niobium films have produced units with low breakdown strength. Further work is being done with these other film-forming metals as well as with tantalum.

Conclusions

A new tantalum capacitor, essentially two-dimensional in structure, has been made and has properties superior to other types of tantalum capacitors.

Capacitances obtained are comparable to the capacitance-area relationships for tantalum electrolytic capacitors formed to the same voltages. Using counter electrodes of 95 mils to 250 mils in diameter and the single-layered structure, capacitors have been prepared with capacitances ranging between about 2000 $\mu\mu$f and 0.25 μf.

DC leakages measured at three-quarters the formation voltages are of the order of 4×10^{-11} a for 250-mil diameter electrodes. Expressed in terms of insulation resistance, this amounts to about 60,000 ohm farads.

Dissipation factors are in the neighborhood of 0.008 at 100 c and increase to between 0.1 and 0.8 at 100 kc. The relatively high losses at the higher frequencies are caused by the high series resistance of the tantalum films. Thicker films of tantalum should reduce these losses.

Room temperature breakdown voltages approximate the formation voltages for these units, and successful models have been formed to 5, 10, 20, 30, 40, 50, 100, 150, and 200 v. A suggested working voltage is one half the formation voltage for temperatures up to 65°C. Voltage derating characteristics for elevated temperature operation have not been determined as yet. The units operate well at very low temperatures, however, even down to -196°C.

This type of capacitor should find many applications in the lower capacitance areas, and seems ideally suited for printed circuit applications.

Acknowledgment

The authors are indebted to D. A. McLean and N. Schwartz for many helpful discussions, and to H. Basseches for aid with the sputtering technique.

Linear Diversity Combining Techniques*

D. G. BRENNAN†, SENIOR MEMBER, IRE

***Summary*—This paper provides analyses of three types of diversity combining systems in practical use. These are: selection diversity, maximal-ratio diversity, and equal-gain diversity systems. Quantitative measures of the relative performance (under realistic conditions) of the three systems are provided. The effects of various departures from ideal conditions, such as non-Rayleigh fading and partially coherent signal or noise voltages, are considered. Some discussion is also included of the relative merits of predetection and postdetection combining and of the problems in determining and using long-term distributions. The principal results are given in graphs and tables, useful in system design. It is seen that the simplest possible combiner, the equal-gain system, will generally yield performance essentially equivalent to the maximum obtainable from any quasi-linear system. The principal application of the results is to diversity communication systems and the discussion is set in that context, but many of the results are also applicable to certain radar and navigation systems.**

* Original manuscript received by the IRE, April 21, 1958; revised manuscript received, January 14, 1959. The research reported in this paper was partly supported by the Army, Navy and Air Force under contract with the Massachusetts Institute of Technology.

† Lincoln Lab., Lexington, Mass., and Dept. of Mathematics, M.I.T., Cambridge, Mass.

I. Introduction

When a steady-state, single-frequency radio wave is transmitted over a long path, the envelope amplitude of the received signal is observed to fluctuate in time. This phenomenon is known as fading, and its existence constitutes one of the boundary conditions of radio system design. It is observed that if two or more radio channels are sufficiently separated in space, frequency, or time, and sometimes in polarization, then the fading on the various channels is more or less independent; *i.e.*, it is then relatively rare for all the channels to fade together. The standard techniques for reducing the effect of fading—known as diversity techniques—make use of this fact. The object of these techniques is to make use of the several received signals to improve the realized signal-to-noise ratio, or to improve some other performance criterion.

Several diversity combining and switching techniques are known, and there have been numerous papers on this subject in recent years. (A sample of these, with comments, is indicated in a Bibliography at the end; these papers will be referenced by numbers in square brackets, running footnotes by superscript.) However, very few of these have provided quantitative comparative data on the relative performance of the various techniques, especially the two significant techniques (maximal-ratio and equal-gain) investigated since 1954. The major exception to this is a paper by Altman and Sichak [8], which is not widely known and even less understood.

Furthermore, there has been little attempt to explain the fundamental concepts and principles involved. For such reasons, therefore, it appeared desirable to provide an expository treatment of a comparative analysis, within a unified framework, of the three most promising diversity techniques presently known. An earlier memorandum[1] aimed at these objectives indicated that such a treatment might be of fairly general interest.

Of course, in an undertaking of this kind, several previously published results are naturally included as individual cases, though the available information will also be rounded out in a number of ways. Specifically, this paper includes the following material that the author has not seen published elsewhere:

1) A careful statement of the idealized circumstances required for canonical performance of coherent combiners (Section II),

2) Simple expressions for the mean signal-to-noise power ratios of various combiners [(18), (28), and (44); Fig. 8; Table I],

3) Probability distribution curves for equal-gain combiners for 3, 4, 6, and 8 channels (Figs. 10–13, Table II),

4) Estimates of the relative performance of three standard combiners for non-Rayleigh fading (Section VII),

5) A discussion of the relative performance of three standard combiners for correlated fading (Section VIII),

6) Estimates of the degradation of the average performance of equal-gain and maximal-ratio combiners caused by correlated noise voltages (Section X),

7) A bound (due to Stein) on the degradation of coherent-type combiners with imperfectly coherent signals (Section XII),

8) Certain aspects of the determination, meaning, and use of long-term distributions (Section XIII).

In addition, some previously published material has been simplified or otherwise clarified.

It should be mentioned that the criteria employed below are expressed entirely in terms of SNR. This has sometimes been taken to mean that the results were principally applicable to continuous signals, although they are also applicable to certain binary or other discrete signals and can be translated into error rates once a suitable detection characteristic is either theoretically or experimentally known. But in the case of binary systems, it is possible to obtain more specific and precise results on error rates for specific systems. Such results have been extensively studied by Pierce [10], [15] and others and are not considered below. Neither is there a discussion of the considerable benefits obtainable by coding or other signal preprocessing techniques designed to counteract fading, several of which are currently under investigation by other workers.

On the other hand, it should be noted that radar and navigation systems in which a repetitive-addition signal-enhancement technique is employed are closely similar, in some respects, to certain diversity systems. Although radar and navigation systems are not discussed in detail below, many of the results and discussions set forth there are directly applicable to such systems.

II. Basic Assumptions and Other Preliminaries

The principal background required of the reader is a basic acquaintance with certain elementary notions of probability and statistics, essentially equivalent to those developed in the first six pages of a highly readable tutorial paper given by Bennett.[2] No advanced techniques are required here. However, we shall make frequent use of a few ideas and techniques that were not particularly emphasized by Bennett, and a brief exposition of these is given in Appendix I. All probability distributions used in this paper will be interpreted as explained there.

We shall be concerned throughout with random variables given as functions of time (waveforms) in various intervals. In this setting, time and distribution averages are one and the same thing so $\bar{f}$ or $\langle f\rangle$ or $\bar{x}$ or $\langle x\rangle$ for such averages will be written interchangeably, but it is important to note at the outset that our averages will refer to intervals of different durations. Specifically, intervals of three different durations will be considered: 1) Short intervals, whose duration will be denoted by T. The requirement for T is that it must be short in comparison to the time required for the fading amplitude to change appreciably, but long in comparison to the period of the lowest frequency of interest in the signal. Specific representative values of T would range from a few microseconds to a few milliseconds. 2) Intermediate intervals, whose duration will be denoted by T_1. The requirements T_1 must satisfy are rather complicated and will be explained at various points below. Specific suitable values of T_1 would range from a few minutes to a few hours. 3) Long intervals, whose duration will be denoted by T_2. Values of T_2 would range from one month to one year or more.

