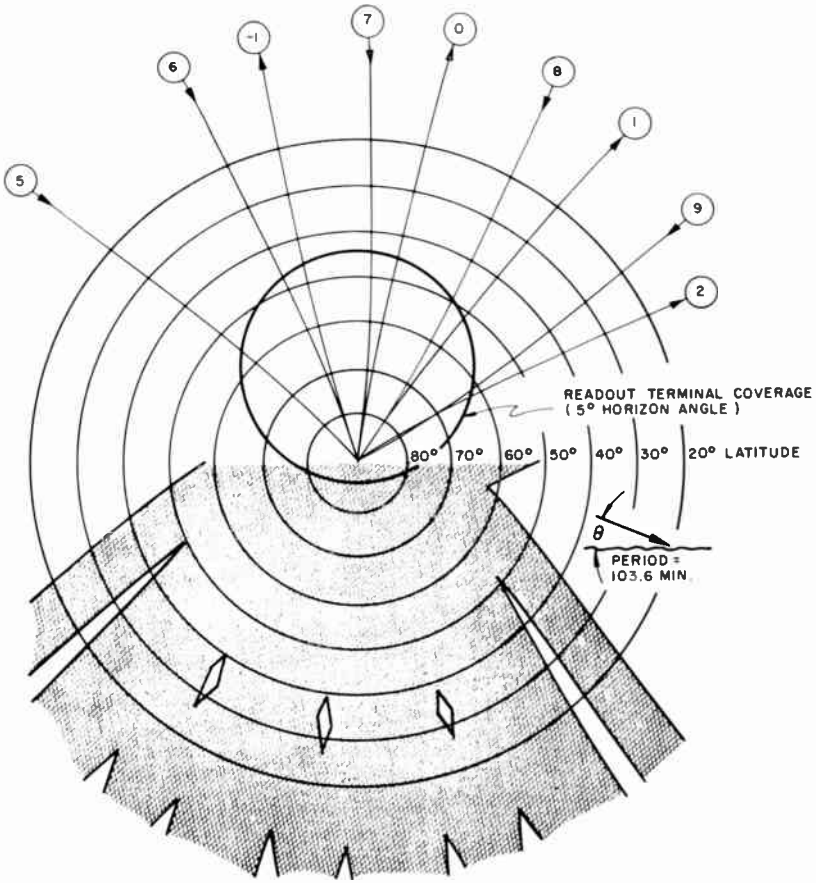


Fig. 16—Ground coverage of 500-mile altitude satellite (polar orbit) $30° < \theta < 60°$.

measures, except for the following few statements pertaining to target camouflage.

Target camouflage, concealment or simulation, is mainly a function of the targets that the system is designed to observe; for example small targets such as tanks will be relatively easy to camouflage or simulate (automobiles and tanks may appear similar to a radar that

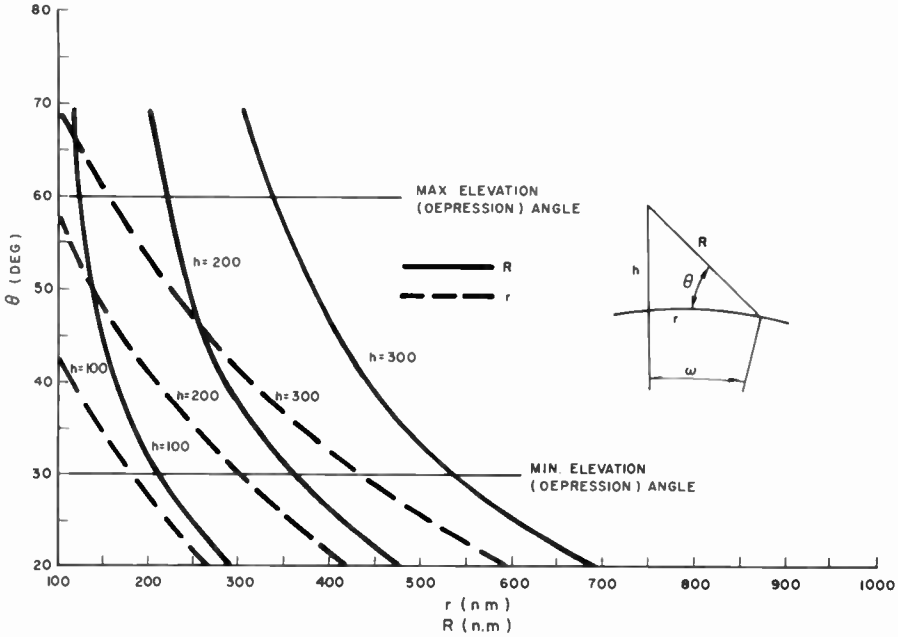


Fig. 17—Slant range and ground range as a function of satellite altitude.

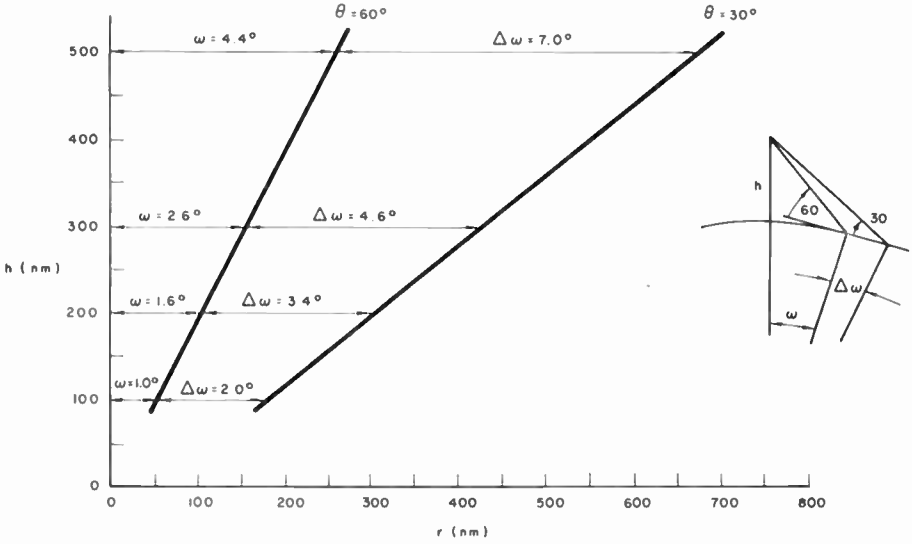


Fig. 18—Field of view as a function of altitude.

cannot resolve details of either) whereas large installations may be relatively difficult to camouflage or simulate. Chaff spread over a field may seriously confuse and degrade the radar performance.

The degree of vulnerability of the side-looking radar system to electronic countermeasures may be roughly assessed in terms of the ratio of peak received signal power to rms received jammer power (S/J). The received signal to jammer power ratio (integrated) is given by

$$\frac{S}{J} = \frac{P_{ave} A \sigma k}{P_{ave_J}\left(\frac{B_{int}}{B_J}\right) A_J 4\pi R^2}\left(\frac{R_J^2}{R^2}\right) \tag{32}$$

where

P_{ave} = average transmitter power (radar) $\doteq$ 1000 watts,

P_{ave_J} = average transmitter power (jammer),

A = satellite real aperture $\doteq$ 40 ft^2,

σ = target echo area $\doteq$ 100 ft^2,

k = ratio of maximum satellite receiver gain to gain in direction of jammer $\geqq 1$,

B_{int} = integration bandwidth and is approximately equal to the reciprocal of the time to traverse the synthetic aperture (see Figure 5) $\doteq$ 10 Hz,

B_J = bandwidth over which jammer spreads its power* $\doteq 100 \times 10^6$ Hz,

A_J = effective aperture of jammer (ft^2),

R = radar slant range $\doteq$ 300 miles,

R_J = jammer slant range.

When the above typical parameters are substituted into Equation (32),

$$\frac{S}{J} = \frac{k}{P_{ave_J} A_J}\left(\frac{R_J}{R}\right)^2. \tag{32A}$$

Equation (32A) implies that relatively modest requirements are imposed upon the jammer even when jamming the side-looking radar through side lobes that may be 40 db ($k = 10^4$) below the main beam gain, and R_J is on the order of R. Anti-jamming techniques should be

* B_J assumed equal to or greater than $2/\tau$.

incorporated where possible, but such techniques as pulse-to-pulse frequency hopping may not be very effective, since relatively wide bandwidths are already required by the system. It should be mentioned that frequency agility may not have great utility, for the specific geometry of the side-looking radar system, against a properly located repeater type jammer. Jamming of the side-looking radar seems technically feasible; however, operational factors may impose some problems upon the potential jammer. For example, if the side-looking radar system is used to spot-check a target area having small dimensions, say on the order of 20 by 20 miles, then the time required for mapping the target area is on the order of 5 seconds. Thus if the potential jammer is not "looking" in the right direction at the right time, the side-looking radar will have served its purpose prior to being jammed.

Discussion

The type of mission for which the side-looking radar may be used will be radically affected by the achievable resolution capability. The resolution capability in the direction of the satellite ground track is determined by the synthetic aperture radiation pattern. To successfully accomplish the formation of a synthetic aperture radiation pattern, a high degree of phase coherence must be maintained in the signals transmitted to and reflected from a target as the real aperture traverses the synthetic aperture. The phase coherence may be perturbed by unpredictable variations in the propagation media, unknown or uncompensated motions of the real aperture, fluctuations or variations in the receiver signals, etc. The loss of phase coherence manifests itself as beam broadening, loss in peak gain, beam canting, and an increase in side-lobe levels.

The phase coherence is disturbed by variations in the propagation delay to and from the targets caused by variations in the electron density or index of refraction in the ionosphere and by temporal and spatial variations in the density, temperature, and relative proportions of the constitutent gases and other particles in the troposphere. The propagation effects on phase coherence are both time and space dependent. Rondinelli and Zeoli[11] have considered the tropospheric effects on a relatively low altitude side-looking radar. Their calculations indicate that system performance will not be materially degraded when apertures of up to several thousand feet are synthesized at X-band fre-

[11] L. A. Rondinelli and G. W. Zeoli, "Evaluation of the Effect of Random Tropospheric Propagation Phase Errors on Synthetic Array Performance," *Eighth Annual Radar Symposium Record*, June 1962, University of Michigan.

quencies. Since tropospheric effects are generally confined to altitudes below about 15,000 feet, their results may be extrapolated to vehicles at higher altitudes.

