RCA REVIEW

a technical journal

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

VOLUME XXVII JUNE 1966 NUMBER 2

CONTENTS

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

RCA REVIEW

MULTIPLE-ACCESS CONSIDERATIONS FOR COMMUNICATION SATELLITES*

By

F. Assadourian and D. L. Jacoby

RCA Communications Systems Division
New York, N. Y.

Summary—This paper discusses special aspects of multiple access for communication satellites. The major classes of multiple-access systems and their characteristics are briefly reviewed, and a system design philosophy is presented for those systems that share satellite power, i.e., frequency-division and spread-spectrum systems. A practical method is described for establishing nominal design values of satellite sharing factors and transmitter power in terms of other ground and satellite parameters and volume of traffic per link.

Power level control is discussed in theoretical terms for the frequency-division multiplex (FDM) and spread-spectrum modes of multiple access. Up-leg and down-leg path-loss calculations for the simultaneous links lead to a set of generally nonlinear equations, with ground effective-radiated power (ERP) levels as unknowns. Simplifications are possible when the satellite noise is negligible compared to the communication signals entering the satellite. Graphical procedures may then be used to obtain approximate solutions. Power-control dynamic ranges are also discussed.

The final section discusses power-level control techniques that employ pilot signals emanating from ground stations, reference tones inserted at satellite repeaters, and comparative measurements of their returns to ground points. Centralized measurements, decentralized measurements, and programmed approaches are all considered.

Introduction

MULTIPLE access has been the subject of considerable interest and study by those concerned with planning and designing communication satellite systems. An effective multiple-access scheme that permits many users to share a single satellite repeater will reduce the cost of communication satellite services. It will also permit the simultaneous participation of small and large ground terminals having a wide diversity of communication requirements.

The sharing systems that have received the greatest attention fall into three categories: time division (TDM), frequency division (FDM), and spread-spectrum or code-division systems (CDM). The differences between these systems in their use of time, spectrum, and

* Presented at MIL-E-CON 9, Washington, D. C., September 24, 1965.

satellite power are illustrated in Figure 1 through the use of a 3-dimensional domain (light lines) that may be apportioned in various ways for assignment to particular links (dark lines). Each system has merits and shortcomings and there appears to be no universally superior approach. Time-division systems have the advantage of re-

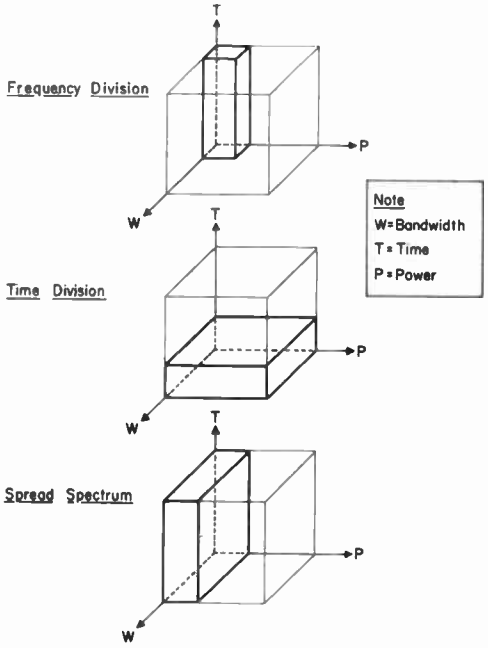

Fig. 1—Multiple-access methods.

quiring no control of ground transmitter power; however, they require accurate synchronization among the participating terminals and are inefficient for small terminals operating through wide-band repeaters. Frequency-division systems are basically the simplest, but they require power control for efficient satellite utilization. Both time-division and frequency-division systems require considerable supervision to permit allocation of time slots or frequency channels on the basis of traffic load. Spread-spectrum systems, by comparison, permit random access and have a nonblocking, graceful overload characteristic. However, they require power control and relatively complex equipment.

Power sharing and the related power-control problem constitute one of the major technical challenges in the implementation of multiple-access systems of the FDM or CDM types. It was therefore

decided to give this subject the greatest emphasis in this study. The systems under consideration are those in which the total available satellite power is split among a number of links. Both frequency-division and spread-spectrum systems fall in this category.

The capacity of a satellite system depends on both the satellite and the ground-terminal parameters. A given satellite may be required to provide many traffic channels when serving a pair of large terminals, but only a few channels when used between small terminals. For the more general case where a mix of small and large terminals are sharing the satellites, the capacity will be related to the number of links between various classes of terminals and the traffic per link. In such a mix, small terminals require a larger share of satellite power than large terminals for the same amount of traffic because of their reduced reception capabilities.

A desirable goal in designing an efficient shared system is that no link should receive a larger or smaller share of the satellite power than is actually required for a given traffic and performance. To accomplish this, a straightforward method is desired for determining the following characteristics for a satellite system:

(a) The share of satellite power required for each receiving terminal in terms of its traffic and required performance.

(b) The power that must be transmitted by the corresponding transmitting terminal.

(c) The capacity of the system in terms of links, terminals, traffic, and performance.

A method will now be described for establishing nominal design values for these factors. In the dynamics of system operation, an active power-control and monitoring system will ensure that the system capacity and performance is maintained under changing conditions. Approaches to power control and system monitoring are discussed later.

Step-by-Step Method

Factors that must be considered are the parameters of the receiving terminal, the traffic and performance level, down-link transmission losses, power sharing at the satellite, the satellite input/output characteristic, up-link transmission losses, and the parameters of the transmitting terminal.

Receiving Terminal

The receiving-terminal capability can be expressed in terms of its sensitivity, S_R, i.e., the effective noise power per cycle of band-

width per unit area at the receiving antenna

$$S_R = -228.6 + 10 \log_{10} T - 10 \log_{10} A \text{ dbw/cycle/ft}^2,$$

where —228.6 is Boltzmann's constant in dbw/cycle/°K, T is over-all noise temperature in °K, and A is effective antenna aperture in ft^2.

S_R is independent of both bandwidth and frequency. For practical purposes a more convenient expression is the ratio D^2/T where D is the antenna diameter in feet and T is again the noise temperature in °K. This ratio may vary from about 0.1 for a small highly mobile terminal to 100 or more for a large fixed installation. It is thus closely related to the size and complexity of the terminal.

Down-Link Transmission Loss

The down-link transmission losses can be expected to vary from link to link and under certain conditions to vary with time over particular links. For the purposes of system design it is desirable to establish a nominal value of transmission loss that does not unnecessarily limit the capacity of the system and is sufficient to assure the desired level of performance under all but extreme conditions. Suggested criteria for nominal transmission loss are:

(a) Assume a free-space loss that corresponds to minimum elevation angle and maximum slant range at which it is proposed to operate.

(b) Include satellite antenna gain reduction corresponding to the conditions of (a).

(c) Include design values of polarization loss, radome loss, and tracking-error loss.

(d) Include atmospheric loss corresponding to clear weather, low humidity, and sea-level at the operating frequency.

Factors such as losses and noise-temperature increases associated with rain, clouds, or fog are not included. It is left for the power-control system to cope with deviations from nominal conditions in the down link, such as reduced losses due to smaller slant range or increased losses due to inclement weather.

An alternative procedure is to base nominal values on conditions that prevail for, say, 50% or more of the time.

Traffic and Performance

We may proceed to specify the received signal power for a given traffic and performance.

$$P_R = -228.6 + 10 \log_{10} T + 10 \log_{10} B + \frac{C}{N},$$

where P_R is in dbw, T is over-all noise temperature in °K, B is predetection bandwidth in cycles/second, and C/N is the carrier-to-noise predetection ratio in db. B and C/N are functions of the modulation technique and the desired traffic and performance.

For the spread-spectrum case, this equation is applicable if we assume that the full processing gain may be applied against the

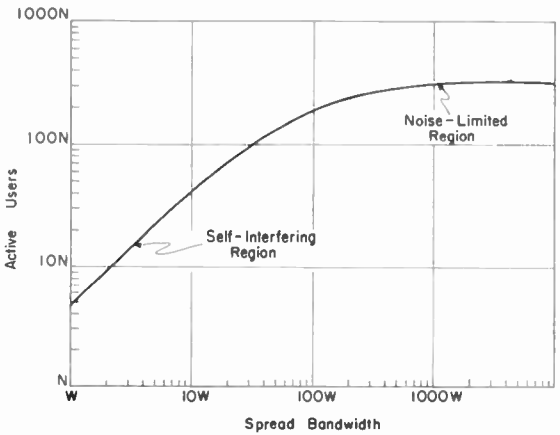

Fig. 2—Spread-spectrum modulation characteristic.

thermal noise and that the interfering signals will have a lesser effect. These conditions tend to prevail for very-wide-band systems with high processing gains. A typical modulation characteristic for a spread-spectrum system is shown in Figure 2. As the spread bandwidth increases, the number of users that can be accommodated levels off, indicating that in this region the thermal noise predominates.

Satellite Power Sharing

We are now ready to specify the required satellite ERP (effective radiated power) in terms of traffic and ground terminal sensitivity as illustrated in Figure 3. If the total available satellite ERP is known, satellite shares can be expressed in percent of the total, and the capacity of the satellite may be specified. Figure 3 assumes a satellite ERP of +30 dbw (20 watts, 17 db antenna gain), a nominal transmission loss of 205 db at 8 gc, and frequency-shift-keyed (FSK) modu-

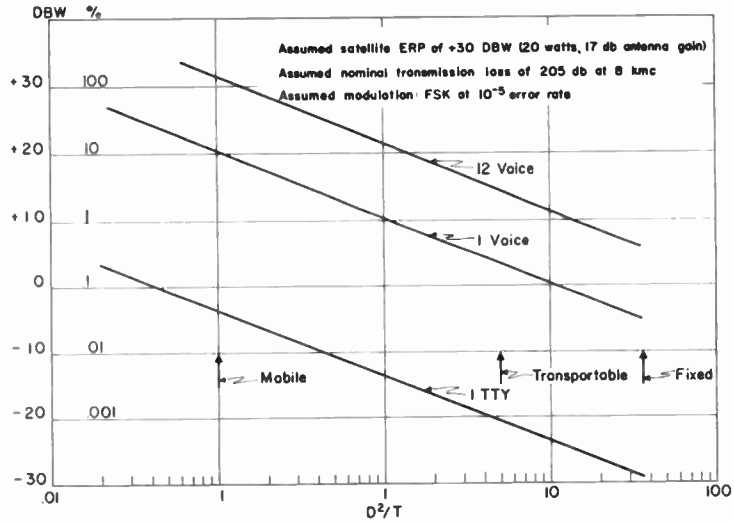

Fig. 3—Satellite sharing factor versus terminal sensitivity and traffic.

lation at 10^{-5} error rate. The resulting satellite ERP shares derived from Figure 3 are given in Table I.

The capacity of the satellite may be expressed in terms of the classes of terminals and the traffic destined for each. Let U and V be the number of TTY (teletype) and voice channels, respectively, destined for mobile terminals. Let W and X be the number of TTY and voice channels, respectively, destined for transportable terminals. Let Y and Z be the number of TTY and voice channels, respectively, destined for fixed terminals.

For the particular satellite we have postulated (ERP of 30 dbw), the condition for a fully loaded satellite occurs when the sum of the

Table I—Satellite ERP Shares

Terminal	D^2/T	Per Voice Channel		Per TTY Channel	
		dbw	%	dbw	%
Mobile	0.12	+20	10	— 4	0.04
Transportable	4.5	+ 4	0.25	—20	0.001
Fixed	36	— 5	0.03	—29	0.00013

individual percentage shares equals 100. The capacity is

$$C = 0.04\,(U + 250V) + 0.001\,(W + 250X) + 0.00013\,(Y + 250Z) = 100\%.$$

For the general case where the links, terminals, and traffic have been specified and we desire to know how large a satellite we need, the required satellite capability is

$$\text{ERP} = 0.4\,(U + 250V) + 0.01\,(W + 250X) + 0.0013\,(Y + 250Z) \text{ watts.}$$

This completes the down-link considerations. As stated previously the shares that have been derived are nominal design values that may be adjusted up or down during operations to reflect deviations from nominal conditions. This problem will be considered as part of the discussion of power control.

Satellite Input/Output Characteristic

For the purposes of this simplified analysis, certain assumptions have been made.

(a) A common wide-band, hard-limiting heterodyne repeater is employed.

(b) A linear power split has been assumed among the signals reaching the satellite. As will be discussed, this is not necessarily realizable, depending on various factors such as relative levels, number of signals, and frequency spacing in FDM systems.

(c) The relative input levels will be determined by the share of satellite power required by the receiving terminal, not by the capability of the transmitting terminal.

(d) The absolute input levels will be sufficient to limit the share of satellite power lost to up-link noise to a tolerable level.

Up-Link Transmission Losses

As in the down-link case, a nominal value of transmission loss is established based on the free-space loss at maximum slant range and with realistic allowances for atmospheric, radome, antenna alignment, polarization and other miscellaneous losses. Under nominal conditions the signal received at the satellite will be at the level required to assure a proper share of the satellite power. At other than nominal conditions, the signal reaching the satellite will be smaller or greater than this level and active power control is necessary to preserve satisfactory system performance.

Transmitting Terminal Parameters

With the transmission loss and up-link noise contributions established, the ground terminal ERP may be determined corresponding to a given share of a fully loaded satellite. The ground-terminal ERP can be satisfied by various combinations of power and antenna gain. Under the assumption that some system of power control is employed, greater than nominal power will be required under adverse conditions (either up or down link) while less than nominal power should be

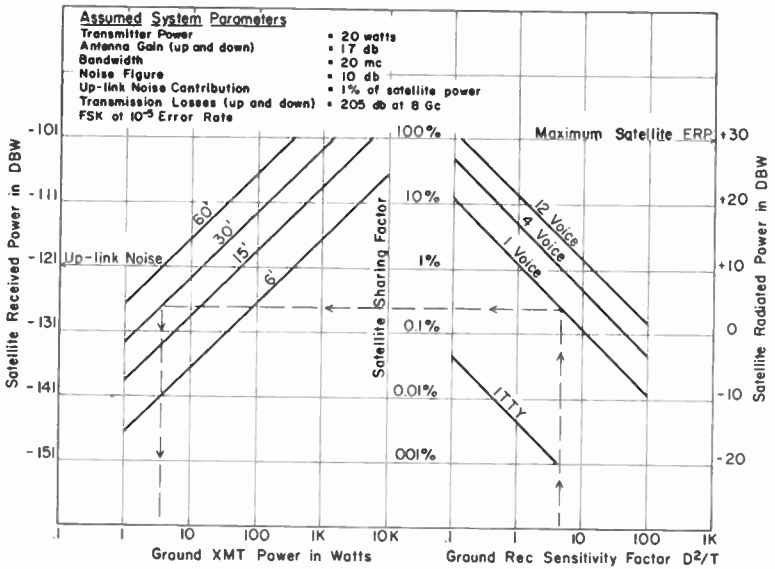

Fig. 4—System design chart.

used under favorable conditions, if the performance and effective utilization of the satellite system are to be maintained.

The transmitting-terminal ERP is directly related to the traffic and to the receiving-terminal sensitivity. If the required ERP exceeds the capability for the class of terminal that is planned, either the traffic must be reduced, a lower level of performance must be accepted, or we can settle for less efficient utilization of the satellite.

The up-link and down-link characteristics may now be combined in a single system design chart. Figure 4 shows the relationships of ground-terminal parameters, satellite power levels, and sharing factors and traffic for a hypothetical satellite system based on the assumptions given in Table II. The assumed up-link noise contribution at

Table II—Assumed Satellite Characteristics

Transmitter Power	= 20 watts
Antenna Gain (up and down)	= 17 db
Bandwidth	= 20 mc
Noise Figure	= 10 db
Up-Link Noise Contribution	= 1% of satellite power
Transmission Losses (up and down)	= 205 db at 8 gc

1% of the satellite power establishes the absolute power levels at which all users of the satellite must transmit. A signal reaching the satellite at a level of —121 dbw will equal the noise and thus receive 1% of the satellite ERP. A composite signal level of —101 dbw would, in effect, fully load the satellite.

As an illustration of use of the chart, consider a requirement for a single voice channel to be transmitted between two transportable terminals. Assume an antenna diameter of 30 feet and a noise temperature of 200°K ($D^2/T = 4.5$). Following the dotted line we see that the satellite ERP required is plus 3 dbw or 0.25% of the available satellite ERP. The corresponding satellite received power is —128 dbw. For a 30-foot transmitting antenna the transmitter power required is 3 watts. Table III illustrates the traffic capabilities of the system between various classes of terminals.

Some of the cases shown have little practical significance, but are included to illustrate the extreme restraint that large stations must exert to permit participation by small stations. For example, a fixed station (60-foot antenna) transmitting to another fixed station would be allowed only 0.1 watt per voice channel, while a mobile terminal

Table III—System Traffic Capabilities

Rec. Term.	Traffic	Sat. Share %	TRANS. TERM POWER 60-foot Antenna	30-foot Antenna	6-foot Antenna
Mobile ($D^2/T = 0.12$)	TTY	0.04	0.15w	0.6w	16w
	1 Voice	10.	40w	150w	4Kw
Transportable ($D^2/T = 4.5$)	1 Voice	0.25	1w	4w	100w
	12 Voice	3.	13w	50w	1.3Kw
Fixed ($D^2/T = 36$)	1 Voice	0.03	0.1w	0.5w	13w
	12 Voice	0.42	1.5w	6w	150w

(6-foot antenna) transmitting to another mobile terminal would be permitted 4 kw per voice channel. The major application of the large terminal would seem to be as the receiving point for transmissions from small mobile terminals so as to reduce the transmission requirements for the small terminal.

In summary, a simple method has been described for determining for any satellite system employing multiple access, the nominal sharing factors as they relate to the traffic and ground terminals, the corresponding ground-terminal transmitting powers, and the capacity of the satellite in terms of participating ground terminals and traffic. The considerations basic to an optimum system design are (1) that satellite sharing factors for each link depend solely on the traffic, performance, and sensitivity of the receiver terminal, and (2) that ground transmitter power be limited to that needed to achieve the required satellite sharing factor for a particular link. A chart similar to that shown here wlil give the system designer at least a first cut at the optimum operating parameters of the system.

We will now proceed to an analysis of power sharing in greater technical depth and to consider the difficult problem of dynamic power control.

Theoretical Power-Level-Control Considerations

The equitable sharing of an active satellite repeater by a number of simultaneous users in the FDM and spread-spectrum modes of multiple access is facilitated by proper control of all signal levels. Control points may be located in the satellite or in ground-station transmitters (in this paper, only the latter case is considered). It is assumed that the various multiple-access participants compete with each other and, possibly, with satellite thermal noise for satellite ERP, but do not have to contend with outside interference.

Reasonable criteria for the fair distribution of satellite ERP among various users, which can be uniformly "small" or "large" or a mixture of both types, would include the following as far as possible:

(1) The satellite ERP intended for a given user should be a function of his channel requirements and should increase or decrease as they increase or decrease, although not necessarily in a linear fashion.

(2) Proper cognizance should be taken of the individual user's channel performance requirements.

(3) For efficient operation, the satellite repeater should be loaded by communication signals to prevent the competing satellite noise from capturing an undue portion of the ERP. An exception is noted later.

Previous thinking has usually simplified the power-control problem by assigning equal signal levels to like users at the repeater input, and in many situations this solution may represent a practical compromise. However, in situations that permit computer-type calculations, the second criterion above suggests that a fairer solution is available. The computer approach is feasible in systems that tend to employ computers for other purposes (such as for satellite acquisition and tracking) at either local sites or control centers.

A step-by-step development is presented for either the spread-spectrum or FDM mode of multiple access. In essence, the usual path-loss analysis for satellite links is applied to all multiple-access (one-way) links to form a set of nonlinear algebraic equations with ground transmitter ERP's as the unknowns. The solutions to these equations represent the correct values of these ERP's that guarantee the desired performances at all receiving sites. Changing the solutions in certain ways, such as multiplying all ERP's by a common scale factor, will improve performance. Since such improvement may not be required, nominal ERP's corresponding to nominal system conditions will generally be adequate, with provision for varying these ERP's in accordance with changing path losses over some dynamic range.

FDM Mode of Multiple Access

Link calculations must take into account the following effects at the satellite repeater:

(1) sharing of satellite ERP among various users, with loss of power to satellite thermal noise as one of the users in case of inefficient operation,

(2) loss of linear power due to out-of-band frequencies that are generated by repeater nonlinearity,

(3) enhancement of differences in incoming user signal levels, with satellite thermal noise included as a user, produced by repeater nonlinearity (weak signal suppression),

(4) noise increase in user communication channels due to intermodulation distortion products resulting from repeater nonlinearity.

FDM here refers to the assignment at any given time of separate r-f frequency slots in the satellite repeater bandwidth to the various simultaneous users. The spread-spectrum (SS) mode here refers to the simultaneous assignment of the entire repeater bandwidth to several users for communications via pseudorandum trains of pulsed signals (FSK, PSK, etc.). Of the effects listed above, the first and third are most important for the SS mode; in the case of FDM, the fourth effect is also important.

Up-leg and down-leg calculations are now tabulated for the FDM case, with the up-leg results applicable to the SS case as well.

FDM Up-Leg Calculations for i-th (One-Way) Link

The ERP from the ground station is

$$\alpha_i = P_{Ti}\, G_{Ti}\,.$$

The signal level at the satellite input, therefore, would be $a_i\alpha_i$, where a_i is a factor that includes up-leg path loss and satellite antenna gain.

The satellite ERP for the i-th link si given by

$$\gamma_i = \frac{k_i\, a_i\, \alpha_i\, W_s}{N_s + k_i a_i \alpha_i + \sum\limits_{i \neq j} a_j \alpha_j}$$

where k_i is the signal degradation (or improvement) due to repeater nonlinearity according to effect (3) noted previously; N_s is the satellite thermal noise in the bandwidth, B_s; and W_s is the total satellite ERP (P_sG_s) given by

$$W_s = \sum_{1}^{n} \gamma_j + \gamma_N$$

γ_N (with associated k_N) represents noise ERP with k_N determined by the signal k's.

Intermodulation distortion noise ERP, γ_{IM}, for the i-th link is also a factor in the FDM case, but can be made negligible if repeater bandwidth is sufficient to permit appropriate spacing of user channels.

Satellite repeater gain for the i-th link is given by

$$g_i = \left(\frac{1}{a_i\alpha_i}\right)\left(\frac{\gamma_i}{G_s}\right).$$

FDM Down-Leg Calculations for i-th (One-Way) Link

The signal level at the ground input is $\gamma_i b_i$, where b_i is a factor that includes down-leg path loss and ground antenna gain.

The up-leg thermal noise in the i-th link is given by

$$N_{ui} = b_i \gamma_N \frac{B_{Ri}}{B_s},$$

where B_{Ri} is the r-f bandwidth of the receiver, which, in the case of binary PSK transmission, is set equal to the bit rate, R_i. The intermodulation noise from the satellite repeater is $\gamma_{IM} b_i$.

The down-leg noise in the i-th link is given by

$$N_{di} = n_{di} B_{Ri},$$

where n_{di} includes external noise (sky, atmospheric, and ground reflections), internal noise (pre-receiver losses, front-end noise from parametric amplifier and front-end stages), and other noise (e.g., coupling of transmitting power into receiver).

The total receiver noise in the i-th link, therefore, is

$$N_i = N_{ui} + b_i \gamma_{IM} + N_{di},$$

and the required carrier-to-noise ratio for the i-th link is

$$\rho_i = \frac{b_i \gamma_i}{N_i}.$$

Specification of C/N performance for n links leads to n nonlinear equations that are to be solved for the various ERP's from the ground stations. For efficient satellite repeater operation, the power sacrificed to satellite noise should be small. For this case, the n nonlinear unhomogeneous equations approach n linear homogeneous equations. This limiting case is consistent with physical intuition, which would predict that, if satellite noise is disregarded, the multiplication of each ground ERP by the same constant should not alter the power-control balance.

Spread-Spectrum Mode of Multiple Access

The following major effects must now be observed in link calculations:

(1) sharing of satellite ERP among various users,

(2) enhancement of differences in incoming signal levels (and satellite noise, if not negligible) produced by repeater nonlinearity (e.g., hard limiting),

(3) processing degradation at ground receiver, reducing the processing gain from its design value.

Up-leg calculations for the SS mode are the same as those for the FDM case (except that intermodulation distortion need not be considered). Down-leg calculations are as follows,

SS Down-Leg Calculuations for the i-th (One-Way) Link

As before, the signal level at ground input is $b_i\gamma_i$. The up-leg thermal noise in the i-th link is

$$N_{ui} = b_i\gamma_N + \sum_{j \neq i} b_j\gamma_j,$$

where

$$\gamma_N = W_s - \sum_{1}^{n} \gamma_j.$$

The down-leg noise in the i-th link (effect (2)) is

$$N_{di} = n_{di}B_R,$$

where B_R is the receiver bandwidth, which is equal to both the satellite bandwidth and to the unprocessed (fast) bit rate in the example of binary PSK.

The spread-spectrum processing gain of the receiver, $k_{PG}(PG)$, is the product of the ratio of the SS to information bandwidths, (PG), and a degradation factor k_{PG} (effect (3) noted previously). The SS processing relationship of the receiver, then, is

$$\left(\frac{C}{N}\right)_{\mathrm{BW}=r} = k_{PG}(PG)\left(\frac{C}{N}\right)_{\mathrm{BW}=R},$$

where r is the information rate.

From the above, total receiver noise in the i-th link is

$$N_i = N_{ui} + N_{di}.$$

The unprocessed CN ratio for the i-th link is $b_i\gamma_i/N_i$, but the condition to be satisfied for the i-th link, i.e., processed C/N ratio, is

$$\rho_i = k_{PG}(PG)\,\frac{b_i\gamma_i}{N_i}$$

As in the FDM case, specified performances on n links yield n nonlinear equations to be solved for the various α_i.

Implications of Theoretical Calculations

The previous calculations have shown that rigorous approaches to the power-control problem for the FDM and SS modes of multiple

access result in sets of nonlinear equations in the various ground ERP levels. Although no attempt has been made to solve these equations, it is believed that approximate procedures described elsewhere in this paper represent practical solutions. Furthermore, with the theoretical formulations as guides, a number of features may be observed, as noted below.

Special Case of Two Signals

In the special case of a single two-way link (or its equivalent of two one-way links) where the path loss is different for the two halves of the link, two courses of action are considered. First, power control may be omitted. In this case, with identical terminals assumed at the ends of the link, the weaker signal arriving at the satellite is further reduced by, say, hard limiter or other repeater nonlinear degradation, but regains lost ground on the superior down-leg. Second, let power control guarantee equal levels at the satellite by increasing the ERP for the poorer up-leg. This approach unjustly favors the signal on the better down-leg. A more equitable form of power control, perhaps along the lines suggested by the previous theoretical treatment, would seem to be preferable. The more general case of two or more two-way links leads intuitively to similar conclusions.

Power-Control Dynamic Ranges

One of the practical questions to be settled in any power-control implementation is the selection of operating regions defined by dynamic ranges around selected "nominal" values. A reasonable approach might be to determine these values from conditions (satellite range, atmosphere losses, satellite antenna and repeater parameters, antenna polarization and tracking losses, ground receiver noise, number of channels and required performance etc.) that can be expected to prevail for, say, 50% or more of the time. As noted earlier, nominal values were previously determined in a different manner. Unsymmetrical dynamic ranges may now be inserted on either side of the nominal values, with a compressed lower range, to accommodate changes in transmission conditions and parameters. The upper end of the dynamic range is fixed by the extent of degradation (anticipated during some percentage of time) under which satisfactory performance is desired. It is clear that an extensive range may be costly in that, most of the time, a transmitter may be compelled to operate at fractions of its peak capability.

The influence of the number of multiple-access signals on the selection of a dynamic range for the power control of a particular signal

can be explained with the aid of Figure 5. First, a few facts are recalled from the earlier theoretical discussion. Assuming that the number of multiple access signals may vary from time to time, the i-th link must be prepared to operate when it is the only user or when it is one of 2 to n users. It will be seen below that there is a problem of efficiency of operation.

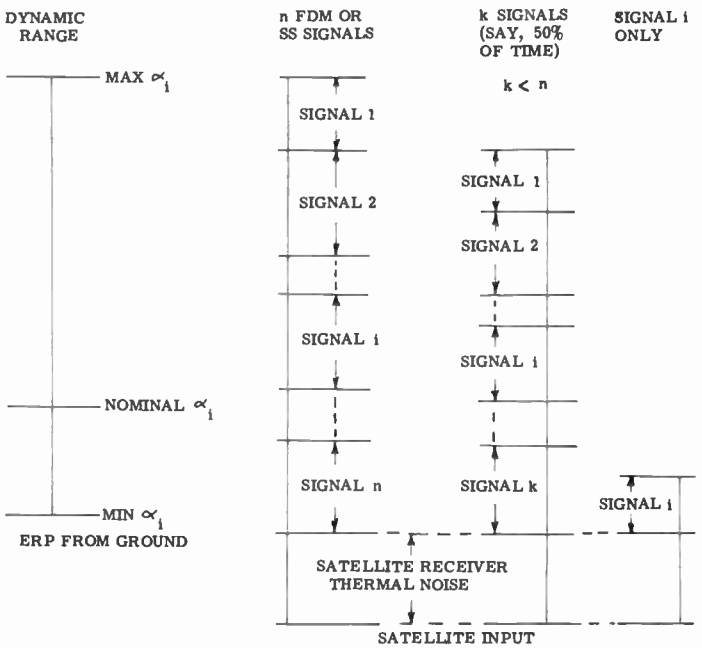

Fig. 5—Power-control dynamic range considerations.

To start with, let the satellite repeater be fully loaded by signal inputs from all n (one-way, or $n/2$ duplex) links, and let their sum be much larger than the satellite thermal noise. The relative signal levels are assumed to lead to corresponding portions of satellite ERP that are satisfactory for link operation considering both capacity and performance. With hard limiting, the relative signal levels will not change greatly unless some signals are very large compared to others. Now assume that "nominal" operation (say for 50% of time or more) involves k links ($k < n$), each with an associated nominal number of channels. Then the i-th signal level may be used in the setting of the nominal ground ERP of α_i.

The influence of the number of multiple-access signals on the

selection of dynamic ranges for power control is now illustrated by the simple case of Table IV for equal signals and noise. Here n is 10 (maximum) and k is 6 (nominal). It is seen that the presence of all n links calls for an increase in ground ERP by a factor of 11/7. Also, when only the i-th link is present, the ground ERP can be decreased by a factor of 7/2. Of course, an alternative in the latter case would be to leave ground ERP unaltered, and to use the increased share of satellite ERP for more traffic. This approach may be more practical in that it may limit the dynamic-power-control problem. Table IV

Table IV—Dynamic Range Considerations

	Maximum	Nominal	i-th
Number of equal satellite input signals	10	6	1
Satellite noise equal in level to any signal	1	1	1
Approximate portion of satellite ERP per signal	1/11	1/7	1/2
Factor needed to restore nominal portion of sat. ERP	11/7 increase		7/2 decrease

must be corrected if limiter degradation effects are considered. Also, in the case of hard limiting, it is assumed that full output (past the "knee" of the characteristic) is obtained when the input consists only of satellite thermal noise, and hence when at least one input signal is present in addition.

Figure 5 suggests an observation about power-control accuracy and discipline. When only a few links are sharing the satellite repeater simultaneously, their relative input signals into the repeater must be controlled more closely than when there are many links. This is so because, when there are many signals, large variations around correct levels can be tolerated for a few signals (or even for many signals if the sum of the variations is small) without appreciably affecting the portion of satellite ERP intended for a given link. This conclusion should remain approximately correct even when nonlinear repeater effects on weak signals are considered, so long as the total input satellite power does not change appreciably with respect to a desired input signal.

Rate of Change of Power-Control Requirements

Since practical power-control schemes, no matter how fine-grained, may depend on discrete data received periodically, the required frequency of such data and required system reaction time to these data may be important parameters. Both types of data obviously depend upon the intended application and may typically range from a fraction of a second to several minutes.

Power-Level-Control Techniques

Techniques for implementing power-level control can be realized with the use of two types of pilot signals. One type serves as a monitor of link transmission conditions. For this purpose a pilot signal can be transmitted toward the satellite from each ground station participating in multiple access. The pilot signals arriving at the satellite have relative levels that convey information concerning relative up-link losses. When retransmitted to a ground measurement point this information undergoes uniform attenuation. After its detection, relative information collected for the up-links is now available. When applied in reverse and combined with known system parameters, the same information enables the determination of all relative down-link conditions. If each pilot signal also contains information on the number of channels and performance requirements for its link, then the total information at hand should be sufficient for the setting of relative ground ERP levels, at least to within a constant factor when satellite noise is negligible.

The second type of pilot signal is inserted at the satellite as a reference against which the other pilot signals can be compared at the ground measurement point. Information gained in this fashion is more complete than previous data derived from only the ground pilot signals and should be useful in fixing ground ERP levels. There are several approaches for the insertion of the reference. For example, in the case of a hard-limiting repeater, the reference can be added to the ground pilots and can then proceed through the limiter with them and the communication signals. Another approach is to provide a separate path, with or without its own hard limiter, through the repeater for just the ground and reference pilots. Any method adopted must yield information on how the hard-limiting affects the multiple access signals, since this data is needed for the estimation of correct ground ERP levels.

An illustrative procedure is to introduce a satellite reference tone that is high compared to the other pilot signals and noise. With a separate hard limiter for all pilot tones, the ratio of the satellite refer-

ence to any other pilot signal will then increase by around 6 db. After ground detection and subtraction of these 6 db, the results can be compared to "nominal" values and, after theoretical or approximate computations, used to set the proper values of α_i.

Centralized Power-Control Measurements

Based on the previous discussion, a technique for centralized power-control measurements is illustrated on Figure 6. To repeat, pilot signals are transmitted from each multiple access ground terminal and

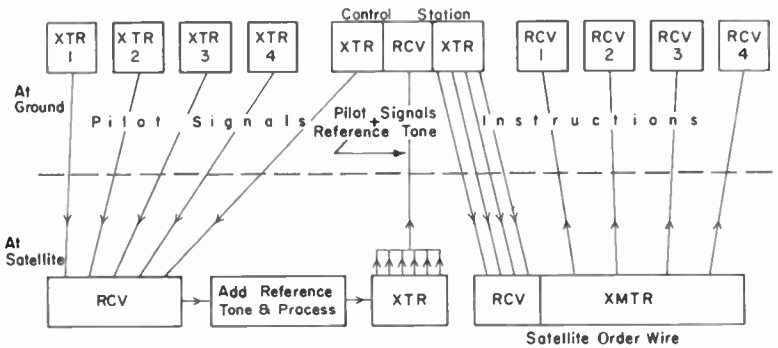

Fig. 6—Centralized power-control measurements.

the control station, with the former containing data on numbers of channels occupied and required performance over each link. The control station collects this information plus relative measurements derived from the reference pilot tone and the pilot signals. Approximate or rigorous computer calculations are made to predict correct values of ground station ERP levels. These settings are sent to each ground station via satellite order wire.