These intervals will be employed as follows. The short intervals of length T will be used to form "local" statis-

[1] D. G. Brennan, "Linear techniques in diversity communication," March, 1956. (Unpublished memorandum.)

[2] W. R. Bennett, "Methods of solving noise problems," Proc. IRE, vol. 44, pp. 609–638; May, 1956.

tics. For example, suppose $e_1(t)$ is the instantaneous signal voltage and $e_2(t)$ is the instantaneous noise voltage on some circuit. Then

$$x(t) = \left[\frac{1}{T}\int_{t-T}^{t} [e_1(\tau)]^2 d\tau\right]^{1/2} = \sqrt{\langle e_1^2\rangle} \qquad (1)$$

and

$$y(t) = \left[\frac{1}{T}\int_{t-T}^{t} [e_2(\tau)]^2 d\tau\right]^{1/2} = \sqrt{\langle e_2^2\rangle} \qquad (2)$$

would be the local rms signal and local rms noise, respectively, and x^2 and y^2 would be the local mean-square signal and noise voltages. Letting R denote the circuit resistance, x^2/R would be the local average signal power at time t, obtained by averaging e_1^2 over the last T seconds to find $x^2(t)$. This averaging could be performed, for example, by feeding e_1^2 into a suitable linear filter. Alternatively, one could determine the distribution of e_1 in the interval $[t-T, t]$ and obtain $x^2(t)$ as the second moment of the distribution, though distributions in intervals of length T will not actually be of concern here.

Local statistics such as (1) and (2) will generally fluctuate in time because of fading and other effects. For example, the local rms SNR $x(t)/y(t)$ and the local signal-to-noise power ratio

$$p(t) = \frac{x^2(t)}{y^2(t)} \qquad (3)$$

will usually vary over wide limits, though they will be much better behaved than the (meaningless) instantaneous ratio $e_1(t)/e_2(t)$. The behavior of variables such as the local statistic (3) in intervals of length T_1, where $T_1 \gg T$, will be studied. In particular, various distributions and averages relative to intervals of length T_1 will be considered. Such T_1-distributions and T_1-averages will also change with time, in ways discussed in Sections VII and XIII. Performance relative to T_1-intervals under standard conditions is summarized in Section VI.

Finally, the variability of certain T_1-averages will be considered in intervals of length T_2, where $T_2 \gg T_1$. This is done in Section XIII. It is usually assumed in system design that, for suitable values of T_2, all distributions for the system in question will be essentially the same in every corresponding interval of length T_2. (A suitable value might be one year, for example.) This is in marked contrast to the situation for T_1-distributions. However, it is found experimentally that this assumption is a reasonable first approximation; moreover, if this assumption were not satisfied, there would be no method available for predicting the performance of the system, at least at the present time.

By concentrating on system behavior relative to such prescribed lengths of intervals, it is possible to keep the relation between theory and experiment clearly visible, including, in particular, the practicable experiments required to verify theoretical predictions. This procedure is therefore vital to a complete and realistic analysis of communication systems in general and diversity systems in particular.

In general, the term "diversity system" refers to a system in which one has available two or more closely similar copies of some desired signal. For example, certain radar systems operate by storing the signal received during one scan and adding this to the signal received during the next scan. If $f_1(t)$ is written for the output of the storage device and $f_2(t)$ for the signal currently being received, then the composite signal is simply $f_1(t)+f_2(t)=f(t)$. Now, $f_1(t)$ may consist of a desired message component $s_1(t)$ and an undesired additive noise component $n_1(t)$, so that $f_1(t)=s_1(t)+n_1(t)$, and similarly $f_2(t)=s_2(t)+n_2(t)$. Hence, the composite signal may be written

$$f(t) = [s_1(t) + s_2(t)] + [n_1(t) + n_2(t)], \qquad (4)$$

i.e., in the form of a resultant message component (s_1+s_2) plus a resultant noise component (n_1+n_2). If the message components s_1 and s_2 are closely similar, their sum s_1+s_2 will simply approximate an enlarged copy of either s_1 or s_2. On the other hand, the noise components n_1 and n_2 may be quite dissimilar; one may be negative part of the time the other is positive, and vice versa, so they may partially cancel for part of the time. The sum (4) may then be a better signal than either f_1 or f_2 alone; in particular, $f(t)$ may have a higher local SNR $p(t)$, defined as in (1)–(3), with $e_1=s_1+s_2$, $e_2=n_1+n_2$ than either f_1 or f_2 alone. Thus, one way of using two similar or suitably related copies, f_1 and f_2, may be simply to add them together. Certain navigation systems in which a periodic signal is transmitted also use this storage-and-addition principle.

More generally, one may have N such copies $f_1(t)$, $f_2(t), \cdots, f_N(t)$, each of the form $f_j(t)=s_j(t)+n_j(t)$, and one may form the sum

$$f(t) = f_1(t) + f_2(t) + \cdots + f_N(t) = \sum_{j=1}^{N} f_j(t), \qquad (5)$$

which may outperform, in some sense, the individual f_j. However, in view of the fact that the f_j will have fluctuating local statistics, it will be convenient to consider weighted sums of the f_j; that is, the general linear combination will be considered:

$$f(t) = a_1 f_1(t) + \cdots + a_N f_N(t) = \sum_{j=1}^{N} a_j f_j(t), \qquad (6)$$

in which each f_j is weighted by a combining coefficient a_j, which is proportional to the channel gain and may be allowed to vary in accordance with the fluctuating local statistics of the $f_j(t)$. However, the cases to be considered will be those in which the a_j are locally constant, or at least approximately so. The adjective "linear" in the title of this paper stems from (6). Since the a_j may be allowed to vary, depending on the f_j, one should perhaps speak of (6) as locally linear or "quasi-linear." Evidently (4) is simply the case of (6) in which $N=2$, $a_1=a_2=1$.

In diversity communication systems, there are several known methods of obtaining two or more signals f_j, and several known methods of combining these to obtain an improved signal. However, all of the latter methods in current practical use are special cases of (6). Let us first consider briefly methods of obtaining several suitable f_j. The simplest of these is that in which a single transmitting antenna furnishes a signal to several well-separated remote receiving antennas; this method is called space diversity. A variant of this, suitable for use in systems operating at UHF and above, uses two separated transmitting antennas, one of which transmits vertically polarized radiation and the other of which transmits horizontally polarized radiation, and a single receiving reflector with two feed horns or dipoles to separate the vertical and horizontal received signals. By combining these two methods, Altman and Sichak [8] obtained a fourth-order, bidirectional, full duplex space diversity system that requires only two reflectors at each end, as indicated in Fig. 1. (However, it should be added that recent experimental evidence indicates that the fading on the crossed pair of paths is more highly correlated than on the other pairs of paths.) In one form or another, space diversity has been the most commonly used form of diversity communication.

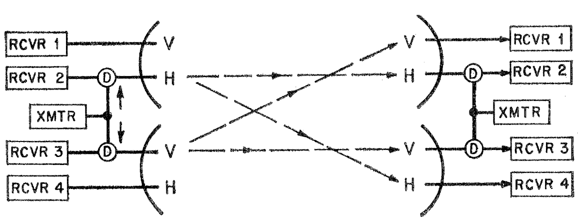

Fig. 1—Four-channel bidirectional space diversity system suitable for UHF and SHF systems. Signal paths are indicated for one direction only. The circles marked *D* denote diplexing filters. The transmitters are on different frequencies.