Little data is available as to the fluctuation of the index of refraction in the ionosphere over paths separated by feet or tens of feet. Thus, the phase fluctuations across the aperture caused by spatial fluctuations in the index are not accurately known. However, estimates have been made that range from achievable resolution on the order of feet (if one is an optimist) to on the order of many tens of feet (if one is a pessimist). The implication of the above is that only relatively gross (on the order of possibly 50 feet or more) target characteristics may be distinguishable. Thus, vehicles on a road may be detected and perhaps identified as vehicles by their presence on a road (not by distinguishing an image of a vehicle). It will probably not be possible to directly determine type of vehicle (i.e., automobile, tank, truck, etc.). On the other hand it will probably be possible to distinguish between roads, buildings, ships, etc.

The satellite-borne side-looking radar is basically limited to mapping narrow strips of the earth's surface (see Figure 4). Because of PRF, data-processing, and power considerations, the mappable swath width will probably be limited to several tens of miles. The field of view from a 200-mile-altitude satellite is, however, several hundred miles ($30° < \theta < 60°$). Thus, to fully realize the maximum capability of the satellite, the radar system should be designed so that the mapped swath can occur, as commanded from the ground terminals, at any desired position within the field of view. The requirements for this may be fully appreciated by viewing Figure 15. In order to provide this capability, phased array or frequency scan antennas seem desirable (rather than physically rotating an antenna or the satellite) for positioning the mapped swath in the desired position.

Because of the inherently narrow mappable swath width, the radar reconnaissance system is at a severe disadvantage if it is required to map a large portion of the surface of the earth. To efficiently utilize the radar reconnaissance system it would seem desirable to combine it with an optical surveillance system, where the radar plays a subordinate role. That is, the radar is used when conditions are such (cloud cover and darkness) that reconnaissance in the visible, or near visible, portion of the spectrum is not possible. The radar is thus required to obtain information on those targets that can change during the time the optical sensors may be inoperative. Since the radar can only map a narrow strip, it must be told where the targets of interest are and where to point. The number of satellites required will depend upon the type of targets to be observed and the rates of change in the

characteristics of these targets. It is clear though, that as long as the radar system is limited to mapping narrow strips on the surface of the earth, accurate orbit predictions are required and accurate attitude stabilization systems are necessary (on the order of several tenths of a degree).

Side-looking radar systems may be jammed. The difficulty of jamming seems to be an operational one—the jammer must look at the correct place at the correct time at the correct frequency. Since the reconnaissance mission over a certain area may last for a few seconds, the operational problems of the jammer may be very difficult. An item of concern is the use of passive deception techniques, such as spreading chaff on the ground, using corner reflectors, etc., to confuse the radar picture. The judicious use of chaff and/or corner reflectors might seriously degrade the utility of the radar reconnaissance system.

THE SIMULATION OF TIME-DISPERSED FADING CHANNELS

By

L. W. Martinson and J. E. Courtney

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

Summary—The simulation of highly dispersive, Rayleigh fading communication channels can be achieved by combining digital, analog, and optical techniques. The basic implementation employs a digital shift-register delay line whose output taps are modulated by individual independent fading noise sources. Independent noise modulation of each tap is accomplished electro-optically using a noise film projected on a photoresistor matrix. Tests were conducted on an initial channel simulator with 196 time-dispersive elements with pseudorandom code, single sideband, and frequency-diversity modulation. Results from the simulation program followed both theoretical and experimental predictions.

Recent advances in integrated circuitry make improvements in both capacity and bandwidth of the channel simulator practical. Amplitude encoding of the input-signal samples avoids the clipping losses associated with the single-level shift register of the initial simulator implementation. Parallel multi-level shift-register delay lines are now a practical approach with high-speed digital integrated circuits. Achievable time-dispersion lengths encompass virtually all known communication channel characteristics. Noise modulation of each channel tap can be done optically or with independently derived digital noise sequences. A number of techniques already exist for generating sequences with given probability distributions.

Introduction

EVALUATION of anti-multipath communication techniques in time-dispersed, fading channels is difficult from both an experimental and analytical viewpoint. If the channel parameters can be encompassed and selectively controlled in a channel simulator, the relative performance of particular communication techniques can be accurately determined. Selective control of the channel simulator parameters is important in gathering accurate data for a variety of conditions, and the channel implementations described emphasize this capability.

Mathematical Model of a Time-Dispersed Channel

Before discussing time-dispersed channel simulation techniques,

we will briefly consider the general mathematical model of these channels. More complete analyses can be found in the literature.[1-4]

A time-variable linear filter is descriptive of the channel characteristic when additive noise effects are ignored. The filter is sufficiently general to include the multipath dispersion and fading that occurs. This model provides an expression for the received signal $z(t)$ from any transmitted signal $x(t)$ by means of the convolution integral

$$z(t) = \int_0^\infty A(\lambda,t)\, x(t-\lambda)\, d\lambda \,. \tag{1}$$

The function $A(\lambda,t)$ represents the response of the channel filter at time t to an impulse transmitted at time $(t-\lambda)$. The variable λ is a memory parameter. Assuming the channel filter to be physically realizable implies that $A(\lambda,t)$ equals zero for $\lambda < 0$; hence the limits of integration are from 0 to ∞.

The meaning of the impulse response function $A(\lambda,t)$ can be clarified by considering some special cases. For example, if the channel filter is time invariant, then $A(\lambda,t) = A(\lambda)$. This would occur if the medium were motionless and perfectly homogeneous and if the transmitter and receiver were stationary. The time-spreading effect of the multipath propagation is expressed by $A(\lambda)$; the function $A(\lambda)$ is a delta function for a single propagation path. It is sometimes useful to express $A(\lambda)$ as

$$A(\lambda) = \sum_k A_k(\lambda - \lambda_k) \,. \tag{2}$$

Equation (2) expresses the possible occurrence of a discrete set of multipath components with characteristic delay λ_k. Each response function A_k might refer to a particular ray path. In the case of flat boundaries and infinite bandwith in the channel,

---

[1] T. Kailath, *Communication Via Randomly Varying Channels*, D. Sc. Thesis, Massachusetts Institute of Technology, June 1961.

[2] G. L. Turin, *Communication Through Noisy Random Multipath Channels*, Technical Report No. 116, Lincoln Laboratory, Massachusetts Institute of Technology, 14 May 1956.

[3] J. Salz, *Communication Through Time Variable Random Channels*, Ph. D. Thesis, University of Florida, June 1961.

[4] S. Stein, "Theory of a Tapped Delay Line Fading Simulator," *First IEEE Annual Communications Convention Record*, Boulder, Colorado, June 7-9, 1965.

$$A(\lambda) = \sum_k a_k \delta(\lambda - \lambda_k) \ . \tag{3}$$

That is, $A(\lambda)$ is a linear combination of Dirac delta functions with characteristic path delays λ_k. The coefficients a_k express the differing transmission losses of the various paths. In the case of irregular but time-invariant propagation boundaries, we do not have simple specular reflections, since the boundaries act like extended sources. As long as the transmitter and receiver are motionless, however, the response functions will be time-invariant. For each transmitter–receiver location, we will generally obtain a different set of response functions, A_k. For an ensemble of locations, the A_k must be regarded as an ensemble of nonstationary random functions.

When considering moving transmitters and receivers, moving propagation boundaries, and a time-variable nonhomogeneous medium, it is clearly inadequate to consider the response $A(\lambda)$ to be time-invariant. Furthermore, $A(\lambda,t)$ must be statistically defined in accordance with the statistical nature of the turbulent medium and the propagation boundaries. The following model for $A(\lambda,t)$ can be employed:

$$A(\lambda,t) = \sum_k b_k(t) A_k(\lambda - \lambda_k,t) \tag{4}$$

The A_k are an ensemble of time-variable nonstationary random functions representing the multipath smearing effects, and the $b_k(t)$ are a set of independent complex random functions,

$$b_k(t) = R_k(t) \exp \{j\theta_k(t)\} \ , \tag{5}$$

representing envelope and phase fluctuations of the sinusoid for the k^{th} transmission mode. The envelope fluctuations $R_k(t)$ are often referred to as fading. However, the term fading will be used here to refer to variations in both R_k and θ_k. For the case where $A(\lambda,t)$ is an aperiodic band-limited function, the sampling theorem can be applied and the channel impulse response can be separated into varying and fixed components. In this case,

$$A(\lambda,t) = \sum_k b_k(t) A_k(\lambda - \lambda_k) \ . \tag{6}$$

For a single transmitted frequency $\exp \{j\omega_o t\}$, the application of

Equations (1), (4), and (5) shows that the received signal is of the form

$$z(t) = \exp\{j\omega_o t\}\left[\sum_k R_k(t)\exp\{-j[\omega_o\lambda_k - \theta_k(t)]\}H_k(j\omega t)\right]. \quad (7)$$

The $H_k(\omega,t)$ are the Fourier transforms of the impulse response functions $A_k(\lambda,t)$ and thus represent the frequency response functions

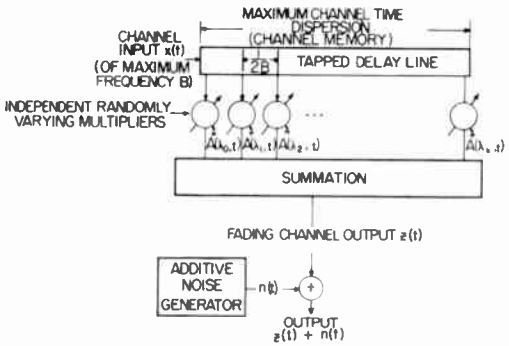


Fig. 1—Tapped-delay-line model of the fading channel.

for the k^{th} transmission mode. The time variations of H_k will generally be much slower than those of R_k and θ_k and can be assumed to be constant for purposes of the analysis. The channel response to a single frequency can be expressed as

$$z(t) = E(t)\exp\{j[\omega_o t + \phi(t)]\}, \quad (8)$$

or, expressing Equation (8) as a real function,

$$z(t) = R(t)\cos[\omega_o t + \phi(t)], \quad (9)$$

where the envelope fluctuation R and phase fluctuations ϕ correspond to the magnitude and phase of the complex function in Equation (7). The dependence of the envelope and phase fluctuation on ω in Equation (7) demonstrate the selective fading characteristic of the channel.