Decentralized Power-Control Measurements

Decentralization in the previous approach can be achieved by receiving the pilot signals and satellite reference tone at all ground sites (Figure 7). Local computations with the data derived from these signals can then be used to adjust local ERP levels. This approach makes sense if approximate rather than large-scale computer calculations are to be made, since the latter would favor the centralized technique typified in Figure 6.

Ideal decentralization would permit each ground station to make measurements (say with only its own pilot signal, a pilot signal from

the opposite end of the link, and a satellite reference) and calculations and then set its ERP without resorting to measurements involving other ground stations. Such an approach is not possible in the case illustrated by Figure 5, which assumes interdependence among the various links. However, ideal decentralization may be approached if it is possible to make a particular ground station share of satellite ERP nearly independent of the other links. In this case, the illustrative quantities in Table IV indicate that such a share might range from 1/11 to 1/2 of the satellite ERP. Hence, if the former were to yield acceptable link performance, the latter should yield superior performance.

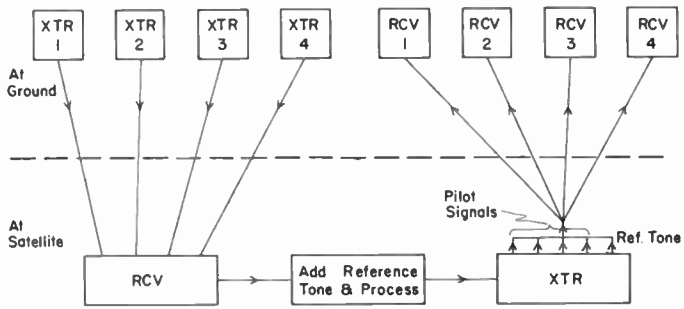

Fig. 7—Decentralized power-control measurements.

Programmed Power Control

The previous approaches involve data collection, measurement, computation, distribution (in the centralized case) and application in ERP adjustments, all in fairly short times, say not longer than a few minutes. In the centralized case, if a computer is to be employed for long-term ephemeris calculations and network scheduling, it may be feasible to add the function of preparing long-term power-control schedules for multiple-access applications. These might then be entered in satellite books (along with satellite schedules, acquisition, and tracking data) and stored at the various satellite link terminals. The power-control schedules would cover situations involving nominal system parameters, but could not cope directly with short-term weather and other factors. However, it might be possible to make provision for such factors locally at the various ground sites during the course of system operation.

MICROWAVE POWER GENERATION USING OVERLAY TRANSISTORS

By

H. C. Lee

RCA Electronic Components and Devices
Somerville, N. J.

Summary—The high-power capability of the overlay transistor has been extended into the microwave region. With developmental transistors, cw power output of 6.5 watts at 1 gc and over 1 watt at 2 gc has been obtained. This paper describes the state-of-the-art performance of microwave power transistors. Device considerations and circuit design approaches at microwave frequencies are presented. Circuit details and performance data are given for practical circuits. The relative merits of common-base and common-emitter configurations of microwave power transistors are compared.

Introduction

CONTINUED DEVELOPMENT of the overlay transistor has extended its high-power capability into the microwave region. Commercial microwave transistors, such as the 2N4012, are now available for use at frequencies up to 1.4 gc. The typical cw power output of this device varies from 3 watts at 1 gc to 1.5 watts at 1.3 gc. With developmental transistors, cw power outputs of 5 watts at 1 gc and 1 watt at 2 gc have been obtained. Future prospects are even more promising.

Microwave power can be generated by operation of overlay transistors in either the fundamental-frequency mode or the harmonic-frequency mode. The former includes straight-through amplifiers and fundamental-frequency oscillators, and the latter includes amplifier–multipliers and oscillator–multipliers. Both modes of operation are important in microwave applications. Amplifiers and amplifier–multipliers are used where crystal-controlled stability is required. Long-term stability up to 1 part in 10^7 is possible in an amplifier–multiplier chain driven by a crystal-controlled oscillator. Fundamental-frequency oscillators and oscillator–multipliers are used in applications where cost is a prime consideration. In these circuits, stability is reduced to the order of 1 part in 10^3.

Operation of the overlay transistor in the harmonic-frequency mode can extend the upper limit of the frequency range far beyond that possible from the same transistor operating in the fundamental-

frequency mode. A further advantage of the harmonic mode is that frequency multiplication and power amplification can be realized simultaneously.[1,2] An overlay transistor operating in this mode provides power amplification at the fundamental frequency of the input-drive power, and the nonlinear capacitance of the collector-to-base junction, acting as a varactor, generates harmonics of the input-drive frequency. It is possible, therefore, to use a single transistor to replace a transistor power amplifier and a varactor-diode frequency multiplier.

This paper describes the use and capabilities of overlay transistors in these circuits, which are the essential building blocks for solid-state microwave, radiosonde, and new telemetry equipments. Device considerations and circuit-design approaches with overlay transistors at microwave frequencies are presented, and some basic problems that may be encountered in such circuits are discussed. Circuit details and state-of-the-art performance data are given for practical circuits using commercial and developmental devices. The relative merits of common-base and common-emitter configurations of overlay transistors in such circuits at microwave frequencies are compared.

Device Consideration

The important performance criteria in microwave power-amplifier and frequency-multiplier circuits are power output, power gain, and efficiency. Transistors suitable for power amplification must be capable of delivering power efficiently, with sufficient gain at the frequency of interest. Transistors suitable for frequency-multiplication must also be capable of efficiently converting the power from the fundamental frequency into power at the desired harmonic frequency.

The power output that can be obtained from a transistor is determined by the current- and voltage-handling capabilities of the device at the frequency of interest. Because high-voltage operation of transistors is usually not practical, high power output requires a transistor that has a high current capability. The overlay transistor is suitable for high-power operation at ultra-high and microwave frequencies, because its construction provides a substantial increase in the ratio of emitter periphery to emitter area and, therefore, high current-handling capability.[3]

---

[1] M. Caulton, H. Sobol, and R. Ernst, "Generation of Microwave Power by Parametric Frequency Multiplication in a Single Transistor," *RCA Review*, Vol. XXVI, p. 217, June 1965.

[2] H. C. Lee and R. Minton, "Designing Transistor Multipliers," *Microwaves*, Vol. 4, p. 18, Nov. 1965.

[3] D. R. Carley, P. L. McGeough, and J. F. O'Brien, "The Overlay Transistor, Part 1: New Geometry Boosts Power," *Electronics*, Vol. 38, p. 71, Aug. 23, 1965.

A simple expression for power gain, G, of a transistor power amplifier is

$$G = \frac{|Y_{21}|^2 R_L}{4R_e(Z_{in})} \tag{1}$$

where Y_{21} is the dynamic current-transfer ratio (current gain), R_L is the real part of the collector load impedance determined by the required power output, and Z_{in} is the dynamic input impedance when the collector is loaded by a complex load impedance, Z_L. This expression shows that, for high-gain operation, a transistor should have high and constant current gain with current level.

For efficiency considerations, the transistor is usually operated under the signal-bias condition at which the collector-to-base junction is reverse-biased and the emitter-to-base junction is forward-biased partially with the input-drive signal. The collector efficiency of a transistor amplifier is defined as the ratio of the r-f power output at the frequency of interest to the d-c input power. Therefore, high efficiency implies that (1) circuit loss should be minimum, (2) the ratio of the real part of the dynamic output admittance of the transistor and its collector load conductance should be maximum at the frequency of interest, and (3) the circulation of harmonic current in the circuit should be minimized. For efficient operation, a transistor with small dynamic output conductance should be used in a low-loss circuit that can block the flow of all harmonic currents. However, transistor output conductance increases very rapidly with frequency. If a high ratio of output conductance of the transistor and its collector load conductance is to be maintained for efficient operation, the transistor should be operated under the conditions of a distorted-pulse collector-voltage waveform and a nearly-sinusoidal current waveform. This mode of operation can be considered as the opposite of vacuum-tube operation, in which the plate voltage is sinusoidal and the plate current is in pulses.

The efficiency of a transistor frequency-multiplier circuit can be considered as the product of the collector efficiency of the amplifier portion discussed previously and the r-f conversion efficiency that results from the varactor action of the collector-to-base capacitance. The latter is limited by the loss associated with the nonlinear capacitance. Overlay transistors are suitable for frequency multiplication, because the loss in the nonlinear collector-to-base capacitance is minimized as the result of the emitter structure. The cross-sectional view shown in Figure 1 indicates the loss distribution in an overlay tran-

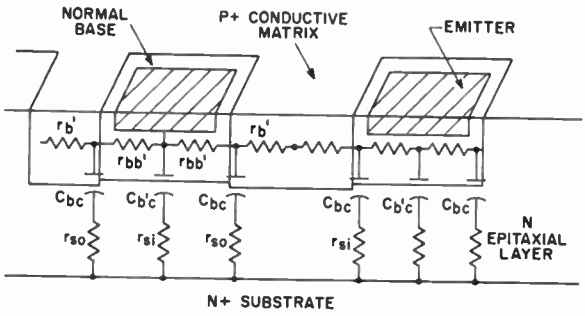

Fig. 1—Cross-sectional view of an overlay transistor.

sistor. Figure 2 shows how the varactor portion separates from the intrinsic transistor portion. The active portion of the varactor, C_{bc}, is made up of the capacitance formed by the collector-to-base junction not opposite the emitter sites. The capacitance $C_{b'c}$ consists of the part of the collector-to-base junction opposite the emitter-base junction. C_{bc} is much larger than $C_{b'c}$ and is a much more efficient varactor. Capacitance C_{bc} can charge and discharge only through the extrinsic base-spreading resistance $r_{b'}$ and the series resistance r_{so}, not through the intrinsic base-spreading resistance $r_{bb'}$. Because $r_{bb'}$ is much greater than the combination of $r_{b'}$ and r_{so} in the overlay transistor,[4] the varactor loss is minimized and, therefore, high conversion efficiency can be achieved.

The inherent varactor frequency-multiplication ability of the collector-to-base junction capacitance, added to the excellent frequency capability of overlay transistors, has made possible the use of overlay devices as efficient frequency multipliers, as well as power amplifiers.

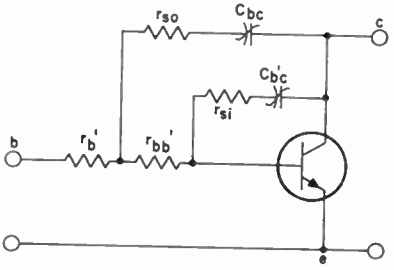

Fig. 2—Simplified circuit of an overlay transistor.

[4] H. C. Lee and G. J. Gilbert, "Overlay Transistors Move into Microwave Region," *Electronics*, Vol. 39, p. 93, March 21, 1966.

Circuit Design Approach

The design of transistor microwave-power circuits involves two steps: (1) the determination of load and input impedances under dynamic operating conditions, and (2) the design of properly distributed filtering and matching networks required for optimum circuit performance. For design of the input circuit, it is necessary to know the input impedance at the emitter-to-base terminals of the packaged transistor at the drive-power frequency under operating conditions. To design the output circuit, it is necessary to know the load impedance presented to the collector terminal at the fundamental frequency for amplifiers and at the harmonic frequency for multipliers. These dynamic impedances are difficult to calculate at microwave frequencies, because values of transistor parameters such as Y_{21} in Equation (1) vary considerably under large-signal operation from those for small-signal operation, as well as changing with power level. Small-signal equations which might serve as useful guides for transistor design cannot be applied rigorously to large-signal circuits. Because an equivalent circuit for large-signal operation of microwave transistors has not yet been developed, transistor dynamic impedances are best determined experimentally by use of slotted-line measurement techniques.

Determination of Transistor Dynamic Impedances

The system required to determine transistor impedances under large-signal operating conditions is shown in Figure 3. As is shown, the signal generator is well isolated by a pad. The input directional coupler is used to monitor the input reflected power, and the output directional coupler to monitor the output waveform or frequency. At a given frequency and input-power level, the input and output tuners are adjusted for maximum power output and minimum input reflection power. Once the system has been tuned properly, the impedance across terminals 1,1 in Figure 3, without the transistor in the system, is measured at the same frequency in a slotted-line set-up. The conjugate of this impedance equals the dynamic impedance of the transistor. Similarly, the impedance across terminals 2,2, without the transistor in the system, is the dynamic load impedance presented to the collector of the transistor. Such measurements can be performed at each power level.

In addition to determining the dynamic input impedance and the load impedance, the system of Figure 3 can also be used to determine

the power output, power gain, and efficiency of a transistor. When the system is being set up to test transistor performance characteristics, careful consideration must be given to the selection of the line length and the characteristic impedance of the input and output line sections. For the input section, the low-characteristic-impedance line should be used to reduce the VSWR because the real part of the transistor input impedance is very small at microwave frequencies. A line length is

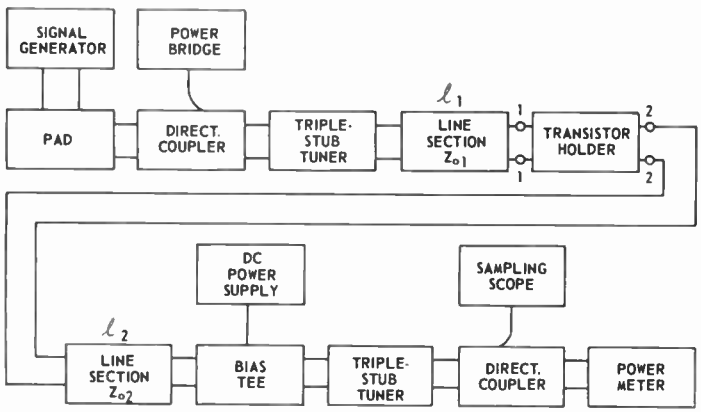

Fig. 3—Set-up to measure transistor impedances.

chosen that will transform the small input resistance of the transistor to a value closer to that of the driving source resistance. For the output section, a line with a proper length and characteristic impedance (Z_0 close to the load resistance) should also be used to reduce the VSWR.

For frequency multipliers, an additional arrangement must be incorporated to provide circulation of fundamental-frequency current in the collector of the transistor. An idler circuit might be required, depending on the order of multiplication.[2,4] The output circuit is designed to pass the desired harmonic frequency.

Designing Circuits

Once the dynamic input impedance and the load impedance of a packaged transistor are established, the input and output circuits can be designed with the aid of a Smith chart. Although various forms

of circuits using either air-line or strip-line arrangements can be used for microwave applications, this paper discusses only the simple design shown in Figure 4. In this figure, two input circuits using direct coupling are shown at (a) and (b). Output circuit (d) provides

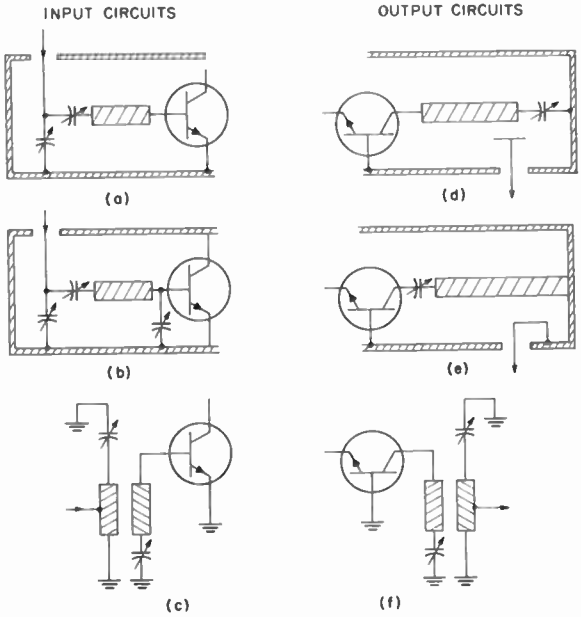

Fig. 4—Various forms of input and output circuits.

capacitive-probe coupling and output circuit (e) uses inductive coupling. The input circuit at (c) and the output circuit at (f) utilize two sections of resonant line. In all these circuits, either air-line or strip-line arrangements can be used; the former are superior in performance, while the latter are preferred for economy.

The design techniques for air-line and strip-line circuits are very similar. The design of an amplifier circuit at a frequency of 2 gc will be described here as an example. The dynamic input impedance Z_{in} and collector load admittance Y_{CL} of the transistor to be used (measured in the system of Figure 3 at 2 gc) are

$$Z_{in} = 7.5 + j\,10.2 \tag{2}$$

$$Y_{CL} = 8.3 \times 10^{-3} - j\,15.3 \times 10^{-3}. \tag{3}$$

must be modified by a factor of $1/\sqrt{e}$, where e is the dielectric constant of the insulator material of the strip-line.

Microwave Power Amplifiers

The complete circuit diagram of a 2-gc power amplifier using an RCA developmental coaxial transistor is shown in Figure 7. The coaxial transistor is placed in series with the center conductor of the line, or cavity, and its base is properly grounded to separate the input

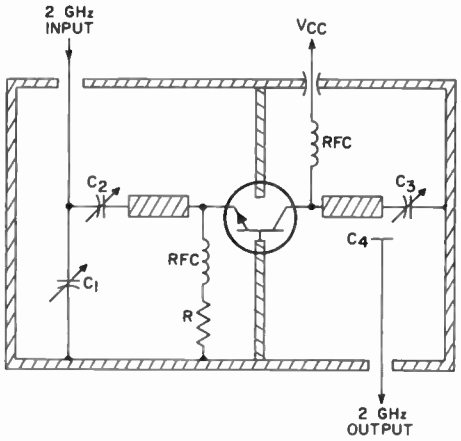

Fig. 7—Circuit diagram of a 2-gc power amplifier.

and output cavities. The input section consists of capacitances C_1 and C_2 and a line section. Direct coupling is used in the input. The output section consists of a line section, capacitance C_3, and a capacitive probe C_4 through which the output power is obtained. The two r-f chokes and the resistance R form the bias circuit.

A c-w output power of 1.0 watt with 4 db gain has been obtained in this circuit, with the grounded-base coaxial transistor operated at a collector voltage of 28 volts. The collector efficiency measured at an output-power level of 1.0 watt is better than 25%. The 3-db bandwidth measured at the same power level is greater than 6%.

A c-w power output of 1.0 watt was also obtained at 2 gc in this amplifier circuit with a grounded-emitter coaxial transistor. However, the measured power gain was somewhat less (3 db). The collector efficiency of the common-emitter amplifier is comparable to that of a common-base circuit. Figure 8 shows the output power as a function of the input power for the 2-gc amplifier circuit operated at a collector supply voltage of 28 volts.

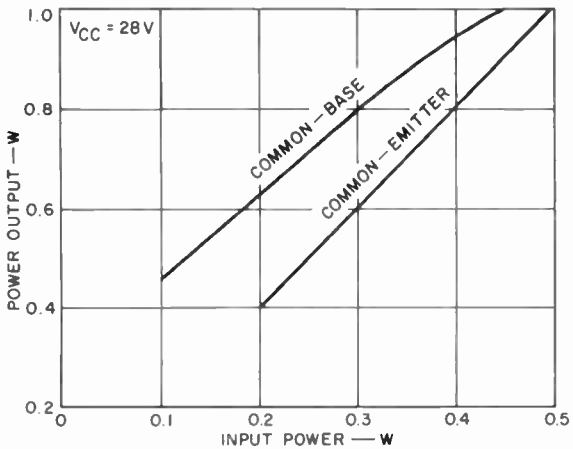

Fig. 8—Power output versus power input for the 2-gc amplifier.

The variation of output power with frequency at power levels of 0.2 and 0.45 watt is shown in Figure 9 for the common-base coaxial transistor operated at a collector voltage of 28 volts. The device delivers 2.65 watts of power at 1 gc with a collector efficiency of 45%.

A transistor capable of higher power output at 1 gc is now in

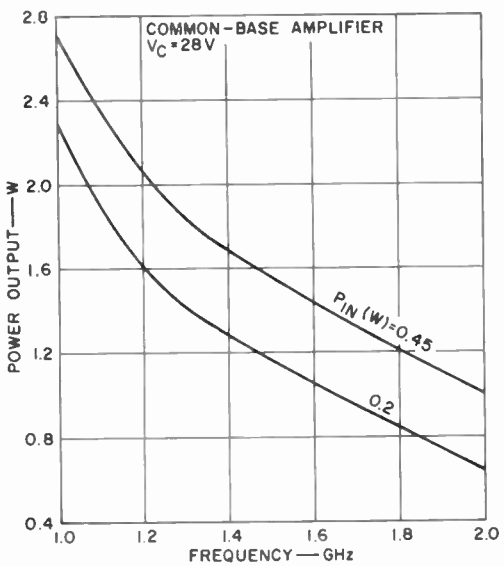

Fig. 9—Power output versus frequency for the common-base coaxial transistor.

development. A power output of 5 watts has been measured in an amplifier circuit similar to that of Figure 7 for a common-emitter transistor in a TO-60 package with a drive of one watt.

Amplifier–Frequency-Multipliers

Figure 10 shows a transistor tripler circuit that uses a selected 2N4012 for operation from 367 mc to 1.1 gc. This circuit uses the

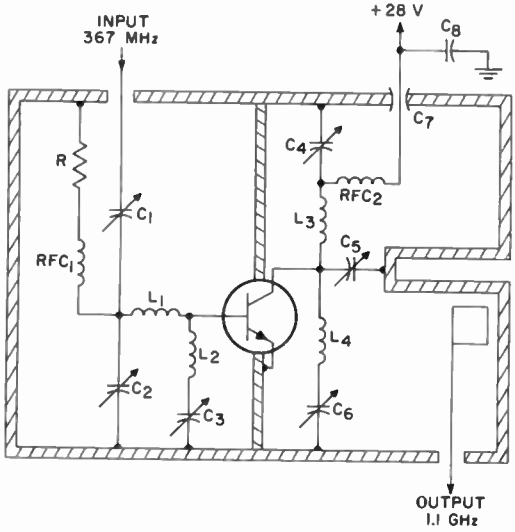

Fig. 10—A 367-mc to 1.1-gc transistor tripler circuit.

air-line output circuit shown in Figure 4(e) and lumped-element input and idler circuits. The transistor is placed inside the cavity, and its emitter is grounded to the chassis. A pi section is used in the input (C_1, C_2, L_1, L_2, and C_3) to provide impedance matching at 367 mc between the driving source and the base–emitter terminals of the transistor. L_2 and C_3 provide the necessary ground return for the nonlinear capacitance of the transistor. L_3 and C_4 form the idler loop for the collector at 367 mc. L_4 and C_6 form the second-harmonic idler circuit for the collector at 734 mc. The output circuit consists of a foreshortened coaxial cavity. A lumped capacitance C_5 in series with a hollow-center conductor of the cavity near the open end permits adjustment of the electrical length. Output power at 1.1 gc is obtained by loop coupling at a point near the shorted end of the cavity.

Figure 11 shows the output power of this tripler at 1.1 gc as a function of the input power at 367 mc for a selected 2N4012 operated at a collector supply voltage of 28 volts. The solid-line curve is the power output obtained when the circuit is retuned at each input-power level. The dashed-line curve is obtained with the circuit tuned at an

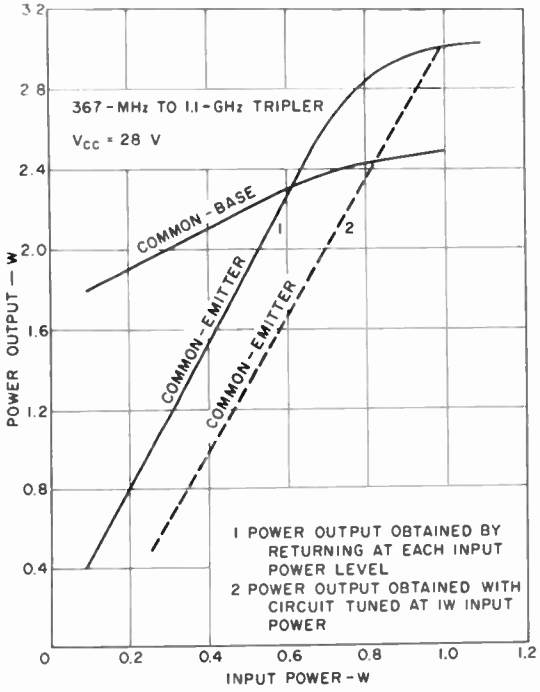

Fig. 11—Output power versus input power for the transistor tripler circuit.

output level of 3.0 watts. An output power of 3.0 watts at 1.1 gc is obtained with a drive of 1.0 watt at 367 mc. The 3-db bandwidth measured at this power level is 2.3%. The spurious-frequency components measured at the output are down more than 20 db. The variation of output power and collector efficiency with collector supply voltage at an input-drive level of 1.0 watt is shown in Figure 12. These curves are obtained with the circuit tuned when the collector voltage is 28 volts and the power output is 3.0 watts. Figures 11 and 12 indicate that the transistor amplifier–multiplier circuit is capable of amplitude modulation.

A 367-mc amplifier that used the same circuit configuration and

components shown in Figure 10 was constructed to compare amplifier and tripler performance. The conversion efficiency for a large number of transistors was measured. (The conversion efficiency is defined as the power at 1.1 gc obtained from the tripler, divided by the power at 367 mc obtained from the amplifier with the same input power of 1.0 watt.) The efficiency varies between 60 and 75%, comparable to that of a good varactor in this frequency range.

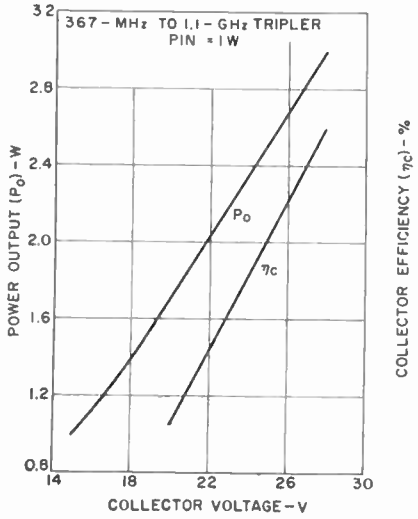

Fig. 12—Power output and collector efficiency versus collector voltage for the tripler circuit.

In comparison with varactor frequency-multiplier circuits, the transistor multiplier is simpler, less costly, and equally efficient. It is anticipated that this mode of operation will permit extension of the available frequency spectrum for overlay transistors by a factor of two.

Common-Emitter Versus Common-Base Operation

The choice of transistor configurations at microwave frequencies is based on both performance and stability. From the performance point of view, common-base amplifiers and amplifier–multipliers provide higher gain, and common-emitter circuits produce higher output-power capability. Figures 8 and 11 show the performance comparison between a common-base and a common-emitter configuration for a 2-gc amplifier and for a 367-mc to 1.1-gc amplifier–tripler, respectively.

It is apparent that output powers of the common-emitter tripler saturate at a much higher input-power level than those of common-base circuit. Bandwidth and collector efficiency for common-base and common-emitter circuits at microwave frequencies are found to be about the same.

From the stability point of view, it has been generally accepted that a common-emitter configuration will provide a more stable circuit up to the VHF range. This assumption arises from the linear analysis of transistors, in which parasitic elements are not included. However, for high-power operation at ultra-high and microwave frequencies, transistor parasitics contributed by the package must be considered. Moreover, r-f power transistors must be considered as nonlinear devices rather than as four-terminal linear devices.

Certain forms of instability can be incurred in both common-emitter and common-base circuits, e.g., hysteresis, parametric oscillations, and low- and high-frequency oscillations. Usually, most of these difficulties can be eliminated or minimized by careful design of the bias circuit, by proper location of the transistor ground connections, and by use of packages with minimum parasitic inductance and capacitance.[2] Stable operation has been obtained at 2 gc with both common-base and common-emitter coaxial transistors.

Fundamental-Frequency Oscillator and Oscillator-Frequency Multiplier

The high-frequency oscillations discussed previously usually occur at a frequency close to the output frequency of the amplifier or the amplifier–multiplier when the input power is removed. This form of instability can be attributed to the parasitic elements of the package that set up the frequency of oscillation with the intrinsic transistor. Such parasitic elements can be used to form an economical transistor oscillator operated at microwave frequencies, provided the frequency of oscillation can be controlled.

Figure 13(a) shows a Colpitts transistor oscillator suitable for microwave application. The inductance L and the capacitances C_1 and C_2 can be considered as the parasitic elements of the package. The transistor in this oscillator circuit can be grounded in high-frequency operation at the collector, the base, or the emitter without effect on its performance. For example, a useful oscillator circuit can be derived from the basic Colpitts oscillator by the use of a TO-39 transistor. In Figure 13(b), the collector of such a transistor is returned to ground through the collector parasitic inductance L. This connection is a convenient method of applying a heat sink to the collector, which is

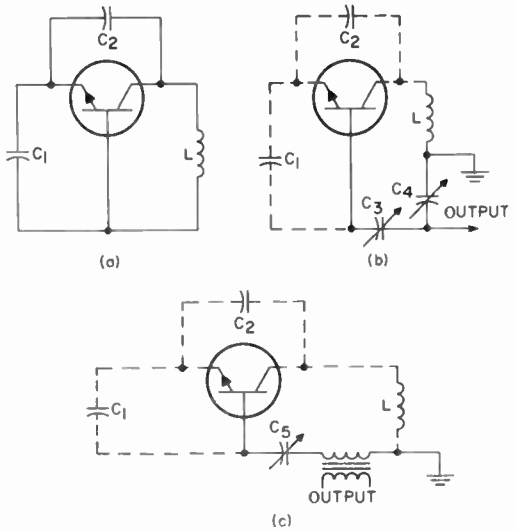

Fig. 13—Transistor oscillator circuits.

connected to the case in a TO-39 package. Output power is obtained from the base through capacitances C_3 and C_4. Figure 13(c) shows another method of coupling power output from the oscillator.

Figure 14 shows the complete circuit diagram of a 1.68-gc fundamental-frequency oscillator. An output power of 0.45 watt has been obtained in this circuit using a developmental transistor; the efficiency

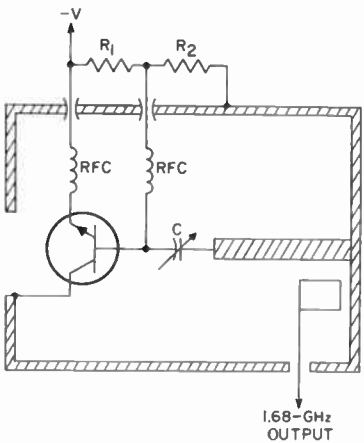

Fig. 14—1.68-gc fundamental-frequency transistor oscillator circuit.

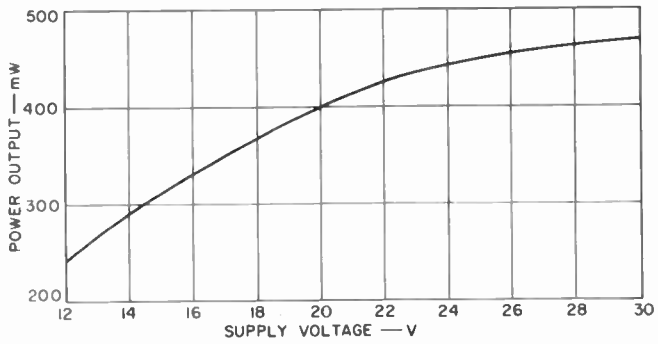

Fig. 15—Power output versus collector voltage for the 1.68-gc oscillator circuit.

is 20% at a collector supply voltage of 25 volts. The variation of output power with collector supply voltage is shown in Figure 15. This oscillator circuit can be operated from 1.5 gc to 2.2 gc, and can deliver 150 milliwatts at 2.2 gc.

The inherent varactor frequency-multiplication ability in overlay transistors, as discussed previously, has also made it possible to use these devices as oscillator–multipliers. Figure 16 shows an oscillator–quadrupler circuit that uses a selected 2N3866 transistor. This circuit can deliver more than 300 milliwatts at 1.68 gc. The two r-f chokes and R_1 and R_2 form the bias circuit. The fundamental frequency of the oscillator is 420 mc, as determined by C_0, L_1, and C_1. L_2 and C_2 form the second-harmonic idler. The second-harmonic component produced by this idler circuit beats with the fundamental-frequency component to generate additional fourth-harmonic components. A series-tuned circuit that consists of L_3 and C_3 completes the output circuit.

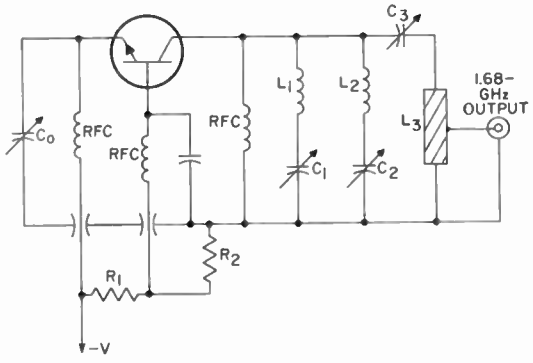

Fig. 16—A 420-mc to 1.68-gc oscillator–quadrupler circuit.

THE LANGMUIR CURRENT LIMIT FOR DIFFERING AXIAL AND RADIAL ELECTRON BEAM TEMPERATURES IN HIGH-RESOLUTION IMAGE DEVICES*

BY

JOAN LURIE

RCA Astro-Electronics Division
Princeton, N. J.

Summary—The Langmuir current limit was originally derived for the case of a beam in zero magnetic field and for a single beam temperature. However, realistic electron beams are characterized by distinct axial and radial temperatures and are usually focused by magnetic fields. In this paper, an analogous current-density limit is derived for a solid electron beam characterized by two temperatures. Both uniform and spatially varying magnetic fields are considered for unintercepted beams and for beams that pass through a limiting aperture. A typical result, for the case of a beam with no limiting aperture is

$$j = \frac{j_0 \alpha}{T_z - T_r}\left\{ T_z - \frac{\alpha T_r T_z}{T_z + (\alpha - 1)T_r} \exp\left(-\frac{eV_0}{k}\left[\frac{1}{\alpha T_r} - \frac{1}{\alpha T_z} \right] \right) \right\}$$

where α is the ratio of magnetic field at the target to that at the gun and is greater than unity, j_0 is the emitted density, and T_r and T_z are the radial and axial temperatures, respectively.