Another method, called frequency diversity, involves transmitting the same information on two or more carrier frequencies. If these are sufficiently separated, the fading on the various channels is approximately independent, as in the case of space diversity. This method is economical in terms of antennas and real estate, but expensive in terms of transmitters and required bandwidth. It has been discussed more often than used. (However, there are circumstances in which it is useful and has actually been used.) This is also true of the method called time diversity, so far as communication systems are concerned; however, it is not true of radar and navigation systems, as the method discussed in the opening paragraph of this section is essentially time diversity, although this terminology has not been much used in the radar field. In radio communication systems, time diversity involves transmitting the same information two or more distinct times. When this is instrumented for automatic operation, its chief disadvantage is equipment complexity; however, the simple practice of sending each word twice, as used by many commercial CW stations, is actually a primitive but useful form of time diversity. At the other extreme, a very sophisticated communication system, currently under development,[3] which is designed to eliminate effects due to multiple transmission paths between fixed antennas, actually sorts out the various multipath contributions and recombines them with suitable delays and may be regarded as a form of time diversity in which the diversity is provided by the transmission medium itself.

A method that will sometimes yield two approximately independent fading signals is called polarization diversity. In normal ionospheric transmission at frequencies of a few megacycles, it is found that the received signal includes both vertically and horizontally polarized components, and the fading of these components is approximately independent.[4,5] However, in tropospheric transmission at UHF and above, the polarization of the transmitted signal is quite well preserved[6] and very little effect of this type takes place. Furthermore, even if both horizontal and vertical components are transmitted and separately received, the fading of the two components is far from independent if only a single transmission path is involved.[6]

Another method that has been used (infrequently) in the high-frequency region involves the combination of signals arriving with different angles of arrival (the Musa system).[7] A somewhat similar approach at UHF and above is currently under investigation by several workers,[8–10] but the efficacy of this technique is not yet firmly established.

Whichever of these methods is used, the signals obtained will initially be at radio frequency. The diversity combining techniques employed subsequent to this stage may be classed in two groups: predetection combining

[3] R. Price and P. E. Green, Jr., "A communication technique for multipath channels," Proc. IRE, vol. 46, pp. 555–570; March, 1958.

[4] J. L. Glaser and L. P. Faber, "Evaluation of polarization diversity performance," Proc. IRE, vol. 41, pp. 1774–1778; December, 1953.

[5] G. L. Grisdale, J. G. Morris, and D. S. Palmer, "Fading of long-distance radio signals and a comparison of space- and polarization-diversity reception in the 6–18 mc range," *Proc. IEE*, pt. B, vol. 104, pp. 39–51; January, 1957.

[6] J. H. Chisholm, P. A. Portmann, J. T. deBettencourt, and J. F. Roche, "Investigations of the angular scattering and multipath properties of tropospheric propagation of short radio waves beyond the horizon," Proc. IRE, vol. 43, pp. 1317–1335; October, 1955. See especially Fig. 20, p. 1331.

[7] F. E. Terman, "Radio Engineers Handbook," McGraw-Hill Book Co., Inc., New York, N.Y., pp. 660–661; 1943. See also papers by Polkington and Friis and Feldman cited therein.

[8] R. Bolgiano, Jr., N. H. Bryant, and W. E. Gordon, "Diversity Reception in Scatter Communication with Emphasis on Angle Diversity," Cornell Univ., Ithaca, N. Y., Elec. Engrg. Res. Rep. 359; January, 1958.

[9] A. T. Waterman, Jr., "A rapid beam-swinging experiment in transhorizon propagation," IRE Trans. on Antennas and Propagation, vol. AP-6, pp. 338–340; October, 1958.

[10] J. H. Chisholm, L. P. Rainville, J. F. Roche, and H. G. Root, "Angular diversity reception at 2290 mcps over a 188-mile path," presented at Symp. on Extended Range and Space Communications, George Washington Univ., Washington, D. C.; October 6-8, 1958. (Published in the Symp. Rec.)

methods and postdetection combining methods. In those methods in which, at any given time, only one of the a_j in (6) is different from zero, *i.e.*, a switch of some kind, the distinction is basically unimportant. However, important differences arise when the combining method is one in which two or more of the a_j may be different from zero at the same time. For example, it is clear that the simple addition scheme (4) can fail grievously if the message components $s_1(t)$ and $s_2(t)$ are not in the same phase, and RF or IF diversity signals will not usually be in the same phase unless special measures are taken to insure this. Consequently, such combining methods require special phase-control provisions when used in predetection applications, while this is not always the case in postdetection applications. An even more important difference arises in the case of FM or other bandwidth-exchange systems, where predetection combining can lead to substantial improvement over postdetection methods, as will be seen.

Once the method of providing a multiplicity of signals is decided, the basic problem confronting the designer of a diversity system becomes one of choosing the most appropriate method of combining these signals on the basis of reasonably accurate quantitative estimates of the performance of the various techniques. The balance of this paper is principally devoted to this problem. Instrumentation problems as such are not considered here; however, papers which describe certain instrumentation techniques are indicated.

We shall find it most economical to consider first a particular class of circumstances, and then indicate the way in which the results are modified by other circumstances or, in some cases, indicate where such modifications are treated elsewhere in the literature. The circumstances initially considered are those often applicable to postdetection combining in an AM system, or a single-sideband system in which provision is made for maintaining coherence of the postdetection signals.[11] These conditions are as follows: assume that N simultaneous functions, $f_1(t)$, $f_2(t)$, $\cdots$, $f_N(t)$ represent the signals received in N different diversity channels as corrupted by noise and fading; each f_j, $j=1, 2, \cdots, N$ represents the corrupted signal in the jth channel containing the originally transmitted signal $m(t)$. For convenience, suppose that $m(t)$ is a steady test tone at a representative midband frequency, or some other steady test signal with a constant local mean square $\overline{m^2}=1$. That the following conditions are approximately satisfied is also assumed:

(**A**) The noise in each channel is independent of the signal, and additive: $f_j(t)=s_j(t)+n_j(t)$ where s_j and n_j are the signal and noise components, respectively, in the jth channel.

(**B**) The signals $s_j(t)$ are locally coherent; *i.e.*, $s_j(t) = x_j m(t)$, where the x_j are positive real numbers that change with time because of fading, but at a rate that is very slow in comparison to the instantaneous variations of $m(t)$. More precisely, assume that the x_j do not change appreciably within any period of duration T, where T is the duration of the interval employed for the local averages. Then, since $\overline{m^2}=1$,

$$\overline{s_j^2} = \frac{1}{T}\int_{t-T}^{t}[s_j(\tau)]^2 d\tau = \frac{1}{T}\int_{t-T}^{t} x_j^2[m(\tau)]^2 d\tau$$
$$= x_j^2 \cdot \frac{1}{T}\int_{t-T}^{t}[m(\tau)]^2 d\tau = x_j^2, \tag{7}$$

so that $x_j = x_j(t)$ is simply the local rms value of s_j, taken over the last T seconds before the present time, t. It is clear that T must be short in comparison to the time required for the fading amplitude to change appreciably, but long in comparison to the period of $m(t)$.

(**C**) The noise components $n_j(t)$ are locally incoherent (*i.e.*, uncorrelated) and have zero means: $\overline{n_i n_j} = \overline{n_i}\ \overline{n_j}$ if $i \neq j$, where the duration of the averages is the same as in (7). We shall also assume that the local mean square noises $\overline{n_j^2}$ are slowly varying, or, sometimes, constant.