In mechanizing a channel simulator, the tapped-delay-line model [1] as expressed by Equation (6) was used directly. A diagram of this concept is shown in Figure 1.

A statistical model for the fading was assumed in developing a channel simulator. This model assumes that the polar coordinates

(R_k,ϕ_k) correspond to the envelope and phase of statistically independent stationary Gaussian random processes with zero mean value, with variances σ_k^2, and identical normalized spectral densities.

Channel Multipath Parameters

The values of time dispersion, fading bandwidth, attenuation, and doppler effect encountered in communication channels vary widely. A maximum time delay of 5 milliseconds has been considered sufficient for most radio-wave simulation purposes [5] for carrier frequencies below 50 MHz. However, these delays consist of gross multipath propagation, and the time dispersion or smearing of a single path is much less. Tests on pulse phase-change signaling indicated time smearing as high as 0.1 millisecond [6] in the high-frequency range. Beyond the horizon, propagation measurements at 3000 MHz exhibited typical pulse broadening of 0.1 and 0.2 microsecond with peaks as high as 0.5 microsecond.[7] Meteor-burst propagation channels introduce doppler effects, and measurements at 41 MHz over a 1000-km path have shown pulse broadening of up to 10 microseconds.[8] The West-Ford Channel [8a] was an extreme example of doppler and time dispersion for radio propagation. The fading bandwidth of these channels will vary from a fraction of a hertz to a few hertz.

Even more severe characteristics are found in underwater acoustic channels [9] where the total delay of gross arrival paths may typically be from two to three seconds and the doppler spread may be 1 to 2% of the carrier. A single arrival path will have a time dispersion up to several hundred milliseconds with typical values of tens of milliseconds.

The fading statistics in almost all situations will yield a Rayleigh

---

[5] K. W. Otten, "Radio Wave Propagation Simulator," *Electromagnetic Wave Propagation*, Page 63, Ed. by M. Desirant and J. L. Michiels, Academic Press, London and New York, 1960.

[6] S. G. Lutz, et al., "Pulse Phase-Change Signaling in the Presence of Ionospheric Multipath Distortion," *Electromagnetic Wave Propagation*, Page 357, Ed. by M. Desirant and J. L. Michiels, Academic Press, London and New York, 1960.

[7] G. Carlson, "Beyond-the-Horizon Propagation Characteristics at 3000 MHz," *Electromagnetic Wave Propagation*, Page 459, Ed. by M. Desirant and J. L. Michiels, Academic Press, London and New York, 1960.

[8] R. J. Carpenter and G. R. Ochs, "High Resolution Pulse Measurements of Meteor-Burst Propagation at 41 Mc/s over a 1295 km Path," *J. Research NBS*, Vol. 66D, pp. 249-263, May-June 1962.

[8a] I. L. Lebow, et al, "The West-Ford Belt as a Communications Medium," *Proc. IEEE*, Vol 52, No. 5, p. 543. May 1964.

[9] A. W. Ellinthorpe and A. H. Nuttall, "Theoretical and Empirical Results on the Characterization of Undersea Acoustic Channels," First IEEE Annual Communications Convention, Boulder, Colorado, June 7-9, 1965.

distribution of the envelope amplitude with a uniform phase distribution,[10] but situations have been found where this has not been the case.

Another channel characteristic that may be encountered is frequency dispersion, as typified in ionospheric propagation. In this case, the time delay is not a linear function of frequency, and a sharp transmitted pulse will be received extended in time with a steadily increasing frequency.

Channel Implementation

The Initial Optical-Digital Channel Simulator

The initial time-dispersed channel simulator was designed to encompass the characteristics of underwater acoustic channels. The simulator's characteristics have been reported previously [11] but will be included here as background for extensions and improvements to be discussed. The simulated parameters that were deemed typical for underwater acoustic channels were:

(1) Number of transmission modes, k—as many as 200,

(2) Gross multipath delays—up to 1 or 2 seconds,

(3) Time dispersion—up to 50 milliseconds,

(4) Bandwidth of multiplicative noise function $A(\lambda,t)$—0.15 to 15 Hz.

The signal time dispersion in the channel is accomplished by digitizing the transmitted signal and feeding it to a 196-stage shift register. Each stage provides a delay increment determined by the digital clock rate. The fading effects of the channel are introduced by recording a Gaussian noise signal on film and projecting the noise onto a matrix of photoresistors. Each photoresistor is connected to a stage of the shift register in a manner effecting a multiplication of the output of the stage with the noise projected from the film. The spacing of the photoresistors is sufficient to assure statistical independence of the noise at each tap. A diagram of the system is shown in Figure 2.

The input signal is clipped, sampled at a rate at least two times the Nyquist sampling rate, and sent down the shift register as a series of ones and zeros. Complementary outputs from a given stage of the shift register feed a pair of resistors. One is a photoresistor that multiplies the signal by the value of the noise function projected on it.

---

[10] E. J. Baghdady, *Lectures on Communication System Theory*, Chapter 7, p. 133, McGraw-Hill Book Company, Inc., New York, 1961.

[11] L. W. Martinson, "A Time-Dispersed Rayleigh Fading Channel Simulator," Ninth National Communications Conference, Rome, N. Y., Oct. 1963.

The other resistor is a potentiometer adjusted to balance out variations in the average value of the photoresistor and give a bipolar output from the film photoresistor combination, which would normally be unipolar.

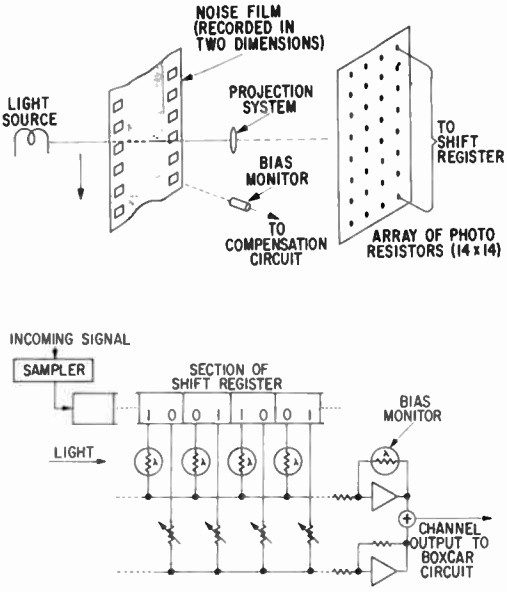


Fig. 2—Schematic of modulating optics (top) and time dispersed channel delay line (bottom).

By adding the outputs in groups of 14 adjacent stages, as shown in Figure 3, it is possible to weight the channel response in 14 sections. For example, it may be desired to weight the earliest arrivals strongest, with successive arrivals growing weaker, section by section.

The operational amplifier used to sum the outputs of the photoresistors has, as a part of its feedback loop, a photoresistor of the same type use in the noise multipliers. This photoresistor sees the total light coming through the particular section of noise film being projected and balances out variations in projection lamp intensity and average film density.

The total of the entire channel output is fed to a box-car circuit that holds the value of each output pulse until the next one arrives.

Additional large multipath delays encountered in underwater sound transmission were simulated by using a multitrack tape recorder with tracks connected in cascade.

The fading bandwidth of the channel is controlled by the film drive speed, the number and arrangement of transmission modes is controlled by the number of shift register stages and the tape-recorder parameters, and the time dispersion is controlled by the number of stages and the shift register clock rate.

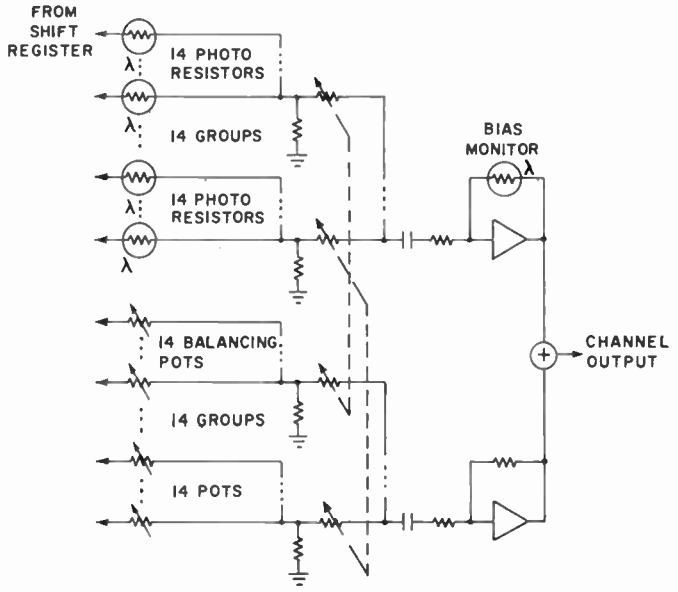


Fig. 3—Method of adding channel-tap outputs.

The Appendix includes derivations of the individual photoresistor modulator operation and shows that for a large number of delay stages the channel output is a valid representation of the requirements presented by Equations (7) and (9).

Extensions To Optical Digital Channel Simulator

Since the construction of the original simulator, many new components have become available that make the method employed even more attractive. Recent developments in integrated-circuit techniques make elaborate extensions and improvements to the simulator practical. These advances are coupled with a decline in the cost per function.

The principal limitations in the initial optical digital channel simulator implementation were clipping losses, which affected its narrowband signal-to-noise capability, and a clock rate limitation imposed by the rather ancient circuitry employed.

The problem of clipping losses can be handled in either of two ways. The first method would employ sample-and-hold circuits to form an analog shift register. The bandwidth capability of these devices has been steadily improving, and recently a 120-MHz sample-and-hold circuit was announced.[12] The delay-line-length capability with sample-and-hold circuits is determined by the inherent noise in each amplifier stage and the attenuation per stage. Even with an attenuation factor of 0.001, a loss of 1 db in 39 stages or about 25 db in 1000 stages is

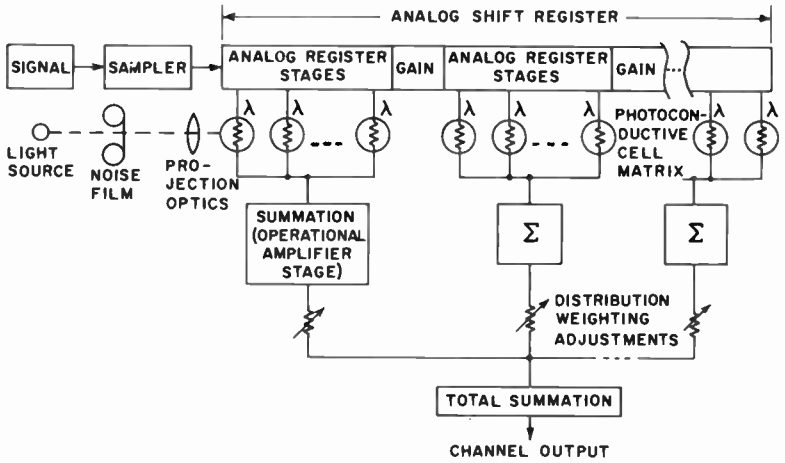


Fig. 4—Analog shift-register mechanization of the channel.

obtained. This loss most critically affects the signal-to-additive noise ratio in the channel, especially if a high signal dynamic range is desired.