INTRODUCTION

THE presence of a thermal velocity distribution in an electron or ion beam imposes an upper limit on the ratio of current density that can be collected at an anode to the emitted density. This result follows from statistical mechanics alone and is due to the randomness of the velocity distribution. The upper limit perists even in the absence of space-charge effects and in the best conceivable focusing systems. An expression for the maximum current density obtainable at a point whose potential is V volts above cathode was first derived by D. B. Langmuir[1] in 1937 for the case of a beam characterized by one temperature and a Maxwellian velocity distribution. For a Maxwellian distribution of speeds, the differential axial current density is

* This work was supported by the Air Force Avionics Laboratory, Air Force Systems Command, Aeronautical Systems Division, Wright-Patterson Air Force Base, Chio under Contract No. AF33(657)-11485.

[1] D. B. Langmuir, "Theoretical Limitations of Cathode-Ray Tubes," *Proc. I.R.E.*, Vol. 25, Aug. 1937.

$$dj_z = N_0 ev \cos\theta \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) \overrightarrow{dv}$$

where $\overrightarrow{dv}$ is the differential volume element in velocity space. Using this expression and integrating over speeds and angles, Langmuir obtained

$$j_m = j_0 \left[1 + \frac{eV}{kT}\right] \sin^2\theta, \tag{1}$$

where dj_z is the current density contributed to the beam by electrons emitted by the cathode (at temperature T) that have speeds v in the range $\overrightarrow{dv}$, N_0 is the number density of the emitted electrons, j_m is the maximum current density that can be collected in a cone of semi-vertical angle θ at the anode at potential V, k is Boltzmann's constant, e is the charge on an electron, m is the mass of an electron, and j_0 is the current density emitted by the cathode. The limit given by Equation (1) cannot be reached in practice. It is attainable only in an aberrationless focusing system in the absence of space charge.

In realistic electron-beam devices, it is often found that the beam must be characterized by two temperatures, one in the axial and one in the radial direction. The relation of these two temperatures to the velocity distributions in the two directions is described more completely in the Appendix. In many devices (e.g., high-resolution television camera tubes or dielectric tape cameras), the radial temperature is typically about equal to the cathode temperature, but the axial temperature may be up to several times larger. The cause of this condition has not been definitely established, but it is probably due, at least in part, to the "patchiness" of the work function on the cathode surface.

The electron-beam devices with which we have been concerned employ magnetic fields to focus the beam. In cases where a single beam temperature can be used to describe the velocity distribution of the electrons, the focusing field does not affect the exponent in the expression for dj_z. Thus Equation (1) is valid whether or not magnetic focusing is employed, although the factor $\sin^2\theta$ will be different for different focusing conditions. However, in the presence of two velocity distributions, the focusing field plays an important role in determining the limiting conditions.

In the following sections, we will consider the effect of the two beam temperatures and of constant and spatially varying focus fields on the limiting current density. The results have practical significance for relatively low voltage (~500 volts or less) high-resolution devices where thermal velocities cannot be considered negligible. The problem of limiting current density is important for high-resolution devices because, in addition to the small spot sizes required, a relatively large current is necessary to obtain the required sensitivity. This paper deals only with magnetically focused systems. However, it has been pointed out to the author* that the results may be readily generalized to apply to any type of focusing field.

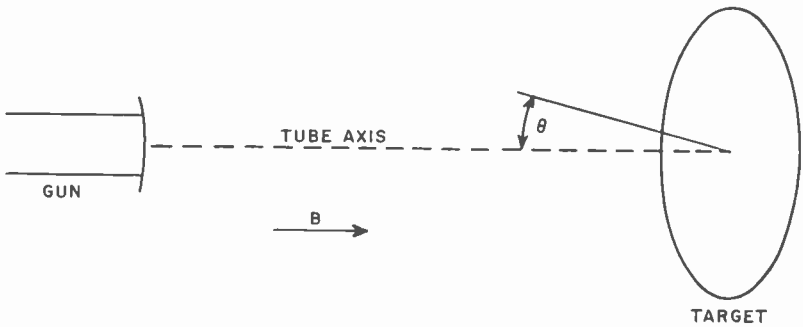

Fig. 1

Constant Magnetic Focus Field

The first case to be considered is the current density limit for an electron beam focused by a constant magnetic field. The geometry is cylindrical with B parallel to the z axis. The electrons are emitted from the cathode and accelerated by a uniform electric field. The geometry is shown in Figure 1. Note that electrons are accepted at the target from a cone of semivertical angle θ.

In a cylindrically symmetrical system with unequal radial and axial temperatures, T_r and T_z, the differential axial current density emitted by the cathode is given, in spherical coordinates, by

$$dj_z(v) = 2\pi eCv^3 \sin\theta \cos\theta \exp\left\{-\frac{mv^2}{2k}\left(\frac{\cos^2\theta}{T_z}+\frac{\sin^2\theta}{T_r}\right)\right\} dv d\theta, \quad (2)$$

* E. G. Ramberg, private communication.

where the integration over the azimuthal angle ϕ has been carried out, C is related to the total cathode current density, and the cathode potential is equal to zero. Spherical coordinates are used because the target landing angle θ enters more naturally than in cylindrical coordinates. The total cathode current density j_0 is obtained by integrating Equation (2) over all angles and all speeds of emission. Then

$$j_0 = \frac{2\pi eC}{m^2} kT_z kT_r,$$

or

$$C = \frac{m^2 j_0}{2\pi e kT_z kT_r}.$$

The differential current density collected at a point at potential $V > 0$ is, using the expression for C,

$$dj_z(v) = \frac{m^2 j_0}{kT_r kT_z} \exp\left\{-\frac{mv^2}{2k}\left(\frac{\cos^2\theta}{T_z} + \frac{\sin^2\theta}{T_r}\right) + \frac{eV}{kT_z}\right\} v^3 \sin\theta \cos\theta \, d\theta dv. \tag{3}$$

As in the case of a single beam temperature, the magnetic field does not appear in the expression for the differential current density. The electron speeds appear only in the radial and axial energies and the total particle energy is unaffected by a magnetic field.

The total current density is obtained by integrating Equation (3) over all speeds to which a point at potential V is accessible. The result for the axial current density j_z is

$$j_z(V) = \int_0^{\theta} \int_{\sqrt{2eV/m}}^{\infty} dj_z(v)$$

(where electrons are accepted from a cone of semivertical angle θ) or, performing the indicated integrations,

$$j_z(V) = \frac{j_0}{T_z - T_r}\left[T_z - \frac{\exp\left\{\frac{eV}{k}\left(\frac{1}{T_z} - \frac{1}{T_r}\right)\sin^2\theta\right\}}{\frac{(1-\sin^2\theta)}{T_z} + \frac{\sin^2\theta}{T_r}}\right]. \tag{4}$$

Equation (4) is valid for all values of the angle θ. In practice, however, θ is held to very small values, because the exiting beam is intercepted by a small limiting aperture. $\mathrm{Sin}^2\theta$ is given by the ratio of the largest radial energy accepted at the target to the potential V. For a constant magnetic field the target radial energy is the same as the radial energy at the limiting aperture of the gun. For the devices studied here the radial energy has been determined from a computer analysis of the trajectories in the gun. The resulting values of $\sin^2\theta$ are $10^{-4} < \sin^2\theta < 10^{-3}$. A Taylor expansion of Equation (4) can be performed to obtain a simpler form of the current-density limit in those cases where $\sin^2\theta$ is small. The result, retaining terms to first power in the expansion parameter $\sin^2\theta$, is

$$j = j_0\left[\frac{T_z}{T_r} + \frac{eV}{kT_r}\right]\sin^2\theta. \tag{5}$$

It is interesting to compare this expression with the corresponding limit calculated by Langmuir for the case of a single beam temperature. If, in Langmuir's result, we take $T = T_r$ in accordance with the observation that T_r is approximately the cathode temperature,

$$\frac{j}{j_{\mathrm{Lang}}} \approx 1.$$

Thus, for typical low-current, high-resolution devices where $eV \sim 5000\ kT_r$, this calculation predicts the same limiting current density as Langmuir's.

In devices in which V, the target voltage, approaches zero (e.g., the image orthicon), $eV < kT_z$ and

$$\frac{j}{j_{\mathrm{Lang}}} = \frac{T_z}{T_r},$$

so that this calculation predicts a higher limiting current density.

Spatially Varying Magnetic Focus Field

The next case to be considered is the axial-current-density limit in a device with an axial magnetic focus field that increases in the space between the gun and the target. This type of field is often used in

camera tubes and dielectric-tape cameras because it demagnifies the image on the target of the electron-beam crossover in the gun.* Since a varying magnetic field also correlates the axial and radial energies (which are independent in the constant-field case), the magnetic field enters explicitly in the calculation of the current-density limit. This complication does not arise if one beam temperature is assumed, because then only the total electron energy enters the expression for current density.

All of the calculations in the following section have been carried out in the adiabatic approximation. It is assumed that the flux linked by a charged-particle orbit is constant, or, equivalently, that v_r^2/B is constant, where v_r is the total electron speed in the plane perpendicular to the tube axis (z direction). (The magnetic field is parallel to the z axis except for a small component required to satisfy $\vec{\nabla} \cdot \vec{B} = 0$.) This approximation implies rather slow variations in B and will give better results for larger magnetic fields or for beams with smaller thermal velocity distributions.

The current density at the target or other collecting plane will be calculated in terms of the electron speeds and landing angles at that plane. If α is the ratio of the magnetic field at the target to the magnetic field at the gun exit, then, from the adiabatic approximation, the square of an electron's radial velocity at the target is α times the square of the radial speed at the gun exit,† i.e. $v^2\sin^2\theta = \alpha v_0^2\sin^2\theta_0$, where the zero subscripts refer to cathode variables. Then, from conservation of energy, the axial component of emission velocity is related to the target speed by

$$v_0^2\cos^2\theta_0 = v^2\left(1 - \frac{\sin^2\theta_0}{\alpha}\right) - \frac{2eV}{m},$$

where v_0 is the magnitude of the thermal emission velocity, V is the target potential, and θ_0 is the angle of emission at the gun exit.

The current density at the target, contributed by electrons with speeds v in dv and landing angles θ in $d\theta$ is

---

* As seen below, this demagnification does not necessarily imply an increase in current density.

† The quantity α is related to magnification; in the adiabatic approximation, the increasing magnetic field demagnifies the beam according to $B(Z_1)/B(Z_2) = \text{Area}(Z_2)/\text{Area}(Z_1)$. Thus, α is the reciprocal of the square of the linear magnification and the calculation could be simply extended, following Langmuir, to cases where focusing is achieved by using electric fields or combinations of E and B.

$$dj_z(v) = \frac{m^2 j_0}{kT_r kT_z} \exp\left\{ -\frac{1}{2}\frac{mv^2 \sin^2\theta}{\alpha kT_r} - \frac{1}{2}\frac{mv^2}{kT_z}\left[1 - \frac{\sin^2\theta}{\alpha}\right] + \frac{eV}{kT_z}\right\} v^3 dv \sin\theta\cos\theta d\theta. \quad (6)$$

The form of the exponent is an extension of a result of kinetic theory. If a beam is emitted with a Maxwellian distribution at a temperature T and is then accelerated through a potential V, the resulting distribution function is

$$n(v)\vec{dv} \propto \exp\left\{ -\frac{1}{2}\frac{mv^2}{kT} + \frac{eV}{kT}\right\}\vec{dv}.$$

This result and the relation $dj_z(v) = n(v)v\cos\theta(\vec{dv}/2\pi)$ leads to Equation (6).

An electron will land at the target if its energy in the target plane is at least equal to eV, where V is the potential of the target relative to the cathode. If there are no apertures in the system, electrons will be accepted from all angles θ between zero and $\pi/2$. Thus the total current density at the target is

$$j_z(V) = \int_0^{\pi/2} \int_{\sqrt{2eV/m}}^{\infty} dj_z(v) \quad (7)$$

where $dj_z(v)$ is given by Equation (6). The integral over the angle θ is of the form $\int e^{-ax}dx$, where $x = \sin^2\theta$, and is easily performed, giving

$$j_z(V) = \exp\left\{\frac{eV}{kT_z}\right\} \frac{mj_0\alpha}{k(T_z - T_r)} \int_{\sqrt{2eV/m}}^{\infty} vdv \left(\exp\left\{-\frac{mv^2}{2kT_z}\right\} \left[1 - \exp\left\{-\frac{mv^2}{2k}\left(\frac{1}{\alpha T_r} - \frac{1}{\alpha T_z}\right)\right\}\right]\right)$$

or, performing the integration,

$$j_z(V) = \frac{j_0\alpha}{T_z - T_r}\left[T_z - \frac{\alpha T_r T_z}{T_z + (\alpha - 1)T_r} \exp\left\{-\frac{eV}{k}\left(\frac{1}{\alpha T_r} - \frac{1}{\alpha T_z}\right)\right\}\right]. \quad (8)$$

In the limit $\alpha \rightarrow 1$, this result is identical with the result of the constant-field calculation, Equation (4).

If Equation (8) is evaluated in the limit $T_r \rightarrow T_z$ (without letting $\alpha \rightarrow 1$), one obtains

$$j = j_0\left(1 + \frac{eV}{kT}\right),$$

the result obtained by Langmuir. Physically, this result is not unexpected. If an increasing focus field is used, the increase in current density due to the decrease in beam area is compensated by a decrease in axial speed at the expense of the increased radial speed. An alternative method of explaining this result is to note, in the exponent in Equation (8), that the factor α appears in the denominator. Thus the increasing magnetic field leads to an apparent temperature increase.

Equation (8) is chiefly of academic interest. In all high-resolution devices, the beam is intercepted in the gun–target space by one or more apertures that limit the landing angles to very small values. Typical values of the landing angle were quoted in the preceding section. To obtain the density limit as a function of angle, the upper limit in Equation (7) is changed from $\pi/2$ to θ. In this case,

$$j_z(V) = \frac{\alpha j_0}{T_z - T_r}\left[T_z - \frac{\alpha T_r T_z}{\beta T_z + (\alpha - \beta)T_r}\exp\left\{-\frac{eV\beta}{k\alpha}\left(\frac{1}{T_r} - \frac{1}{T_z}\right)\right\}\right] \tag{9}$$

where $\beta = \sin^2\theta$. In the limit of small β, a Taylor expansion in powers of β can be performed. The resulting current density is

$$j_z(V) = j_0\beta\left[\frac{T_z}{T_r} + \frac{eV}{kT_r}\right], \tag{10}$$

so that to the first power in β, the expression for the current density is not changed by the increasing magnetic field.

In typical high-resolution devices $eV/(kT_r)$ is about 400 times as large as T_z/T_r, so that radial properties of the beam dominate the device performance.

Conclusions

One of the striking features of these expressions is their relative independence of the axial-beam properties, as is apparent in Equations

(5) and (10), where only the very small term T_z/T_r depends on the axial-beam properties. This is particularly fortunate for the designers of high-resolution camera tubes and other devices, since they can control the radial properties of the beam fairly easily by using apertures, but relatively little can be done about axial properties. An independent calculation of focusing properties using individual electron trajectories* has confirmed this result.

From Equation (10), it appears that the use of an increasing magnetic-focus field in a practical device with a small aperture does not lead to an increase in the attainable current density. However, the maximum value of β increases as the ratio of radial energy at the target to total energy. For a fixed magnetic field at the gun exit, an increase in target magnetic field (higher α) will, therefore, lead to higher current density. In practice, the use of an increased focus field in high-resolution devices does indeed lead to higher resolution at constant beam current, indicating that the current density has been increased.

Acknowledgment

The author is grateful to E. G. Ramberg for his interest and assistance in the course of this research.

Appendix

The beam temperature is a measure of the width of the velocity distribution. Quantitatively, the definition of beam temperature is not complete unless the fraction of electrons emitted with energies less than or equal to kT is known. In the present case there will be two numbers, one defining the radial and one the axial temperature.

The fraction of electrons emitted with radial energies less than or equal kT_r is given by

$$n_{kT_r} = \frac{N_{\leq kT_r}}{N_{\text{total}}} = \frac{\displaystyle\int_0^{\sqrt{2kT_r/m}} v_r dv_r \exp\left(-\frac{mv_r^2}{2kT_r}\right)}{\displaystyle\int_0^{\infty} v_r dv_r \exp\left(-\frac{mv_r^2}{2kT_r}\right)}$$

* S. M. Siskind, private communication.

or

$$\frac{N_{\leq kT_r}}{N_{\text{total}}} = 1 - \frac{1}{e} = 63.2\%.$$

Thus, 63.2% of the electrons emitted from the cathode have radial energies less than or equal to the beam's radial temperature.

The equivalent number for the axial direction is

$$n_{kT_z} = \frac{\int_0^{\sqrt{2kT_z/m}} v_z dv_z \exp\left(-\frac{mv_z^2}{2kT_z}\right)}{\int_0^{\infty} v_z dv_z \exp\left(-\frac{mv_z^2}{2kT_z}\right)} = 63.2\%.$$

DEMODULATOR THRESHOLD PERFORMANCE AND ERROR RATES IN ANGLE-MODULATED DIGITAL SIGNALS*

BY

J. KLAPPER

RCA Communications Systems Division
New York, N. Y.

Summary—A theory is developed that provides a link between the threshold mechanism in analog FM signals and error rates in digital FM reception. A main result is a formula that predicts error rates for binary FM signals with limiter–discriminator reception and integrate–dump decision. Unlike the work of earlier investigators, this formula is also applicable to large deviation indices, is relatively simple, and facilitates the inclusion of center-frequency shifts. Several important illustrations are included.

INTRODUCTION

AVAILABLE formulas for predicting error rates in digital FM reception do not include the effect of post-detection processing of the digital signals.[1-3] The difficulty is due to the non-gaussian distribution of the post-detection noise. However, some form of post-detection low-pass filtering is essential for all except small deviation indices that, in turn, are associated with appropriately narrow-band predetection filtering. System constraints, such as those due to Doppler shifts and frequency instabilities, often dictate the use of predetection bandwidths that are substantially wider than the bit rate. A main result in this paper is a formula that permits the prediction of the probability of error in such systems. In addition, this formula is simple to use and permits the ready inclusion of center-frequency shifts, such as those due to the Doppler effect.

Additive gaussian (not necessarily white) noise is assumed to be the sole source of interference. The derivation is based on the work

* The work reported herein was sponsored by the AF Avionics Laboratory, Wright Patterson AFB, Ohio, under Contract AF 33 (615) 2426.

[1] A. A. Meyerhoff and W. M. Mazer, "Optimum Binary FM Reception Using Discriminator Detection and IF Shaping," *RCA Review*, Vol. 22, p. 698, Dec. 1961.

[2] P. D. Shaft, "Error Rate of PCM-FM Using Discriminator Detection," *Trans. IEEE PGSET*, Vol. SET-9, p. 131, Dec. 1963.

[3] W. R. Bennett and J. Salz, "Binary Data Transmission by FM over a Real Channel," *Bell Syst. Tech. Jour.*, Vol. 42, p. 2387, Sept. 1963.

of Rice[4] and on experimental evidence that the noise at the output of a limiter–discriminator for additive gaussian-noise contamination of the input is comprised mainly of small, nearly gaussian, noise with superimposed, randomly occurring threshold impulses. These impulses are all approximately of unity area in the output-versus-time plane if the ordinate is calibrated in terms of frequency deviation. For the limiter–discriminator, however, they are not necessarily of identical shape. Rice has postulated that an impulse is generated each time the vector resulting from the addition of instantaneous noise to the instantaneous signal encircles the origin. The speed of encirclement is a function of the instantaneous signal and noise conditions, and generally differs from case to case. Thus the shapes of the impulses are not expected to be identical. The encirclement produces a 2π jump in the phase-versus-time plane, and differentiation with respect to time yields a spike of unity area in the frequency-versus-time plane. Since these spikes are all of essentially the same area and are relatively sharp, they are referred to herein as impulses. According to present theory these impulses are the cause of the well-known FM threshold phenomenon for analog signals.[4] When the rate at which these impulses occur is small, their effect on the output signal-to-noise power ratio is negligible. The analog FM threshold is caused by a certain rate of these impulses, and thus they are referred to herein as threshold impulses (TI).

In angular feedback demodulators, the TI rate is considerably reduced by the compressive and filtering actions of the demodulator. However, a different phenomenon occurs, "loss-of-lock", that also results in 2π phase jumps and impulsive post-detection noise. This noise is referred to as loss-of-lock impulses (LLI). In a well-designed angular feedback demodulator, the combined effect of TI and LLI is less than that of TI in a limiter–discriminator (LD). Thus, the noise picture analysis for the phase-locked loop and FM feedback types of low-threshold demodulators is similar, although a new mechanism of impulse creation appears. The LLI of the phase-locked demodulator (PLD) are more uniform in that they depend mainly on the relock mechanism, and can be made sharper than those of the discriminator by circuit design. The extension of Rice's noise analysis to the PLD was proposed by Schilling et al,[5] who used it in the prediction of error

[4] S. O. Rice, "Noise in FM Receivers," Chapter 25, *Time Series Analysis*, John Wiley, 1963.

[5] D. L. Schilling, J. Billig and D. Kermish, "Error Rates in FSK Using the Phase Locked Loop Demodulator," 1st IEEE Annual Communications Convention, Boulder, Colorado, June 1965; also, Research Report No. PIB MRI-1254-65, Polytechnic Institute of Brooklyn.

rates. Their results differ from those in this paper in that their work accounts only for the effect of a single loss-of-lock impulse and is thus applicable only over a certain range of deviation indices. Also, their system uses a simple low-pass filter instead of the integrate–dump circuit used here and thus requires a different analytical approach. In the system described here, the probability of error is shown to be mainly a function of the input-carrier-to-noise ratio and the deviation

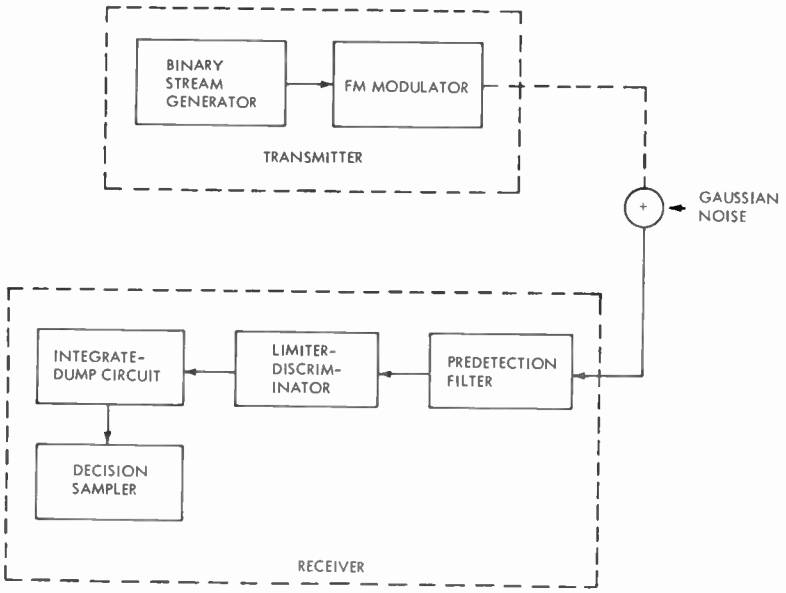

Fig. 1—Communication system.

index, where the deviation index is defined as the ratio of the peak-to-peak deviation to the bit rate;

$$D = \frac{2\Delta f}{B_R} \text{ cycles per bit,} \qquad (1)$$

where Δf is the peak deviation in cps and B_R is the bit rate in bits/sec.

The System

Figure 1 is a block diagram of the system under consideration. At the transmitter end, a binary stream of essentially rectangular transitions frequency modulates a carrier. The only admitted con-

tamination by the channel is additive gaussian noise. The receiver input signal is filtered to reject as much noise as possible consistent with the bandwidth requirements of the signal. The filter output is fed to a limiter–discriminator for demodulation. The integrate–dump circuit integrates the output of the limiter–discriminator for the duration of the bit. At the end of each bit, the decision sampler decides whether a mark or a space was received, depending upon the polarity of the integrator output at that instant. The contents of the integrator are then dumped to prepare for the next bit period. Perfect timing in the integrate–dump and decision-sampler circuits is assumed. In practice, the timing would be extracted from the received signal. For sufficiently low error rates, the timing so extracted will usually not materially affect the probability of error. The optimum shape and width of the predetection filter is not treated here, and the carrier-to-noise ratio in the discussion that follows refers to the value measured at the output of this filter. Although the actual carrier level of interest is that at the filter input, the carrier level at the output was taken in order to facilitate the discussion. In any particular situation it will be necessary to calculate the signal power loss, if any, in the predetection filter. The basic difference between this system and that of earlier investigators is the presence of the integrate–dump circuit.[1-3] Earlier investigators utilized the formulas for the exact distribution of the noise at the output of a limiter–discriminator derived by Rice.[6] Since this noise is nongaussian, difficulty arose in obtaining its distribution after filtering. The lack of low-pass filtering is permissible only for small deviation indices, since the predetection filter can then be made sufficiently narrow to do the prime filtering. For large deviation indices or large predetection bandwidths, the error rates without baseband filtering become prohibitive.

The integrate–dump circuit results in improved system performance[7] and, in addition, makes possible the "equivalent area" analytical approach, which facilitates the inclusion of the effect of a plurality of threshold impulses.

Error Rates Due to Gaussian Noise

Consider the post-detection noise with the threshold impulses removed. According to Rice,[4] this noise has a nearly gaussian distribution, an assumption that works well for analog signal S/N ratio calculations. With digital signals, the tails of the distribution (or the

[6] S. O. Rice, "Statistical Properties of a Sine-Wave Plus Noise," *Bell Syst. Tech. Jour.*, Vol. 27, Jan. 1948.

[7] J. Klapper, "The Effect of the Integrator-Dump Circuit on PCM/FM Error Rates," *Trans. IEEE PGCT*, June 1966.

infrequent events) are of primary interest, and therefore the assumption that this noise is gaussian needs further evaluation. It will be shown that if this noise can be assumed to be gaussian, then it is negligible as a cause of errors except for very small deviation indices, provided that appropriate filtering follows the discriminator.

The integrate–dump circuit is a sin x/x type of filter for gaussian noise. The formula for the probability of error for rectangular binary video signals immersed in additive gaussian noise is given by[8]

$$P_e = \frac{1}{2} \operatorname{erfc} \frac{A/2}{\sqrt{2}\,\sigma}, \tag{2}$$

where A is the peak-to-peak amplitude of the video pulse, σ^2 is the power of the gaussian noise, and

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-y^2}\, dy. \tag{3}$$

Utilizing the FM improvement factor and writing in terms of the carrier-to-noise ratio at the limiter–discriminator input, ρ, one obtains[5]

$$P_e = \frac{1}{2} \operatorname{erfc} \sqrt{3 \left(\frac{\Delta f}{B_v} \right)^2 \left(\frac{B_i/2}{B_v} \right) \rho}. \tag{4}$$

where Δf is the peak deviation of the input signal from the center frequency, B_v is the video bandwidth, and B_i is the input predetection bandwidth.

Since rectangular filters are assumed in Equation (4) and since the integrate–dump circuit is a sin x/x type of filter, the results obtained are not exactly applicable. Assuming that the relation between the input predetection and the video bandwidths is given by Carson's rule,

$$\frac{B_i}{2} = (1 + D) B_v, \tag{5}$$

the error rate can be written

[8] M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw-Hill, New York, 1959.

$$P_e = \frac{1}{2}\operatorname{erfc}\sqrt{3D^2\frac{E}{N_0}}, \tag{6}$$

where E is the input energy per bit, N_0 is the input noise power density, and D is defined by Equation (1). If the video bandwidth is taken as one-half the bit rate, which is consistent with the Nyquist signalling rate,[9] then

$$\rho = \frac{E}{N_0}\frac{1}{B_i T}, \tag{7}$$

where T is the bit duration in seconds.

A practical value of E/N_0 is 12 db. The error probability is plotted in Figure 2 as a function of the deviation index, D, for this value of E/N_0. Two points are marked on the curve; one is the minimum error rate obtained by Meyerhoff and Mazer[1] for any deviation index, and the other is the minimum error rate available in a coherent phase-shift-keyed (PSK) system. It is evident that the contribution of this gaussian noise is negligible except for very small deviation indices.

The Rate of Opposing and Aiding Threshold Impulses

This section shows that threshold impulses appear mainly in the direction opposing the frequency deviation of the signal from the center. When the carrier is at the center frequency of a symmetrical noise-power density, then, as expected, the probability of threshold impulses of either polarity is the same.

Let N_+ be the number of aiding threshold impulses per second and similarly, let N_- be the rate of opposing threshold impulses. Rice's formulas* may be put in the form

$$N_+ = \frac{\Delta f}{2}\left\{\left[1+\left(\frac{r}{\Delta f}\right)^2\right]^{1/2}\operatorname{erfc}\left[\rho\left(1+\left(\frac{\Delta f}{r}\right)^2\right)\right]^{1/2} - e^{-\rho}\operatorname{erfc}\left[\frac{\Delta f}{r}\sqrt{\rho}\right]\right\} \tag{8}$$

and

$$N_- = N_+ + \Delta f e^{-\rho}, \tag{9}$$

[9] H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Trans. AIEE*, Vol. 47, p. 617, April 1928.

* See Equation (71) in Reference (4).

where r is the radius of gyration of the power spectrum of the noise. Two basic assumptions were made in the derivation of Equations (8) and (9). First, the noise power density is assumed to have arithmetic

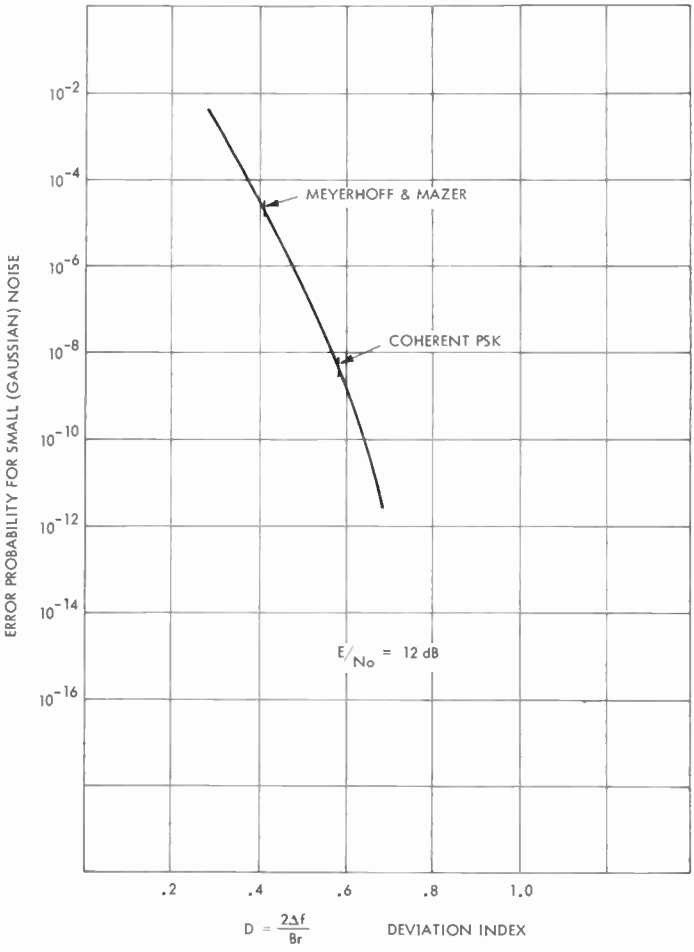

Fig. 2—Error probability for small noise.

symmetry about its center frequency. This assumption was also made by Meyerhoff and Mazer[1] and by Shaft.[2] Second, the signal is assumed to be a sine wave existing for all time. The frequency-shift-keyed signal is a wave in which the frequency is being switched. For this analysis, however, it is assumed (as it was by Shaft[2]) that the signal may be represented as being in steady state.

For large values of its argument, the complementary error function can be closely approximated by[4]

$$\operatorname{erfc}(x) \cong \frac{e^{-x^2}}{\sqrt{\pi}\, x}. \tag{10}$$

It is interesting to note that, when Equation (10) is inserted into Equation (8), N_+ becomes identically zero. How readily can the aiding threshold impulses be ignored in practice? The region over which the approximation can be made is given, from Equations (8) and (9), by

$$\frac{N_+}{\Delta f} \ll \frac{N_-}{\Delta f}, \tag{11}$$

or

$$\frac{1}{2}\left[\left(\frac{r}{\Delta f}\right)^2 + 1\right]^{1/2} \operatorname{erfc}\left[\rho\left(1 + \left(\frac{\Delta f}{r}\right)^2\right)\right]^{1/2} - e^{-\rho} \operatorname{erfc}\left[\frac{\Delta f}{r}\sqrt{\rho}\right] \ll e^{-\rho}. \tag{12}$$

The approximation is dependent on two parameters, $\Delta f/r$ and $\sqrt{\rho}$. The first quantity is usually less than unity; the value of the second quantity is mainly dependent upon the quality of service and the deviation ratio. One cannot say, in general, that Equation (10) holds; thus a test on the basis of Relationship (11) is required. Figure 3 presents plots of comparative values of N_+ and N_- for practical values of the parameters $\Delta f/r$ and ρ. One concludes from the curves that for the usual operating values of the normalized deviation, $\Delta f/r$, the aiding threshold impulses may be neglected. In what follows we shall therefore assume the rate of threshold impulses to be given by

$$N_+ = 0 \tag{13}$$

and

$$N_- = \Delta f e^{-\rho} = N. \tag{14}$$

This fact, which has been shown experimentally, results in important simplifications in the prediction of error rates. For example, in the usual case where the mark and space frequencies are spaced sym-

metrically about the center, one can ignore the effect of the "aiding" threshold impulses. Also, for an extreme shift in the signal frequencies (due to the Doppler effect, oscillator instabilities, etc.), there will be very few errors in the symbol that is closest to the center frequency.