(**D**) The local rms values of the signals, $x_j(t) = \sqrt{\langle s_j^2 \rangle}$, are statistically independent. Note that this assumption automatically implies that at least two intervals are considered: first, the period T [of (7)] involved in the definition of the x_j; and second, an interval of duration T_1 in which we observe the $x_j(t)$ as new random variables. Evidently $T_1 \gg T$; in practical cases, T might be a few milliseconds and T_1 approximately 30 minutes. A discussion of the requirements on T_1 is provided in Section XIII. It is important that T_1 cannot be too long. Assumption (**D**) is that, when observed in intervals with a duration on the order of T_1, the variables $x_j(t)$ are statistically independent.

The circumstances characterized by assumptions (**A**)–(**D**) are illustrated in exploded fashion in Fig. 2 for $N=2$. By "exploded," we mean that the actual signals given would be $f_j(t)=s_j(t)+n_j(t)$, $j=1, 2$, while the s_j and n_j are shown separately. The meaning of the locally coherent assumption (**B**) is that, over periods of length T, the signals s_1 and s_2 are essentially identical except in amplitude, which is approximately constant over such periods. The local rms values $x_j(t)$ are indicated by the dashed curves. Note that the assumption $s_j(t)=x_j m(t)$ implies that the $s_j(t)$ have the same zero crossings, and are in phase. If the s_j are RF or IF signals, the period T might be several microseconds or more, in which case no variation of the x_j would be perceptible within the scale of Fig. 2. If the s_j represent base-band signals, T might be a few milliseconds.

In contrast to the s_j, it will often be required that the noises $n_j(t)$ be essentially different; this is the meaning of (**C**), as suggested by the waveforms $n_1(t)$ and $n_2(t)$ of Fig. 2. In particular, it will often be assumed that $\langle n_i n_j \rangle = 0$ (if $i \neq j$) over every interval of length T. In addition, however, it will sometimes also be as-

[11] W. E. Morrow, Jr., C. L. Mack, Jr., B. E. Nichols, and J. Leonhard, "Single-sideband techniques in UHF long-range communications," Proc. IRE, vol. 44, pp. 1854–1873; December, 1956.

sumed that the noises n_j have constant local average power, *i.e.*, that

$$\overline{n_j^2} = \frac{1}{T}\int_{t-T}^{t} [n_j(\tau)]^2 d\tau \tag{8}$$

is a constant, independent of t and j. This would be at least approximately true of the waveforms n_1 and n_2 of Fig. 2.

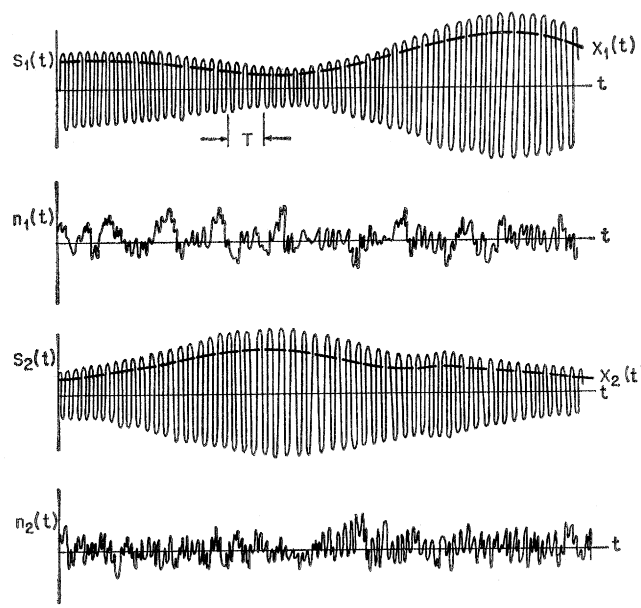

Fig. 2—Signals and noises in two diversity channels.

Assumption (**D**) is not particularly illustrated in Fig. 2, and could not be successfully illustrated there because the period T_1 required for approximate independence of the x_j is much greater than the total scale length of Fig. 2. If the x_j were plotted throughout an interval of length T_1 and the graph were then compressed to the length of Fig. 2, the resulting curves would resemble the waveforms illustrated for n_1 and n_2, except that the x_j would be non-negative and would not usually be symmetric about their mean values. In particular, the x_j are not locally coherent in the sense of (**B**), where this "locally" refers to intervals of length T_1. Note that distributions or averages of the x_j or quantities derived therefrom, *e.g.*, $\langle x_j^2 \rangle$, refer to intervals of length T_1. Such averages could be distinguished by suitable notation, *e.g.* $\langle \quad \rangle_1$, but it will simplify the appearance of various expressions if the context is relied upon to make clear whether a short-term or intermediate-term average or distribution is meant.

Most of the work below is concerned with signal-noise-ratios, and from here on the word "ratio" is to mean "SNR." This will be qualified as an amplitude ratio or a power ratio as the context requires. $pj = x_j^2/\overline{n_j^2}$ will be written for the local power ratio in the jth channel, and $x_j/\sqrt{\langle n_j^2 \rangle}$, is, similarly, the local amplitude ratio. We shall often take $\overline{n_j^2} = 1, j = 1, 2, \cdots$, in which case the local amplitude ratio is simply x_j numerically, and $p_j = x_j^2$.

It will frequently be assumed that the variables x_j follow a Rayleigh distribution with density and distribution functions

$$p(x_j) = 2x_j e^{-x_j^2}, \qquad P(x_j) = 1 - e^{-x_j^2}, \tag{9}$$

respectively. A plot of the Rayleigh density function is given in Fig. 15. All distributions considered in this paper are zero for negative values, and expressions such as (9) are to be understood as referring to positive values only. Writing the Rayleigh distribution in the form (9) implies a particular choice of scale; in particular, it implies that $\langle x_j^2 \rangle = 1$. The Rayleigh distribution is often written with an arbitrary scale factor, say

$$P(y) = 1 - e^{-(y/R)^2}, \qquad p(y) = \frac{2y}{R^2} e^{-y^2/R^2}, \tag{10}$$

in which case $\langle y^2 \rangle = R^2$. However, the data below are given in a form that is completely independent of such scale factors, until Section XIII. This saves considerable cluttering of the landscape below. Similarly, $\overline{n_j^2} = 1$ will often be taken instead of $\overline{n_j^2} = n_0^2$, for example. For, when $\overline{n_j^2} = 1$, then x_j is exactly the local amplitude ratio, which has the distribution (9), and $p_j = x_j^2$ has the simple distribution

$$G(p_j) = 1 - e^{-p_j}, \qquad g(p_j) = e^{-p_j}. \tag{11}$$

(Distribution functions will always be written with upper case letters, density functions with lower case letters.)

There are four principal types of diversity combining systems in practical use. Many of the combiners in actual use are not pure examples of one of these types; *i.e.*, they involve approximations to, or modifications of, one of these types. However, the effect of such modifications can often be estimated, at least roughly. (The terminology used here is not entirely standard—indeed, there is no generally accepted standard terminology—but is the result of careful consideration and discussion with several colleagues.) The four "pure" techniques are as follows:

1) Scanning Diversity. This technique is of the switched type; *i.e.*, at any time, only one of the a_j in (6) is different from zero, and that one is equal to 1. A selector device scans the channels in a fixed sequence until finding a signal above a preset threshold, uses that signal only until it drops below threshold, and then scans the other channels in the same fixed sequence until it again finds a signal above threshold. It is often applied to the case of two antennas supplying a single receiver through the switch, which is why it is sometimes called antenna selection diversity. It does not require a separate receiver for each channel, but at least one of the following techniques will always outperform it. We shall not consider this type in the present paper, confining our attention to the next three. Th

has been analyzed (for $N=2$) by Hausman [4] and most recently and most extensively by Henze [13].