Figure 4 illustrates an analog shift-register channel mechanization. The balancing network that provides bipolar tap outputs is not shown. The attenuation factor through the channel shown is compensated by periodic insertion of gain. An alternative method is to store separate gross delays of signals in a tape recorder and fill in the region between tape recorder pickoffs with an analog shift register. This latter method has the advantage that signal-to-additive-noise margin is not adversely affected by the analog-shift-register attenuation.

The second method of eliminating clipping losses is to use several levels of amplitude quantization. The current state of the art of integrated-circuit digital devices and the prospective improvements in them

[12] R. E. Fisher, "VHF Sample and Hold," 1967 International Solid-State Circuits Conference, Philadelphia, Pa., Feb. 15-17, 1967.

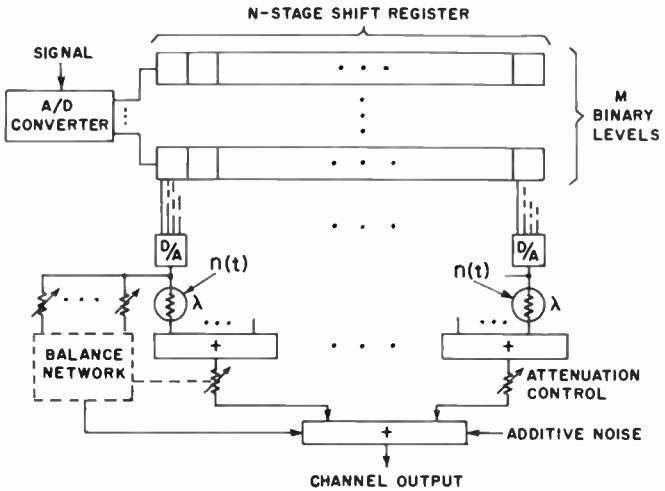


Fig. 5—Multilevel shift-register implementation of channel.

make a multilevel shift-register delay line with analog or digital noise generation practical. The general form of a multilevel shift-register delay-line channel employing optical modulators is shown in Figure 5. The signal is sampled and amplitude encoded with an analog/digital converter. The dynamic range and fidelity requirements set the number of bits and parallel digital channels. Noise levels in the digital circuits will set the allowable number of stages, but the noise levels should not be a problem in the simulation of communication channels with the longest time dispersions encountered.

Doppler effects on broad-band signals can conceivably be handled within limits in a long digital delay line through the use of a switching matrix at the output taps. The principle is depicted in Figure 6. Relative motion of the source and receiver and/or channel is simulated by placing what is essentially a multiple-position switch between the shift-

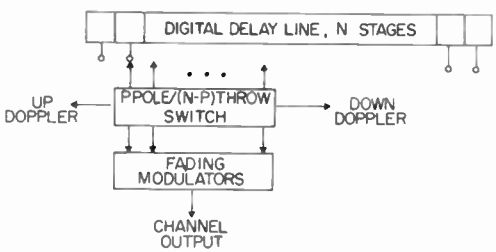


Fig. 6—Limited doppler simulation.

register outputs and the channel-noise modulators. The degree of doppler sets the rate of movement in one or the other direction along the shift register. Movement toward the input yields up-doppler and movement away from the input yields down-doppler. The technique is obviously limited by the required doppler rate in relation to the signal bandwidth and the length of time that the doppler must last in a given simulator test. Its advantage is that a true doppler effect represented by a time dilation is simulated, rather than a frequency shift.

Another area of application is in the mechanization of the tap multipliers. It should now be economical to mechanize these in digital

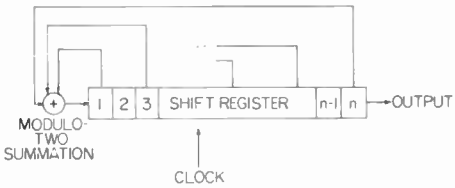


Fig. 7—Feedback shift-register method of generating pseudo-random number sequence.

form to replace the light-modulated taps used originally. Of course this would require the generation of random numbers in digital rather than analog form. This, of course, is a well-developed technique in the simulation field; users of the Monte Carlo method have developed numerous programs for the generation of pseudo-random number sequences. These sequences usually have a uniform probability distribution when generated, and are then processed further to give the desired distribution. We will not attempt to review here the various number generation methods, but we will point out one that is particularly useful for special-purpose equipment. The feed-back shift register method, shown in Figure 7, consists of adding (modulo two) the contents of certain stages of the register to form the input to the register. By selecting the proper stages to feed back, an n-stage register will generate a sequence $2^n - 1$ in length before beginning to repeat itself. (The generator produces all possible n-bit numbers except 0.) This sequence also has a uniform probability distribution.

Changing the distribution from uniform to normal has also been programmed in various ways by users of general-purpose computers. One method is to solve the equation

$$u_t = \int_{-\infty}^{U_t} \exp\left\{-\frac{1}{2}x^2\right\} dx$$

for U_i for each input value of u_i from the uniformly distributed sequence. This gives an accurate result, but is somewhat cumbersome for a special-purpose machine, especially in our application, where a very accurate distribution is not needed for the multipliers. Here the use of a piece-wise linear transfer function, as shown in Figure 8,

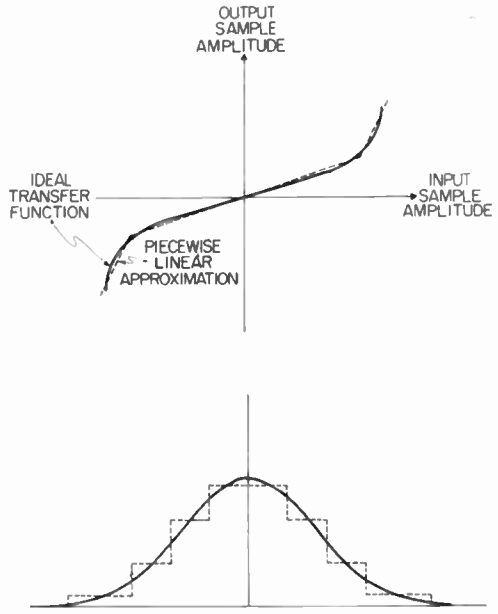


Fig. 8—(Top) Transfer function used to generate a desired probability distribution function from a uniform distribution and (bottom) approximation of a normal distribution function by application of a piecewise linear transfer function to a uniform distribution.

will suffice. This method is adaptable to special-purpose equipment, as shown in Figure 9. Each incoming variable (x) is multiplied by a linear transfer function to yield an output $y = m_i x + b_i$. The particular function used depends on the range (i) in which the input variable falls.

The problem of summing the outputs of all the tap multipliers is probably still best done with operational amplifiers, as in the original version, which used optical tap multipliers. The digital output of each tap multiplier could be applied to a set of correctly weighted resistors, which in turn are fed to a common operational amplifier summer, as shown in Figure 10. It would be convenient to use several amplifiers, so that gross weights could be applied to various groups of taps.

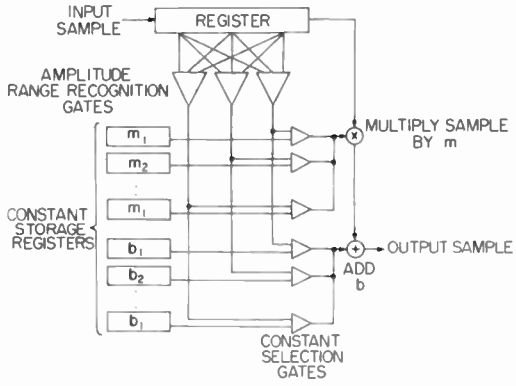


Fig. 9—Mechanization of an approximate probability distribution.

An Application of the Channel Simulator

One of the techniques tested on the channel simulator was multiple-tone binary frequency-shift keying (FSK), which was being investigated as part of a study on methods of improving communications in underwater acoustic channels.

Due to the time-dispersive and fading nature of underwater channels, the use of various forms of phase and amplitude modulations presents special difficulties, and it was only natural to consider some form of frequency modulation. In a digital system, this takes the form of frequency-shift keying (FSK). Since the fading encountered is frequency selective, it was thought that some form of frequency diversity should be useful, and it was to determine the effectiveness of this method of communication that the experiment was conducted on

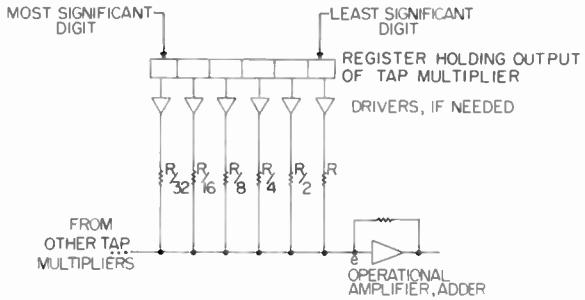


Fig. 10—Implementation of tap summation from digital multipliers.

the simulator. In order to keep the number of variables at a minimum while obtaining data in a useful form, a multiple-tone binary FSK system was simulated. Once the performance of this type of system is understood, the data may be extended to other systems, such as n-ary digital systems, or frequency-hopping systems, and it is possible to compute the trade-offs between alphabet size, data rate, and probability of message error.

Several aspects of communication in a fading channel that bear on this particular problem have been investigated by Lieberman and others.[13-15] To handle the FSK problem analytically, however, it has frequently been necessary to assume either of the extreme cases of a nonselective channel, or one in which the tones used fade completely independently.

For generating the tones used in the experiment, a set of 20 oscillators and corresponding filters were available, spread at 10-Hz intervals. The mark and space frequencies were selected from this set, translated into the desired transmission band by single-sideband modulation, clipped, sampled, and injected into the simulator. The transmission band was centered at about 1.6 kHz.