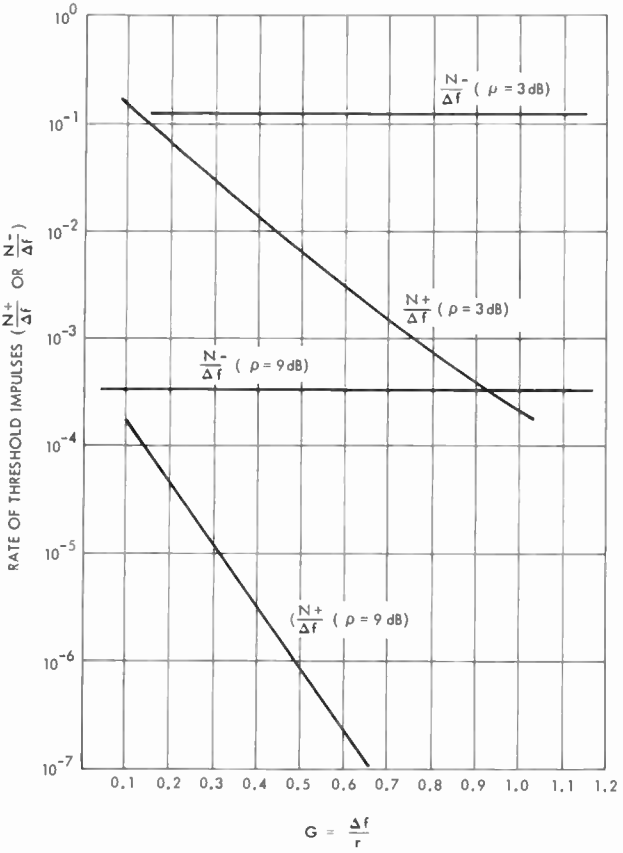

Fig. 3—Comparative rates of threshold impulses.

The Error-Rate Formula

The assumed decision method is basic to our discussion. The output of the demodulator is integrated over a bit period. At the end of the integration period the output of the integrator is sampled and a binary decision is made on the basis of the polarity of the output. After the sampling, the integrator content is rapidly dumped and a new integration period begins. The binary decision is thus based on the

polarity of the area under the demodulator output wave, comprising both signal and noise, over a bit period. It is crucial to note here that the time integral of frequency is phase, and, therefore, the decision method stated above can be restated in terms of phase as follows—the binary decision is based on the polarity of the difference between the final and initial phases of the resultant; the resultant is the sum of the signal and noise waves at the demodulator input.

The noise superimposed on the carrier causes the resultant's phase to deviate from that of the noise-free carrier. If the phase difference between the beginning and the end of the bit period introduced by the noise cancels the phase difference introduced by the modulation, then an error in decision takes place. Note that in this decision the phase modulations within the bit do not matter; only the phase difference between the beginning and the end of the bit counts. We can begin to appreciate the relative significance of the impulsive and nonimpulsive components of the post-detection noise on the error rates. The nonimpulsive noise has undulations in both directions, and the probability of a large value at the sampling instant is small. The impulsive noise, however, introduces a phase step, mainly in the direction to oppose the modulation, and generally no return of this step takes place.

There is yet another important limitation on the nonimpulsive noise. It was stated previously, from Rice's theory,[4] that a threshold impulse occurs nearly every time the additional angle due to the noise exceeds $\pm\pi$. In this manner, the nonimpulsive noise can contribute phase undulations no greater than 2π. We will now compare this effect to the angle introduced by the modulation. The angle due to the modulation at the sampling instant, θ_s, is given by

$$\theta_s = 2\pi \int_0^T (f - f_0)\, dt$$

where $f - f_0$ is the instantaneous frequency deviation and T is the bit duration. For a frequency step $f - f_0 = \Delta f$, and from the definition of the deviation index, Equation (1),

$$\theta_s = 2\pi \Delta f T = \pi D.$$

Thus, for deviation indices greater than two, an error must be accompanied by the occurrence of a threshold or loss-of-lock impulse. Intersymbol interference is neglected in this development. The major cause

of errors seems to be the impulses, whether they are of the threshold (phase jumps in the received wave) or loss-of-lock type (phase jumps in the voltage-controlled-oscillator wave). The nonimpulsive noise appears to be of importance for very small deviation indices or as a correction term at the higher deviation indices. Only the impulsive noise will be considered in the development of the error-rate formula that follows.

In the plane of frequency deviation versus time, the uncorrupted signal area is[7] ΔfT. For the same scaling, the threshold impulse area is —1. Thus an error occurs in a particular bit if

$$i \geqq \Delta fT, \tag{15}$$

where i is the number of opposing threshold impulses in the bit. It is assumed that the threshold impulses are so narrow in the time scale that only an integral number of these appear in a bit. In other words, the area of a threshold impulse is never shared by adjoining bits but is fully contained in a single bit. Since T is the reciprocal of the bit rate, Equation (15) can be written in terms of the deviation index (Equation (1)) as

$$i \geqq \frac{D}{2}. \tag{16}$$

It is assumed that the threshold impulses have a Poisson distribution. Such a distribution was assumed by Rice[4] and was later verified experimentally.[10] The probability of exactly k threshold impulses in a time interval T is given by[11]

$$P(k) = \frac{(NT)^k e^{-NT}}{k!}, \tag{17}$$

where N is the number of threshold impulses per second given by Equation (14). Since the events are mutually exclusive, the probability of k or more threshold impulses in T seconds is given by the sum

---

[10] I. Ringdahl and D. L. Schilling, "Cn the Distribution of the Spikes Seen at the Output of an FM Discriminator Below Threshold," *Proc. IEEE*, Vol. 52, p. 1756, Dec. 1964.

[11] E. Parzen, *Modern Probability Theory and Its Applications*, John Wiley and Sons, New York, 1960.

$$P(i \geq k) = \sum_{i=k}^{\infty} \frac{(NT)^i e^{-NT}}{i!} = 1 - \sum_{i=0}^{k-1} \frac{(NT)^i e^{-NT}}{i!} . \quad (18)$$

For our application, Equation (17) yields a sufficiently close value for the probability of k or more threshold impulses in time T. It can

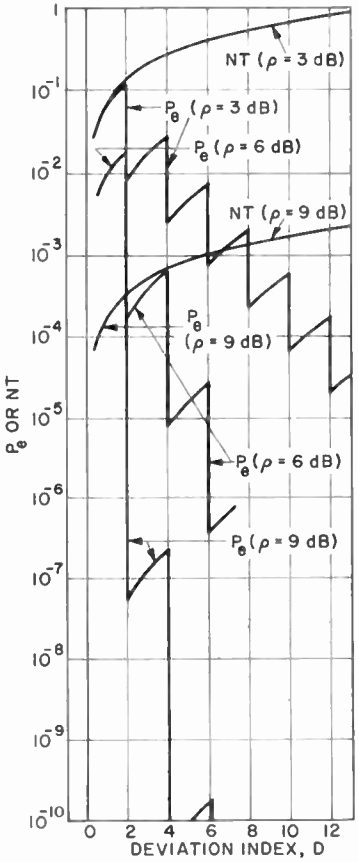

Fig. 4—P_e or NT versus deviation index.

be readily shown that

$$\frac{P(k+1)}{P(k)} = \frac{NT}{k+1} . \quad (19)$$

Typical values of NT are given in Figure 4. NT is small for small deviation indices where k is not large, and may be near unity for large

deviation indices where k is large. In either case, the ratio $NT/(k+1)$ is small compared to unity. Since, in addition, the Poisson distribution monotonically decreases in the region of our interest, the assertion that Equation (17) is sufficiently close in value to Equation (18) is justified. Thus

$$P(i \geq k) \cong \frac{(NT)^k e^{-NT}}{k!} = P(k). \tag{20}$$

In the same manner we shall ignore the "or greater" sign and write Equation (16) as

$$k = D/2 \quad \text{(next higher integer)}. \tag{21}$$

It takes at least one opposing threshold impulse to cause an error for deviation indices of less than two. Two or more opposing threshold impulses are required to cause an error if the deviation index is between two and four, three or more are required between four and six, and so on. This sudden jump in the required number of threshold impulses at deviation indices that are multiples of two yields some interesting changes in slope in the curves of error rates versus deviation index.

Thus, the probability of error, P_e, is given from Equations (20), (1) and (14) by the relatively simple expressions

$$P_e = \frac{I^k e^{-I}}{k!}, \tag{22}$$

where

$$I = NT = \frac{D}{2} e^{-\rho} \text{ impulses per bit.} \tag{23}$$

As written in Equations (22) and (23), the probability of error is only a function of the deviation index and the carrier-to-noise ratio (CNR) at the limiter input. The average number of threshold impulses per bit is directly proportional to the deviation index and decreases exponentially with increasing CNR.

It is of interest to amend the error-rate formulas so that center frequency shifts of the signal can be readily incorporated. It is assumed that the d-c value associated with the center frequency shift is cancelled out at the output of the discriminator. Since the area of the

signal bit has not changed, Equation (21) still gives the number of threshold impulses that cause an error. Also, since the Poisson distribution still holds, Equation (22) is correct. The only equation that will differ is Equation (23), since the threshold impulse rate for the mark and space symbols will be different. Let d be the shift in the signal center frequency in cycles per second and let

$$\delta = \frac{d}{\Delta f}. \tag{24}$$

Equation (23) can now be generalized to our case using Equations (1) and (14):

$$I = \frac{D}{2}(1 \pm \delta)e^{-\rho}. \tag{25}$$

Equation (25) would indicate that there are no threshold impulses when the signal frequency is at the center. This is clearly incorrect and is due to the failure of Relationship (11). The errors increase for the symbol whose effective deviation from the center has increased, and vice versa.

Illustrations

The foregoing formulas will now be used in some important examples that will illustrate properties of the binary FM receiver, some of which were unknown heretofore. The data for the plots were obtained on a digital computer.

Error Rates for Constant Carrier-to-Noise Ratio

Suppose we are constrained to keep the predetection bandwidth at a certain value. Practical reasons for this may be the required accommodation of a Doppler shift, oscillator frequency uncertainty, or the lack of sufficiently narrow-band filters. With the noise-power density, N_0, a constant, the noise-power level at the input cannot be reduced, and for a certain carrier power the maximum CNR is fixed. Suppose further that we are free to choose the deviation index in order to achieve a specified probability of error. Since there is a limit to the value of the frequency deviation before excessive signal power losses are suffered in the predetection filter, the maximum information rate is also limited. In any case, the appropriate plots of error rates (P_e) versus deviation index are given in Figure 4 for CNR of 3, 6, and 9 db using Equations (22) and (23).

The general trend of the curves is, as expected, towards increased reliability of transmission with an increasing deviation index or, equivalently, with a decreased information rate. The surprising part, however, is the zigzag behavior of the plot, which indicates that a number of relative minima are available to the designer. This behavior is due to the consideration of only an integral number of threshold impulses per bit and the neglect of the nonimpulsive component of the noise. For example, at $D = 2$ we jump from the probability of one threshold impulse per bit to the probability of two, resulting in a large jump in error rates.

The curves marked NT represent the average number of threshold impulses per bit. For deviation indices of less than two, the probability of a threshold impulse per bit, the average number of threshold impulses per bit, and the probability of an error are numerically nearly the same. NT is actually increasing linearly with the deviation index, as predicted by Equation (23).

Error Rates for a Constant Energy Ratio

We shall again predict the probability of an error as a function of the deviation index. However, the parameter will be E/N_0, i.e., the ratio of the energy per bit to the noise power in a 1-cps bandwidth. This parameter is used extensively in the literature in judging the quality of a digital communication system.

To use Equations (22) and (23), we must convert from ρ to E/N_0 using Equation (7). Also, B_iT must be replaced by known parameters. An optimum relation for digital FM is not known, so we shall use Carson's rule (see Equation (5)). Then, from Equations (5) and (7),

$$\rho = \frac{E}{N_0} \frac{1}{1+D}. \tag{26}$$

The probability of error obtained using Equations (22), (23) and (23) is plotted in Figure 5 for E/N_0 of 9 and 15 db. The general behavior is similar to that in Figure 4, except that here the probability of error increases with D, verifying the generally known fact that, if the system has no constraints preventing such operation, the deviation index should be kept small. There are, however, systems where a small predetection bandwidth is not possible (e.g., due to Doppler shifts), and for such cases these curves indicate the error rates obtained and penalty suffered. It is observed from Equation (7) that ρ and E/N_0 are the same for $B_iT = 1$.

It would appear in Figure 5 that the error rates decrease limitlessly as the deviation index is decreased. However, as discussed earlier, the nonimpulsive noise begins to predominate at small deviation indices. If the small noise were truly gaussian, its effect would be described by Equation (4) (also see Figure 2). Since it deviates from the

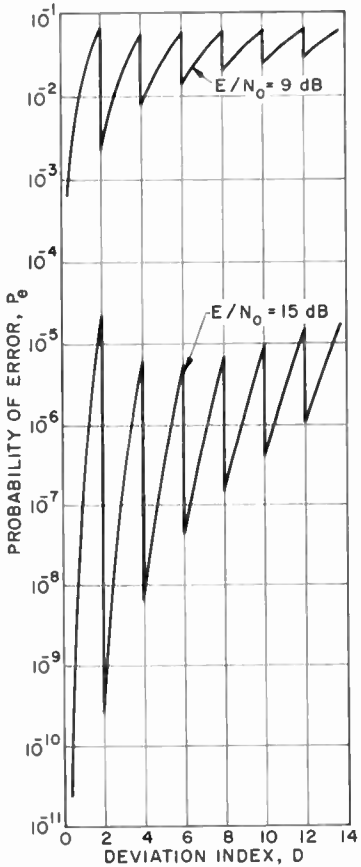

Fig. 5—Probability of error versus deviation index.

gaussian distribution, the exact value of D at which it becomes important is not known. At deviation indices of 0.7 it is still sufficient to consider only the threshold impulses, since the results are the same as those obtained by other analytical methods and experimentally.[2]

Center Frequency Shift

To illustrate the variation in error rates as the signal center frequency is shifted away from the center frequency of the noise-power

density, we utilize Equations (25) and (22) and obtain the curves given in Figure 6. The CNR is kept at 6 db and the curves are plotted for two deviation indices: 0.75 and 3. Three curves are given for each deviation index:

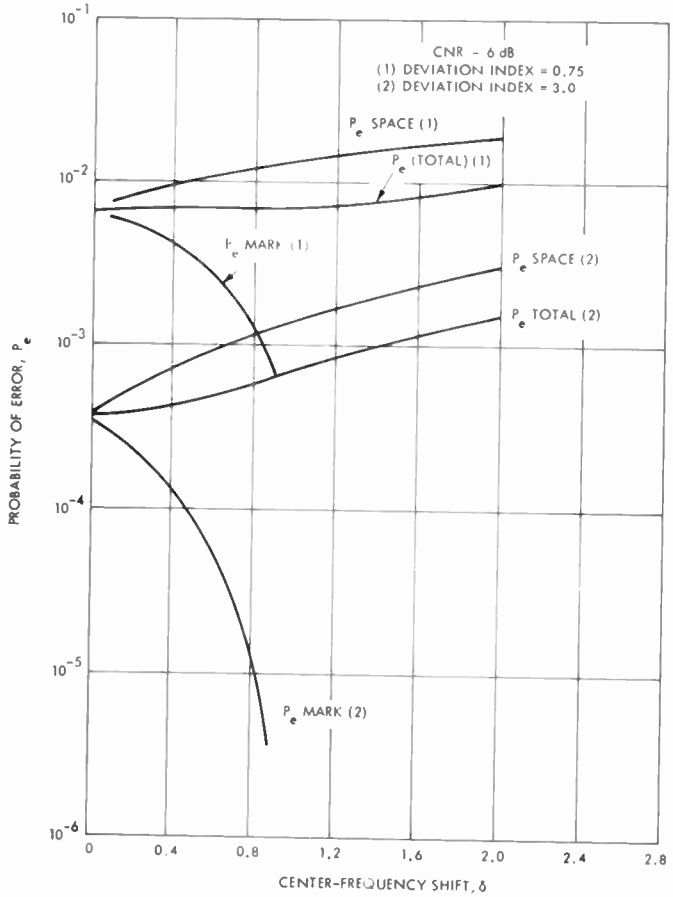

Fig. 6—Probability of error versus center-frequency shift.

P_e(space), which is the probability of error for the symbol being moved further away from the center frequency of the filter;

P_e(mark), which is the probability of error for the symbol being brought closer to the center frequency of the filter; and

P_e(total), which is the probability of error per bit transmitted. In general,

$$P_e(\text{total}) = \frac{1}{2}\left[P_e(\text{space}) + P_e(\text{mark}) \right]. \qquad (27)$$

The shift in the d-c level at the discriminator output due to this center frequency shift is cancelled out, as explained in the derivation of Equation (25). An important conclusion from Figure 6 is that, if the signal frequency is shifted, the symbol that is further removed from the filter center frequency carries the brunt of errors, and careful centering is required for equal error rates. The total error rate, however, does not change appreciably. These conclusions were also found experimentally.

Conclusion

In addition to giving further insight into the error-generation process, the theory presented here results in an error-rate formula with the following advantages: it predicts error rates for large deviation indices; it gives the effect of video processing by an integrate–dump circuit; it is relatively simple to use; and it permits easy inclusion of center-frequency shifts. Further, the essential features of the theory also apply to the recently developed low-threshold demodulators. The only additional information required is the prediction or measurement of the rate of the TI and LLI of these demodulators. The theory can also be extended to multilevel FM reception. The system designer will note the various tradeoffs involved, especially when doppler shifts or other constraints do not permit a narrow-band predetection filter and small deviation indices. When narrow-band operation is possible, the theory verifies the conclusions of earlier researchers that small deviation indices are desirable.

Experimentally, it was found that the error rates of a limiter–discriminator receiver are in reasonable agreement with the theory, but the zigzag behavior is not pronounced. Experimental data available on the phase-locked demodulators do show these predicted rapid changes in slope, and the lack of the zigzag pattern for the limiter–discriminator is believed to be due to the consideration of only integral numbers of threshold impulses and the neglect of the non-impulsive noise. In view of the greater sharpness of these impulses in the phase-locked demodulators, the shape of the error-rate curve versus deviation index is closer to that predicted.

Figure 6 is of interest in showing the behavior of the system when the signal and noise center frequencies do not coincide. While the probability of error on the total-number-of-bits basis does not change considerably with a center-frequency shift, one of the two symbols bears the brunt of the errors.

It is assumed in Figure 2 that the small noise is gaussian. Since the distribution is not exactly gaussian, the effect of the nonimpulsive noise will be felt at different deviation indices than those shown in Figure 2. It appears, however, from a comparison of our results with those of earlier investigators and from available experimental data, that the nonimpulsive noise can still be ignored at $D = 0.7$.

It is worthwhile to list the assumptions made in the development:

(1) If post-detection filtering (e.g., integrate–dump circuit) is used, the threshold impulses are the main source of digital FM errors. This assumption was justified analytically for the case where the nonimpulsive noise can be assumed to be gaussian.
(2) The signal is of rectangular shape.
(3) The signal dwells sufficiently long at each frequency so that an analysis based on a cw signal applies.
(4) The input noise is gaussian.
(5) The noise-power density is of arithmetic symmetry.
(6) The parameters of the system are such that the aiding threshold impulses can be ignored.
(7) Only an integral number of threshold impulses appear in a bit.
(8) The probability of k or more threshold impulses per bit is the same as the probability of k impulses per bit.
(9) All threshold impulses are of unity area.
(10) The threshold impulses follow a Poisson distribution in time.
(11) Intersymbol interference can be ignored.

Some of these assumptions are not essential, and were made for the sake of analytical simplicity.

The error rates predicted herein do not give any bound on optimum demodulation, but they are indicative of what can be obtained with the system described. For example, it may be possible to obtain lower error rates even with a limiter–discriminator if nonrectangular bits or different filtering is used.

ACKNOWLEDGMENT

The author is indebted to G. Aaronson, J. Frankle, M. Masonson, S. J. Mehlman, and A. Newton for helpful comments and discussions.

THE GENERALIZED TRANSFER FUNCTION AND POLE-ZERO MIGRATIONS IN SWITCHED NETWORKS

By

A. Acampora

RCA Communications Systems Division
New York, N. Y.

Summary—If the inputs and outputs of M two-port networks, each with transfer function H(s) (where $s = \sigma + j\omega$), are commutated or switched sequentially and periodically at ω_0 radians per second, then a specific result is that the over-all transfer function (including switches and networks) is $T(s) = K\,[H(s - j\omega_0) + H(s + j\omega_0)]$, for all $M \geq 3$. The majority of the past analyses and applications of this "arithmetic-shift" property deal with the case where H(s) is a specific low-pass filter for which T(s) becomes a band-pass filter. This paper treats the general case of arbitrary H(s) wherein the transfer function T(s) is derived in closed form as the ratio of two polynomials. Included in the derivation is the unique relationship of the roots of these polynomials (i.e., the poles and zeros of T(s)) and the switching frequency ω_0. It is seen that these poles and zeros migrate in the s plane and form loci as a function of ω_0 analogous to the loci obtained in feedback systems for forward-gain variations. In this context, root-locus methods used in feedback systems are completely applicable.

Introduction

A GREAT DEAL has been written on the design of band-pass filters using commutated low-pass structures. Such band-pass filters have been referred to as modulated filters,[1-3] comb filters,[4] commutated filters,[5] quadrature function filters,[6,7] sampled-

[1] N. F. Barber, "Narrow Band-Pass Filter Using Modulation," *Wireless Engineer*, Vol. 24, p. 132, May 1947.

[2] G. B. Madella, "Single-Phase and Polyphase Filtering Devices Using Modulation," *Wireless Engineer*, Vol. 24, p. 310, Oct. 1947.

[3] I. F. MacDiarmid and D. G. Tucker, "Polyphase Modulation as a Solution of Certain Filtration Problems in Telecommunication," *Proc. IEE*, Vol. 97, Pt. 3, p. 349, Sept. 1950.

[4] W. R. LePage, C. R. Cahn, and J. S. Brown, "Analysis of a Comb Filter Using Synchronously Commutated Capacitors," *AIEE Trans.*, Pt. 1, Vol. 72, p. 63, March 1953.

[5] B. D. Smith, "Analysis of Commutated Networks," *Trans. IRE PGAE*, Vol. 10, p. 21, 1953.

[6] D. K. Weaver, Jr., "A Third Method for Generation and Detection of Single-Sideband Signals," *Proc. IRE*, Vol. 44, p. 1703, Dec. 1956.

[7] H. B. Paris, Jr., "Utilization of the Quadrature Functions as a Unique Approach to Electronic Filter Design," *IRE Int. Convention Record*, Part 9, p. 204, 1960.

data filters,[8] digital filters,[9] N-path filters,[10] and switched filters.[11] The end objective in all cases has been the analysis of band-pass (or comb of band pass) structures made up of an array of low-pass filters and suitable switches.

This paper discusses the unique properties of such a commutated device when networks other than true low-pass structures are involved. It attempts to answer questions such as: If switched low-pass filters yield band-pass filters, do switched high-pass filters yield band-stop filters? What happens when band-pass filters or all-pass networks are switched?

Over-All Transfer Function of The Switched Networks

In Figure 1(a), an arbitrary, realizable network $H(s)$ (where $s = \sigma + j\omega$) is placed in an M-path array. The derivation of the over-all transfer function, given in Appendix A, assumes that the network is preceded and followed by switches that are actually product devices. The "carrier" applied is a sinusoidal function of radian frequency ω_0, phase shifted in accordance with the leg position as shown in Figure 1(b). Such an assumption is warranted; the same over-all transfer function is derived for any set of identically shaped sequential waveforms with periodicity $T = 2\pi/\omega_0$ by expanding such waveforms in a Fourier series and considering the multiplication process with the fundamental sinusoidal component alone. One such waveform might be a sequence of rectangular pulses of width T/M, which are indicative of switch closures and commutation.

The presence of other harmonic frequencies in the switching waveforms naturally invites over-all transfer responses at these harmonics in a manner *identical* to that of the fundamental, thus we have a comb structure. The derivation is no less generalized if only a single sinusoidal multiplier is considered. The over-all transfer function is

$$T(s) = \frac{U(s)}{V(s)} = K\left[H(s - j\omega_0) + H(s + j\omega_0) \right], \qquad (1)$$

[8] L. E. Franks and F. J. Witt, "Solid State Sampled Data Band-Pass Filters," *Proc. Solid State Circuits Conf.*, 1960.

[9] J. Thompson, "RC Digital Filters for Microcircuit Bandpass Amplifiers," *EEE Magazine*, p. 44, March 1964.

[10] L. E. Franks and I. W. Sandberg, "An Alternative Approach to the Realization of Network Transfer Functions: The N-Path Filter," *Bell System Tech. Jour.*, Vol. 39, p. 1321, Sept. 1960.

[11] A. Acampora, B. Rabinovici, and C. Renton, "Generation of Bandpass Filters by Switching Techniques," *Proc. IEEE*, Vol. 51, p. 256, Jan. 1963.

where K is some conversion loss factor (equal to 0.25 in the present sinusoidal case). Of interest is the arithmetic shift of the complex frequency variable s in the j direction by ω_0.

Equation (1) represents a band-pass filter centered at ω_0 when $H(s)$ is a low-pass filter with cutoff ω_c much less than ω_0, as can be

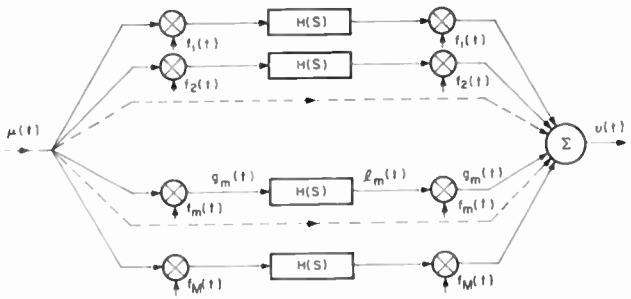

Fig. 1(a)—M-Network array.

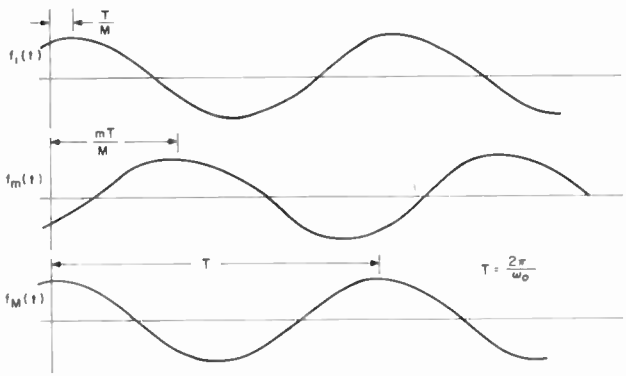

Fig. 1(b)—Multiplying waveforms: General $f_m(t) = \cos\left(\omega_0 t + \dfrac{2\pi m}{M}\right)$.

seen by considering real frequency operation where $s = j\omega$. Further, let the input frequency be different from ω_0 by some small amount $\Delta\omega < \omega_c$, i.e., $\omega = \omega_0 \pm \Delta\omega$. Then, Equation (1) reduces to

$$T(j\omega) = K\,[H(\pm j\Delta\omega) + H(j2\omega_0 \pm j\Delta\omega)]. \qquad (2)$$

For the case $\omega_0 \gg \omega_c$, the low-pass filter response at $2\omega_0$ can be neglected compared to the in-band response at $\Delta\omega$, and

$$T(j\omega) = K\,H(\pm j\Delta\omega). \tag{3}$$

This form indicates that the resulting band-pass structure is equivalent to folding the one-sided low-pass filter response (in mirror-image fashion) about the switching carrier ω_0. This creates an arithmetically symmetrical band-pass filter with the amplitude and phase response of the two-sided original low-pass filter.

The above heuristic argument relies entirely on the premise that $H(2\omega_0)$ is nonexistent (or, at least, negligibly small). There are entire classes of transfer functions $H(s)$, including low-pass filters, where this is not the case, and it is necessary to determine the overall transfer function $T(s)$.

Generalized Closed Form of T(s)

If it is assumed that $H(s)$ can be written as the ratio of two polynomials, $N(s)$ and $D(s)$, both of which are of finite degree, then

$$\begin{aligned} T(s) &= K\left[\frac{N(s-j\omega_0)}{D(s-j\omega_0)} + \frac{N(s+j\omega_0)}{D(s+j\omega_0)}\right] \\ &= K\left[\frac{N(s-j\omega_0)\,D(s+j\omega_0) + N(s+j\omega_0)\,D(s-j\omega_0)}{D(s-j\omega_0)\,D(s+j\omega_0)}\right] = K\frac{P(s)}{Q(s)}. \end{aligned} \tag{4}$$

The poles of $T(s)$ (the roots $Q(s)$) are uniquely described in terms of the poles $H(s)$ (the roots of $D(s)$) by the lineal shift of the $H(s)$ poles in the $\pm j\omega$ direction by an amount ω_0. The zeros of $T(s)$ (the roots of $P(s)$), however, are not so simply related to the zeros of $H(s)$ (the roots of $N(s)$). It is shown in Appendix B that $P(s)$ can always be written in the form:

$$P(s) = 2[G_0(s) + \omega_0^2\,G_1(s) + \omega_0^4\,G_2(s) + \cdots + \omega_0^{2k}G_k(s)], \tag{5}$$

where k is finite. Furthermore, the general function $G_k(s)$ is related to $N(s)$ and $D(s)$ by

$$G_0(s) = N^0(s)\,D^0(s)$$

$$G_1(s) = \frac{-1}{2!}\,[N^2(s)\,D^0(s) - 2N^1(s)\,D^1(s) + N^0(s)\,D^2(s)]$$

$$G_2(s) = \frac{1}{4!}\,N^4(s)\,D^0(s) - 4N^3(s)\,D^1(s) + 6N^2(s)\,D^2(s) - 4N^1(s)\,D^3(s) + N^0(s)\,D^4(s)]$$

$$G_k(s) = \frac{(-1)^k}{(2k)!}\left[N^{2k}(s)\,D^0(s) - 2kN^{2k-1}(s)\,D^1(s) + \frac{2k(2k-1)}{2!}N^{2k-2}(s)\,D^2(s) + \ldots + N^0(s)\,D^{2k}(s)\right] \tag{6}$$

where the general superscript j on the $N(s)$ and $D(s)$ functions denotes the j derivative with respect to s, i.e.,

$$N^j(s) = \frac{d^j}{ds^j}[N(s)].$$

The superscript "zero" is, of course, the function itself and is used to preserve symmetry.

The general $G_k(s)$ function is so reminiscent of a binomial expansion that a short-hand notation can be used where $G_k(s)$ is defined as

$$G_k(s) \overset{\text{DEF}}{=} \frac{(-1)^k}{(2k)!}\left[N(s) - D(s)\right]^{(2k)}. \tag{7}$$

The bracketed exponent is initially treated as a true exponent, and the expansion is carried out using the binomial theorem. The exponents in the resulting expanded function are then treated as derivatives, as shown in Equation (6). In all that follows, superscripts on $N(s)$ and $D(s)$ denote derivative and the short-hand notation will be used.

Notice that when both $N(s)$ and $D(s)$ are of finite degree, the numerator of the commutated transfer function given by $P(s)$ in Equation (5) is naturally truncated (that is, composed of a finite number of $G(s)$ functions) at a point where higher order derivatives of $N(s)$ and $D(s)$ are zero. Table I shows the subscript of the last nonzero $G(s)$ function for various degrees of the polynomials $N(s)$ and $D(s)$. If the degrees of these polynomials are l and m, respectively, then the subscript k of the last nonzero $G(s)$ function is

$$k = \begin{cases} \dfrac{l+m}{2}; & l+m \text{ even} \\ \dfrac{l+m-1}{2}; & l+m \text{ odd.} \end{cases} \tag{8}$$

Of particular interest are those $H(s)$ networks that fall in the bounded region of Table I. For this case, the over-all transfer function is

$$T(s) = 2K \frac{G_0(s) + \omega_0^2 G_1(s)}{D(s - j\omega_0) D(s + j\omega_0)}$$

$$= 2K \frac{N(s)D(s) - \frac{\omega_0^2}{2}[N(s) - D(s)]^{(2)}}{D(s - j\omega_0) D(s + j\omega_0)}. \quad (9)$$

Table I—Subscript of Last Nonzero $G(s)$ for Various Degrees of $N(s)$ and $D(s)$.

DEGREE OF N(S) \ DEGREE OF D(S)	0	1	2	3	4	5	6	7	8	9	10
0	X	0	1	1	2	2	3	3	4	4	5
1	0	1	1	2	2	3	3	4	4	5	5
2	1	1	2	2	3	3	4	4	5	5	6
3	1	2	2	3	3	4	4	5	5	6	6
4	2	2	3	3	4	4	5	5	6	6	7
5	2	3	3	4	4	5	5	6	6	7	7
6	3	3	4	4	5	5	6	6	7	7	8
7	3	4	4	5	5	6	6	7	7	8	8
8	4	4	5	5	6	6	7	7	8	8	9
9	4	5	5	6	6	7	7	8	8	9	9
10	5	5	6	6	7	7	8	8	9	9	10

The poles and zeros of $T(s)$ are neatly defined for all ω_0. The poles originate (at $\omega_0 = 0$) as double roots of $D(s)$, and they proceed along straight lines in the $\pm j\omega$ direction as a direct function of ω_0. The zeros, however, originate at the combined roots of the product $N(s)D(s)$ (at $\omega_0 = 0$), and migrate along calculable loci as a function of ω_0. They finally terminate at the roots of $G_1(s)$ as ω_0 approaches infinity. Some examples involving basic forms of the $H(s)$ networks, which fall in the bounded region, will be shown. Established root-locus techniques[12] are utilized.

[12] J. G. Truxal, *Automatic Feedback Control System Synthesis*, McGraw-Hill Book Co., Inc., New York, p. 221, 1955.

Examples of the Application of Root Locus

Simple High-Pass Filter

Consider a simple normalized high-pass filter such that $H(s) = s/(s+1)$. The pole–zero location for this is shown in Figure 2(a).

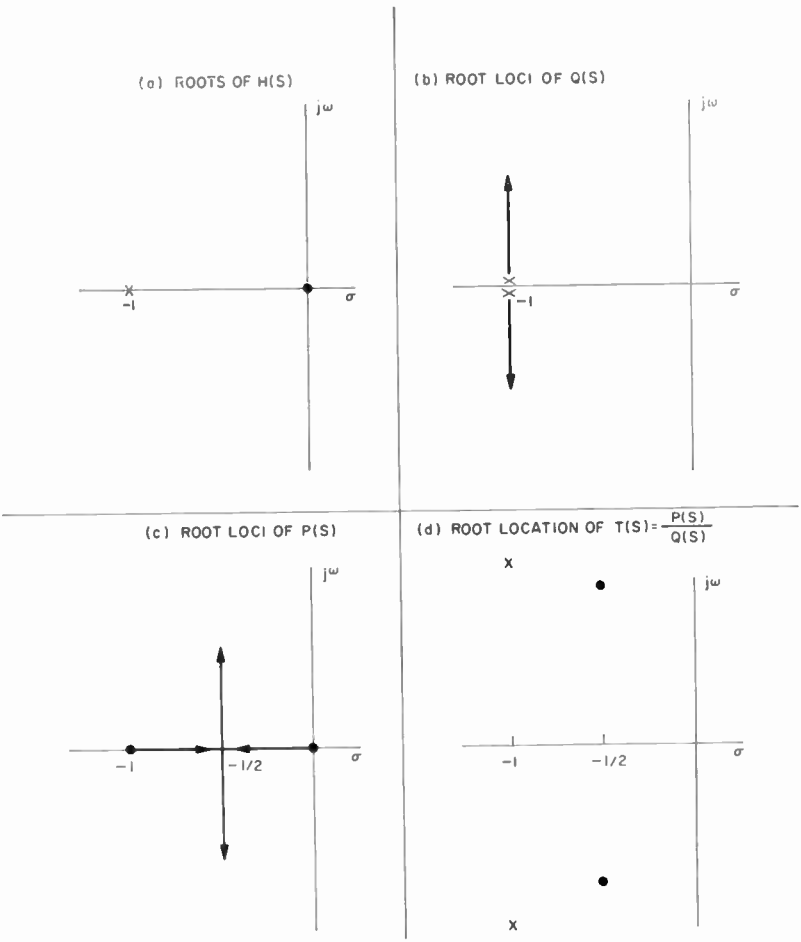

Fig. 2—Pole–zero location and locus for example $H(s) = \dfrac{s}{s+1}$.