2) Selection Diversity. This is also a switched technique, but of a more sophisticated sort. The design criterion here is that, at any given time, the system simply picks out the best of the N noisy signals $f_1, f_2, \cdots, f_N$, and uses that one alone; the others do not then contribute to $f(t)$. More precisely, let k denote the index of a channel for which $p_k \geq p_j$, $j=1, 2, \cdots, N$; then this type of system is characterized by the design criterion

$$a_j = \begin{cases} 1, & \text{for } j = k, \\ 0, & \text{for } j \neq k, \end{cases} \tag{12}$$

in terms of (6). This is essentially the classical form of diversity communication [1], [2]. Very often, the selection in such systems is by electronic means (*e.g.*, by using a common detector in such a way that the strongest signal cuts the others off) and is not quite as sharp as (12) would indicate; however, (12) is often a good approximation to such cases. A three-channel selection diversity system is depicted in Fig. 3.

3) Maximal-Ratio Diversity. This system is defined by the property that, among all systems of the type (6), it yields the maximum SNR of the output signal $f(t)$, provided assumptions (**A**), (**B**), and (**C**) are satisfied. More precisely, let p denote the local power ratio of $f(t)$. Then a maximal-ratio system realizes

$$p = \sum_{j=1}^{N} p_j; \tag{13}$$

i.e., the maximum power ratio realizable from any linear combination (6) is equal to the sum of the individual power ratios. Furthermore, the result (13) is equivalent to the requirement that the coefficients in (6) be proportional to

$$a_j = x_j/\overline{n_j^2}; \tag{14}$$

i.e., the maximum output ratio (13) is realized if and only if the gain of each channel is proportional to the rms signal and inversely proportional to the mean square noise in that channel, with the same proportionality constant for all channels. This will be proven below.

(This result has several times been quoted in the literature as requiring the weighting to be proportional to the amplitude ratio. It should be noted that this is correct only in the case where the local noise powers are all equal, in which case it would be less misleading to speak of weighting proportional to the rms signal.)

It is clear that (13) is a definite improvement over either scanning diversity or selection diversity, which can yield only one of the terms in the sum $\sum p_j$ as the output power ratio. This observation is essentially due to Kahn [5], although the form stated here can be traced to [6], and closely similar results have been used in radar systems for some time. (It is quite possible that the diversity system discussed by Peterson, *et al.* [2] was actually a maximal-ratio system. However, the authors made it clear that they were thinking in terms of selection diversity, whatever their actual instrumentation may have realized.) Maximal-ratio diversity has sometimes been called ratio squarer diversity, optimum diversity, and combiner diversity. Radar systems of the type discussed in connection with (4) and which employ square-law detection are essentially maximal-ratio systems. The general arrangement of a two-channel maximal-ratio system suitable for postdetection combining is shown in Fig. 4; a predetection combiner would require the addition of phase-control circuitry to satisfy assumption (**B**).

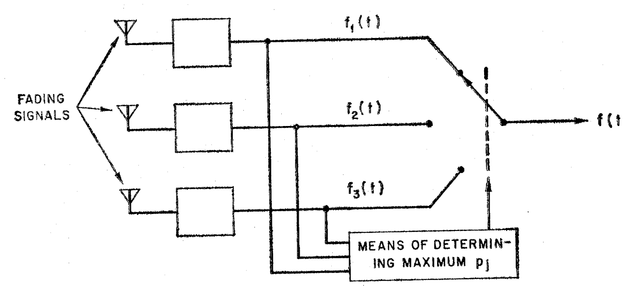

Fig. 3—Selection diversity. The f_j may be predetection or postdetection signals.

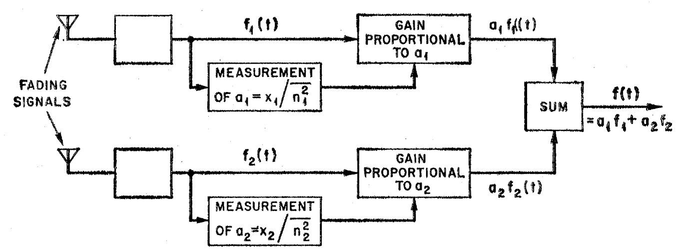

Fig. 4—Maximal-ratio diversity.

4) Equal-Gain Diversity. This is probably the simplest possible linear diversity technique; it is characterized by the property that all channels have exactly the same gain. Thus, in terms of (6),

$$a_j = 1, j = 1, 2, \cdots, N, \tag{15}$$

i.e., the noisy signals $f_j(t)$ are simply added together. In applications of this technique, the channel gains can be made to vary in such a way that the resultant signal level is approximately constant; however, this is irrelevant to the performance of the system. The important feature is that the channel gains are all equal. Note: it is important to observe that this is not the case with conventional common-detector type diversity systems; a common-detector combiner is essentially a selection diversity system, and an equal-gain system is decidedly different both in instrumentation and performance. However, an equal-gain system may well use a common AGC detector, but not a common signal detector of the usual type, as is also the case with maximal-ratio systems [5]. A basic two-channel equal-gain system is illustrated in Fig. 5. Note that the blank boxes representing receivers must have the same gains, including conversion and detection gains, which, therefore, must be fixed; they could not include separate, independent AGC systems. Also, they could not be conventional FM (or similar) receivers, as the detection gain of an FM receiver depends on the signal level. However, it is possible to instrument unconventional FM detectors for postdetection equal-gain combining.[12] An arrangement suitable for use with AGC is shown in Fig. 6. As in the case of Fig. 4, the application of an equal-gain combiner before detection would require the addition of phase control provisions to Figs. 5 and 6.

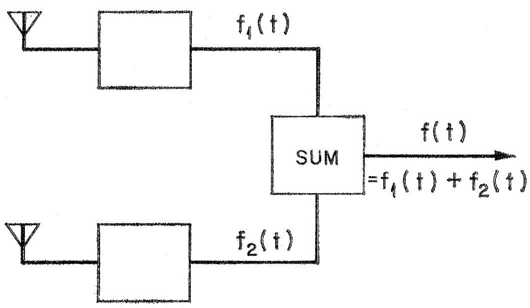

Fig. 5—Basic equal-gain diversity.

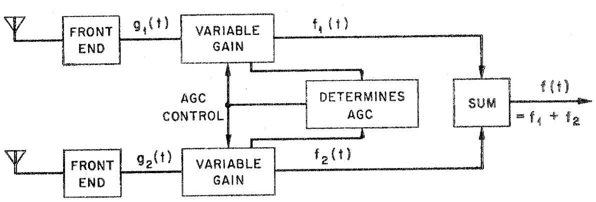

Fig. 6—Equal-gain diversity. The boxes "variable gain" must have the same gain, which may include conversion and detection gain.