The signal emerging from the channel was demodulated to recover the original set of frequencies and passed through a set of band-pass filters corresponding to the original frequencies transmitted. These were active filters that used as a reference the same oscillators used to transmit the original signal. In this respect, ideal conditions were assumed, since in an actual system an identical but separate set of oscillators would have to be used at the receiver, and the possibility of misalignment arises.

The channel fading bandwidth was varied from 0.12 to 12 Hz, and the dispersion from about 1.65 to 23 milliseconds. While the range of parameters was limited by the characteristics of the simulator, it was felt that they were representative of the conditions to be expected in the channel being studied. The baud length (or time over which the signal was integrated before making a decision) was set at 0.5 second. The data from the receiver was adjusted to compensate for the effects of clipping noise.

One portion of the experiment was devoted to determining the

[13] A. B. Glenn and G. Lieberman, "Performance of Digital Communications Systems in an Arbitrary Fading Rate and Jamming Environment," *IEEE Trans. Communications Systems*, p. 57, Vol. CS-11, No. 1, March 1963.

[14] J. N. Pierce, "Theoretical Diversity Improvement in Frequency-Shift Keying," *Proc. IRE*, p. 903, Vol. 46, No. 5, Part 1, May 1958.

[15] G. Lieberman, "Effect of Fading on Quadrature Reception of Orthogonal Signals," *RCA Review*, p. 353, Vol. 23, No. 3, Sept. 1962.

effect of time spacing on system performance. Figure 11 shows the results of changing the tone spacing by a factor of two. The results suggest that for even the closest spacing of tones possible in the system being tested, the tones fade independently. This conclusion is important for this system, since in the particular channel being simulated the possibility of relatively high doppler shifts due to ship motion

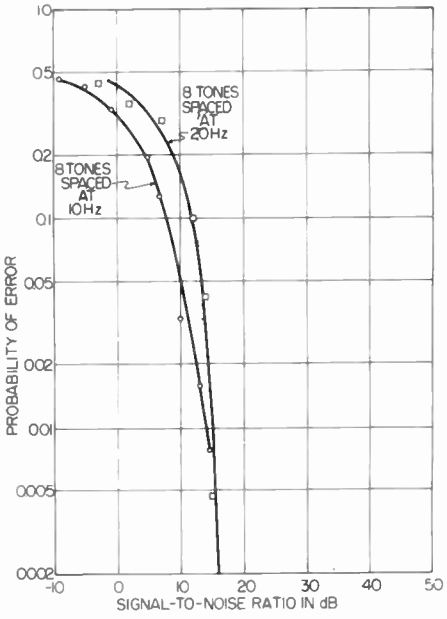


Fig. 11—Effects of varying tone spacing on the multiple-tone FSK system.

would make the use of more closely spaced tones difficult. Hence, analytic predictions based on the assumption of independently fading tones would seem to be reasonably safe.

Figures 12 and 13 illustrate the effect of fading rate and time dispersion on the system. It may be seen that, as hoped, the system was very insensitive to changes in these two parameters.

One final characteristic of the system that was shown by the simulation was the fact that the overall effect of the channel under any combination of events (including the distortions of the signal required to get it into the simulator) has the same effect as additive noise. Knowledge of this characteristic is important, since it means that the multiplicative effects of the channel may be overcome by the use of more power at the transmitter, more receiver array gain, or some

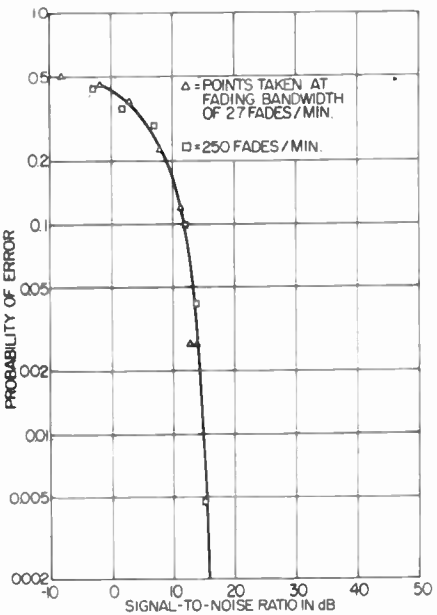


Fig. 12—Effects of fading rate on the multiple-tone FSK system.

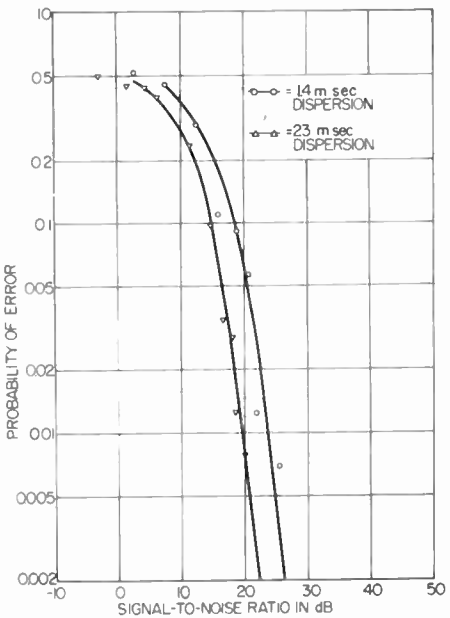


Fig. 13—Effects of time dispersion on the multiple-tone FSK system.

similar technique. This is not possible with some communication systems; for example, once distortion is introduced into the signal in a straight voice system, no amount of signal level will make the voice more intelligible.

Use of the Simulator with a "Canned" Channel

The "canned" channel is an area of potential use with the channel simulator. The term refers to the use of tap-multiplier functions that have been recorded directly from a real channel, thus eliminating the requirement for estimating the fading bandwidth and the statistics of the tap-multiplier functions. The simulator would contain a sample of the channel as it actually existed at some specific time and place, and many systems could be tested under identical and real conditions. The main problem here is that a suitable probing signal must be found, and this may become quite difficult for some channels due to extremes of fading bandwith, time dispersion, or additive noise. However, efforts spent on field trials of communication systems might be better spent in making general probing measurements of the channel, with most of the system testing being done in the laboratory.

Conclusion

A fading-channel simulator based on a tapped delay-line model of the channel provides a means of separating and controlling channel parameters such as time dispersion, fading bandwidth, and fading statistics, thus permitting the evaluation of the effects of these parameters on a given communication system. While the mechanization described made use of clipped signals, advances in component design have made mechanizations practical that carry along amplitude information. The on-line operation afforded by this type of simulator provides a means for testing new system concepts or changes in existing systems rapidly, and at an economical breadboard stage, rather than on a full-scale model in the field. It is also possible to record the characteristics of a particular channel by means of probing signals, so that the channel may be "canned" for use in the simulator.

Acknowledgment

The authors wish to acknowledge the contributions of G. Lieberman to the mathematical modeling of the channel and of H. M. Halpern and W. G. Anderson to the implementation and simulation programs.

APPENDIX—SHIFT-REGISTER–PHOTOCONDUCTIVE-CELL MODULATOR

Figure 14 shows the basic form of the photoconductive-cell modulator using a $(+1, -1)$ logic in the shift register with a current-sensitive amplifier as the adder. A resistor with conductance G_R and a photocell with mean conductance $\overline{G}_\lambda = G_R$ are connected to complementary outputs of a shift-register stage.

The conductance of the photocell is $G_\lambda = \overline{G}_\lambda\,(1 + n(t))$, where (1

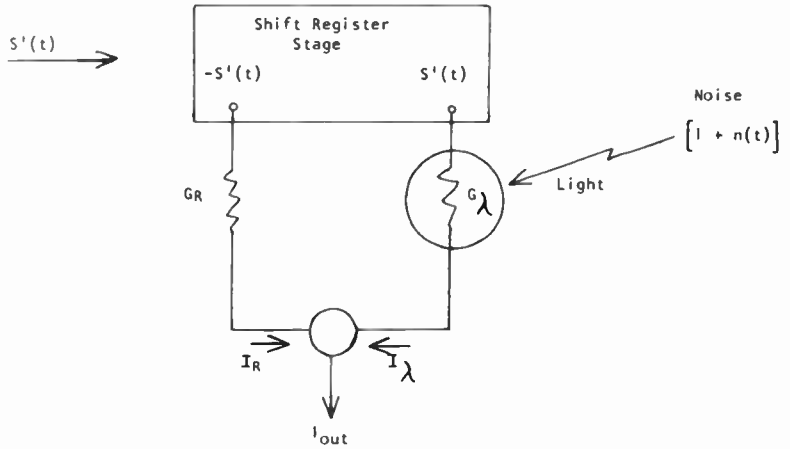


Fig. 14—Photoconductive-cell modulator with current adder.

$+ n(t))$ is a light modulation with $|n(t)| < 1$ projected on the photocell; $n(t)$ is in most cases a band-limited white noise with a Gaussian amplitude distribution. The output current of the modulator with the sample signal $S'(t)$ at the shift register is

$$I_{out} = I_R + I_\lambda$$
$$= -S'(t)G_R + S'(t)G_\lambda .$$

Substituting the values for G_R and G_λ,

$$I_{out} = -S'(t)\overline{G}_\lambda + S'(t)\overline{G}_\lambda\,(1 + n(t))$$
$$= \overline{G}_\lambda S'(t)n(t) . \qquad (10)$$

Equation (10) indicates that the output is a constant modifying the product of the sampled signal and the noise function recorded on the film.

The effect of the mean conductance modifying the amplitude of the modulator output can be avoided by using a voltage-sensitive adder as shown in Figure 15. Assume, as before, that the conductance of the photocell is a linear function of light; the voltage e_{out} will then be

$$e_{out} = S'(t) - \frac{2S'(t)G_R}{G_\lambda + G_R} = \frac{S'(t)(G_\lambda - G_R)}{G_\lambda + G_R}.$$

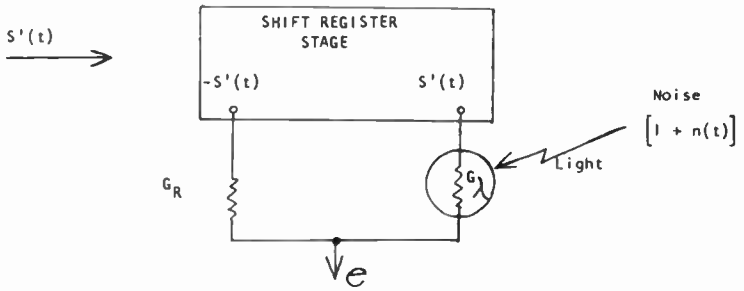


Fig. 15—Photoconductive-cell modulator with voltage adder.