For this case, $N^2(s)D^0(s) = 0$, $-2N^1(s)D^1(s) = -2$, and $N^0(s)D^2(s) = 0$, so that from Equation (9),

$$G_1(s) = \frac{-1}{2} \quad (2)$$

and

$$T(s) = 2K \frac{s(s+1) + \omega_0^2}{(s+1-j\omega_0)(s+1+j\omega_0)}.$$

The pole migrations for increasing ω_0 are straightforward loci, as shown in Figure 2(b). The zero migration, however, takes the form shown in Figure 2(c). The zeros merge as a double root (at $s = -\frac{1}{2}$) for $\omega_0 = \frac{1}{2}$, and then split into complex conjugates. The over-all pole–zero diagram of $T(s)$ for some $\omega_0 > \frac{1}{2}$ is shown in Figure 2(d). Notice that a commutated high-pass filter does not lead to a band-stop filter. In fact, Figure 2(d) depicts an attenuation of at most 6 db for some high ω_0 where the pole and zero are almost horizontally aligned.

First-Order All-Pass Section

Consider next the simple all-pass form $H(s) = (s-1)/(s+1)$, as shown in Figure 3(a). Again, $N^2(s)D^0(s) = N^0(s)D^2(s) = 0$; $-2N^1(s)D^1(s) = -2$, so that

$$T(s) = \frac{(s-1)(s+1) + \omega_0^2}{(s+1-j\omega_0)(s+1+j\omega_0)}.$$

The pole migration of $T(s)$ is shown in Figure 3(b). The zero migration, shown in Figure 3(b), is rather interesting. A coalescing occurs at the origin for $\omega_0 = 1$, after which purely imaginary zeros are formed. The combined pole–zero location of $T(s)$ is shown in Figure 3(d). Unexpectedly, the first-order equalizer provides a notch filter when placed in a commutated device.

Simple Band-Pass Filter

Suppose $H(s)$ is of the form $s/(s^2+2s+2) = N(s)/D(s)$ with pole–zero locations shown in Figure 4(a). Then, $N^2(s)D^0(s) = 0$; $-2N^1(s)D^1(s) = -4s-4$; $N^0(s)D^2(s) = 2s$; so that

$$G_1(s) = \frac{-1}{2}(-2s-4)$$

and

$$T(s) = \frac{s(s^2+2s+2) + \omega_0^2(s+2)}{D(s-j\omega_0)D(s+j\omega_0)}.$$

The root loci for the poles are straightforwardly shown in Figure

4(b). The loci for the zero of $T(s)$ are calculated by root-locus method and are shown in Figure 4(c). Notice that the zero on the negative real axis terminates at $s = -2$ and that the complex zeros have an angle of departure of zero degrees and asymptotically approach the

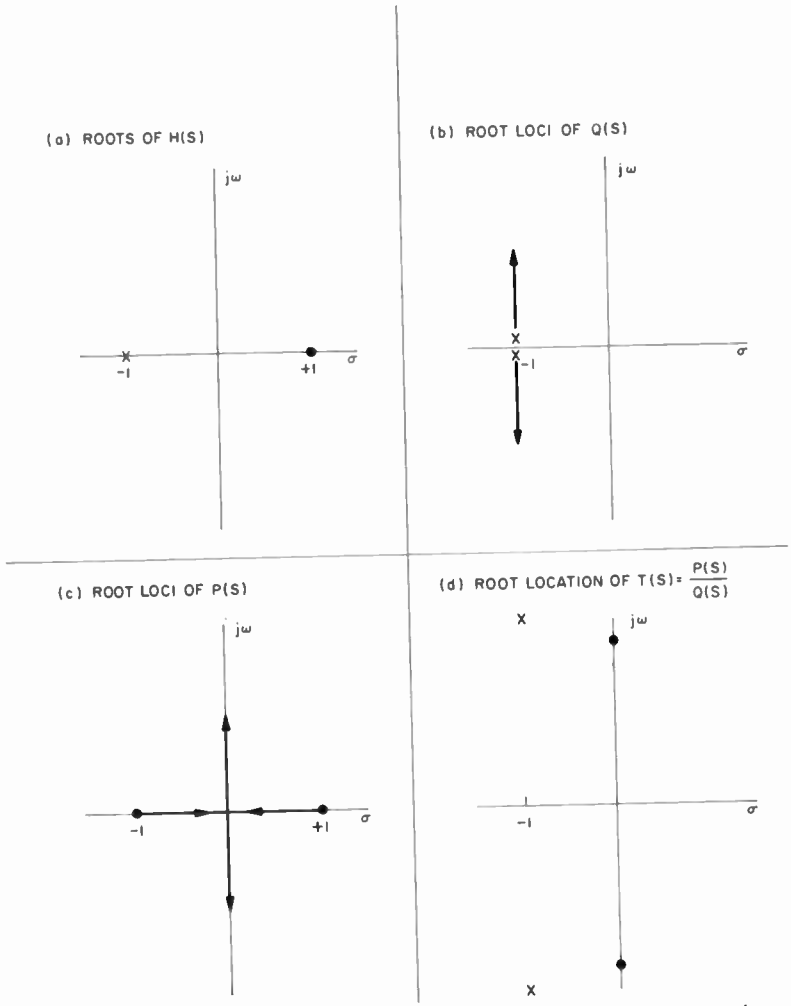

Fig. 3—Pole–zero location and locus for example $H(s) = \dfrac{s-1}{s+1}$.

$j\omega$ axis. The resultant pole–zero locations for some fairly high ω_0 is shown in Figure 4(d). The end result is a pair of adjacent band-pass filters with a notch between them that grows deeper as ω_0 increases and the complex zeros approach the $j\omega$ axis. Such a filter is useful, for example, in double-sideband suppressed-carrier generation.

Extension to More Complex $H(s)$

The transfer functions in the foregoing examples were simple enough to derive directly from Equation (4), since the last nonzero $G(s)$ function was $G_1(s)$. The form indicated by Equation (5) is

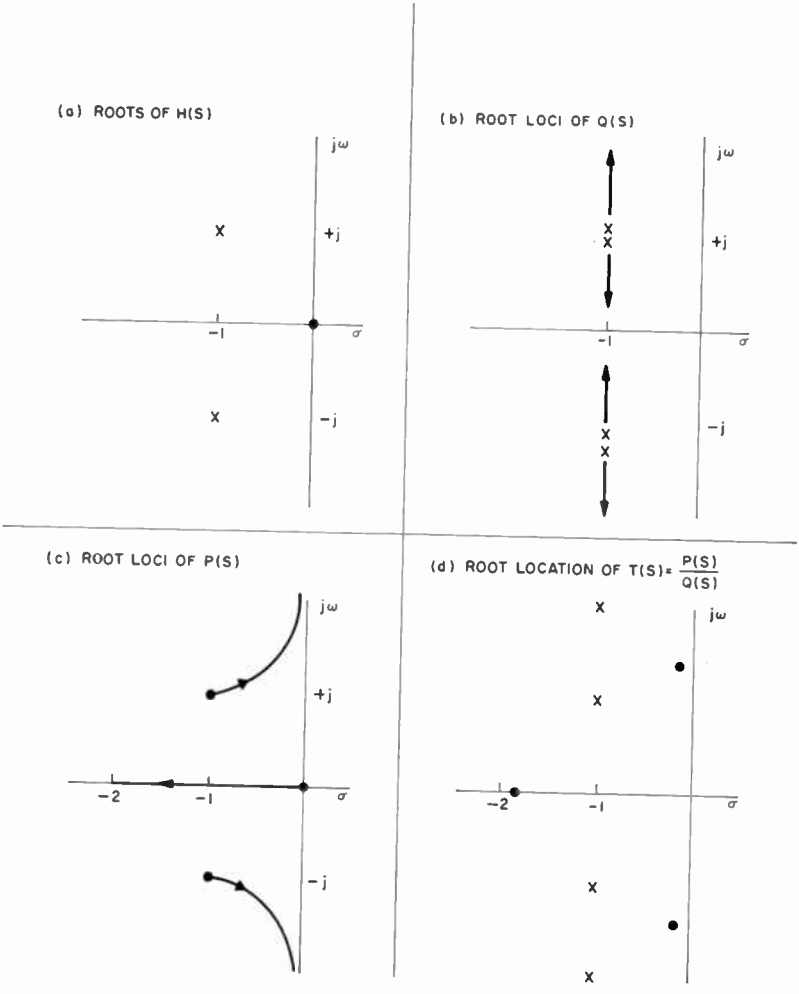

Fig. 4—Pole–zero location and locus for example $H(s) = \dfrac{s}{s^2 + 2s + 2}$.

easiest to use for more complex $H(s)$ functions. With it, the roots, especially the zeros, of the overall transfer function can be characterized as a function of ω_0 for the more complex $H(s)$.

Consider the cases where $H(s)$ is of a form where the last nonzero $G(s)$ is $G_2(s)$, so that

$$P(s) = 2\{G_0(s) + \omega_0^2 G_1(s) + \omega_0^4 G_2(s)\}.$$

Rewriting gives

$$P(s) = 2\{G_0(s) + \omega_0^2[G_1(s) + \omega_0^2 G_2(s)]\} = 2\{G_0(s) + \omega_0^2 P_1(s)\}.$$

The term $P_1(s)$ can be thought of in terms of root loci that begin at the roots of $G_1(s)$ and terminate at the roots of $G_2(s)$. However, these migrating roots of $P_1(s)$ are, as seen, the terminal points of the root of the over-all numerator $P(s)$. The terminal points in root-locus methods are instrumental in calculating the nature of the resulting loci. Thus, the loci of $P(s)$ are under a continual state of change, since their terminal-points (the roots of $P_1(s)$) are on loci of their own. This approach can be extended to more complex $H(s)$ situations, but then the simplicity of the root-locus method is quickly lost. For complex cases, the locus of zeros can best be determined by insertion of particular ω_0 into Equation (5) and direct factorization.

In any event, the transfer function form of Equation (5) provides insight into the pole–zero location of $T(s)$ as a function of ω_0. It can be used to analyze the over-all transfer function in terms of a given $H(s)$ and ω_0 or, conversely, to choose a particular $H(s)$ and ω_0 for a desired end result.

APPENDIXES

The following Appendixes first derive the over-all transfer function in terms of $H(s)$, and then extend the derivation to the $G(s)$ function form. The two-sided Laplace transform was used, primarily to limit consideration to the steady-state response. In this regard, the Fourier transform could have been used, but solutions in terms of the complex frequency variable s, were sought. Strip convergence of the two-sided transform which encompassed the $j\omega$ axis was tacitly assumed because of the consideration of realizable $H(s)$.

A—DERIVATION OF THE TRANSFER FUNCTION IN TERMS OF $H(s)$

Consider the M-path array of identical networks with transfer function $H(s)$, as shown in Figure 1. Let the m^{th} leg multiplier be supplied with a "carrier"

$$\cos\left(\omega_0 t - \frac{2\pi m}{M}\right). \tag{10}$$

Define the following time-domain signals and their Laplace transform pair:

input signal: $u(t)$, $U(s)$

m^{th} leg multiplying signal: $f_m(t)$, $F_m(s)$

m^{th} leg multiplier output (at input side): $g_m(t)$, $G_m(s)$

impulse response of m^{th} leg network: $h(t)$, $H(s)$

m^{th} leg network output: $l_m(t)$, $L_m(s)$

m^{th} leg multiplier output (at output side): $q_m(t)$, $Q_m(s)$

output signal: $v(t)$, $V(s)$

Using two-sided Laplace transform notation,

$$F_m(s) = \int_{-\infty}^{\infty} f_m(t) \exp\{-st\} ds, \tag{11}$$

where $f_m(t)$ is given by Expression (10), which can be more conveniently written

$$f_m(t) = \frac{1}{2}\left[\exp\left\{ j\left(\omega_0 t - \frac{2\pi m}{M} \right) + \exp\left\{ -j\left(\omega_0 t - \frac{2\pi m}{M} \right) \right\} \right]. \tag{12}$$

Then,

$$F_m(s) = \frac{1}{2} \int_{-\infty}^{\infty} \left[\exp\left\{ -j\frac{2\pi m}{M} \right\} \exp\{-(s - j\omega_0) t\} + \exp\left\{ j\frac{2\pi m}{M} \right\} \right.$$
$$\left. \exp\{(s + j\omega_0) t\} \right[ds, \tag{13}$$

or

$$F_m(s) = \frac{1}{2}\left[\delta(s - j\omega_0) \exp\left\{ -j\frac{2\pi m}{M} \right\} + \delta(s + j\omega_0) \exp\left\{ j\frac{2\pi m}{M} \right\} \right], \tag{14}$$

where the delta function is the unit impulse.

The output of the m^{th} multiplier is given by the product relationship

$$g_m(t) = u(t) f_m(t). \tag{15}$$

Since time domain multiplication corresponds to convolution in the frequency domain, Equation (15) can be written

$$G_m(s) = U(s) * F_m(s), \tag{16}$$

or by the convolution integral

$$G_m(s) = \int_{-\infty}^{\infty} U(p) F(s-p)\, dp. \tag{17}$$

Substituting for $F(s)$ from Equation (14),

$$\begin{aligned} G_m(s) = &\frac{1}{2} \int_{-\infty}^{\infty} U(p) [\delta(s - j\omega_0 - p)] \exp\left\{-j\frac{2\pi m}{M}\right\} dp \\ &+ \frac{1}{2} \int_{-\infty}^{\infty} U(p) [\delta(s + j\omega_0 - p)] \exp\left\{j\frac{2\pi m}{M}\right\} dp, \end{aligned} \tag{18}$$

from which

$$\begin{aligned} G_m(s) = \frac{1}{2}\Bigg[&U(s - j\omega_0) \exp\left\{-j\frac{2\pi m}{M}\right\} \\ &+ U(s + j\omega_0) \exp\left\{j\frac{2\pi m}{M}\right\}\Bigg]. \end{aligned} \tag{19}$$

The two-sided Laplace transform of the output signal from the network, $H(s)$, is

$$L_m(s) = G_m(s) H(s) \tag{20}$$

or

$$\begin{aligned} L_m(s) = \frac{1}{2}\Bigg[&U(s - j\omega_0) H(s) \exp\left\{-j\frac{2\pi m}{M}\right\} \\ &+ U(s + j\omega_0) H(s) \exp\left\{j\frac{2\pi m}{M}\right\}\Bigg]. \end{aligned} \tag{21}$$

As in the input multiplier case, the output multiplier performs a time product,

$$q_m(t) = l_m(t) f_m(t), \tag{22}$$

which can be written as a convolution integral,

$$Q_m(s) = \int_{-\infty}^{\infty} L_m(p) F_m(s-p)\, dp, \tag{23}$$

or, again substituting for $F_m(s)$ and $L_m(s)$,

$$\begin{aligned} Q_m(s) = \frac{1}{4} \int_{-\infty}^{\infty} \Bigg[& U(s - j\omega_0) H(s) \exp\left\{-j\frac{2\pi m}{M}\right\} \\ & + U(s + j\omega_0) H(s) \exp\left\{j\frac{2\pi m}{M}\right\} \Bigg] \\ & \Bigg[\delta(s - j\omega_0 - p) \exp\left\{-j\frac{2\pi m}{M}\right\} \\ & + \delta(s + j\omega_0 - p) \exp\left\{j\frac{2\pi m}{M}\right\} \Bigg] dp, \end{aligned} \tag{24}$$

or

$$\begin{aligned} Q_m(s) = & \frac{1}{4} \Big[U(s) H(s - j\omega_0) + U(s) H(s + j\omega_0) \Big] \\ & + \frac{1}{4} \Bigg[U(s - 2j\omega_0) H(s - j\omega_0) \exp\left\{-j\frac{2\pi m}{M}\right\} \\ & + U(s + 2j\omega_0) H(s + j\omega_0) \exp\left\{j\frac{2\pi m}{M}\right\} \Bigg]. \end{aligned} \tag{25}$$

Now, the output signal is the sum of the M leg outputs,

$$v(t) = \sum_{m=1}^{M} q_m(t), \tag{26}$$

or equivalently

$$V(s) = \sum_{m=1}^{M} Q_m(s). \tag{27}$$

Therefore,

$$V(s) = \frac{U(s)}{4}\left[H(s - j\omega_0) + H(s + j\omega_0)\right]$$

$$+ \frac{1}{4} U(s - 2j\omega_0) H(s - j\omega_0) \sum_{m=1}^{M} \exp\left\{-j\frac{2\pi m}{M}\right\}$$

$$+ \frac{1}{4} U(s + 2j\omega_0) H(s + j\omega_0) \sum_{m=1}^{M} \exp\left\{j\frac{2\pi m}{M}\right\}. \tag{28}$$

For $M \geq 3$, the last two terms of $V(s)$ cancel vectorially, and

$$V(s) = \frac{U(s)}{4}\left[H(s - j\omega_0) + H(s + j\omega_0)\right]. \tag{29}$$

Defining the over-all transfer function as

$$T(s) = \frac{V(s)}{U(s)} \tag{30}$$

gives

$$T(s) = \frac{1}{4}\left[H(s - j\omega_0) + H(s + j\omega_0)\right]. \tag{31}$$

B—Derivation of the Transfer Function in Terms of the $G_k(s)$ Functions

Before deriving the general switched network transfer function in terms of the $G_k(s)$ functions, some basic notations involving the two sided Laplace transform pair relationships will be reviewed. These are presented in the following Equations (32) through (37):

$$h(t) \longleftrightarrow H(s), \tag{32}$$

$$t^n h(t) \longleftrightarrow (-1)^n H^n(s), \text{ where } H^n(s) = \frac{d^n}{ds^n}\left[H(s)\right], \tag{33}$$

$$\int_{-\infty}^{\infty} h(\tau)\, g(t - \tau)\, d\tau \longleftrightarrow H(s) G(s), \tag{34}$$

$$\int_{-\infty}^{\infty} h(\tau)g(t-\tau)\tau^n(t-\tau)^m d\tau \longleftrightarrow (-1)^n(-1)^m H^n(s)G^m(s); \quad (35)$$

furthermore,

$$H(s+j\omega_0) \longleftrightarrow h(t)\exp\{-j\omega_0 t\}, \quad (36)$$

$$H(s-j\omega_0) \longleftrightarrow h(t)\exp\{j\omega_0 t\}. \quad (37)$$

Now, if

$$H(s) = \frac{N(s)}{D(s)}, \quad (38)$$

then the transfer function of the switched network is given by

$$T(s) = \frac{1}{4}\left[\frac{N(s-j\omega_0)}{D(s-j\omega_0)} + \frac{N(s+j\omega_0)}{D(s+j\omega_0)}\right], \quad (39)$$

as shown in Appendix A, or

$$T(s) = \frac{1}{4}\left[\frac{N(s-j\omega_0)D(s+j\omega_0) + N(s+j\omega_0)D(s-j\omega_0)}{D(s-j\omega_0)D(s+j\omega_0)}\right] = \frac{1}{4}\frac{P(s)}{Q(s)}. \quad (40)$$

We write the numerator products as the convolved time domain functions; that is, we use the transform pair correspondence,

$$P(s) \longleftrightarrow \int_{-\infty}^{\infty}\Big\{ d(\tau)\,[\exp\{-j\omega_0\tau\}]\,n(t-\tau)\,[\exp\{j\omega_0(t-\tau)\}] + d(\tau)\,[\exp\{j\omega_0\tau\}]\,n(t-\tau)\,[\exp\{-j\omega_0(t-\tau)\}]\Big\}\,d\tau \quad (41)$$

or

$$P(s) \longleftrightarrow \int_{-\infty}^{\infty} d(\tau)n(t-\tau)\,[\exp\{j\omega_0(t-2\tau)\} + \exp\{-j\omega_0(t-2\tau)\}]\,d\tau \quad (42)$$

$$P(s) \longleftrightarrow 2 \int_{-\infty}^{\infty} d(\tau) n(t-\tau) \cos \omega_0 (t-2\tau)\, d\tau. \tag{43}$$

Expanding $\cos x$ as

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$$

gives

$$P(s) \longleftrightarrow 2 \int_{-\infty}^{\infty} d(\tau) n(t-\tau) \left[1 - \frac{\omega_0{}^2 (t-2\tau)^2}{2!} + \frac{\omega_0{}^4 (t-2\tau)^4}{4!} + \cdots \right.$$
$$\left. + \frac{(-1)^k |\omega_0 (t-2\tau)|^{2k}}{(2k)!} + \cdots \right] d\tau. \tag{44}$$

This expression can be rewritten

$$P(s) \longleftrightarrow 2 \int_{-\infty}^{\infty} d(\tau) n(t-\tau) \left[1 - \frac{\omega_0{}^2}{2!} |(t-\tau) - \tau|^2 + \frac{\omega_0{}^4}{4!} [(t-\tau) - \tau]^4 + \cdots \right.$$
$$\left. + \frac{(-1)^k \omega_0{}^{2k}}{(2k)!} [(t-\tau) - \tau]^{2k} + \cdots \right] d\tau. \tag{45}$$

Use of the binomial theorem expansion of $|(t-\tau) - \tau|^{2k}$, which (as will be seen) accounts for the "binomial expansion" of derivative of $N(s)$ and $D(s)$, gives

$$P(s) \longleftrightarrow 2 \left\{ \int_{-\infty}^{\infty} d(\tau) n(t-\tau)\, d\tau - \frac{\omega_0{}^2}{2!} \int_{-\infty}^{\infty} d(\tau) n(t-\tau) |(t-\tau)^2 - 2(t-\tau)\tau \right.$$
$$+ (t-\tau)\tau^2 |d\tau + \cdots + \frac{(-1)^k \omega_0{}^{2k}}{(2k)!} \int_{-\infty}^{\infty} d(\tau) n(t-\tau) \left[(t-\tau)^{2k} \right.$$
$$\left. \left. - 2k(t-\tau)^{2k-1}\tau + \frac{2k(2k-1)}{2!} (t-\tau)^{2k-2}\tau^2 + \cdots + \tau^{2k} \right] d\tau + \cdots \right\} \tag{46}$$

or, finally,

$$P(s) = 2\left\{N^0(s)D^0(s) - \frac{\omega_0^2}{2!}\left[N^2(s)D^0(s) - 2N^1(s)D^1(s) + N^0(s)D^2(s)\right] + \frac{\omega_0^4}{4!}\left[N^4(s)D^0(s) - 4N^3(s)D^1(s) + 6N^2(s)D^2(s) - 4N^1(s)D^3(s) + N^0(s)D^4(s)\right] + \cdots + \frac{(-1)^k\omega_0^{2k}}{(2k)!}\left[N^{2k}(s)D^0(s) - 2kN^{2k-1}(s)D^1(s) + \frac{2k(2k-1)}{2!}N^{2k-2}(s)D^2(s) + \cdots + N^0(s)D^{2k}(s)\right] + \cdots\right\}. \quad (47)$$

This may be written more conveniently as

$$P(s) = 2\{G_0(s) + \omega_0^2G_1(s) + \omega_0^4G_2(s) + \cdots + \omega_0^{2k}G_k(s) + \cdots\}, \quad (48)$$

where

$$G_0(s) = N^0(s)D^0(s),$$

$$G_1(s) = \frac{-1}{(2)!}[N^2(s)D^0(s) - 2N^1(s)D^1(s) + N^0(s)D^2(s)],$$

$$G_k(s) = \frac{(-1)^k}{(2k)!}[N^{2k}(s)D^0(s) - 2kN^{2k-1}(s)D^1(s) + \cdots + N^0(s)D^{2k}(s)],$$

$$\stackrel{\text{DEF}}{=} \frac{(-1)^k}{(2k)!}[N(s) - D(s)]^{(2k)}. \quad (49)$$

Where the subscripts on the $N(s)$ and $D(s)$ functions denote derivative with respect to s, a short-hand notation, as defined in the text, can be used. When both $N(s)$ and $D(s)$ are finite degree polynomials, Equation (48) is truncated at a point where all higher order derivatives of $N(s)$ and $D(s)$ are zero.

MICROWAVE GENERATION FROM PHOTOCONDUCTIVE MIXING OF AMPLIFIED SPONTANEOUS RADIATION

By

M. C. Steele

RCA Laboratories
Princeton, N. J.

Summary—A model is developed for the generation of microwaves by photoconductive self-mixing of amplified spontaneous radiation in solids. The model is then used to propose an explanation for the microwave emission that has been observed from InSb in strong electric and magnetic fields.

Introduction

ABOUT TWO YEARS ago Larrabee and Hicinbothem[1] observed broad-band microwave emission from InSb subjected to strong electric and magnetic fields. Since that time many other groups have confirmed these experiments. The first attempts to explain the origin of such radiation[2] were based upon a possible two-stream instability postulated by Bok and Nozieres.[3] However, as more experimental data became available,[4] it became clear that some other mechanism was needed to explain this phenomenon. This paper proposes a new model that more closely fits the experimental situation. The model is based on the self-photoconductive mixing of the amplified spontaneous radiation that exists within the InSb when the electron–hole density produces the population inversion required for amplification of the band-gap radiation.

Model

We consider a bar of InSb with the dimensions shown in Figure 1 and with a large current passing along the length. The current con-

[1] R. D. Larrabee and W. A. Hicinbothem, Jr., "Observation of Microwave Emission from Indium Antimonide," *Proc. Symposium on Plasma Effects in Solids*, p. 181, Academic Press, Paris, 1964.

[2] M. C. Steele, "Possible Explanation for Microwave Emission from In-Sb in Magnetic Fields," *Proc. Symposium on Plasma Effects in Solids*, p. 189, Academic Press, Paris, 1964.

[3] J. Bok and P. Nozieres, "Instabilities of Transverse Waves in a Drifted Plasma," *Jour. Phys. Chem. Solids*, Vol. 24, p. 709, 1963.

[4] R. D. Larrabee and W. A. Hicinbothem, Jr. (private communication).

sists of electrons and holes drifting at velocities v_n and v_p, respectively. It is assumed that the applied electric field is sufficiently large to have caused band-gap ionization and that the added electron–hole density far exceeds the original extrinsic carrier concentration. Therefore, the electron and hole densities are assumed to be equal. Under these conditions, the model applies equally well to n-type or p-type InSb, the only difference being in the magnitude of the required electrostatic drifting field, E_0.

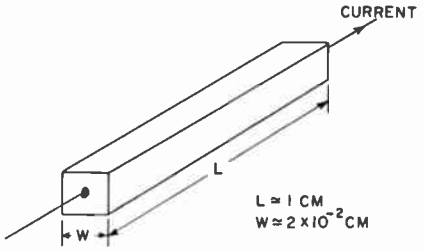

Fig. 1

At first we shall neglect the magnetostatic field, although we shall show later that such fields play a critical role in the model. When the excess electron–hole density is sufficient to cause the difference of the quasi-Fermi levels to exceed the band-gap energy, the crystal can amplify radiation of the proper frequency. We know that there is much spontaneous recombination radiation in the body as a result of excess electron–hole pairs. With the geometry of Figure 1 and the assumption of a very poor Q for the system, such spontaneous radiation can be amplified in a single pass along the length of the crystal. The situation is analogous to that of superradiance[5] in lasers. Superradiance can occur in a lossy resonator that has poor mode selectivity.

The frequency bandwidth of the spontaneous band-gap radiation present in the body will be approximately ξ_n/h, where ξ_n is the quasi-Fermi level of the electrons measured from the bottom of the conduction band. It is assumed here that ξ_p, the quasi-Fermi level of the holes, is anchored near the top of the valence band due to the much greater density of states in that band.

The total amount of power, P_T, in the amplified spontaneous radiation[6] is

[5] R. H. Dicke, "Coherence in Spontaneous Radiation Processes," *Phys. Rev.*, Vol. 93, p. 99, Jan. 1, 1954.
[6] G. Birnbaum, *Optical Masers*, p. 34, Academic Press, New York, 1964.

$$P_T \cong \frac{h\nu_0 nV}{\tau'}, \tag{1}$$

where ν_0 = frequency of band-gap radiation,

V = active volume of system,

n = density of electrons,

τ' = enhanced spontaneous emission rate.

In light of the assumption of single-pass amplification, we can define a coherence time, τ_K, for this radiation;

$$\tau_K = \frac{L\sqrt{\epsilon_L}}{c}, \tag{2}$$

where ϵ_L is the relative dielectric constant of the medium ($\simeq$18 for InSb). For $L = 1$ cm, $\tau_K \simeq 10^{-10}$ second.

Reabsorption of Radiation—Uniform Densities

If the entire volume of the crystal is uniformly filled with electrons and holes, the amplified radiation should *not* be strongly reabsorbed. The absence of reabsorption results from the position of the quasi-Fermi level of the electrons. For reabsorption to occur, the radiation must have a frequency greater than the maximum of the amplified radiation spectrum. Since a uniform density gives no significant reabsorption, there can be no photoconductive mixing of the various components in the spectrum.

Effect of Density Gradients on Reabsorption—Role of Magnetic Field

If we can establish a strong density gradient in the body, reabsorption becomes possible only in the portion of the body that has the lower density. Such density gradients can be established by several means:

(1) Self-pinching of the electron-hole plasma[7] in the absence of any applied magnetostatic fields,

[7] M. Glicksman and M. C. Steele, "Plasma Pinch Effects in Indium Antimonide," *Phys. Rev. Letters*, Vol. 2, p. 461, June 1, 1959.

(2) The Lorentz force exerted by a transverse magnetostatic field, which will concentrate the electron–hole plasma in a very narrow surface layer. This effect is often called the Suhl effect and is discussed in detail by Shockley.[8]

Figures 2(a) and 2(b) show the density distributions in a cross section of the material for these situations. The two conditions are really more similar than they appear. The Suhl-effect distribution is really one half of the pinch-effect distribution that has been moved to the surface because of the Lorentz force.

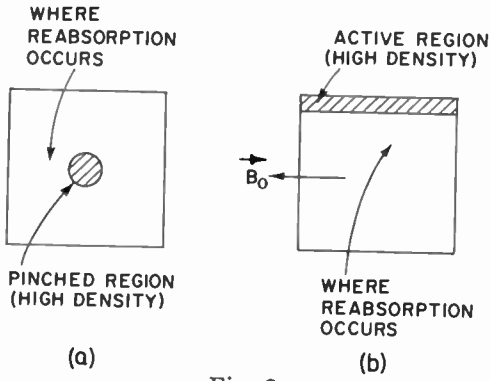

Fig. 2

The active role of a longitudinal magnetic field is more complicated. It clearly does not aid in producing the needed density gradients. In fact, it dilutes the effect of pinching (although this dilution is not a strong effect for the conductivities being considered) by spreading the pinch.[7] The effect of a longitudinal magnetic field is considered in more detail in the next section.

Other Role of Magnetic Field—Landau Levels

Any orientation of magnetic field will produce Landau levels that assist the establishment of the population inversion needed for amplification. For InSb, the Landau levels in the conduction band play the important role. With a magnetic field strong enough so that

$$\frac{3}{2}\hbar\omega_c \gtrsim \xi_n(\simeq kT), \tag{3}$$

[8] W. Shockley, *Electrons and Holes in Semiconductors*, p. 325, Van Nostrand Co., Inc., New York, 1950.

where ω_c is the electron-cyclotron frequency, most of the electrons will be in the lowest Landau level so that there can be a significant narrowing of the spontaneous emission spectrum (see Figure (3)). If Equation (3) is satisfied, the frequency bandwidth should be $\simeq 1/\tau$ (τ is the scattering time) instead of ξ_n/h. For the temperatures we will be concerned with ($T \simeq 77°\text{K}$), $1/\tau < \xi_n/h$. It then follows that magnetic fields sufficiently strong to satisfy Equation (3) will have a

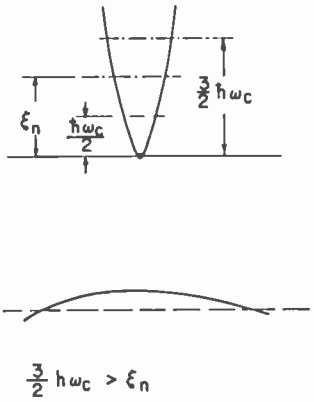

Fig. 3

significant effect on the process with which we are concerned. With $T \simeq 77°\text{K}$, the magnetic field needed to satisfy Equation (3) is $\vec{B}_0 \gtrsim 3000$ gauss.

This role of the magnetic field is closely related to the lowering of the threshold current needed to make an InSb p-n junction injection laser operate.[9]

Photoconductive Mixing of Amplified Spontaneous Radiation

Having established the necessary density gradients to get reabsorption, we will proceed to develop an expression for the power available at the mixed (microwave) frequencies. It has already been demonstrated[10] that a photoconductor can mix two coherent laser frequencies,

[9] R. J. Phelan, A. R. Calawa, R. H. Rediker, R. J. Keyes, and B. Lax, "Infrared InSb Laser Diode in High Magnetic Fields," *Appl. Phys. Letters*, Vol. 3, p. 143, 1 Nov. 1963.

[10] M. DiDomenico, Jr., R. H. Pantell, O. Svelto, and J. N. Weaver, "Optical Frequency Mixing in Bulk Semiconductors," *Appl. Phys. Letters*, Vol. 1, p. 77, 1 Dec. 1962.