It has been pointed out by Sichak [8][13] that, under conditions often occurring in practice, equal-gain systems will outperform selection diversity, and will perform almost as well as maximal-ratio systems. In view of the simplicity of the instrumentation required for (15) as compared to (14) (equivalently, see Figs. 5 and 4), this fact is of great practical importance. The conditions required are that assumptions (**A**)–(**D**) must be satisfied; in addition, the x_j must be Rayleigh-distributed, and the local mean square noises $\overline{n_j^2}$ must be approximately constant. Under other conditions, it may not perform as well as selection diversity, as will be seen. Since, however, these conditions are often (approximately) satisfied, it follows that equal-gain diversity should be more widely known than is presently the case.

In the following three sections, the principal features of diversity combiners of types 2)–4) above will be developed. In particular, distribution functions will be obtained for the local power ratio p of the composite signal $f(t)$, and mean values of p, under the conditions discussed. Then, in Sections VII to XII the results will be compared and evaluated, and the way in which the results are altered by various modifications of the conditions as they occur in practice will be indicated.

[12] R. T. Adams, private communication; December, 1958.
[13] W. Sichak, Fed. Telecommun. Labs., Nutley, N. J., private communication; August 19, 1955.

III. Selection Diversity

The distribution function for an N-channel selection diversity system is particularly simple to obtain, provided the local noise powers $\overline{n_j^2}$ are constant. Let $\overline{n_j^2}=1$, $j=1, 2, \cdots, N$, and assume that the x_j are Rayleigh-distributed. Then the individual channel power ratios p_j have the distribution $G(p_j)$ of (11). By (12), the output power ratio of the combiner is simply the largest of the individual p_j. Now, if the largest power ratio is $\leq p$, then the power ratio of every channel is $\leq p$; conversely, if the power ratio of every channel is $\leq p$, then so is the power ratio of f. Hence, the probability of having the power ratio of f be $\leq p$ is precisely the probability of having the individual channel ratios all $\leq p$ simultaneously. Since the x_j are independent, by (**D**), so are the $p_j = x_j^2$, and, hence, the probability that all channels have power ratio $\leq p$ is simply the product of the separate probabilities that each channel individually has a power ratio $\leq p$. Thus,

$$\begin{aligned} S_N(p) &= G(p)\cdot G(p) \cdots G(p) = [G(p)]^N \qquad (16) \\ &= (1 - e^{-p})^N \end{aligned}$$

is the distribution function of p, the realized local power ratio, for an N-order selection diversity system.

The average value $\bar{p}$ of p will be required for this system. In this case, it is most easily obtained from the distribution (16). Thus,

$$\bar{p}(N) = \int_{-\infty}^{\infty} p dS_N(p) = \int_0^{\infty} pN(1 - e^{-p})^{N-1} e^{-p} dp. \qquad (17)$$

This integral is evaluated in Appendix VI, where it is shown to reduce to the remarkably simple form

$$\bar{p}(N) = \sum_{k=1}^{N} \frac{1}{k} \cdot \qquad (18)$$

Thus, $\bar{p}(2) = 1 + \frac{1}{2} = \frac{3}{2}$, $\bar{p}(3) = 1 + (\frac{1}{2}) + (\frac{1}{3}) = 11/6$, etc.; these values will be used below. It is clear at once from (18) that increasing the number of channels in a selection diversity system yields rapidly diminishing returns; adding the Nth channel increases $\bar{p}$ by only $1/N$. It will be seen that the next two systems to be considered can perform much better in this respect, in consequence of (**B**) and (**C**). However, it should be noted here that neither the functioning of a selection diversity system nor the statistics developed in this section depend on assumptions (**B**) or (**C**), which are not required here. The significance of this fact will be discussed.

IV. Maximal-Ratio Diversity

The first order of business is to establish (13) and (14). In order to do this, it will be convenient to use a mathematical device known as the Schwarz inequality. This is not specifically related to statistics, but is a general result of great importance in many fields of pure and applied mathematics. One form of this states that if $u_1, u_2, \cdots, u_N$ are any N real numbers and $v_1, v_2, \cdots, v_N$ are any N real numbers, then

$$\left[\sum_{j=1}^{N} u_j v_j\right]^2 \leq \left[\sum_{j=1}^{N} u_j^2\right]\left[\sum_{j=1}^{N} v_j^2\right]. \qquad (19)$$

The proof of this, which is quite short, is given in Appendix II. Note that if $u_j = a_j\sqrt{\langle n_j^2\rangle}$, $v_j = x_j/\sqrt{\langle n_j^2\rangle}$, then (19) takes the form

$$\left[\sum_{j=1}^{N} a_j x_j\right]^2 \leq \left[\sum_{j=1}^{N} a_j^2\overline{n_j^2}\right]\left[\sum_{j=1}^{N} (x_j^2/\overline{n_j^2})\right], \qquad (20)$$

which, since $p_j = x_j^2/\overline{n_j^2}$, can be written

$$\left[\sum_{j=1}^{N} a_j x_j\right]^2 \leq \left[\sum_{j=1}^{N} a_j^2\overline{n_j^2}\right]\left[\sum_{j=1}^{N} p_j\right]. \qquad (21)$$

Now, in (6), let us write

$$s(t) = \sum_{=1}^{N} a_j s_j(t), \qquad n(t) = \sum^{N} a_j n_j(t), \qquad (22)$$

so that $f(t) = s(t) + n(t)$, and $p = \overline{s^2}/\overline{n^2}$ is the local power ratio of f. But (all sums from 1 to N)

$$\overline{s^2} = \langle [\sum a_j s_j]^2 \rangle$$

and by assumption (**B**)

$$= \langle m^2[\sum a_j x_j]^2 \rangle$$
$$= \langle m^2 \rangle \cdot [\sum a_j x_j]^2$$

and since $[\sum a_j x_j]^2$ is locally constant and can be taken outside the average, and since $\overline{m^2} = 1$,

$$\overline{s^2} = [\sum a_j x_j]^2. \qquad (23)$$

Furthermore,

$$\overline{n^2} = \langle [\sum a_j n_j]^2 \rangle$$
$$= \sum a_j^2 \overline{n_j^2}, \qquad (24)$$

by (61) and assumption (**C**). Thus, using (21),

$$p = \frac{\langle s^2 \rangle}{\langle n^2 \rangle} = \frac{[\sum a_j x_j]^2}{\sum a_j^2 \overline{n_j^2}} \leq \sum p_j, \qquad (25)$$

which proves that p cannot exceed $\sum p_j$. On the other hand, if $a_j = x_j/\overline{n_j^2}$, then

$$p = \frac{[\sum x_j^2/\overline{n_j^2}]^2}{\sum x_j^2/\overline{n_j^2}} = \frac{[\sum p_j]^2}{\sum p_j}$$
$$= \sum p_j, \qquad (26)$$

so that $p = \sum p_j$ if $a_j = x_j/\overline{n_j^2}$, and similarly if $a_j = k(x_j/\overline{n_j^2})$ for any $k \neq 0$, thus proving (13) and (14). [Readers acquainted with the Schwarz inequality for complex numbers will recognize that it may be used to include the case of positive or negative x_j, or even complex x_j. This is, however, only a more formal way of including assumption (**B**).]

It is interesting to note that the only purely statistical fact used in this development is that averaging is a linear operation, as discussed in Appendix I. No use was made of distribution functions or any similar apparatus. However, it is important to observe that each of the assumptions (**A**)–(**C**) entered in a very vital way.