Substituting $G_\lambda = \overline{G}_\lambda\,(1 + n(t))$ and $G_R = \overline{G}_\lambda$

$$e_{out} = \frac{S'(t)n(t)}{2 + n(t)} \tag{11}$$

In normal modulation on photographic film the linear region of light transmission varies from about 0.8 to 0.2 for normal analog recording. Thus $|n(t)| \leq 0.3$ with a film bias transmission of 0.5, and $|n(t)| < 0.6$ when referred to the bias of unity as employed in Equation (11). With $|n(t)| \leq 0.6$ in Equation (11), there is very little deviation from linearity in the output.

The Combined Modulator Outputs

The sum of the N modulator stages, each modulated in accordance with Equation (10) or (11), is

$$E_n = \sum_{i=0}^{N-1} C_i S'(t - \tau_i) n_i(t)\,, \tag{12}$$

where the n_i are independent, band-limited white-noise functions with Gaussian amplitude distributions, C_i is the proportionality constant for

a particular modulator, and $S'(t-\tau_i)$ is the delayed time and amplitude quantized signal. The modulation of $S'(t-\tau_i)$ by $n_i(t)$ is a linear process. Therefore, for the purpose of analysis, the low-pass filter at the output of the channel can be considered to occur prior to summation. This gives $S'(t) = S(t)$. For a single-frequency signal,

$$E_n = \sum_{i=0}^{N-1} C_i n_i(t) \cos [\omega_o(t-\tau_i)] , \tag{13}$$

which can be expressed as

$$E_n = \sum_{i=0}^{N-1} C_i n_i(t) [\cos \omega_o t \cos \omega_o \tau_i + \sin \omega_o t \sin \omega_o \tau_i] . \tag{14}$$

For the case, $n = 2$, $\tau_i = i/(4f_o)$, and $C_o = C_i = 1$

$$E_2 = n_o(t) \cos \omega_o t + n_i(t) \sin \omega_o t , \tag{15}$$

where $n_o(t)$ and $n_i(t)$ are independent low-pass Gaussian processes that can be expressed as [16]

$$\begin{aligned} n_o(t) &= R(t) \cos \phi(t) , \\ n_i(t) &= R(t) \sin \phi(t) . \end{aligned} \tag{16}$$

The term $R(t)$ is a random amplitude function with a Rayleigh probability density

$$p(R) = \frac{R}{\sigma^2} \exp \left\{ -\frac{R^2}{2\sigma^2} \right\} \tag{17}$$

and $\phi(t)$ is a random phase function with probability density

$$p(\phi) = \frac{1}{2\pi} ; \qquad 0 \leq \phi \leq 2\pi .$$

Thus,

$$\begin{aligned} E_2 &= R(t) \cos \omega_o t \cos \phi(t) + R(t) \sin \omega_o t \sin \phi(t) \\ &= R(t) \cos [\omega_o t - \phi(t)] . \end{aligned} \tag{18}$$

[16] S. O. Rice, "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, Vol. 23, p. 282, and Vol. 24, p. 46, 1945.

Therefore under the limited conditions specified, a Rayleigh fading channel coinciding with Equation (18) is derived.

For the case of an arbitrary sampling frequency greater than the Nyquist rate, Equation (13) can be written

$$E_n = I_c \cos \omega_o t + I_s \sin \omega_o t \,, \tag{19}$$

where

$$I_c = \sum_{i=0}^{N-1} n_i(t) \cos \omega_o \tau_i \,, \quad \text{and} \quad I_s = \sum_{i=0}^{N-1} n_i(t) \sin \omega_o \tau_i \,.$$

Following Rice [17], if I_s and I_c are normal, have equal variances, and are uncorrelated, then E_n will be Gaussian with a Rayleigh amplitude distribution of the envelope and random phase.

I_c and I_s have normal distributions, since $n_i(t)$ is normally distributed and the sum of any number of normally distributed functions is itself normal. The cross correlation of I_c and I_s is

$$E[I_c I_s] = \frac{1}{N} \sum_{i=0}^{N-1} \cos \omega_o \tau_i \sin \omega_o \tau_i = \frac{1}{2N} \sum_{i=0}^{N-1} \sin 2\omega_o \tau_i \,. \tag{20}$$

With a large number of stages, the signal phases $\omega_o \tau_i$ in the modulator stages will tend to distribute themselves uniformly. Then

$$E[I_c I_s] \approx \frac{1}{2\pi} \int_0^{2\pi} \sin x \, dx = 0 \,. \tag{21}$$

The variance of I_c is

$$E[I_c^2] = \sum_{i=0}^{N-1} \cos^2 \omega_o \tau_i = \frac{1}{2} \sum_{i=0}^{N-1} (1 + \cos 2\omega_o \tau_i) \,. \tag{22}$$

For a uniform distribution of $\omega_o \tau_i$, $E[I_c^2] = N/2$. Similarly $E[I_s^2] = N/2$. Hence, for a large number of stages, E_n will have the desired statistical characteristics.

---

[17] S. O. Rice, "Statistical Properties of a Sine Wave Plus Random Noise," *Bell System Technical Journal*, p. 109, Vol. 27, Jan. 1948.

CORRECTION

The following corrections should be noted in the paper, "The Resolving-Power Functions and Quantum Processes of Television Cameras," by O. H. Schade, which appeared in the September 1967 Issue of *RCA Review*, pp. 461-535.

1. On page 501, the last two sentences of the first full paragraph should read as follows: "For calculation purposes, a daylight illuminant (5000° K) and a spectral sensitivity close to that of an ASOS photoconductor will be assumed. With the value $\lambda/P_\lambda = 1$ in Equation (63) of the Appendix, the conversion gain is"

2. In the table on page 534, the last figure in the column headed λ/P_λ should read 0.55 instead of 1.0. The following note of explanation should also be added below the table: "The factor λ/P_λ is closely equal to unity for an ASOS photoconductor."

RCA Technical Papers

Third Quarter, 1967

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Application of Dual-Gate MOS Field-Effect Transistors in Practical-Radio Receivers," H. M. Kleinman, *Trans. IEEE PTGBTR* (July) 1967

"Automated Mask Production for Semiconductor Technology," H. O. Hook, *SCP & Solid State Tech.* (July) 1967

"Cryogenic Random-Access Memories," A. R. Sass, W. C. Stewart, and L. S. Cosentino, *IEEE Spectrum* (July) 1967

"CW Three-Terminal GaAs Oscillator," K. G. Petzinger, A. E. Hahn, and A. Matzelle, *Trans. IEEE PTGED* (Correspondence) (July) 1967

"Evaporated Metallic Contacts to Conducting Strontium Titanate Single Crystals," J. E. Carnes and A. M. Goodman, *Jour. Appl. Phys.* (July) 1967

"Improving Engineering Reports and Talks," I. M. Seideman, *Trans. IEEE PTGEWS* (July) 1967

"Restoration of High Frequencies in Photomask Line Patterns by Photographic Non-Linearity," H. Johnson, *SCP & Solid State Tech.* (July) 1967

"The Revolution in Electrical Energy Sources—A State-of-the-Art Review," P. Rappaport, *Trans. IEEE PTGBC* (July) 1967

"Small-Signal High-Frequency Transistors," T. J. Robe, *Electronics World* (July) 1967

"Small-Signal Power Flow and Energy Density for Streaming Carriers in the Presence of Collisions," B. Vural and S. Bloom, *Trans. IEEE PTGED* (July) 1967

"SMPTE Color Television Subjective Reference Test Films and Slides —Issue No. 3," E. P. Ancona, Jr. and Coauthor, *Jour. SMPTE* (July) 1967

"Solid-State Optically Pumped Microwave Masers," E. S. Sabisky and C. H. Anderson, *IEEE Jour. Quantum Electronics* (July) 1967

"Synthesis of Ambiguity Functions for Prescribed Responses," W. Blau, *Trans. IEEE PTGAES* (July) 1967

"Measurement of Recombination Lifetimes in Superconductors," A. Rothwarf and B. N. Taylor, *Phys. Rev. Letters* (3 July) 1967

"Historical Postscript," G. H. Brown, *Electronics* (Readers Comment) (July 10) 1967

"Optically Induced Tm^{2+} Ions in Trigonal Sites in BaF_2," E. S. Sabisky and C. H. Anderson, *Phys. Rev.* (10 July) 1967

"Energy Gap in B-Ag_2Se," R. Dalven and R. Gill, *Phys. Rev.* (15 July) 1967

"Analysis of the Effect of Charge Storage on the Frequency Response of SCR's," H. J. Kuno, *Proc. IEEE* (Letters) (August) 1967

"Gain Saturation in Crystal Mixers," P. Torrione, *Proc. IEEE* (Letters) (August) 1967

"High-Power Pulsed GaAs Laser Diodes Operating at Room Temperature," H. Nelson, *Proc. IEEE* (August) 1967

"Improving Automatic Sensitivity Performance in Color TV Film Cameras," R. R. Brooks, *Jour. SMPTE* (August) 1967

"Measurement of Film Thickness Using Infrared Interference," D. J. Dumin, *Rev. Sci. Instr.* (August) 1967

"Reference Chart for Lasers and Detectors," D. Blattner and R. Wasserman, *The Electronic Engineer* (August) 1967

"Seebeck Coefficient of Thin-Film Germanium," W. L. C. Hui and J. P. Corra, *Jour. Appl. Phys.* (August) 1967

"Ultrasonic Attenuation in Ge-Si Alloys," K. R. Keller, *Jour. Appl. Phys.* (August) 1967

"Validity of Rytov's Approximation," D. A. DeWolf, *Jour. Opt. Soc. Amer.* (Letters to the Editor) (August) 1967

"The 'Vee-Zee' Panel as a Side-Mounted Antenna," R. N. Clark and A. L. Davidson, *Broadcast News* (August) 1967

"Vidicons for Space Applications," R. E. Hoffman, *Jour. SMPTE* (August) 1967

"Energy Gap of Superconducting Nb_3Sn," D. R. Bosomworth and G. W. Cullen, *Phys. Rev.* (10 August) 1967