ω_1 and ω_2, to give microwave power at the difference frequency $(\omega_1 - \omega_2)$. For two coherent laser beams of intensity I_1 and I_2 photons/sec, the photoconductive current developed is:[10]

$$i_p(\Omega) = \frac{q}{\tau_{tr}} \frac{\tau_0 \sqrt{I_1 I_2}}{(1 + \Omega^2 \tau_0^2)^{1/2}} \cos(\Omega t), \tag{4}$$

where τ_0 is the lifetime of added carriers, τ_{tr} the transit time of electrons, and Ω is the difference frequency of the two laser beams $(\omega_1 - \omega_2)$. In Equation (4) we have assumed, for simplicity, that the quantum efficiency for the generation of electron–hole pairs is unity. Since we are interested only in mixed frequencies satisfying the condition $\Omega\tau_0 \gg 1$, Equation (4) can be rewritten

$$I_p(\Omega) = \frac{q}{\tau_{tr}} \frac{\sqrt{I_1 I_2}}{\Omega} \cos(\Omega t). \tag{5}$$

Now let us take the total intensity of amplified spontaneous radiation derived from Equation (1),

$$I_T = \frac{P_T}{h\nu_0} = \frac{nV}{\tau'} \tag{6}$$

and distribute it equally over the spectrum extending from ω_0 (corresponding to the band-gap radiation) to ω_u, where $\omega_u = \omega_0 + \xi_n/h$. This simplification of the spectrum will not seriously alter the major conclusions we draw from the model. The photoconductive current at frequency Ω_0 and bandwidth $\Delta\Omega$ is then given by

$$i_p(\Omega_0) = \frac{q}{\tau_{tr}} \frac{I_T}{(\omega_u - \omega_0)} \int_{\Omega_0}^{\Omega_0 + \Delta\Omega} \frac{\cos(\Omega t)}{\Omega} d\Omega. \tag{7}$$

In Equation (7) it is understood that $\Omega_0 \ll (\omega_u - \omega_0)$; we are mainly interested in Ω_0 in the microwave region. For $\Delta\Omega/\Omega_0 \ll 1$, the mean-value theorem can be used to evaluate the integral in Equation (7). Squaring that current and taking the time average of τ_K gives

$$\overline{i_p(\Omega_0)^2} = \frac{q^2}{\tau_{tr}^2} \frac{I_T^2}{(\omega_u - \omega_0)^2} \left(\frac{\Delta\Omega}{\Omega_0}\right)^2 \left[\frac{1}{2} + \frac{\sin 2\Omega_0 \tau_K}{4\Omega_0 \tau_K}\right] \tag{8}$$

In Equation (8) the model is valid only for $\Omega_0\tau_K \gtrsim 1$. If this current is fed into a resistance matched to the source resistance, R, the useful average power at Ω_0 will be:

$$\overline{P(\Omega_0)} \simeq \frac{R}{4}\frac{q^2}{{\tau_{tr}}^2}\frac{{I_T}^2}{(\omega_u - \omega_0)^2}\left(\frac{\Delta\Omega}{\Omega_0}\right)^2. \tag{9}$$

To estimate I_T we first find the active volume of the solid. Previous experiments[1] have shown an average density of about 10^{15}/cm^3 at the time of emission. Since the density must be $\gtrsim 4 \times 10^{16}$/cm^3 for stimulated band-gap emission to occur, the active volume must be about 1/40 of the total volume of the body. Next, we must determine a value for τ' in Equation (6). Once stimulated emission occurs, the limiting rate will be determined by how fast the electrons can be pumped from the valence band back to the conduction band. This time can be estimated in several ways and comes out to about 10^{-11} second. With these parameters and the dimensions assumed in Figure 1, $I_T \simeq 4 \times 10^{22}$ photons/second. Typical resistance values in the experiments[1] are about 100 ohms. The electron drift velocity is 10^6 meters/second and $\xi_n \simeq 8 \times 10^{-3}$ ev. With these values, the estimate of the microwave power from Equation (9) is

$$\overline{P(\Omega_0)} \simeq 6 \times 10^{-2}\left(\frac{\Delta\Omega}{\Omega_0}\right)^2 \text{ watt}. \tag{10}$$

If we take $\Omega_0 = 10$ gc and $\Delta\Omega = 5$ mc, Equation (10) gives a power level of about 2×10^{-8} watt. This is the order of magnitude of values observed experimentally[1] at this frequency and bandwidth. The predicted $1/\Omega_0^2$ frequency spectrum is also in broad agreement with the observations.[1,4] Larrabee and Hicinbothem[4] have made interference measurements of the observed radiation at 15 gc and found that the coherence time is $\simeq 10^{-10}$ second. This is an important agreement between theory and experiment.

Discussion

In principle, the model described here would also be valid without any applied magnetostatic field if a sufficiently large population inversion is provided. Experiments should be continued to higher currents to test this prediction. However, for such a case the necessary density gradients to obtain reabsorption would have to be provided by the self-pinching of the electron–hole plasma. Some additional aspects of

ANALYSIS OF NONCOHERENT FSK SYSTEMS WITH LARGE RATIOS OF FREQUENCY UNCERTAINTIES TO INFORMATION RATES

By

A. B. Glenn

RCA Defense Electronic Products
Moorestown, N. J.

Summary—This paper discusses noncoherent binary and nonbinary frequency shift keyed (FSK) systems in which the predetection filter bandwidth is very much greater than the information bandwidth. A predetection bandwidth of at least ten times the post-detection matched-filter bandwidth is assumed. Expressions for the information bit-error probability as a function of the channel input (predetection) signal-to-noise ratios (SNR) are derived for both linear and square-law envelope detectors.

In Part I, wide-band binary FSK systems are considered. It is shown that the square-law detector is superior for low SNR (less than 0 db) whereas the linear detector is best for high SNR. Preliminary experimental data verifies the theoretical analysis to within 1/2 db at error probabilities of 10^{-4}. Part II is concerned with wide-band nonbinary FSK systems. The results of the analysis shows the same trends as for a nonbinary FSK system that uses predetection matched filters. It is shown that, for the same word-error probability, the required channel SNR decreases as the information per word increases. For example, a comparison of a 5-bit word for the 32-ary (or 32-frequency) wide-band noncoherent FSK system with a 5-bit word for a binary (or 2-frequency) wide-band noncoherent FSK system showed that, for the same error probability, the 32-ary system required about 6 db less SNR than the binary system.

PART I — BINARY SYSTEMS

Introduction

THIS PART IS concerned with noncoherent binary FSK (frequency shift keyed) systems in which, due to frequency uncertainty, the predetection filter bandwidth is very much greater (a factor of 10 or more) than the information bandwidth. With post-detection filters matched to the transmitted signal, expressions for bit-error probability as a function of channel (mark or space) input-to-noise ratios were derived for square-law and linear envelope detectors. The νth-law detector is analyzed insofar as it acts as a square-

law detector for very low signal-to-noise ratios (SNR) and as a linear detector for very large SNR's. Because the relationships for a νth-law detector between these extremes of SNR are mathematically very complex, they were analyzed on the basis of experimental detector characteristics.

Wideband FSK Using Envelope Detectors Followed by Matched Filters

In this section, linear envelope detectors followed by post-detection match filters are analyzed. One of the two channels in this type of

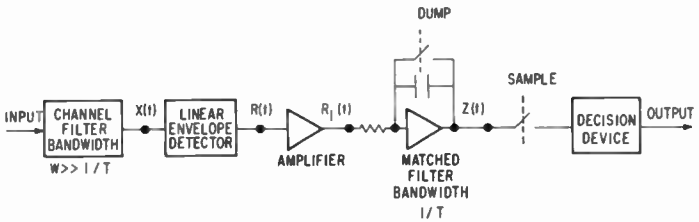

Fig. 1—Single-channel wide-band FSK detector.

detector is shown in Figure 1. The detector input $x(t)$ when a signal and narrow-band noise are present is

$$x(t) = A \cos \omega_c t + I_n(t), \tag{1}$$

where[1]

$$I_n(t) = I_c(t) \cos \omega_c t - I_s(t) \sin \omega_c t \tag{2}$$

$$\text{signal} = A \cos \omega_c t, \quad 0 \leq t \leq T$$

$$= 0, \quad \text{otherwise.}$$

Both $I_c(t)$ and $I_s(t)$ are slowly changing random variables with a normal distribution. Both have zero average value and the same second moment, or mean-squared value. Thus,

$$\overline{I_c} = \overline{I_s} = 0 \tag{3}$$

$$\overline{I_c^2} = \overline{I_s^2}. \tag{4}$$

[1] S. O. Rice, *Mathematical Analysis of Random Noise*, Selected Papers on Noise and Stochastic Processes by N. Wax, Dover Publication.

If a signal with an input-noise-power spectral density η is passed through a filter of transfer function $H(j\omega)$ centered at frequency f_c, then the output-power spectral density of the filter is[2]

$$G_H(\omega) = \eta \, |H(j\omega)|^2. \tag{5}$$

The total noise output of the filter is

$$N = \int_0^\infty G_H(\omega)\, df. \tag{6}$$

Therefore, the variance of the noise $I_n(t)$ is

$$\overline{I_c^2(t)} = \overline{I_s^2(t)} = N. \tag{7}$$

Equation (1) may then be written

$$\begin{aligned} x(t) &= (A + I_c)\cos\omega_c t - I_s \sin\omega_c t \\ &= x_1(t) - x_2(t). \end{aligned} \tag{8}$$

The probability density distributions of $x(t)$ are

$$P(x_1) = \frac{\exp\left\{\dfrac{-(x_1 - A)^2}{2N}\right\}}{\sqrt{2\pi N}} \tag{9}$$

$$P(x_2) = \frac{\exp\left\{\dfrac{-x_2^2}{2N}\right\}}{\sqrt{2\pi N}}. \tag{10}$$

An equivalent form of $x(t)$ is

$$x(t) = R(t)\cos[\omega_c t + \theta(t)] \tag{11}$$

where $R(t)$ is the envelope of $x(t)$. Thus,

$$R^2(t) = R^2 = (A + I_c)^2 + I_s^2. \tag{12}$$

---

[2] Y. W. Lee, *Statistical Theory of Communication*, John Wiley & Sons, Inc., Chapter 13, Section 5.

The linear detector will now operate upon the input voltage such that the output voltage will be directly proportional to the envelope of the input voltage, i.e., the probability density function (pdf) of the output voltage is the same as that of the envelope of the input voltage.

Rice[1] has shown that the output of the envelope detector is given by

$$P(R) = \frac{R}{N} \exp\left[-\frac{R^2 + A^2}{2N}\right] I_0\left[\frac{RA}{N}\right], R \geq 0$$
$$= 0, \quad R < 0 \tag{13}$$

for the case of signal plus noise at the input, and by

$$P(R) = \frac{R}{N} \exp\left[-\frac{R^2}{2N}\right], \quad R \geq 0$$
$$= 0, \quad R < 0 \tag{14}$$

for the case of noise alone at the input. $I_0(z)$ is the modified Bessel function of the first kind and zero order.

The energy of a signal of duration T seconds is

$$E = \int_0^T A^2 \cos^2 \omega_c t \, dt$$
$$= \frac{A^2}{2} T. \tag{15}$$

The single-sided noise-power spectral density at the output of the channel filter of bandwidth W is

$$N_0 = \frac{N}{W}. \tag{16}$$

Then Equations (13) and (14) may be rewritten

$$P(R) = \frac{R}{N_0 W} \exp\left[-\frac{R^2 + \frac{2E}{T}}{2N_0 W}\right] I_0\left[\sqrt{\frac{2E}{T}} \frac{R}{N_0 W}\right], \quad R \geq 0$$
$$= 0, \quad R < 0, \tag{17}$$

and

$$P(R) = \frac{R}{N_0 W} \exp\left[-\frac{R^2}{2N_0 W}\right], \quad R \geq 0$$
$$= 0, \quad R < 0. \tag{18}$$

It is convenient to let the post-detection filter consist of an amplifier with a voltage gain of $\sqrt{T}$ and an ideal integrator with an impulse response $h(t)$. The output of the amplifier is

$$R_1(t) = \sqrt{T}\, R(t). \tag{19}$$

Thus,* for the case of signal plus noise at the input,

$$P(R_1)\, dR_1 = P(R)\, dR$$
$$P(R_1) = \frac{R_1}{N_0 WT} \exp\left[-\frac{R_1^2 + 2E}{2N_0 WT}\right] I_0\left[\frac{\sqrt{2E}\, R_1}{N_0 WT}\right], \quad R_1 \geq 0$$
$$= 0, \quad R_1 < 0, \tag{20}$$

and for the case of noise alone at the input,

$$P(R_1) = \frac{R_1}{N_0 WT} \exp\left[-\frac{R_1^2}{2N_0 WT}\right], \quad R_1 \geq 0$$
$$= 0, \quad R_1 < 0. \tag{21}$$

The output of a filter with an impulse response $h(t)$ is†[3]

$$z(t) = \int_{t-T}^{t} h(t) R_1(t-u)\, du$$
$$= \int_{0}^{T} du h(t) R_1(t-u), \tag{22}$$

where the impulse response of an ideal integrator is

---

* See Reference (2), Chapter 6, Section 4.

† See Reference (2), Chapter 6, Section 1.

[3] W. W. Harman, *Principles of the Statistical Theory of Communications*, McGraw-Hill Company, Sections 2-4, 3-3, and 6-5.

$$h(t) = \frac{1}{T}, \quad 0 \leq t \leq T$$
$$= 0, \quad \text{otherwise.} \tag{23}$$

To determine the probability density function of $z(t)$ when the pdf of $R(t)$ is known to a close approximation, the integral of Equation (22) may be replaced by the sum of a number of independent samples:*[4]

$$z(t) \doteq \frac{1}{T} \sum_{m=1}^{n} \Delta T \, R_1(t_m), \tag{24}$$

where $R(t_m)$ is the sample of $R(t)$ taken at time t_m, ΔT is the incremental time interval centered on t_m, and $n\Delta T = T$. Thus,

$$z(t) \doteq \frac{1}{n} \sum_{m=1}^{n} R_1(t_m). \tag{25}$$

Equation (25) is the sampled mean of $R_1(t_m)$. The number of independent samples for a rectangular channel band-pass characteristic of W cycles per second is W samples per second.[5] Thus, in a signal of duration T seconds the number of independent samples is

$$n = WT. \tag{26}$$

The central limit theorem states that the probability distribution of the sum of n statistically independent random variables approaches a Gaussian distribution as n increases without limit, regardless of the probability distribution of the random variables as long as each random variable has a finite mean and a finite variance.[4] For a finite n, however, the normal estimate of the density of the sum is only an approximation.[6] It is necessary to determine the error between the actual distribution and the normal distribution as a function of the number of independent samples or random variables. This subject is discussed in detail by Marcum[7] for square-law and linear detectors. Expressions

---

* See Reference (2), Chapter 11.

[4] W. B. Davenport and W. L. Root, *Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Co., Chapter 5.

[5] M. Schwartz, *Information Transmission Modulation and Noise*, McGraw-Hill Book Co., Chapter 4, Section 4-5.

[6] A. Papoulis, *Probability Random Variables and Stochastic Processes*, McGraw-Hill Book Co., Chapter 8, pp. 266 to 272.

[7] J. I. Marcum, "Statistical Theory of Target Detection by Pulse Radar," *IRE Trans. on I.T.*, p. 174, April 1960.

are given for the density and cumulative distribution functions, which depend upon the number of samples n and the SNR.* Figures 1 to 7 in Reference (7) give the probability density for the envelope and sine wave plus noise for the square-law detector, and Figures 36 to 41 show the density distributions for the linear detector. A comparison of Figures 5 and 39 in Reference (7), which are for noise only, shows that as n increases from 1 to 9 the density distribution approaches the Gaussian distribution more rapidly for the linear detector than for the square-law detector. Figure 42 in Marcum's paper shows a comparison of the probability of detection for the square-law and linear detectors as a function of the number of samples and SNR. For a probability of detection (P_D) of 99% and for $n = 10$, the linear detector requires a SNR of 3.8 db, while the square-law detector requires a SNR of 4.2 db. As n increases, the difference between the two types of detectors decreases and the SNR required for a given P_D decreases. Marcum's equations for the probability density distribution permit estimations of the error between the actual and Gaussian distributions at any point along the tails of the distribution. As a first approximation, it is seen from his curves that for $n = 10$, the actual distribution closely approaches the Gaussian distribution to about the 3σ point. As n increases, the approximation rapidly improves along the tails of the distribution.

Thus, for $WT \gg 1$ the pdf of the output of the matched filter or the input of the decision device is

$$P(z) \doteq \frac{1}{\sqrt{2\pi}\,\sigma_z} \exp\left\{-\frac{(z-m_z)^2}{2\sigma_z^2}\right\} \tag{27}$$

where[4]

$$m_z = \frac{1}{WT} \sum_{m=1}^{WT} \overline{R_1(t_m)} \tag{28}$$

$$= m_{R_1}$$

$$\sigma_z^2 = \overline{(z-m_z)^2} = \frac{\sigma_{R_1}^2}{WT}. \tag{29}$$

Equations (28) and (29) state that the sample mean is equal to the mean of the sampled function and the variance of the sample mean is

* See Reference (7), Equation (37) for the square-law detector and Equation (117e) for the linear detector.

equal to the variance of the function divided by the sample size for independent discrete samples.

m_{R_1} and $\sigma_{R_1}^2$ may be found from Rice's results* and Equations (20) and (21).

$$\overline{R_1^n} = (2\psi_0)^{n/2}\Gamma\left(\frac{n}{2}+1\right) {}_1F_1\left(-\frac{n}{2}\,;\,1\,;\,\frac{-P^2}{2\psi_0}\right), \tag{30}$$

where

${}_1F_1(a, c, z)$ is the hypergeometric function,[7†]

$\psi_0 = N_0WT$

$P^2 = 2E$

$\Gamma(n) = \int_0^\infty t^{n-1}e^{-t}dt$ is the gamma function.

Thus, when signal and noise are present the first and second moments are

$$\begin{aligned}\overline{R_1} &= (2N_0WT)^{1/2}\,\Gamma\left(\frac{3}{2}\right) {}_1F_1\left(-\frac{1}{2}\,;\,1\,;\,\frac{-E}{N_0WT}\right)\\ &= \left(\frac{N_0WT\pi}{2}\right)^{1/2} {}_1F_1\left(-\frac{1}{2}\,;\,1\,;\,\frac{-E}{N_0WT}\right)\end{aligned} \tag{31}$$

and

$$\overline{R_1^2} = 2E + 2N_0WT. \tag{32}$$

Therefore,

$$\begin{aligned}\sigma_{R_1}^2 &= \overline{R_1^2} - (\overline{R}_1)^2\\ &= 2E + 2N_0WT - \frac{N_0WT\pi}{2}\,{}_1F_1^2\left(-\frac{1}{2}\,;\,1\,;\,\frac{-E}{N_0WT}\right).\end{aligned} \tag{33}$$

When noise alone is present the first and second moments are

$$R_1 = \left(\frac{N_0WT\pi}{2}\right)^{1/2} \tag{34}$$

* See Reference (1), pp. 238-239.

and

$$\overline{R_1^2} = 2N_0WT. \tag{35}$$

Therefore,

$$\sigma_{R_1}^2 = 2N_0WT\left(1 - \frac{\pi}{4}\right) \tag{36}$$

The two conditional pdf of z for signal plus noise present and only noise present are, respectively,

$$P(z|S+N) = \frac{1}{\sqrt{2\pi}\,\sigma_{z_{S+N}}} \exp\left\{-\frac{(z - m_{z_{S+N}})^2}{2\sigma_{z_{S+N}}^2}\right\} \tag{37}$$

$$P(z|N) = \frac{1}{\sqrt{2\pi}\,\sigma_{z_N}} \exp\left\{-\frac{(z - m_{z_N})^2}{2\sigma_{z_N}^2}\right\} \tag{38}$$

where, since from Equations (28) and (29) $m_z = m_{R_1}$ and $\sigma_z^2 = \sigma_{R_1}^2/WT$,

$$m_{z_{S+N}} = \left(\frac{N_0WT\pi}{2}\right)^{1/2} {}_1F_1\left(-\frac{1}{2}\,;1;-\frac{E}{N_0WT}\right) \tag{39}$$

$$m_{z_N} = \left(\frac{N_0WT\pi}{2}\right)^{1/2} \tag{40}$$

$$\sigma_{z_{S+N}}^2 = 2N_0\left[\frac{E}{N_0WT} + 1 - \frac{\pi}{4}\,{}_1F_1^2\left(-\frac{1}{2}\,;1;-\frac{E}{N_0WT}\right)\right] \tag{41}$$

$$\sigma_{z_N}^2 = 2N_0\left(1 - \frac{\pi}{4}\right). \tag{42}$$

Error Probability

The above results may now be used to determine the error probability. The FSK signal may be represented as

$$x(t) = x_1(t)\cos(\omega_1 + \Delta\omega)t + x_0(t)\cos(\omega_0 + \Delta\omega)t + I_n(t) \tag{43}$$

where $\Delta\omega$ is the frequency uncertainty and, during each binary digit

of duration T, either $x_1(t) = A$ and $x_0(t) = 0$ or $x_0(t) = A$ and $x_1(t) = 0$. The energy of either binary digit is $A^2T/2$. Figure 2 shows the basic wide-band noncoherent FSK detector.

The bandwidth of the predetection mark and space filters is W, where $WT \gg 1$. Thus, as previously discussed, the pdf of z closely approximates a Gaussian distribution. For a $WT = 10$, this approxi-

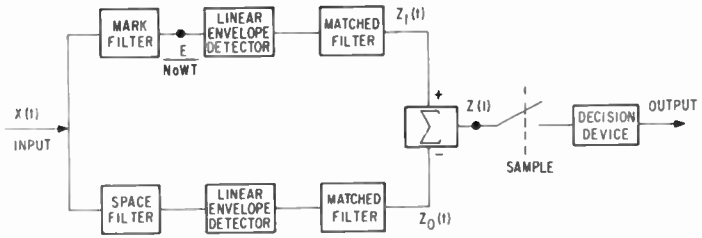

Fig. 2—Basic noncoherent wide-band FSK detector.

mation is good to about the 3σ point. The difference $z = z_1 - z_0$ will also be Gaussian with a mean

$$|m_z| = |m_{z_{S+N}} - m_{z_N}|, \tag{44}$$

and a variance

$$\sigma_z^2 = \sigma_{z_{S+N}}^2 + \sigma_{z_N}^2. \tag{45}$$

It will be assumed that the receiver makes the following hard decision:

If $z > 0$, the binary symbol corresponding to x_1 is called out;

If $z < 0$, the binary symbol corresponding to x_0 is called out.

Then the probability of error P_e (i.e., the probability that the receiver reproduces the wrong symbol) is

$$[P(z > 0 \,|\, x_0 \text{ sent})]\,[P(x_0 \text{ sent})] + [P(z < 0 \,|\, x_1 \text{ sent})]\,[P(x_1 \text{ sent})], \tag{46}$$

where

$P(z > 0 \,|\, x_0 \text{ sent})$ is the conditional probability that $z > 0$ given that x_0 was sent, or that the receiver decided x_1 was sent under the condition that x_0 was sent.

$P(z < 0 \,|\, x_1 \text{ sent})$ is the conditional probability that $z < 0$ given that x_0 was sent, or that the receiver decided that x_0 was sent under the condition that x_1 was sent.

$P(x_0)$ is the a priori probability that x_0 was sent

$P(x_1)$ is the a priori probability that x_1 was sent.

Figure 3 shows the pdf of $P(z|x_0)$ and $P(z|x_1)$. From Equations

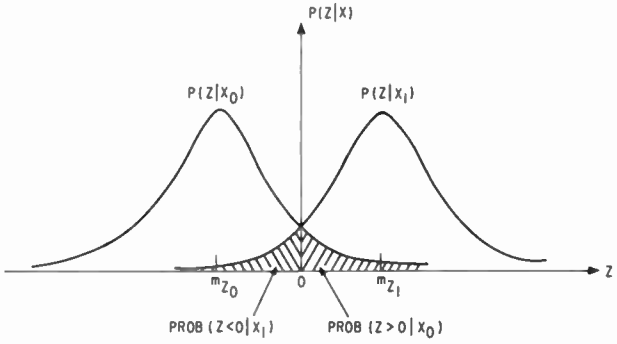

Fig. 3—Matched filter output probability density distributions.

(37) and (38)

$$P(z>0|x_0) \doteq \frac{1}{\sqrt{2\pi}\,\sigma_z}\int_0^{\infty} \exp\left\{-\frac{(z+m_{z_0})^2}{2\sigma_z{}^2}\right\} dz \qquad (47)$$

$$P(z<0|x_1) \doteq \frac{1}{\sqrt{2\pi}\,\sigma_z}\int_{-\infty}^{0} \exp\left\{-\frac{(z-m_{z_1})^2}{2\sigma_z{}^2}\right\} dz \qquad (48)$$

$$P(x_0) = P(x_1) = \frac{1}{2} \qquad (49)$$

where $$\sigma_z{}^2 = \sigma_{z_{S+N}}{}^2 + \sigma_{z_N}{}^2 = \sigma_{z_{11}}{}^2 + \sigma_{z_{01}}{}^2 = \sigma_{z_{00}}{}^2 + \sigma_{z_{10}}{}^2 \qquad (50)$$

$$m_{z_0} = m_{z_N} - m_{z_{S+N}} = m_{z_{10}} - m_{z_{00}}$$

$$m_{z_1} = m_{z_{S+N}} - m_{z_N} = m_{z_{11}} - m_{z_{01}}. \qquad (51)$$

$m_{z_{10}}$ is the mean value in the mark channel when a space (x_0) was sent $= m_{z_N}$

$m_{z_{00}}$ is the mean value in the space channel when a space was sent $= m_{z_{S+N}}$

$m_{z_{11}}$ is the mean value in the mark channel when a mark (x_1) was sent $= m_{z_{S+N}}$

$m_{z_{01}}$ is the mean value in the space channel when a mark was sent $= m_{z_N}$

m_{z_0} and m_{z_1} are the mean values of the signal at the sampling point.

But
$$m_{z_{10}} = m_{z_{01}} \tag{52}$$

$$m_{z_1} = -m_{z_0}. \tag{53}$$

Therefore, from Equations (46)-(53),

$$P_e \doteq \frac{1}{2\sqrt{2\pi}\,\sigma_z}\left[\int_{-\infty}^{0} \exp\left\{-\frac{(z-m_{z_1})^2}{2\sigma_z^2}\right\} dz + \int_{0}^{\infty} \exp\left\{-\frac{(z+m_{z_0})^2}{2\sigma_z^2}\right\} dz\right]. \tag{54}$$

If we let,

$$w = \frac{z-m_{z_1}}{\sqrt{2\sigma_z^2}} = \frac{z+m_{z_0}}{\sqrt{2\sigma_z^2}}, \tag{55}$$

then

$$P_e \doteq \frac{1}{2\sqrt{\pi}}\left[\int_{-\infty}^{-m_{z_1}/\sqrt{2\sigma_z^2}} e^{-w^2}\,dw + \int_{m_{z_0}/\sqrt{2\sigma_z^2}}^{\infty} e^{-w^2}\,dw\right] \tag{56}$$

$$\doteq \frac{1}{\sqrt{\pi}} \int_{m_{z_0}/\sqrt{2\sigma_z^2}}^{\infty} e^{-w^2}\,dw. \tag{57}$$

But,

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-t^2}\,dt = 1 - \operatorname{erf}(z). \tag{58}$$

Thus, since m_{z_0} is always negative,

$$P_e \doteq \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{|m_{z_0}|}{\sqrt{2}\,\sigma_z}\right)\right], \quad WT \geq 10. \tag{59}$$

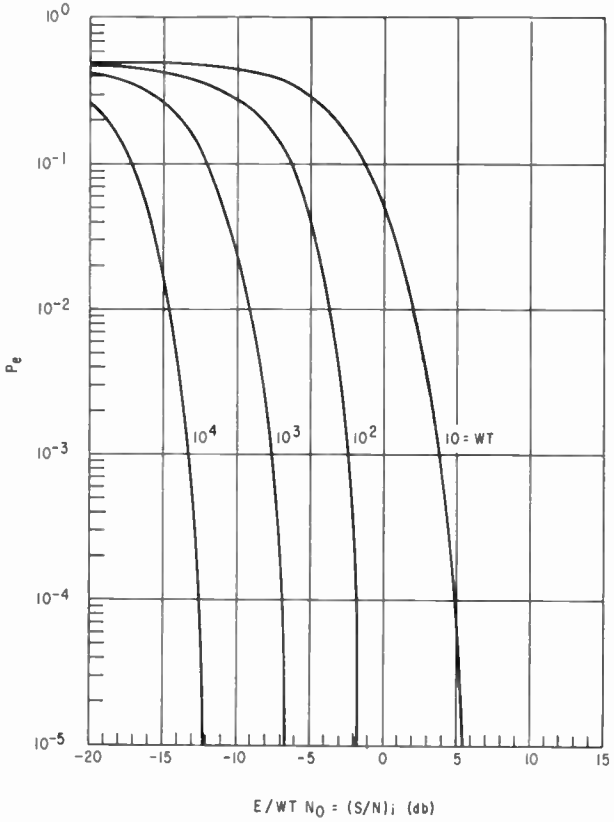

Fig. 4—Bit-error probability versus input SNR for wide-band FSK using linear envelope detection.

This equation is plotted in Figure 4. In Equation (59),

$$\sigma_z^2 = \sigma_{z_{S+N}}^2 + \sigma_{z_N}^2$$
$$= 2N_0\left[\frac{E}{N_0WT} + 2 - \frac{\pi}{4}\left\{1 + {}_1F_1{}^2\left(-\frac{1}{2}\,;1;-\frac{E}{N_0WT}\right)\right\}\right], \tag{60}$$

$$m_{z_0} = m_{z_N} - m_{z_{S+N}}$$
$$= m_{z_{10}} - m_{z_{00}}$$
$$= \left(\frac{N_0 WT\pi}{2}\right)^{1/2} \left[1 - {}_1F_1\left(-\frac{1}{2}\,;1;-\frac{E}{N_0 WT}\right)\right]. \tag{61}$$

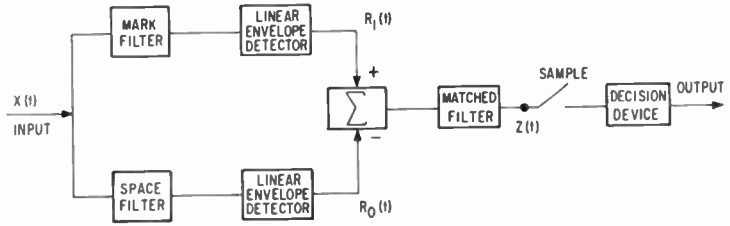

Fig. 5—Noncoherent wide-band FSK detector.

Therefore, since m_{z_0} is always negative, the SNR at the decision device is

$$\frac{|m_{z_0}|}{\sqrt{2}\,\sigma_z} = \frac{\left(\frac{N_0 WT\pi}{2}\right)^{1/2}\left[{}_1F_1\left(-\frac{1}{2}\,;1;-\frac{E}{N_0 WT}\right)-1\right]}{\sqrt{2}\left\{2N_0\left[\frac{E}{N_0 WT}+2-\frac{\pi}{4}\left(1+{}_1F_1{}^2\left(-\frac{1}{2}\,;1;-\frac{E}{N_0 WT}\right)\right)\right]\right\}^{1/2}}$$

$$= \frac{1}{2}\left(\frac{\pi}{2}\right)^{1/2} \frac{\sqrt{WT}\left[{}_1F_1\left(-\frac{1}{2}\,;1;-\frac{E}{N_0 WT}\right)-1\right]}{\left\{\frac{E}{N_0 WT}+2-\frac{\pi}{4}\left(1+{}_1F_1{}^2\left(-\frac{1}{2}\,;1;-\frac{E}{N_0 WT}\right)\right)\right\}^{1/2}} \tag{62}$$

An equivalent form of the noncoherent FSK detector is shown in Figure 5. If the impulse response of the post-detection filters in Figures 2 and 5 is $h(t)$, then the signal at the sampling point for Figure 2 can be written

$$z = z_1 - z_0$$
$$= \int_{t-T}^{t} R_1 h(t-\tau)\,d\tau - \int_{t}^{t} \cdot_{\tau} p(t-\tau)\,h_0 R\,d\tau \tag{63}$$

Since both integrals are convergent,

$$z = \int_{t-T}^{t} (R_1 - R_0) h(t-\tau) d\tau. \tag{64}$$

Equation (64) also represents the signal at the sampling point in Figure 5; thus, the detectors shown in Figures 2 and 5 are equivalent.

Bounds of Error Probabilities for Extremes of SNR

We return to Equation (62). The hypergeometric function ${}_1F_1(-1/2; 1; -z)$ is given by Rice† as

$${}_1F_1\left(-\frac{1}{2}; 1; -z\right) = e^{-z/2}\left[(1+z)\, I_0\left(\frac{z}{2}\right) + zI_1\left(\frac{z}{2}\right)\right] \tag{65}$$

Thus,

$${}_1F_1\left(-\frac{1}{2}; 1; -\frac{E}{N_0WT}\right) = \exp\left\{-\frac{E}{2N_0WT}\right\}$$
$$\left[\left(1+\frac{E}{N_0WT}\right) I_0\left(\frac{E}{2N_0WT}\right) + \frac{E}{N_0WT} I_1\left(\frac{E}{2N_0WT}\right)\right] \tag{66}$$

For small S/N ratio or small values of the arguments of the Bessel functions[8] with ν fixed and $z \to 0$,

$$I_\nu(z) \doteq \frac{\left(\frac{1}{2}z\right)^\nu}{\Gamma(\nu+1)} \tag{67}$$

$$\Gamma(\nu+1) = \nu! \quad \text{for} \quad \nu = 0,\ 1,\ 2,\ \cdots \tag{68}$$
$$I_0(z) \doteq 1, \quad z \to 0$$
$$I_1(z) \doteq 0, \quad z \to 0.$$

† See Reference (1), Appendix (4B).

[8] *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series 55, Chapter 9, p. 375, June 1964.

Thus, for $E/(WTN_0) \ll 1$,

$$
{}_1F_1\left(-\frac{1}{2}\,;1;-\frac{E}{N_0WT}\right) \doteq \exp\left\{-\frac{E}{2N_0WT}\right\}\left(1+\frac{E}{N_0WT}\right)
$$

$$
\doteq \left(1-\frac{E}{2N_0WT}+\cdots\right)\left(1+\frac{E}{N_0WT}\right)
$$

If we neglect high order terms of $E/(WTN_0)$,

$$
{}_1F_1\left(-\frac{1}{2}\,;1;-\frac{E}{N_0WT}\right) \doteq 1+\frac{E}{2WTN_0}. \tag{69}
$$

Equation (62) then becomes

$$
\frac{|m_{z_0}|}{\sqrt{2}\,\sigma_z} \doteq \frac{\frac{1}{2}\left(\frac{\pi}{2}\right)^{1/2}\sqrt{WT}\left[\frac{E}{2WTN_0}\right]}{\left\{\frac{E}{N_0WT}+2-\frac{\pi}{4}\left(2+\frac{E}{WTN_0}\right)\right\}^{1/2}}.
$$

$$
\frac{|m_{z_0}|}{\sqrt{2}\,\sigma_z} \doteq \frac{\frac{1}{2}\sqrt{WT}\left[\frac{E}{2WTN_0}\right]}{\sqrt{\frac{4}{\pi}-1}}. \tag{70}
$$

Therefore, for $E/(WTN_0) \ll 1$,

$$
P_e \doteq \frac{1}{2}\left[1-\operatorname{erf}\left\{\frac{\pi WT}{16(4-\pi)}\left(\frac{E}{WTN_0}\right)^2\right\}^{1/2}\right]
$$

$$
\text{for} \quad WT \geq 10. \tag{71}
$$

Equation (71) shows the square-law degradation for very low signal-to-noise ratio.