We now consider the statistical properties of the local power ratio p. The first point to be noticed is that $p = \sum p_j$ implies

$$\bar{p} = \sum_{j=1}^{N} \bar{p}_j, \qquad (27)$$

without regard to the distribution of the p_j or the possible dependence of these variables. If, in particular, $\bar{p}_j = 1, j = 1, 2, \cdots, N$, then

$$\bar{p}(N) = N. \qquad (28)$$

This behavior is in marked contrast to the corresponding relationship (18) for selection diversity, which increases much less rapidly with N than (28) does.[14] (It should be clear that the notation $\bar{p}(N)$ is used in a somewhat flexible way; $\bar{p}(N)$ is a different function of N in (18) than it is here.) The average value of the local power ratio of the output of an N-order maximal-ratio system is simply $10 \log_{10} N$ db above a single channel.

In order to obtain an explicit distribution of p, we shall employ the same assumptions used for the selection diversity case, namely, that the x_j are independent Rayleigh variables and that $\overline{n_j^2} = 1, j = 1, 2, \cdots, N$, so that the p_j have the exponential distribution (11). Thus we are interested in the distribution of the sum $p = \sum p_j$ of N independent random variables, each with the distribution (11). This problem can be treated by a simple application of characteristic functions,[2] as indicated in Appendix III. Alternatively, it can easily be solved by using the geometric approach mentioned in Appendix I, without reference to characteristic functions. (One integrates the joint density function

$$\exp\left[-(p_1 + p_2 + \cdots + p_N)\right]$$

over the N-dimensional volume bounded by the hyperplane $p_1 + p_2 + \cdots + p_N = p$ and the coordinate hyperplanes.) In either case, the result, writing $G_N(p)$ for the desired distribution function and $g_N(p)$ for the associated density function, is

$$g_N(p) = \frac{1}{(N-1)!} p^{N-1} e^{-p}, \qquad (29)$$

$$G_N(p) = \frac{1}{(N-1)!} \int_0^p y^{N-1} e^{-y} dy. \qquad (30)$$

[14] For large N, (18) is approximately $\log_e N$.

By using $G_1(p)=1-e^{-p}$ and the recursion relation $G_N(p)=G_{N-1}(p)-g_N(p)$, easily verified by an integration by parts, we have

$$G_2(p) = 1 - (1+p)e^{-p},$$

$$G_3(p) = 1 - \left(1 + p + \frac{p^2}{2'}\right)e^{-p},$$

$$G_4(p) = 1 - \left(1 + p + \frac{p^2}{2!} + \frac{p^3}{3!}\right)e^{-p},$$

and, in general,

$$G_N(p) = 1 - \left(\sum_{k=0}^{N-1} \frac{p^k}{k!}\right)e^{-p}, \tag{31}$$

which can also be written

$$G_N(p) = \left(\sum_{k=N}^{\infty} \frac{p^k}{k!}\right)e^{-p}. \tag{32}$$

The utility of the form (32) is that it indicates the approximations

$$G_N(p) \cong \frac{p^N}{N!}e^{-p} \cong \frac{p^N}{N!} \tag{33}$$

are accurate for sufficiently small p. The distribution (31), known as the gamma distribution, is easily computed for the integral values of N of interest here and has also been tabulated.[15] It can also be identified with the chi-square distribution with $2N$ degrees of freedom.[16]

The origin of the maximal-ratio distribution (31) has sometimes been incorrectly attributed (in Pierce [15] and Packard [14] among others). That (31) is the distribution function of sums of squares of Rayleigh variables has been known in radar circles for quite some time. In the context of maximal-ratio diversity combiners, the result (31) for arbitrary N was first published in March, 1956, by Altman and Sichak.[17] (It also appeared independently in an unpublished memorandum[1] at about the same time.) Curves of (31) for several values of N were subsequently published by Staras [9].

V. Equal-Gain Diversity

Recalling that the relations $\overline{s^2}=[\sum a_jx_j]^2$ of (23) and $\overline{n^2}=\sum a_j^2\overline{n_j^2}$ of (24) did not depend on a choice of the a_j [*i.e.*, they hold for any combiner of the type (6), provided assumptions (A)–(C) are satisfied, and hence hold in particular for $a_j=1$, $j=1, 2, \cdots, N$], we have

$$\overline{s^2} = \left[\sum_{j=1}^{N} x_j\right]^2 \tag{34}$$

and

$$\overline{n^2} = \sum_{j=1}^{n} \overline{n_j^2} \tag{35}$$

for an equal-gain system. The relation $\sqrt{\langle s^2\rangle}=\Sigma x_j$ from (34) is simply the well-known fact that the rms value of a sum of coherent signals is equal to the sum of the individual rms values, while (35) similarly expresses the fact that the average power of a sum of uncorrelated signals is equal to the sum of the individual average powers. Some communication engineers express (34) by saying that coherent signals "add linearly"; however, this language is both formally meaningless and conducive of an imperfect understanding of the situation and is better replaced by "add coherently" if some such expression is necessary.

From (34) and (35), we have

$$p = \frac{[\sum x_j]^2}{\sum \overline{n_j^2}} \tag{36}$$

for an equal-gain system. In order to develop comparative statistics for this, it is again assumed that the $\overline{n_j^2}=1, j=1, 2, \cdots, N$, and that the x_j are independent and Rayleigh-distributed. Since $\overline{n_j^2}=1$, $p=[\sum x_j]^2/N$. Put

$$u = \sqrt{Np} = \sum_{j=1}^{N} x_j. \tag{37}$$

It is clear that the distribution function of p will follow immediately from that for u. The distribution of a sum of N Rayleigh variables, each with the distribution (9), is accordingly of interest. Unfortunately, this problem is not nearly as tractable as in the maximal-ratio case. The characteristic function of a Rayleigh variable is not expressible in an immediately useful form. We are here essentially forced to rely on the geometric approach mentioned in Appendix I. Let $B_N(u)$ denote the distribution function of (37). For $N=2$, say $u=x+y$, it can be seen that $B_2(u)$ is given by the integral of the joint density function $4xye^{-(x^2+y^2)}$ over the region of the $x-y$ plane bounded by the line $x+y=u$ and the coordinate axes (Fig. 7). (We can stay in the first quadrant because the density function is zero in all other quadrants.) It is easy to see that this is

$$B_2(u) = \int_0^u \int_{x=0}^{x=u-y} 4xye^{-(x^2+y^2)}dxdy$$

$$= 2\int_0^u ye^{-y^2}[1 - e^{-(u-y)^2}]dy \tag{38}$$

[15] K. Pearson (ed.), "Tables of the Incomplete Γ Function," Cambridge University Press, Cambridge, Eng.; 1946. In his notation, he tabulates

$$I(u, p) = \frac{1}{p!}\int_0^{u\sqrt{p+1}} t^pe^{-t}dt,$$

so that his p is here $N-1$, and his u is here $p/\sqrt{N}$.

[16] E. S. Pearson and H. O. Hartley (eds.), "Biometrika Tables for Statisticians," Cambridge University Press, Cambridge, Eng., Table 7, p. 122 ff.; 1954. Short tables of the χ^2 distribution are given in many other statistical tables and in most textbooks on statistics. See, for example, H. Cramér, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J.; 1946.

[17] Altman and Sichak, [8], p. 55, middle of right-hand column, with $(aL)^2=p$.

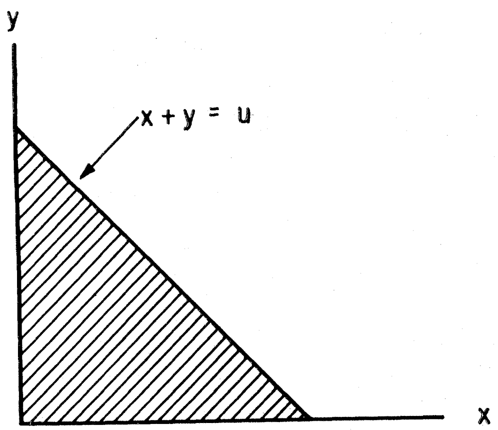

Fig. 7—Region of integration for (38).