"The Effect of Hard Limiting in the Presence of Large Out-of-Band Interfering Signals," R. W. Ottinger, *Trans. IEEE PTGAES* (Correspondence) (September) 1967

"Engineering Manpower Misused," J. M. Fisher, Jr., *EE* (EE Speak Up) (September) 1967

"On the Gaussian Sum of Gaussian Variates, the Non-Gaussian Sum of Gaussian Variates, and the Gaussian Sum of Non-Gaussian Variates," M. Masonson, *Proc. IEEE* (Letters) (September) 1967

"Gigahertz Tunnel-Diode Logic," M. Cooperman, *RCA Review* (September) 1967

"High-Frequency Characteristics of the Insulated-Gate Field-Effect Transistor," J. R. Burns, *RCA Review* (September) 1967

"High Peak Power Epitaxial GaAs Oscillators," S. Y. Narayan and B. E. Berson, *Trans. IEEE PTGED* (Correspondence) (September) 1967

"An Introduction to Computer Graphic Terminals," M. H. Lewin, *Proc. IEEE* (September) 1967

"Manley-Rowe Relations in the Transverse Wave Vector," S. Freeman, *Jour. Opt. Soc. Amer.* (September) 1967

"Measurements of Dipole Antenna Impedance in an Isotropic Laboratory Plasma," K. A. Graf and D. L. Jassby, *Trans. IEEE PTGAP* (September) 1967

"Phase Transformation of As_2Se_3 and Sb_2Se_3 Films," M. D. Coutts and E. R. Levin, *Jour. Appl. Phys.* (September) 1967

"The Resolving-Power Functions and Quantum Processes of Television Cameras," O. H. Schade, Sr., *RCA Review* (September) 1967

"A Self-Scanned Solid-State Image Sensor," P. K. Weimer, G. Sadasiv, J. E. Meyer, Jr., L. Meray-Horvath, and W. S. Pike, *Proc. IEEE* (September) 1967

"The Solar System Environment," S. Miluschewa, *Trans. IEEE PTGAES* (September) 1967

"Speech Processing Techniques and Applications," H. F. Olson, H. Belar, and E. S. Rogers, *Trans. IEEE PTGAU* (September) 1967

"Standard Test-Lamp Temperature for Photosensitive Devices—Relationship of Absolute and Luminous Sensitivities," R. W. Engstrom and A. L. Morehead, *RCA Review* (September) 1967

"Understanding and Using the Dual-Gate MOS FET," R. Dawson, R. Ahrons, and N. Ditrick, *EE* (September) 1967

"Transition of Type-I Superconducting Thin Films in a Perpendicular Magnetic Field: A Microwave Study," J. I. Gittleman, S. Bozowski, and Coauthor, *Phys. Rev.* (10 September) 1967

"Linear IC's: Part 3 Differential Amplifiers at Work," J. P. Keller, *Electronics* (September 18) 1967

AUTHORS

J. BORDOGNA received the BSEE and Ph.D. degrees from the University of Pennsylvania in 1955 and 1964, respectively, and the SM degree from MIT in 1960. Since 1955 he has been associated with RCA, except for a three year period of active duty with the Navy. Currently, he is Winterstein Assistant Professor of Engineering at the University of Pennsylvania and a consultant to RCA. Dr. Bordogna's research includes work on Q multipliers, synchronization problems in FSK receivers, automatic phase control circuits, FM modulators, circuits for a magnetic recording machine used to modify sonar frequencies for video display, very low frequency antennas, PSK communication systems, electronically variable delay lines for correcting random jitter in video tape-recording systems, radar avionics, and communication systems for paralytics. He has completed theoretical studies of FM variable-bandwidth communication systems for trophospheric scatter propagation, applications of lasers, and an analysis of the Ranger moon-probe television subsystem. He also participated in demonstrating the wideband modulation properties of GaAs, the development of the wideband GaAs electro-optic modulator for optical television transmission, and the construction of the Gemini VII laser voice transmitter.

Dr. Bordogna is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Tau, Sigma Xi, I.E.E.E., and ASEE.

JOHN E. COURTNEY received the E.E. degree from the University of Cincinnati in 1957 and MSEE degree from the University of Pennsylvania in 1964. He joined RCA in 1957. His work has been primarily in system engineering and signal processing as applied to radar and sonar. He is a member of the Institute of Electrical and Electronics Engineers, the Acoustical Society of America, and Eta Kappa Nu.

RAYMOND H. DEAN received the B.S. degree from the University of Kansas in 1958. He served two years as a technical liaison officer in the U.S. Navy. In 1961 he received the M.S. degree in electrical engineering from the Massachusetts Institute of Technology. From 1961 to 1964 he was engaged in the engineering and sale of large commercial high velocity air conditioning systems and helped develop a computer program for the automatic design of such systems. He is presently completing requirements for a PhD in electrical engineering at Princeton University. Work on the Gunn Effect was done during summer employment at RCA Laboratories in Princeton, and his current interest is in solid state bulk effect and transit time devices. He is a member of the Institute of Electrical and Electronic Engineers.

JOSEPH F. DIENST was awarded the B.S.E.E. degree by Drexel Institute of Technology in 1955, the M.E.E. by Syracuse University in 1958 and the Ph.D. degree by Rutgers University in 1966. His doctoral dissertation was concerned with the effects of electron collisions and drift velocity on the waves in a plasma loaded waveguide. While a member of the research staff at Syracuse University, Dr. Dienst did work in the fields of magnetic amplifiers and tropospheric scatter propagation. From 1958 to 1963, he was a member of the faculty of Rutgers University, where he taught courses in electronics, circuit theory, and electromagnetic field theory, in addition to doing research on microwave slow-wave structures. From 1961 to 1963, Dr. Dienst was also a part-time member of the research staff of RCA Laboratories where he investigated microwave propagation in plasmas. He joined the RCA Laboratories technical staff full time in 1963. He has worked on diagnostic techniques in gaseous plasmas and microwave amplification and generation using bulk semiconductors. His current interest is in the interaction of drifted carriers and spin waves in solids, such as the magnetic semiconductors $Ag_xCd_{1-x}Cr_2Se_4$.

Dr. Dienst is a member of Tau Beta Pi, Eta Kappa Nu, the Institute of Electrical and Electronics Engineers and is a licensed Professional Engineer in the State of New Jersey.

RONALD E. ENSTROM received the S.B., S.M., and Sc.D. degrees in Metallurgy from the Massachusetts Institute of Technology in 1957, 1962, and 1963, respectively. In 1957 and 1960 he worked at the Metals Research Laboratory of the Union Carbide Metals Company and at Nuclear Metals, Inc. During his graduate career he was a Research Assistant in the Metals Processing Laboratory at M.I.T. His doctoral thesis focussed on superconductivity in the Niobium-Tin system and included research on the superconducting properties, x-ray structure determination, and phase equilibria investigations of the three compounds in the system. Since joining RCA Laboratories in 1963, Dr. Enstrom has investigated the preparation and relationship of growth parameters to the superconducting properties of vapor-deposited Nb_3Sn and Nb_3Sn alloys. Currently he is working on the vapor phase synthesis and characterization of GaAs and GaAs-GaP alloys for high power rectifiers and solid-state microwave oscillators. He is a member of Sigma Xi, the American Institute of Metallurgical Engineers, and the American Physical Society.

HAROLD M. FINN received the BEE degree from the City College of N. Y. 1950, the M.S. degree in E.E. from the University of Maryland in 1955, and the Ph.D. degree from the University of Pennsylvania in 1965. Since 1950, Dr. Finn has been engaged in research and development of industrial control, radar, and communications systems while employed at the Engineering and Research Corp., Melpar, and the Emerson Research Laboratories. He joined the Missile and Surface Radar Division of RCA in 1959 where he is engaged in the development and application of signal detection theory. Dr. Finn is a member of Tau Beta Pi, Eta Kappa Nu, and is a senior member of the Institute of Electrical and Electronics Engineers.

A. B. GLENN received the BEE degree from Polytechnic Institute of Brooklyn in 1938, the MSEE degree from MIT in 1941, and the Ph.D. degree in electrical engineering from Syracuse University in 1952. He is an adjunct professor at Drexel Institute of Technology where he is teaching graduate courses in statistical theory of communications, information theory, and modulation theory. From 1938 to 1949 he worked for the Sperry Gyroscope and Western Electric Companies specializing in radio receiver design and UHF and microwave tube development. From 1949 to 1954 he did advanced development and product design of television receivers for the General Electric Company. Since 1954 he has been with RCA where his primary responsibility is in the synthesis, analysis, and evaluation of communication systems. From 1961 to the present he has worked on survivable satellite communication systems and interplanetary communications. Dr. Glenn is presently staff engineer of the Systems Engineering, Evaluation and Research where he acts as a consultant on communication systems for RCA Defense Electronic Products.

Dr. Glenn is a member of Sigma Xi, Tau Beta Pi, and Eta Kappa Nu and a Senior Member of IEEE. He is listed in the American Men of Science.

JOEL S. GREENBERG received the B.E.E. from the Polytechnic Institute of Brooklyn in 1952 and the M.E.E. from Syracuse University in 1960. From 1952 to 1961, he was employed by Rome Air Development Center, Griffiss AFB, Rome, N.Y. where he was engaged in the general areas of radar development and systems analysis. He worked for several years on guidance and control systems and supervised a systems group responsible for the analysis and synthesis of space surveillance, ballistic missile early warning, and ballistic missile and satellite intercept systems. In 1961 Mr. Greenberg joined the staff of RCA Defense Electronic Products, Advanced Military Systems Group, where he engaged in space system studies, particularly in the analysis of future commercial and military satellite communication systems. Subsequently, Mr. Greenberg joined the Graphic Systems Division as Manager of Customer Systems Engineering. At the end of 1966, he joined the Operations Research Group, where he is presently engaged in the design of mathematical models to be used in new business analysis.

W. J. HANNAN graduated from the RCA Institutes in 1951. He joined the RCA Industrial Products Division, where he worked on the design of an industrial television camera chain and a portable television system. He received the BS in EE from Drexel Institute in 1954, and the MSEE degree from the Polytechnic Institute of Brooklyn in 1955. Mr. Hannan has been employed in RCA's Applied Research department since 1956. His experience includes the design and development of television equipment, digital communication receivers, digital data-processing equipment, wide-band magnetic recorders, laser communications, and radar systems. He taught transistor circuit theory courses at Rutgers University. Recently, he has been working on an electro-optic reading machine, a solar-pumped modulated laser, and an injection-laser communication system.