For large values of the argument* ($|z| \gg 1$), for fixed ν, and $\mu = 4\nu^2$,

$$I_\nu(z) \doteq \frac{e^z}{\sqrt{2\pi z}}\left\{1 - \frac{\mu-1}{8z} + \frac{(\mu-1)(\mu-9)}{2!(8z)^2} - \cdots\right\} \tag{72}$$

$$I_0(z) \doteq \frac{e^z}{\sqrt{2\pi z}}\left\{1 + \frac{1}{8z} + \frac{9}{2!(8z)^2}\right\} \tag{73}$$

$$I_1(z) \doteq \frac{e^z}{\sqrt{2\pi z}}\left\{1 - \frac{3}{8z} - \frac{15}{2!(8z)^2}\right\}. \tag{74}$$

For $z \geqq 10$,

$$I_0(z) \doteq I_1(z). \tag{75}$$

Thus, for large SNR,

$$\begin{aligned} {}_1F_1\left(-\frac{1}{2}\,;1;-\frac{E}{N_0WT}\right) &\doteq \exp\left\{-\frac{E}{2N_0WT}\right\} \\ &\left[\left(1+\frac{E}{N_0WT}\right) I_0\left(\frac{E}{2N_0WT}\right) + \frac{E}{N_0WT} I_0\left(\frac{E}{2N_0WT}\right)\right] \\ &\doteq e^{-z}\left[(1+4z)\frac{e^z}{\sqrt{2\pi z}}\right] \doteq \frac{4z}{\sqrt{2\pi z}} \doteq \frac{2}{\sqrt{\pi}}\sqrt{\frac{E}{N_0WT}}, \end{aligned} \tag{76}$$

and

$$\begin{aligned} \frac{|m_{z_0}|}{\sqrt{2}\,\sigma_z} &\doteq \frac{\frac{1}{2}\left(\frac{\pi}{2}\right)^{1/2}\sqrt{WT}\left[\frac{2}{\sqrt{\pi}}\sqrt{\frac{E}{N_0WT}} - 1\right]}{\left\{\frac{E}{N_0WT} + 2 - \frac{\pi}{4}\left(1 + \frac{4}{\pi}\frac{E}{N_0WT}\right)\right\}^{1/2}} \\ &\doteq \frac{\sqrt{\frac{E}{2N_0}} - \sqrt{\frac{\pi WT}{8}}}{\left(2 - \frac{\pi}{4}\right)^{1/2}}. \end{aligned}$$

* See Reference (8), p. 377.

Thus,

$$P_e \doteq \frac{1}{2}\left[1 - \operatorname{erf}\left(\sqrt{\frac{E}{2N_0}} - \sqrt{\frac{\pi WT}{8}}\right)\right],$$

$$\left(\frac{S}{N}\right)_i = \frac{E}{WN_0T} \gg 1, \quad WT \geq 10. \tag{77}$$

Comparing this expression with the coherent FSK system,[9] $P_e = 1/2\,[1 - \operatorname{erf}\sqrt{E/(2N_0)}]$, we see that for very large E/N_0 and for $E/N_0 \gg WT$, the wide-band FSK system using linear detectors is equivalent to the optimum FSK system. This result is reasonable since the Rician envelope distribution given by Equation (20) approaches a Gaussian distribution.* Also, the contribution of the noise of the channel that does not contain the signal is negligible compared to the total noise at the sampling point for very large SNR.† This is further discussed later.

Analysis of Wideband FSK Systems Utilizing Square-Law Detection Followed by Matched Filtering

In the previous section the wideband FSK system using *linear detection* followed by video matched filtering was analyzed. The derivation of the error probability analysis for a wideband FSK system using *square-law detection* followed by video matched filtering follows very closely the analysis given for linear detection. The block diagram for this system is given in Figure 6.

As discussed in the previous section, under the condition that $WT \gg 1$, the central limit theorem shows that the probability density distribution (pdf) at the output of the matched filter approaches a Gaussian distribution. Thus, the conditional pdf that a mark or x_1 was sent is

$$P(v|x_1) \doteq \frac{1}{\sqrt{2\pi}\,\sigma_v} \exp\left\{-\frac{(v - m_v)^2}{2\sigma_v^2}\right\}, \tag{78}$$

and the conditional pdf that a zero or x_0 was sent is

[9] J. G. Lawton, "Comparison of Binary Data Transmission Systems," *1958 Conference Proc. PGME.*

* See Reference (1), p. 240.

† See Reference (4), p. 263, Equations (12)-(65).

$$P(v \mid x_0) = \frac{1}{\sqrt{2\pi}\,\sigma_v} \exp\left\{-\frac{(v+m_v)^2}{2\sigma_v^2}\right\}, \tag{79}$$

where

$$m_v = m_{S+N} - m_N \tag{80}$$

$$\sigma_v^2 = \sigma_N^2 + \sigma_{S+N}^2 \tag{81}$$

m_{S+N} is the mean value in the mark channel when a mark was sent (and similarly for the space channel)

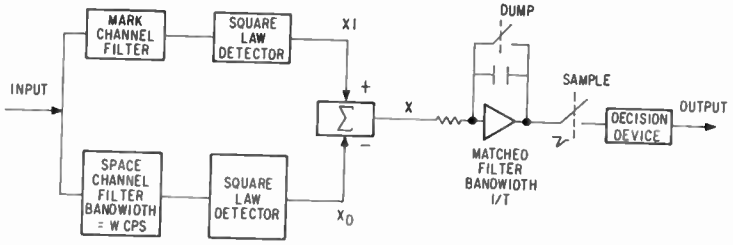

Fig. 6—Wide-band FSK system with square-law detectors.

m_N is the mean value in the mark channel when a space was sent (or vice versa)

σ_{S+N}^2 is the variance in the mark channel when a mark was sent (and similarly for the space channel)

σ_N^2 is the variance in the mark channel when a space was sent (and vice versa).

It can be shown that

$$m_N = 2WTN_0 \tag{82}$$

$$\sigma_N^2 = \frac{(2WTN_0)^2}{WT} \tag{83}$$

$$m_{S+N} = 2WTN_0\left[1 + \frac{E}{WTN_0}\right] \tag{84}$$

$$\sigma_{S+N}^2 = \frac{(2WTN_0)^2\left[1 + \dfrac{2E}{WTN_0}\right]}{WT}. \tag{85}$$

The probability of error is

$$P_e = [P(v > 0 | x_0 \text{ sent})]\,[P(x_0 \text{ sent}] + [P(v < 0 | x_1 \text{ sent})]\,[P(x_1 \text{ sent})]. \tag{86}$$

If $P(x_0) = P(x_1)$ then, as shown in the previous section,

$$P_e \doteq \frac{1}{2}\left[1 - \operatorname{erf}\left(\sqrt{\frac{m_v^2}{2\sigma_v^2}}\right)\right], \quad WT \geq 10, \tag{87}$$

where

$$\frac{m_v^2}{2\sigma_v^2} = \frac{1}{2}\left\{\frac{\sqrt{WT}\,2WTN_0\left[\dfrac{E}{WTN_0}\right]}{(2WTN_0)\sqrt{2}\left[1 + \dfrac{E}{WTN_0}\right]^{1/2}}\right\}^2 \tag{88}$$

$$= \frac{1}{4} WT \frac{\left[\dfrac{E}{WTN_0}\right]^2}{\left[1 + \dfrac{E}{WTN_0}\right]}. \tag{89}$$

Curves of P_e versus E/WTN_0, the input SNR, are plotted in Figure 7 for various values of WT. Let us consider the asymptotic expression for P_e for both very small and very large input SNR.

a. Small SNR

For small SNR, from Equation (89),

$$\frac{m_v^2}{2\sigma_v^2} \doteq \frac{1}{4} WT \left[\frac{E}{WTN_0}\right]^2. \tag{90}$$

Thus

$$P_e \doteq \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{WT}{4}\left\{\frac{E}{WTN_0}\right\}^2\right)^{1/2}\right], \quad WT \geq 10. \tag{91}$$

Thus we see the square-law degradation in P_e for low SNR.

b. *Large SNR*

For large SNR, from Equation (89),

$$\frac{m_v^2}{2\sigma_v^2} \doteq \frac{1}{4}\frac{E}{N_0} \tag{92}$$

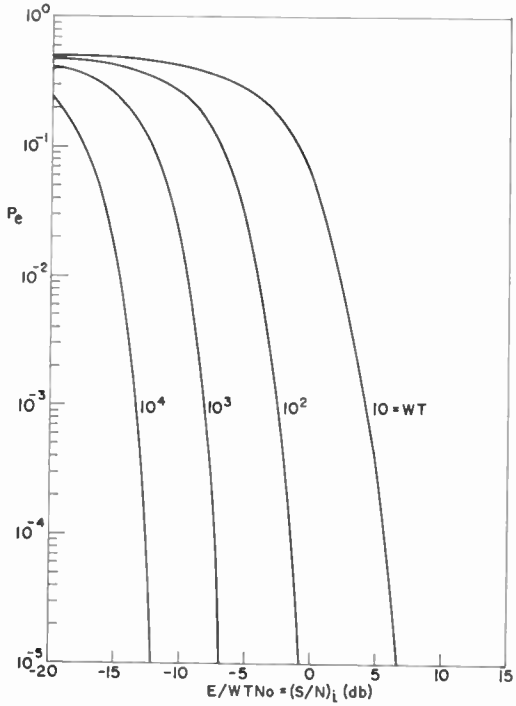

Fig. 7—Bit-error probability versus input SNR for wide-band FSK using square-law detection.

and

$$P_e \doteq \frac{1}{2}\left\{1 - \operatorname{erf}\left(\frac{E}{4N_0}\right)^{1/2}\right\}, \quad WT \geq 10. \tag{93}$$

Equation (93) shows that for high SNR the wideband FSK system using square-law detection depends only upon E/N_0 and is 3 db poorer than the optimum coherent system.

COMPARISON OF LINEAR AND SQUARE-LAW DETECTION

a) Small SNR

$$\frac{\left(\frac{m_{z_0}}{\sqrt{2}\,\sigma_z}\right)_{\text{linear}}}{\left(\frac{m_{z_0}}{\sqrt{2}\,\sigma_z}\right)_{\text{sq law}}} = \frac{\frac{1}{2}\sqrt{WT}\left[\frac{E}{2WTN_0}\right]}{\left(\frac{4}{\pi}-1\right)^{1/2}\frac{1}{2}\sqrt{WT}\left[\frac{E}{WTN_0}\right]} \tag{94}$$

$$= \frac{1/2}{\left(\frac{4}{\pi}-1\right)^{1/2}} = 0.96. \tag{95}$$

Thus, for the same P_e, the square-law detector requires $10\log_{10}[\pi/(4(4-\pi))]$ db, or 0.4 db, less SNR than the linear detector.

b) Large SNR

$$\frac{\left(\frac{m_{z_0}}{\sqrt{2}\,\sigma_z}\right)_{\text{linear}}}{\left(\frac{m_{z_0}}{\sqrt{2}\,\sigma_z}\right)_{\text{sq law}}} = \frac{\sqrt{\frac{E}{2N_0}}-\sqrt{\frac{\pi WT}{8}}}{\sqrt{\frac{E}{4N_0}}} = \alpha \tag{96}$$

$$\alpha = \sqrt{2}-\sqrt{\frac{\pi}{2}}\sqrt{\frac{WTN_0}{E}}. \tag{97}$$

For example,

SNR (db)	α (ratio)	α (db)
10	1.03	0.03
20	1.29	2.2
30	1.42	3.0

As can be seen, when the SNR $\gg$ 1, the linear detector requires 3 db less SNR than the square-law detector for the same P_e.

A comparison of the error probability performance of a wideband noncoherent FSK system using linear and using square-law detectors is shown in Figure 8. At SNR above 0 db, the linear detector is better than the square-law detector. This difference increases to a maximum of 3 db as the SNR increases. At SNR below —2.5 db the square law detector is about 0.5 db better than the linear detector. Figure 9 shows the variation of WT versus the channel input SNR for bit-error probabilities of 10^{-3} and 10^{-4}.

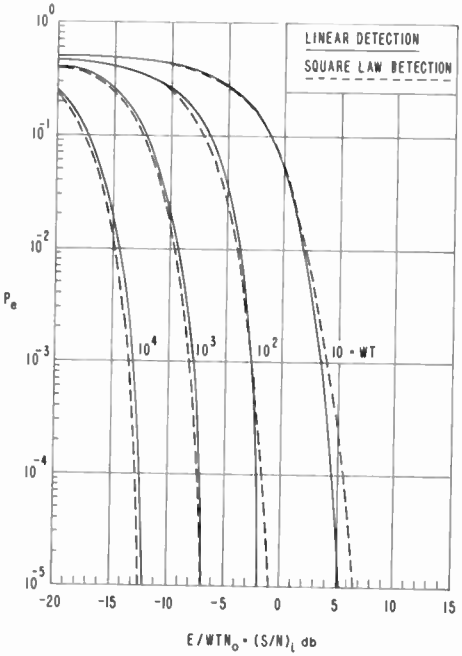

Fig. 8—Comparison of linear and square-law envelope detection for wide-band FSK systems (P_e versus SNR).

Wide-Band Systems Using Experimental Detector Characteristics and the SNR Approach

In the previous sections, consideration was given to the wide-band system utilizing linear envelope detectors and square-law detectors. These detectors were followed by matched filters. Expressions for bit-error probability as a function of the channel (mark and space) input signal-to-noise ratio were derived. Figure 8 shows the performance of these systems. Davenport and Root* showed that for *very*

* See Reference (4), Chapter 13, pp. 305-308.

small input SNR's the output SNR for a νth-law detector is proportional to the *square* of the input SNR for all values of ν. This relationship is the small-signal-suppression effect. They also showed that the output SNR for a νth-law detector is *directly* proportional to the input SNR for *large* values of the latter and for all values of ν. Their analysis was made for either a half-wave ν-law nonlinear device or a full-wave (even) νth-law device followed by a low-pass zonal filter.

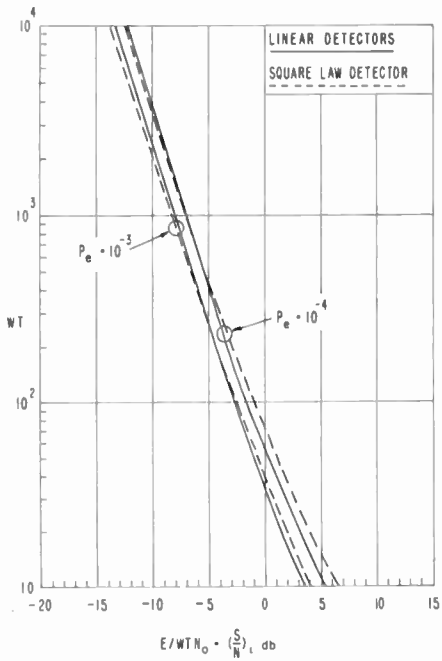

Fig. 9—Comparison of linear and square-law envelope detection for wide-band FSK systems (WT versus SNR).

The input to the νth-law detector is the sum of an amplitude-modulated sine-wave signal and stationary narrow-band Gaussian noise whose spectral density is concentrated about the carrier of the signal. Hence, all νth-law detectors behave in essentially the same way insofar as signal-to-noise-power ratios are concerned.

In the region between the bounds of very large and very small SNR the input–output SNR relationships of the νth-law detector are mathematically very complex. Therefore, an empirical approach is employed. Figure 10 shows the block diagram for the binary wide-band noncoherent FSK system. It is assumed that the ratio of the predetection filter bandwidth W to the post-detection bandwidth $1/T$ is greater than

10. T is the duration of the signal, which is the integration time of the matched filter (ideal integrator). The SNR at point b is determined from the SNR at point a, which is obtained from a link analysis. The SNR at point c is obtained from the envelope detector characteristics[10] shown in Figure 11. This curve is used to determine the SNR at the output of an AM receiver. The curve is valid for sine-wave modulation when the post-detection bandwidth (B_3) is substantially

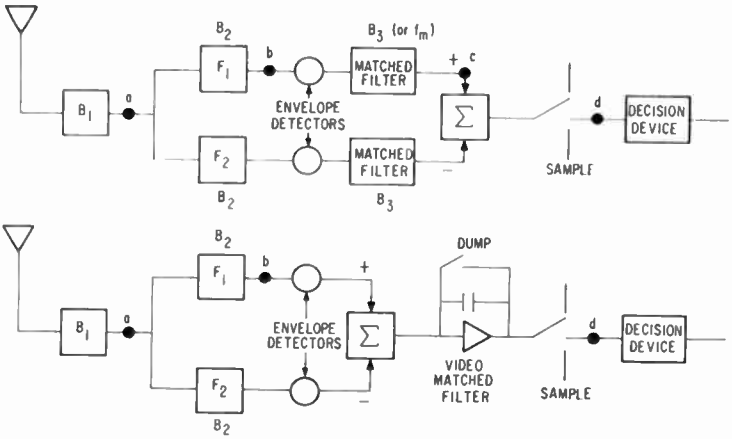

Fig. 10—Receivers for noncoherent FSK binary modulation.

less (at least 6 times) than the predetection bandwidth. The SNR is normalized to take into account the different pre- and post-detection bandwidths. Each ordinate represents the calculated value of SNR at the output in db minus the quantity

$$10 \log_{10} \left(\frac{\text{predetection bandwidth}}{2 \text{ post-detection bandwidth}} \right).$$

The data shown is for 100% modulation.

As previously discussed, use of the central limit theorem shows that the pdf for the channel with signal plus noise approaches a Gaussian distribution with a non-zero mean and the pdf for the channel with noise also approaches a Gaussian distribution. It is assumed that the mean at the output for the channel with noise alone is negligible compared to the mean at the output of the channel containing

[10] E. G. Fubini and D. C. Johnson, "Signal to Noise Ratio in AM Receivers," *Proc. IRE*, Vol. 36, p. 1461, Dec. 1948.

the signal plus noise. The pdf at the sampling point d in Figure 10 is also Gaussian distributed. However, the variance at this point is the sum of the variances from the two channels. Or, the SNR at point

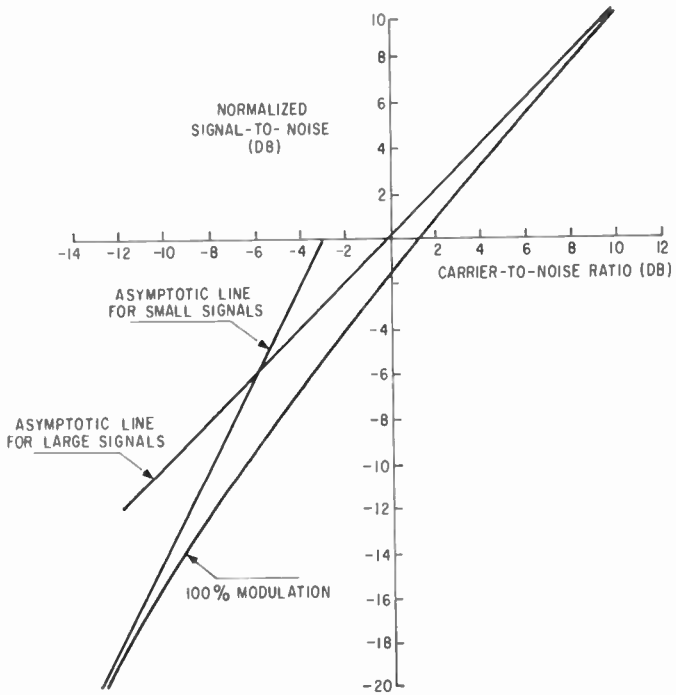

Fig. 11—Envelope detection characteristic.

d must include the contribution of the noise due to the channel that did not contain the signal. The relative increase in noise due to the channel that did not contain the signal* is

$$10 \log_{10} \left(\frac{\sigma_{S\times N}^2 + 2\sigma_{N\times N}^2}{\sigma_{S\times N}^2 + \sigma_{N\times N}^2} \right) \text{db}, \tag{98}$$

where

$$\frac{\sigma_{S\times N}^2}{\sigma_{N\times N}^2} = 2\left(\frac{S}{N}\right)_b, \tag{99}$$

* See Reference (4), p. 263.

$\sigma_{S\times N}^2$ is the variance of the noise due to the interaction of the signal and noise components in the nonlinear detector and $\sigma_{N\times N}^2$ is the variance of the noise due to the interaction of the noise components in the nonlinear detector.

Since the pdf of the noise approaches a Gaussian distribution, the bit-error probability is[9]

$$P_e = \frac{1}{2}\exp\left\{-\frac{E}{2N_0}\right\} = \frac{1}{2}\exp\left\{-\left(\frac{S}{N}\right)_d\right\}. \tag{100}$$

The following SNR exist at the various points in the detector. At point a,

$$\left(\frac{S}{N}\right)_a = \frac{S}{4N_0(B_3 + \Delta F)}, \tag{101}$$

where $B_3 = 1/T$ and ΔF is the frequency uncertainty.

At point b,

$$\left(\frac{S}{N}\right)_b = \frac{S}{2N_0(B_3 + \Delta F)} = \frac{E}{N_0 TW}, \tag{102}$$

where $E = ST$ and $W = 2(B_3 + \Delta F)$.

At point c,

$$\left(\frac{S}{N}\right)_c \text{db} = \left(\frac{S}{N}\right)_0 \text{db} + 10\log_{10}\left(1 + \frac{\Delta F}{B_3}\right), \tag{103}$$

where $(S/N)_0$ is the detector output SNR.

At point d,

$$\left(\frac{S}{N}\right)_d \text{db} = \left(\frac{S}{N}\right)_c \text{db} - 10\log_{10}\left(\frac{\sigma_{S\times M}^2 + 2\sigma_{N\times N}^2}{\sigma_{S\times M}^2 + \sigma_{N\times N}^2}\right) \tag{104}$$

where

$$2\left(\frac{S}{N}\right)_b = \frac{\sigma_{S\times N}^2}{\sigma_{N\times N}^2}. \tag{105}$$

for example, for $(S/N)_b \ll 1$, $\Delta N = 3$ db; for $(S/N)_b = 10$, $\Delta N = 0.4$ db.

Figure 12 shows the SNR at different points in the FSK detector shown in Figure 10. This calculation was done for a bit-error prob-

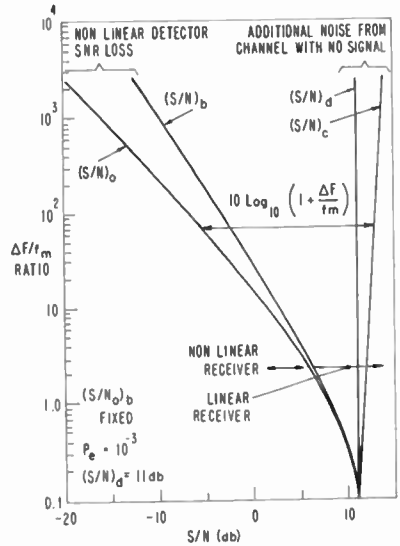

Fig. 12—SNR at different points in FSK detector for fixed input S/N_0 and $P_e = 10^{-3}$.

ability of 10^{-3}, where the required $(S/N)_d$ is 11 db. The effect of changing the input SNR versus $\Delta F/B_3$ shows that both the detector loss and the contribution of the noise from the channel that does not contain the signal decreases as $(S/N)_b$ increases. For $(S/N)_b > 5$ db, the detector is practically linear with respect to the SNR.

Figure 13 shows the variation of information bit rate R versus $\Delta F/R$ where the bandwidth required for R bits per second is B_3 cycles

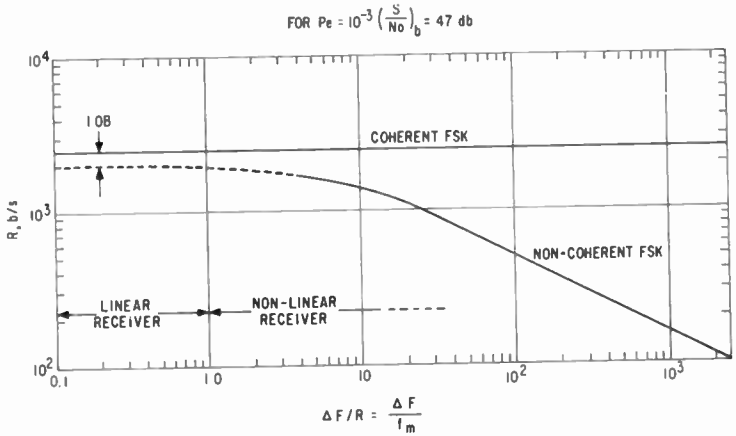

Fig. 13—Data rate versus frequency uncertainty for a fixed input S/N_0 and $P_e = 10^{-3}$.

per second. The curves show the relative performance of noncoherent and coherent FSK. The curves are analyzed for an information bit error probability of 10^{-3} and for a $(S/N_0)_b$ of 47 db. It is seen that for large ratios of $\Delta F/R$ (greater than 10), significant improvement in performance may be achieved by using coherent FSK. But for $\Delta F/R < 1$, only about 1 db improvement may be obtained. When the

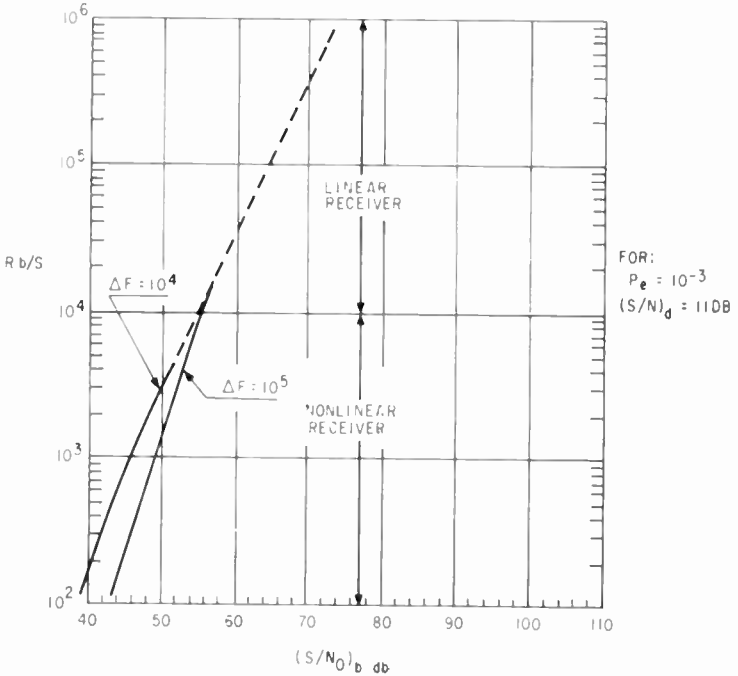

Fig. 14—Data rate versus input S/N_0 for $\Delta f = 10^4$ and 10^5 cps at $P_e = 10^{-3}$.

$(S/N)_b$ is above the threshold of the detector (greater than 5 db), the performance is insensitive to changes in frequency uncertainty ΔF or predetection bandwidth B_2 (or W). The solid curves represent the condition in which the analysis is valid. When $\Delta F/R < 0.1$ the predetection filter may be closely approximated by a matched filter. The dotted curve shows the approach to this limitation.

Figure 14 shows the information rate R in bits per second versus the signal-power-to-noise-power spectral density at the detector input. The curves are plotted for frequency uncertainties ΔF of 10^4 and 10^5 cycles per second. For predetection $S/N_0 > 55$ db, the receiver is linear (to SNR) and is therefore insensitive to changes in ΔF. The solid curves represent the valid conditions of the analysis whereas the

dotted curves show the approach to the condition when the predetection filter may be closely approximated by a matched filter.

Figure 15 shows the input channel SNR at point b versus the WT product for a bit error probability of 10^{-3}. As previously discussed the analysis is valid for $WT \geq 10$. An extension of the curve (dotted) shows the required SNR for predetection matched filters (or approxi-

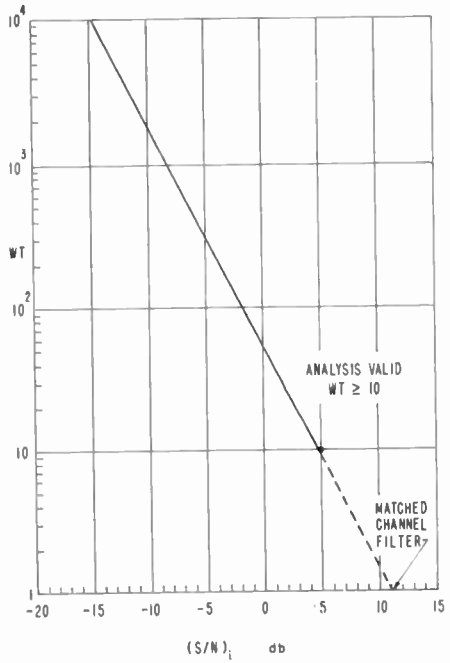

Fig. 15—Channel input SNR $(S/N)_b$ versus WT for $P_e = 10^{-3}$ wide-band noncoherent FSK.

mately for $WT = 1$) and linear detectors. This required SNR for a P_e of 10^{-3} agrees with the more sophisticated analysis[9] to within 0.5 db.

Conclusions (Part I)

Analysis was made for wide-band noncoherent binary FSK systems when the predetection bandwidth was primarily determined by the system frequency uncertainty. The analysis considered linear, square-law, and measured detector characteristics. Expressions for bit-error probability as a function of channel-input signal-to-noise ratio were derived for the linear and square-law detectors.

Figure 16 shows a comparison of the analysis for the linear and square-law detectors. The other curve shows the results using the actual measured detector characteristics and utilizing the SNR approach. All three analysis assumes $WT \geq 10$. The results are plotted for an error probability of 10^{-3}. It is seen that for channel input SNR (detector input SNR) less than 0 db, the analysis using the

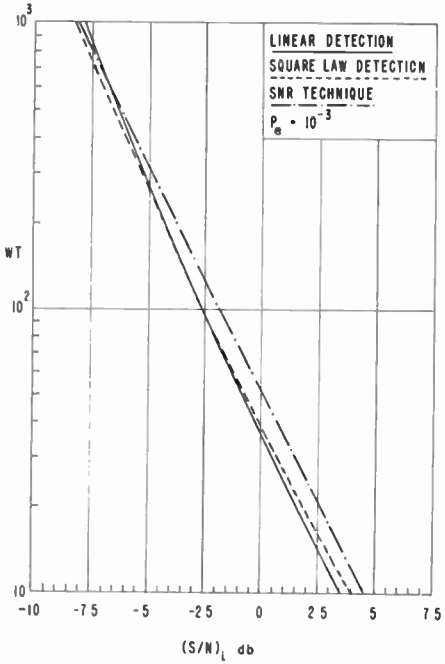

Fig. 16—Error probability performance of wide-band FSK systems using different envelope detectors.

actual detector characteristics and the SNR approach shows a performance characteristic that is between the performance of the linear and square-law detectors. Above 0 db input SNR, this technique showed slightly inferior performance to both the linear and square-law detectors. A comparison of the performance of the linear and square-law detectors showed that for SNR below —5 db, the square-law detector required about 0.4 db less SNR than the linear detector, whereas for SNR above 0 db the linear detector required less SNR than the square-law detector. This difference increases to a maximum of 3 db as the SNR increases. In all cases, the three types of detectors show performance characteristics that are within 1 db of each other.

Figure 17 shows a comparison of theoretical and preliminary experimental* results of bit-error probability versus input channel (predetection) SNR. The data was taken for linear detectors and for WT products of 10 and 100. It is seen that for a WT of 10 the difference between the experimental and theoretical results increases as

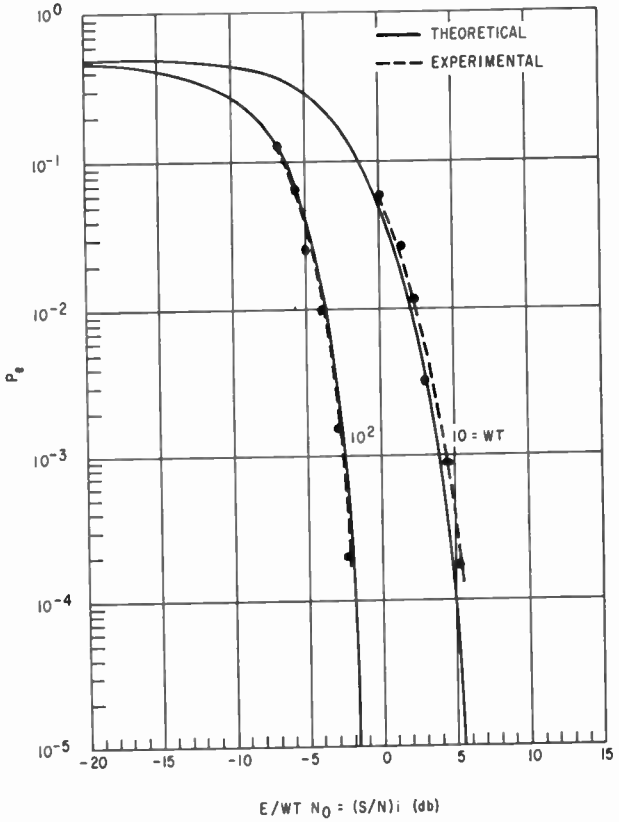

Fig. 17—Comparison of experimental and theoretical results for linear detectors.

the error probability decreases. However, the difference is only about 1/2 db down to error probabilities of 10^{-4}. As predicted in the theoretical discussion, for WT of 100 (and greater) the central limit theorem holds very closely (at least down to error probabilities of 10^{-4}).

* Obtained by H. R. Mathwick of RCA Astro-Electronics Division.

PART II — NONBINARY SYSTEMS

Introduction

In this part, we analyze *nonbinary* FSK systems with large ratios of predetection to post-detection matched filter bandwidths. It has been shown[11-13] that a 5-bit binary noncoherent FSK system requires 6 db more power than a noncoherent 32-bit binary FSK system at an error probability of 10^{-4} and *for matched predetection filters* ($WT = 1$). To communicate n bits of information, there must be 2^n degrees of freedom at the transmitter. Thus, one of 2^n words will be sent over a period of T seconds. Therefore, in an FSK system, each word is represented by a different frequency, which is equivalent to orthogonal coding, i.e., the transmitted waveforms are mutually orthogonal for all relative phase shifts.

The channel will add Gaussian type noise to this transmitted signal. The receiver will compute the conditional probability that one of a possible 2^n words was transmitted over an interval of T seconds, given the received word. It has been shown that if the additive disturbance is white Gaussian noise, the probability computer will consist of 2^n matched-filter detectors. The outputs of these matched filters will go to a decision device that examines all the outputs. If the 2^n words are, *a priori*, all equally likely to be transmitted with the same energy, then the decision device will select the waveform that corresponds to the maximum matched-filter output. This is known as maximum likelihood detection, which can be shown to minimize the error probability.

The transmitted waveforms should be as unlike as possible, so that in a noisy channel there will be the least possible chance of making the wrong selection of the transmitted word. For noncoherent systems, orthogonal waveforms are optimum.

Error Probability Analysis

Figure 18 shows a block diagram of a nonbinary noncoherent FSK system. The input signal is

$$S(t) = S_i(t) + n(t), \quad i = 1, 2, \cdots n, \tag{106}$$

[11] V. A. Kotel'inkov, *Theory of Optimum Noise Immunity*, McGraw-Hill Book Co., 1959.

[13] S. Reiger, "Error Rates in Data Transmission," *Proc. IRE*, Vol. 46, pp. 919 and 920, May 1958.

where $S_i(t)$ is one of a possible 2^n selection of signals transmitted for a period of T seconds:

$$S_i(t) = A \cos(\omega_{c_i} t + \theta_i),$$
$$i = 1, 2, \cdots, 2^n$$
$$0 \leq t \leq T$$

$n(t)$ is white Gaussian distributed noise.

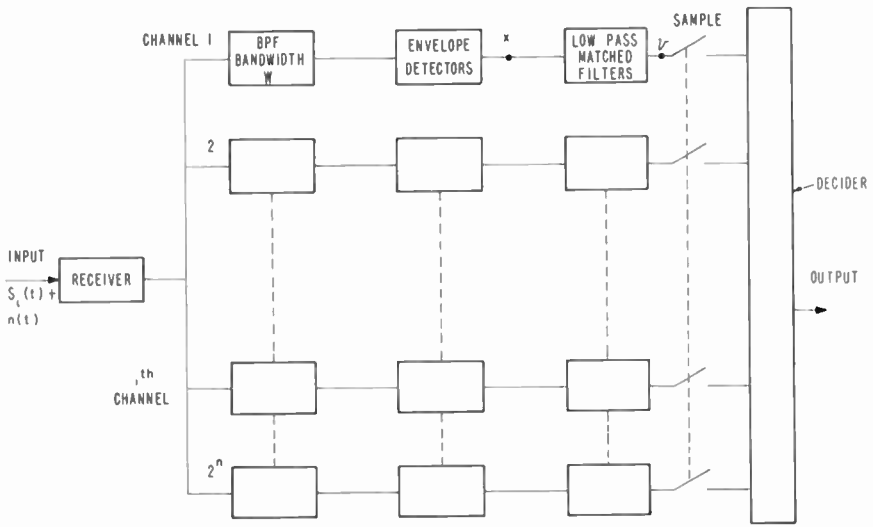

Fig. 18—Noncoherent wide-band nonbinary FSK detector.

Each of the 2^n channels consists of identical predetection filters of bandwidth W and envelope detectors followed by video matched filters. The matched filter may consist of an integrate and dump type device. At the end of the signal duration T, the output of the matched filter is sampled. The decider will then inspect each of the outputs of the 2^n matched filters. The decision device will then select the signal coming from the channel with the largest output.

Figure 19 shows a single channel for the wide-band multiple noncoherent FSK detector. The energy in the ith channel is

$$E_i = \int_0^T |S_i^2(t)|\,dt, \quad i = 1, 2, \cdots 2^n$$

$$E_i = \frac{A_i^2 T}{2}.$$

For a set of 2^n code words, the probability that the correct one will be chosen is

$$P_c(n) = \int_{-\infty}^{\infty} P(v_k)\,dv_k P(y_1, y_2, \cdots y_j, \cdots y_{2^n-1} < v_k). \tag{107}$$

Since the channel noise outputs are mutually independent,

$$P_c(n) = \int_{-\infty}^{\infty} P(v_k)\,dv_k \prod_{j=1}^{2^n-1} P(y_j < v_k), \quad j \neq k, \tag{108}$$

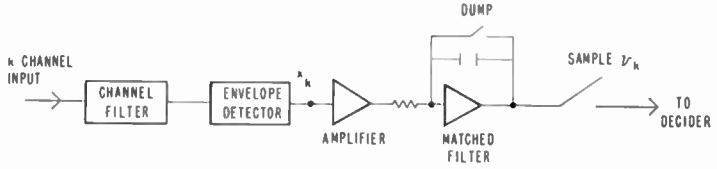

Fig. 19—Single-channel wide-band FSK system.

where $p(v_k)$ is the probability density of the output of the correct channel.

The probability that the output of the j^{th} incorrect channel is less than the output of the correct channel is

$$P(y_j < v_k) = \int_{-\infty}^{v_k} P(y_j)\,dy_j, \quad j \neq k. \tag{109}$$

In terms of the previous notation,

$$P(v|N < v_k) = P(y_j < v_k) = \int_{-\infty}^{v} P(v|N)\,dv \tag{110}$$

and

$$P(v_k) = P(v|S+N). \tag{111}$$

Figure 20 shows the conditional probability density distributions $p(v|S+N)$ and $p(v|N)$. Substituting Equations (37) and (38) into

Equation (108) gives

$$P_c \doteq \int_{-\infty}^{\infty} \frac{\exp\left\{-\dfrac{(v-m_{S+N})^2}{2\sigma_{S+N}^2}\right\}}{\sqrt{2\pi}\,\sigma_{S+N}}\,dv \cdot \left[\int_{-\infty}^{v} \frac{\exp\left\{-\dfrac{(y-m_N)^2}{2\sigma_N^2}\right\}}{\sqrt{2\pi}\,\sigma_N}\,dy\right]^{2^n-1}, \quad WT \geq 10. \quad (112)$$

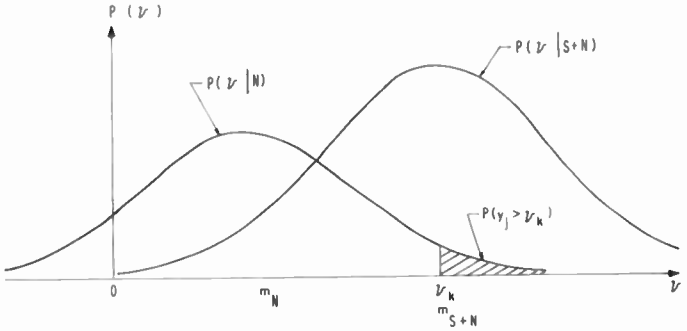

Fig. 20—Probability densities $P(v|N)$ and $P(v|S+N)$.

The probability that a word is in error is

$$P_W(n) = 1 - P_c(n). \quad (113)$$

The receiver will select the correct signal if none of the noise samples exceeds v. If the following substitutions are made

$$u = (y - m_N)/\sigma_N$$

$$z = (v - m_{S+N})/\sigma_{S+N},$$

Equation (112) becomes

$$P_W(n) \doteq 1 - \int_{-\infty}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz \cdot \left[\int_{-\infty}^{L_k} \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,du\right]^{2^n-1}, \quad WT \geq 10 \quad (114)$$

where $L_1 = z(\sigma_{S+N}/\sigma_N) + (m_{S+N} - m_N)/\sigma_N$. For the square-law detector,

$$L_1 = z\left(1 + \frac{2E}{WTN_0}\right)^{1/2} + \frac{\sqrt{WT}\,E}{WTN_0}. \tag{115}$$

Since this integral cannot, in general, be evaluated analytically, numercial integration must be used (see Appendix I).

An examination of Figure 20 and Equation (112) shows that as $m_{S+N} - m_N$ increases, the probability of selecting the correct word $P_c(n)$ increases or $P(y_j < v_k)$ decreases. This difference in the means, as seen by Equations (82) and (84) for a square-law detector, is increased by increasing the input channel signal-to-noise ratio $E/(N_0WT)$. This is further shown by Equations (114) and (115). It is seen from Equations (82)-(85) that increasing $E/(N_0WT)$ will increase σ_{S+N} and m_{S+N} while decreasing the noise power WN_0 will decrease both σ_N and m_N. Thus, any of these combinations will increase the upper limit of Equation (114), which will in turn decrease the word error probability $P_W(n)$.

The integral shown in Equation (114) was solved by numerical computer integration. The results are shown in the next two figures for words containing up to 20 bits of information. Figure 21 shows a plot of the word error probability $P_W(n)$ versus the channel signal-to-noise ratio $E/(N_0WT)$ using square-law detectors.* The factor n indicates that the word contains n bits of information. The data is also plotted for a $WT = 100$.** Figure 23 shows the variation of $P_W(n)$ with WT for $n = 5$. For the binary case, $n = 1$. To appreciate the data shown in Figure 21, one must realize that a comparison of systems should be based on the same information capacity.

For example, at a $P_W(n)$ of 10^{-3}, the $E/(N_0WT)$ for $n = 1$ (or one bit of information) is -2.7 db; while for $n = 5$ (or 5 bits of

* The trend of the data is the same for other types of detectors such as linear envelope detectors.

** The relative results will be similar for other values of WT, as discussed in Part I.

information) the required SNR for the same $P_W(n)$ is —1.5 db. For an increase of 7 db bits of information, the SNR must only be increased by 1.2 db.

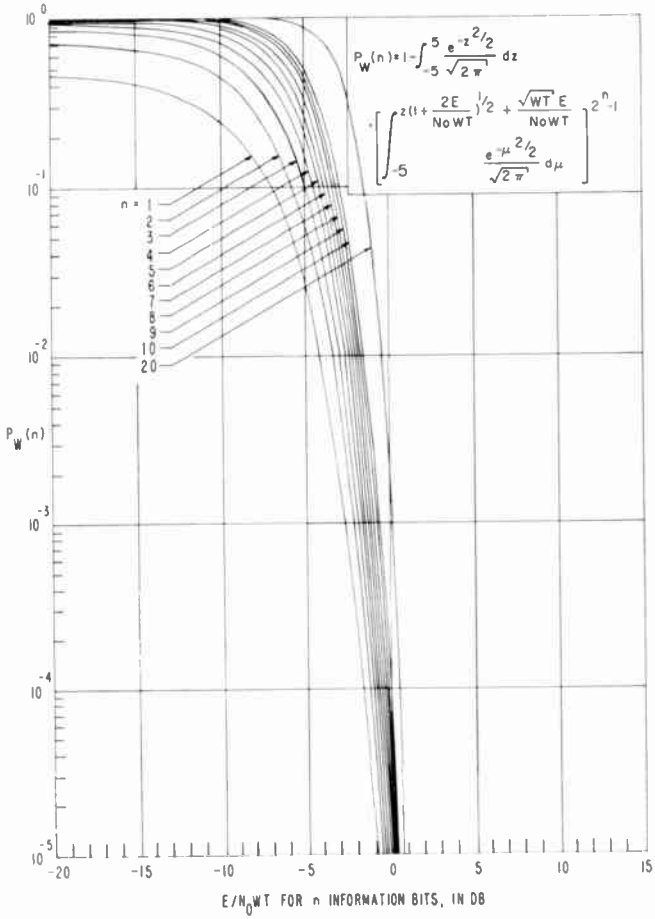

Fig. 21—Word error probability $P_W(n)$ versus channel SNR for n information bit FSK systems using square-law detectors.

In Figure 22 the same data is presented as in Figure 21, but the required $E/(N_0WT)$ is per unit of bit information. Thus, for $n = 1$ and $P_W(n) = 10^{-3}$, the required SNR is —2.7 db; whereas for $n = 5$ and $P_W(n) = 10^{-3}$, the required SNR is —8.5 db. On the basis of the SNR required per unit of bit information and a $P_W(n)$ of 10^{-3}, the system for $n = 5$ requires 5.8 db less SNR than the system for $n = 1$. This checks the results shown in Figure 21.

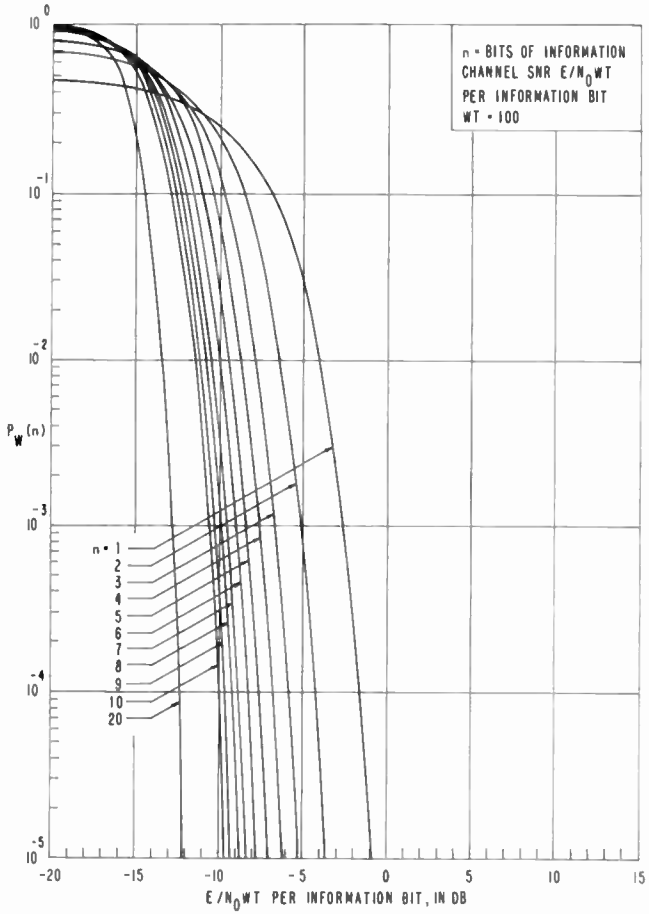

Fig. 22—Word error probability $P_W(n)$ versus channel SNR per information bit for FSK systems using square-law detectors.

Conclusions (Part II)

An analysis of word error probability has been made for nonbinary noncoherent FSK systems operating under conditions of large frequency uncertainties. Under these conditions the bandwidth of the predetection channel filters are primarily determined by the frequency uncertainties rather than the information rates. Thus, it was necessary to use video matched filters. A limit on the analysis is that the ratio of the predetection to post-detection filter bandwidths is 10. The results of the nonbinary word-error probability versus the input-

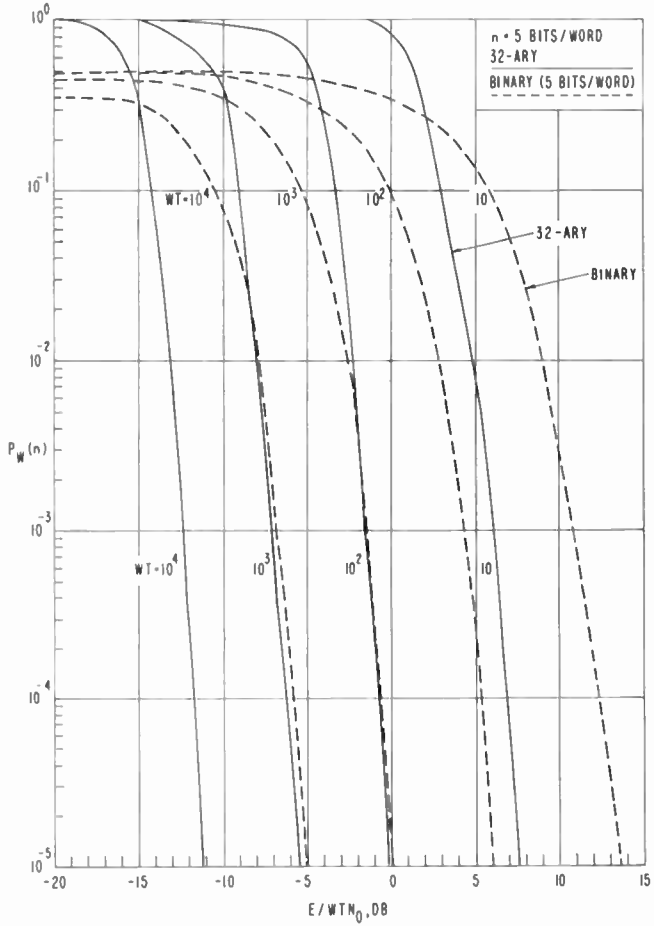

Fig. 23—5-Bit word error probability versus input SNR for wide-band FSK using square-law detection.

channel signal-to-noise ratio showed the same trend as the analysis made for predetection matched filters.*

Comparison of 32-ary and Binary Systems

Figure 23 shows a comparison for a 5-bit information word using 32-ary and binary systems for an FSK system using square-law detectors. The five-bit word for the binary system has five times, or 7 db, more energy than each binary information bit. Since $WT \geq 10$, the channel bandwidth W for all values of n are practically equal. The

* Compare Figures 4 and 5 with Figure 3 in Reference (12).

General Equation $$y = \int_a^b f(x)\,dx \tag{116}$$

$$I = \frac{h}{3}\,(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots 4y_{n-1} + f_n \tag{117}$$

where $h = (b-a)/n$ and n is the number of intervals. Let

A_i = value of inside integral for a given value of z $(i = 0 - 20)$

B = value of outside integral.

$$A_i = \frac{h_a}{2\sqrt{2\pi}}\,(f(-5) + 4f(-5+h_a) + 2f(-5+2h_a) + \cdots + 4f(-5+19h_a) + f(-5+20h_a)) \tag{118}$$

$$h_a = \frac{(f(z)+5)}{20} \tag{119}$$

$$f(z) = z\left(1 + 2\frac{E}{N_0WT}\right)^{1/2} + \sqrt{WT}\,\frac{E}{N_0WT}. \tag{120}$$

Integrating over the entire range of $f(z)$ results in either an over-estimation or under-estimation of the value of A_i, depending on the value of $E/(N_0WT)$. Therefore, when the upper limit of the inside integral $f(z)$ exceeded 5, its value was assumed to be unity.

$n = 1$

$$B = \frac{h_b}{3\sqrt{2\pi}}\,(A_0 f(-5) + 4A_1 f(-5+h_b) + 2A_2 f(-5+2h_b) + \cdots + 4A_{19} f(-5+19h_b) + A_{20} f(5)) \tag{121}$$

$n = 2$

$$B = \frac{h_b}{3\sqrt{2\pi}}\,(A_0{}^3 f(-5) + 4A_1{}^3 f(-5+h_b) + 2A_2{}^3 f(-5+2h_b) + \cdots + 4A_{19}{}^3 f(-5+19h_b) + A_{20}{}^3 f(5)) \tag{122}$$

$$h_b = \frac{5-(-5)}{20} = 0.5. \tag{123}$$

RCA Technical Papers

First Quarter, 1966

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Acoustical-Optical Phonon Scattering in Ge, Si, and III-V Compounds," E. F. Steigmeier and I. Kudman, *Phys. Rev.*, No. 2 (January) 1966

"Asynchronously Multiplexed Binary Channel Capacity," R. C. Sommer, *Proc. IEEE* (Proc. Letters) (January) 1966

"Asynchronously Multiplexed Channel Capacity," R. C. Sommer, *Proc. IEEE* (Proc. Letters) (January) 1966

"A Current-Controlled Negative-Resistance Effect in Indium Antimonide," L. F. Eastman, *Trans. IEEE PTGED* (January). 1966

"Current Oscillations and Microwave Emission in Indium Antimonide," R. D. Larrabee and W. A. Hicinbothem, Jr., *Trans. IEEE PTGED* (January) 1966

"General Three-Resonator Filters in Waveguide," R. M. Kurzrok, *Trans. IEEE PTGMTT* (Correspondence) (January) 1966

"Ion Beam Measurements of Cesium Recombination Cross Sections," J. M. Hammer and B. B. Aubrey, *Phys. Rev.*, No. 1 (January) 1966

"Laser Action in Sulfur Using Hydrogen Sulfide," R. U. Martinelli and H. J. Gerritsen, *Jour. Appl. Phys.* (Communications) (January) 1966

"Learning Systems for Automatic Control," J. Sklansky, *Trans. IEEE PTGAC* (January) 1966

"Multiplication Noise in Uniform Avalanche Diodes," R. J. McIntyre, *Trans. IEEE PTGED* (January) 1966

"A New Modification of the Semiconducting Compound K_3Sb," A. H. Sommer and W. H. McCarroll, *Jour. Appl. Phys.* (January). 1966

"Paramagnetic Resonance of Divalent Holmium in the Cubic Hosts: CaF_2, SrF_2, BaF_2, and $SrCl_2$," E. S. Sabisky, *Phys. Rev.*, No. 1 (January) 1966

"Phenomena Influencing the Temperature Behavior of Stimulated Emission in GaAs p-n Junctions," M. F. Lamorte, T. Gonda, and H. Junker, *IEEE Jour. Quantum Electronics* (January) 1966

"A Plasma Instability Induced by Electron-Hole Generation in Impact Ionization." M. Toda, *Jour. Appl. Phys.* (January) 1966

"The Spectrum of Microwave Radiation from InSb," K. Suzuki, *Trans. IEEE PTGED* (January) 1966

"Streaming Instabilities in Solids and the Role of Collisions," B. Vural and S. Bloom, *Trans. IEEE PTGED* (January)..... 1966

"Surface and Interface Waves in Plasma Gaps," C. Davis and Coauthor, *Jour. Appl. Phys.* (Communications) (January) 1966

"Theory of a Microwave Plasma Instability Due to Transverse Breakdown," M. Toda, *Jour. Appl. Phys.* (January) 1966

Comment on "Time-Saving Method for Partial Fraction Expansion of Function with One Pair of Conjugate Complex Roots," A. A. Clark, *Proc. IEEE* (Proc. Letters) (January) 1966

"Continuous Operation is Near for Uncooled Diode Lasers," M. F. Lamorte, *Electronics* (January 10) 1966

"Tunnel-Diode Oscillator Expands F-M System's Channel Capacity," F. H. Lefrak, *Electronics* (January 10) 1966

"Possible Ferroelectric Effects in Liquid Crystals and Related Liquids," R. Williams and G. Heilmeier, *Jour. Chem. Phys.* (15 January) 1966

"Transient Behavior of Domains in Liquid Crystals," G. H. Heilmeier, *Jour. Chem. Phys.* (15 January) 1966

"Vibrational Levels of the ^2D-E State of Ce^{3+} in CaF_2," C. W. Struck and F. Herzfeld, *Jour. Chem. Phys.* (15 January) 1966

"Direct Evidence of Steady EMF Induced by Flux Motion in Superconductors," J. Pearl, *Phys. Rev. Letters* (17 January) 1966

"Electro-Optic Effect in Reverse-Biased GaAs p-n Junctions," W. L. Walters, *Jour. Appl. Phys.* (Communications) (February).. 1966

"The Influence of Oxidation Rate and Heat Treatment on the Si Surface Potential in the Si-SiO_2 System," A. G. Revesz and K. H. Zaininger, *Trans. IEEE PTGED* (February) 1966

"An Investigation of Instability and Charge Motion in Metal-Silicon Oxide-Silicon Structures," S. R. Hofstein, *Trans. IEEE PTGED* (February) 1966

"Optimum Realizable Transmitter Waveforms for High-Speed Data Transmission," F. Amoroso, *Trans. IEEE PTGCOM* (February) 1966

"Paramagnetic Resonance of Ce^{3+} in Yttrium Aluminum Garnet," H. R. Lewis, *Jour. Appl. Phys.* (February) 1966

"Partial Switching Behavior in Ferroelectric Triglycine Sulfate," G. W. Taylor, *Jour. Appl. Phys.* (February) 1966

"Radar Separation of Closely Spaced Targets," A. Golden, *IEEE Spectrum* (February) 1966

"Surface Potential Measurements of Lithium Drifted Germanium Diodes," D. E. Davies and P. P. Webb, *Trans. IEEE PTGNS* (February) 1966

"Theoretical Efficiency of Pyroelectric Power Converters," E. Fatuzzo, H. Kiess, and R. Nitsche, *Jour. Appl. Phys.* (February) 1966

"Fundamental Optical Absorption in β-Silver Telluride," R. Dalven, *Phys. Rev. Letters* (21 February) 1966

"The Acoustoelectric Effects and the Energy Losses by Hot Electrons —Part I," A. Rose, *RCA Review* (March) 1966

"Adsorption-Type Reservoir for Gas Tubes," K. G. Hernqvist and J. D. Levine, *RCA Review* (March) 1966

"Applications of MOS FET's in Microelectronics," R. D. Lohman, *SCP and Solid State Technology* (March) 1966

"Camera Tubes for Recording Stratoscope II Telescope Images," A. D. Cope and E. Luedicke, *RCA Review* (March) 1966

"Controlled Coverage by Co-Channel Transmission," W. C. Harris and K. W. Angel, *Trans. IEEE PTGVC* (March) 1966

"Extremely-Low-Capacitance Silicon Film MOS Transistors," E. C. Ross and C. W. Mueller, *Trans. IEEE PTGEC* (Correspondence) (March) 1966

"Helping Engineers," C. W. Fields, *Electronic Industries* (Letters) (March) 1966

"How Long-Term Storage Affects Reliability," R. F. Ficchi, *Electronic Industries* (March) 1966

"The Improved Gray Scale and the Coarse-Fine PCM Systems, Two new Digital TV Bandwidth Reduction Techniques," W. T. Bisignani, G. P. Richards, and J. W. Whelan, *Proc. IEEE* (March) 1966

"Low-Power Long-Range Digital Communications System," J. G. Arnold, D. M. Chauvin, J. C. Johnson, and J. K. Oliver, Jr., *RCA Review* (March) 1966

"Microwave Properties of Planar Hexagonal Ferrites," R. A. Braden, I. Gordon, and R. L. Harvey, *Trans. IEEE PTGMAG* (March) 1966

"Monolithic Ferrite Memories," I. Abeyta, M. M. Kaufman, and P. Lawrence, *RCA Review* (March) 1966

"A 180-Stage Integrated Thin-Film Scan Generator," P. K. Weimer, G. Sadasiv, L. Meray-Horvath, and W. S. Homa, *Proc. IEEE* (March) 1966

"An RF Activated Antenna Switch," K. W. Angel, *Trans. IEEE PTGVC* (March) 1966

"SMPTE Color Television Subjective Reference Test and Slides," E. P. Ancona, Jr., *Jour. SMPTE* (March) 1966

"Solar-Pumped Modulated Laser," C. W. Reno, *RCA Review* (March) 1966

"The TIROS IX Wheel Satellite," A. Schnapf, *RCA Review* (March) 1966

"Vidicon Performance Characteristics at Slow Scan Rates," R. E. Johnson, *RCA Review* (March) 1966

"Electrical Transport Properties of the Insulating Ferromagnetic Spinels $CdCr_2S_4$ and $CdCr_2Se_4$," H. W. Lehmann and M. Robbins, *Jour. Appl. Phys.* (1 March) 1966

"Interaction of Spin Waves with Drifted Carriers in Solids," B. Vural, *Jour. Appl. Phys.* (1 March) 1966

"Static and Dynamic Stress Effects in Cylindrical Ferromagnetic Films," H. Weinstein, *Jour. Appl. Phys.* (1 March) 1966

"Wideband Direct Recording with High-Frequency Bias," W. B. Hendershot, III, *Jour. Appl. Phys.* (1 March) 1966

"Selective Absorption of Circularly Polarized Light in Broad Bands by the Zeeman Components of Divalent Thulium in Calcium Fluoride," C. H. Anderson, H. A. Weakliem, and E. S. Sabisky, *Phys. Rev.* (4 March) 1966

"Combined Effects of Sound and Electric Fields upon the Conduction Electrons in Solids," R. Klein and W. Rehwald, *Phys. Rev.* (11 March) 1966

"Energy Gap in B-Ag_2Te," R. Dalven and R. Gill, *Phys. Rev.* (11 March) 1966

"Photo-Hall Effects in Anthracene," J. Dresner, *Phys. Rev.* (11 March) 1966

"Microwave Ultrasonic Attenuation in Semi-Insulating GaAs," K. R. Keller and B. Abeles, *Jour. Appl. Phys.* (Communications) (15 March) 1966

"Photoemission of Electrons from Silicon into Silicon Dioxide. Effects of Ion Migration in the Oxide," R. Williams, *Jour. Appl. Phys.* (15 March) 1966

"It's a Television First — Receivers with Integrated Circuits," J. Avins, *Electronics* (March 21) 1966

"Overlay Transistors Move into Microwave Region," H. C. Lee and G. J. Gilbert, *Electronics* (March 21) 1966

"Advancements in the Facsimile Art During 1965," W. H. Bliss, *IEEE International Convention Record, Part 1, Wire and Data Communication* 1966

"HF Applications to Command Capsule Reentry Surveillance and Tracking," M. H. Lowe, *IEEE International Convention Record, Part 4, Aerospace and Electronic Systems* 1966

"A New Technique for Application of Magnetic Tape to Digital Communications," R. K. Andres and Coauthor, *IEEE International Convention Record, Part 1, Wire and Data Communication* 1966

AUTHORS

A. ACAMPORA received the BEE degree in 1957 and the MEE in 1959 from the Polytechnic Institute of Brooklyn. From 1957 to 1958, he worked in the Microwave Research Institute on a research fellowship. He is currently working toward a doctorate at Brooklyn Polytechnic. Mr. Acampora joined the RCA Communications Systems Division in 1959. He has worked on various aspects of communications systems including BMEWS, PCM-FM troposcatter systems, and FM data communications systems. He is currently engaged in programs involving the detection of FM data signals using advanced demodulators and in the development of a 600-channel FDM FM system. Mr. Acampora is a member of Sigma Xi, Tau Beta Pi, and Eta Kappa Nu.

F. ASSADOURIAN received the BS, MS, and PhD in Mathematics from New York University in 1935, 1936, and 1940, respectively. From 1937 to 1942, Dr. Assadourian was an instructor of mathematics at New York University, and from 1942 to 1944 an Associate Professor of Mathematics at Texas Technological College, Lubbock, Texas. From 1944 to 1946, he was with Westinghouse Research Laboratories and from 1946 to 1956 with the Federal Telecommunication Laboratories. He joined the RCA Communications Systems Division in 1956, where he has worked on communications system analysis and design, modulation, intermodulation distortion analysis in FDM-FM, and microwave and troposcatter systems analyses. Most recently, he has been concerned with satellite communications systems analyses for both active and passive satellites.

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.

D. S. JACOBY received the AB and BSEE degrees from Columbia University in 1940 and 1941, respectively, and the MSEE degree from Rutgers University in 1950. From 1942 to 1963, he was with the U.S. Army Electronics Research and Development Laboratories, Ft. Monmouth, N.J., working on pulse communications, troposcatter and time-division-multiplex development programs, and communications and meteorological satellites. Mr. Jacoby joined the RCA Communications Systems Division in 1963. He has worked on the Army RADAS program, satellite communications systems, and advanced concepts of mobile and personal communications.

J. KLAPPER received the BEE degree from the City College of New York in 1956, the MS (EE) from Columbia University in 1958 and the Eng. Sc. D. degree from New York University in 1965. From 1952 to 1956 Dr. Klapper was an Electronics Engineer with CBS-Columbia working on military communications. From 1956 to 1959 he was a Lecturer in Electrical Engineering at the City College of New York, teaching courses in electronics and analysis. From 1959 to 1960 he was a Project Engineer with the Federal Scientific Corporation on spectrum analysis and related problems. Concurrently, from 1957 to 1960 he was a Staff Engineer and Consultant to the Electronics Research Laboratories of Columbia University, doing research on filtering, speech analysis and synthesis. Dr. Klapper joined the RCA Communications Systems Division in 1960, working on applied research and advanced development of communications techniques and systems study. At present he is concerned with FM and phase feedback techniques and their application to FM demodulation with reduced thresholds.

Dr. Klapper holds the Professional Engineer's license in New York State (1961) and is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi and a Senior Member of the IEEE.

H. C. LEE received the BSEE from Massachusetts Institute of Technology in 1952, and the M.S. degree in Electrical Engineering from Stanford University in 1953. Since joining RCA in 1958, he has had extensive experience in circuit design for UHF power and small-signal transistors and in microwave power generation using varactor diodes. He is responsible for the design of direct frequency-multiplier circuits using overlay transistors, and for the development of amplifier and frequency-multiplier circuits at S-band using overlay transistors.

JOAN LURIE received the B.S. degree in Physics summa cum laude from Brooklyn College in January, 1961 and the M.S. in Physics from Rutgers University in June, 1962. She is currently enrolled in the Ph.D. program at Rutgers, working in the field of theoretical solid-state physics. Her graduate studies have been supported in part by an N.S.F. Cooperative Graduate Fellowship. Since joining the RCA Astro-Electronics Division in June, 1962, Mrs. Lurie has been with the Physical Research Group. She has carried out analytical investigations of high-resolution electron guns for use in television camera tubes and other imaging devices. She has also worked on theoretical analyses of thermal detectors and contributed to studies of photographic image processing. Mrs. Lurie is a member of Phi Beta Kappa and Pi Mu Epsilon and an associate member of Sigma Xi.

MARTIN C. STEELE received the B.Ch.E. (Cum Laude) from Cooper Union Institute of Technology in 1940, and the M.S. and Ph.D. degrees in Physics from the University of Maryland in 1949 and 1952, respectively. He worked as an engineer for the Office of the Chief of Engineers, before the war, and as a research physicist in the Naval Research Laboratory from 1947 to 1955, where he was head of the Cryomagnetic Research Group in the Solid State Division. Since 1955 he has been with RCA, as head of the Solid State Electronics Research Group and for three years (1960-63) as Research Director of RCA's research laboratory in Tokyo, Japan. In August, 1963 he became head of the Plasma Physics Group of the Microwave Research Laboratory at RCA Laboratories, Princeton, New Jersey. His research has covered a wide range of areas in solid-state physics, including superconductivity, magnetic susceptibility, thermometry, theory of magnetic properties of metals, thermal, electric and magnetic properties of semiconductors, and the properties of plasmas in solids.

Dr. Steele is a fellow of the American Physical Society, a member of Sigma Xi and a former member of the Washington Academy of Science.