Mr. Hannan is a senior member of the Institute of Electrical and Electronics Engineers and Sigma Xi.

IMRE J. HEGYI received the B.S. degree in Chemistry from the John Hopkins University in 1942 and the M.S. degree in Structural Inorganic Chemistry from Polytechnic Institute of Brooklyn in 1956. From 1942 to 1943, he was senior chemist with Davis Chemical Company and from 1943 to 1947 was involved in chemical and x-ray diffraction studies of the platinum group metals for Baker and Company, a Division of Englehard Industries. Since 1947, he has been with RCA Laboratories where he has worked on the synthesis of electronically active solids. He has worked on the development of ferrite magnetic memory cores and on new synthesis procedures for the preparation of III-V semiconductor compounds.

Mr. Hegyi is a member of the American Chemical Society, American Physical Society, American Crystallographic Association, Electrochemical Society, and Sigma Xi.

D. KARLSONS received the BSEE degree in 1958 from Drexel Institute of Technology and the MSEE degree in 1960 from the University of Pennsylvania. Mr. Karlsons' experience includes work on early solid-state cavity and traveling-wave microwave masers. He was also instrumental in the development of early working models of the ruby laser. Shortly thereafter he participated in a study evaluating the possibility of the laser's application to ranging and communications. In addition, he developed a Q-controller laser system using a rotating prism. More recently, Mr. Karlsons has participated in the design and construction of a gallium arsenide laser voice-communications system, developing the circuitry needed to drive room temperature laser diodes with very high current, short duration pulses. He has also designed and constructed prototypes, employing the gallium arsenide laser diode, of a radar system, a target indicator, and a proximity sensor. He was also responsible for the development of a portable gallium arsenide laser voice communications system. Presently, he is participating in the development of a three-dimensional laser surveillance system.

Mr. Karlsons is a member of the I.E.E.E., Tau Beta Pi, Phi Kappa Phi, and Sigma Xi.

HERBERT P. MARUSKA received a B.S.E.E. degree, magna cum laude, from New York University in 1965, and an M.S. degree in 1966, also from New York University. He was awarded a National Science Foundation fellowship for his graduate work in the fields of solid-state electronics and computer systems. He joined RCA Laboratories in 1966 under the Research Training Program, during which time he worked on measurement of efficiencies of GaAsP injection lasers, and growth of electroluminescent GaAsP diodes by the vapor-phase-transport technique. He is currently engaged in the preparation of other III-V compound semiconductors by vapor transport.

He is a member of the engineering honor societies Tau Beta Pi and Eta Kappa Nu and of the Institute of Electrical and Electronics Engineers.

H. NELSON received the B.S. degree from Hamline University in 1927 and the M.S. degree in physics from the University of Minnesota in 1929. From 1929 to 1930 he was an engineer with the Westinghouse Lamp Company in Bloomfield, New Jersey. In 1930, Mr. Nelson joined the RCA Electron Tube Division, Harrison, New Jersey, where he engaged in electronic research and development. In 1953 he transferred to the semiconductor research group at RCA Laboratories in Princeton. Mr. Nelson's work has covered almost every important phase of vacuum tube and semiconductor device technology and research. Mr. Nelson is a member of Sigma Xi and of the American Physical Society.

JACQUES I. PANKOVE obtained his B.S. (1944) and M.S. (1948) degrees from the University of California. He received his doctorate from the University of Paris in 1960 for a study of infrared radiation in germanium. Since 1948, when he joined RCA Laboratories, he has made many contributions to the understanding, technology, and evolution of various semiconductor devices, including large-area photocells and transistors. He has worked in the field of superconductivity, studies of silicon carbide, and investigations of the optical properties of degenerate germanium and the electrical properties of tunnel diodes in germanium, as well as in superconductors and in thin oxide layers. Currently, he is concerned with the study of injection luminescence and laser action in gallium arsenide and other compounds.

Dr. Pankove is a member of the American Association for the Advancement of Science, Sigma Xi, and a Fellow of both the American Physical Society and the Institute of Electrical and Electronics Engineers.

ACHILLES G. KOKKAS received the B.S.E.E. degree with honors from Robert College, Turkey, in 1963 and the S.M. degree in Electrical Engineering from MIT in 1966. He joined RCA Laboratories in 1965 as a member of the technical staff research training program. He has worked with silicon-on-sapphire and Gunn-effect devices. At present he is with the integrated electronics group. Mr. Kokkas is an associate member of Sigma Xi.

GILBERT LIEBERMAN received a B.A. degree in Mathematics at New York University in 1948, and an M.A. degree in Mathematical Statistics at Columbia University in 1949. Since then he has done graduate work in Mathematics and Electrical Engineering at University of Maryland, American University, and the University of Pennsylvania. From 1950 to 1955, he was with the U.S. Naval Research Laboratory, Sound Division, Electronics Branch, where he worked on sonar signal processing techniques, analysis of acoustic signal and noise propagation data. From 1955 to 1959, he was with the U.S. Naval Ordnance Laboratory, where he worked on acoustic detection, navigation and guidance systems. During this period, he taught courses in probability and statistics at the University of Maryland. He joined RCA in March of 1959 as a Senior Engineer in the Airborne Systems, Communications Systems Group. His work has involved analysis of digital communication systems, studies of effects of antenna fluctuations and fading on communication system performance, and research on adaptive communication techniques. Mr. Lieberman is a member of the Institute of Radio Engineers, the Institute of Mathematical Statistics, the American Statistical Association, and the Mathematical Association of America.

LLOYD W. MARTINSON received a BSEE degree in 1957 from the University of Wisconsin and an MS in Engineering from the University of Pennsylvania in 1964. Prior to joining the Airborne Systems Division of RCA in 1959 he was a member of the technical staff of the Bell Telephone Laboratories engaged in underwater system development. His work at RCA was initially centered in systems engineering and the development of high capacity signal processing systems with emphasis on optical signal processing. Since 1963 he has been with the Missile and Surface Radar Division engaged in radar, sonar and communications signal processing studies and advanced technique development.

Mr. Martinson is a member of Tau Beta Pi, Eta Kappa Nu, the Institute of Electrical and Electronic Engineers and the AAAS.

CHARLES E. PROFERA received a BSEE and MSEE from Drexel Institute of Technology in 1959 and 1963 respectively. He is currently continuing graduate studies at that institute. Mr. Profera joined RCA Missile and Surface Radar Division, Moorestown, New Jersey in 1959 as a microwave antenna engineer. Since that time he has been concerned with the design and development of various microwave components for antenna feed systems and has been associated with applied research and development programs concentrating on investigations of antenna parameters affecting noise temperature and efficiency. Since 1963, he has been responsible for applied research and development studies of multimode beam-shaping techniques in antenna feeds for high efficiency and low noise temperature, and utilization of these techniques for multiple frequency and very wide-band applications. He is presently responsible for evaluation and development of an original concept to provide a high efficiency array antenna for meteorological measurements. He has also recently been associated with studies for a solid-state array.

ALBERT ROSE received the A.B. degree from Cornell University in 1931 and the Ph.D. degree in Physics in 1935. From 1931 to 1934 he was a teaching assistant at Cornell University. From 1935-1942 he was with the RCA Tube Division in Harrison, N. J., and since 1942 he has been associated with RCA Laboratories at Princeton, N. J. From 1955 to 1957 he directed the research at Laboratories RCA, Ltd., in Zurich, Switzerland. Dr. Rose is a Fellow of the Institute of Electrical and Electronics Engineers and of the American Physical Society.

FRANK V. SHALLCROSS received an A.B. degree in chemistry from the University of Pennsylvania in 1953. He did graduate research in x-ray crystallography and received a Ph.D. in physical chemistry from Brown University in 1958. He was employed as a chemist by M. & C. Nuclear, Inc., Attleboro, Massachusetts, in 1957 and 1958. In 1958 he joined RCA Laboratories in the Electronic Research Laboratory, where he has done research on laser materials, thin-film photoconductors and semiconductors, television camera tubes, thin-film active devices, and integrated circuits. Dr. Shallcross is a member of the American Chemical Society, the American Institute of Chemists, Sigma Xi, and the American Association for the Advancement of Science.

A. H. SOMMER received the Dr. Phil. degree in Physical Chemistry from Berlin University in 1934. From 1935 to 1945, he worked with Baird Television Company, London, in the field of photoelectric and secondary emission. In 1946 he joined EMI Research Laboratories in Hayes, England, to work on photomultipliers and television camera tubes. In 1953 he joined RCA Laboratories, where he is engaged in development and study of new photoemissive materials.

Dr. Sommer is a Fellow of the Institute of Electrical and Electronics Engineers and a member of the Physical Society and of Sigma Xi.

JAMES J. TIETJEN received a B.S. degree, cum laude, from Iona College in 1956 and M.S. and Ph.D. degrees from the Pennsylvania State University in 1958 and 1963, respectively. He was awarded the Speer Carbon Company research fellowship in 1958 and was the recipient of National Science Foundation fellowships in 1959, 1960, and 1961. He has been selected to the honorary societies Pi Mu Epsilon and Phi Lambda Upsilon. His doctoral thesis deals with the theoretical and experimental aspects of thermoelectric effects in graphite and the kinetics of oxygen chemisorption. Since joining the staff of RCA Laboratories in 1963, Dr. Tietjen has worked on the preparation of III-V compound semiconductors, especially GaAs and GaAs-GaP alloys, by vapor transport reactions. In 1966, he was appointed a Research Leader in the Semiconductor Materials Research Group. He is a member of Sigma Xi and the American Chemical Society.

LEONARD H. YORINKS received the degree of Bachelor of Engineering (Electrical) from The Cooper Union in 1965. He is presently completing the requirements for an M.S. degree in electrical engineering at the University of Pennsylvania. Mr. Yorinks joined the Advanced Microwave Technology Group of Missile and Surface Radar Division of RCA in 1965, and has been concerned primarily with the development of high efficiency feed systems for Cassegrain antennas. More recently he has been involved with the analysis and synthesis of a phased-array antenna using large-scale digital computers. Mr. Yorinks is a student member of the Institute of Electrical and Electronics Engineers.