By completing the square in the exponent and making a few other routine manipulations, this becomes [8]

$$B_2(u) = 1 - e^{-u^2} - (\sqrt{\pi/2})ue^{-u^2/2}H\left(\frac{u}{\sqrt{2}}\right), \quad (39)$$

where

$$H(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$$

is the error function and is tabulated.[18] Thus the distribution function of $p = u^2/2$, $A_2(p)$, is $B_2(\sqrt{2p})$, *i.e.*,

$$A_2(p) = 1 - e^{-2p} - \sqrt{\pi p}\,e^{-p}H(\sqrt{p}), \quad (40)$$

and is readily plotted.

Corresponding to (38), it is easy to see that the distribution function of the sum of N independent Rayleigh variables is

$$B_N(u) = 2^N \int_0^u \int_0^{u-x_N} \cdots \int_0^{u-\sum_{k=3}^N x_k} \int_0^{u-\sum_{k=2}^N x_k} x_1 x_2 \cdots x_N\, e^{-(x_1^2+\cdots+x_N^2)} dx_1 \cdots dx_N, \quad (41)$$

which is simply the integral of the joint density function over the N-dimensional volume bounded by the hyperplane $x_1 + x_2 + \cdots + x_N = u$ and the coordinate hyperplanes, as in Fig. 7 for $N = 2$. Unfortunately, the integral (41) is quite as frightful as it appears; numerous workers—going back to Lord Rayleigh himself—have tried to express $B_N(u)$ in terms of tabulated functions, but with no success if $N \geq 3$. However, $B_N(u)$ has recently been tabulated,[19] and curves of $A_N(p) = B_N(\sqrt{Np})$ have been constructed from these tables for the present paper (in Section VI) for $N = 2$, 3, 4, 6, and 8. An outline of the method of computation is sketched in Appendix IV.

We shall next obtain the average values $\bar{p}(N)$ for the equal-gain case. Although these averages depend on the distribution of p and the distribution of p is not given in a particularly explicit form, it is nevertheless easy, following Appendix I, to obtain the $\bar{p}(N)$. Since $\overline{n_j^2} = 1$, (36) becomes

$$p = \frac{1}{N}\left[\sum_{j=1}^N x_j\right]^2 = \frac{1}{N}\left[\sum_{j=1}^N x_j^2 + \sum_{i \neq j} x_i x_j\right]. \quad (42)$$

Since the x_j are independent, $\overline{x_i x_j} = \bar{x}_i \bar{x}_j$ if $i \neq j$, so

$$\bar{p} = \frac{1}{N}\left[\sum_{j=1}^N \overline{x_j^2} + \sum_{i \neq j} \bar{x}_i \bar{x}_j\right]$$
$$= 1 + \frac{1}{N}\sum_{i \neq j} \bar{x}_i \bar{x}_j, \quad (43)$$

using the fact that $\overline{x_j^2} = 1$, $j = 1, 2, \cdots, N$. Let $\bar{x}_j = r$, $j = 1, 2, \cdots, N$. By considering the terms of the sum $\sum_{i \neq j} \bar{x}_i \bar{x}_j$ as the entries from an N by N matrix with the main diagonal deleted, it is seen that there are $N^2 - N = N(N-1)$ such terms, each equal to r^2, and so

$$\bar{p}(N) = 1 + (N-1)r^2, \quad (44)$$

the desired average value. The constant $r^2 = (\bar{x}_j)^2$ depends on the distribution of the x_j. For the Rayleigh distribution, $r^2 = \pi/4 \cong 0.785$. For any distribution, $r^2 = (\bar{x})^2/\overline{x^2}$ is a dimensionless constant between 0 and 1, but values of r^2 much less than 0.785 are relatively infrequent in observed fading distributions.

It is thus seen that $\bar{p}$ increases linearly with N, as was also the case for maximal-ratio systems. The only difference is that (28) increases with slope 1 while (44) increases with slope $r^2 = \pi/4$ for Rayleigh fading. But the absolute maximum by which (28) can exceed (44) is 10 $\log_{10}(4/\pi) = 1.05$ db, and this only in the limit of an infinite number of channels.

VI. Canonical One-Hour Performance

The three systems will first be compared simply on the basis of the average values of the local power ratio p of the output. This is done first in Fig. 8, for $N = 1, 2, 3, \cdots, 10$ channels. The maximal-ratio points are values of $10 \log_{10} N$ from (28), the equal-gain points are $10 \log_{10} [1 + (N-1)(\pi/4)]$ from (44), and the selection diversity values are

$$10 \log_{10}\left[\sum_{k=1}^N (1/k)\right]$$

from (18). Since $\bar{p}_j = 1$, these give the increase in decibels in the average local power ratio over a single channel. The data of Fig. 8 for $N = 2$, 3, 4, 6, and 8 are presented from a different point of view in Table I, which gives

[18] "Tables of the Error Function and its Derivative," Natl. Bur. Stand., Washington, D. C., Appl. Math. Ser. No. 41; 1954. The error function or, for that matter, (39) itself, can also be expressed in terms of the much tabulated normal (Gaussian) probability distribution function. See "Tables of the Normal Probability Functions," Natl. Bur. Stand., Appl. Math. Ser. No. 23; 1953. Brief tables of the normal distribution function appear in most statistical texts, *e.g.*, Cramér, *op. cit.*

[19] W. C. Mason, M. Ginsburg, and D. G. Brennan, "Tables of the distribution functions of sums of Rayleigh variables," to be published.

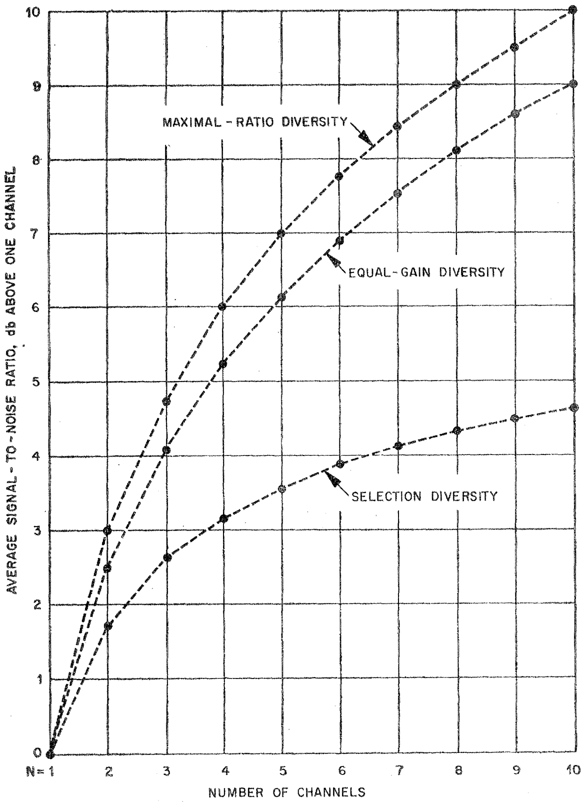

Fig. 8—Diversity improvement (in decibels) in average SNR, for independently fading Rayleigh-distributed locally coherent signals in locally incoherent noises with constant local rms values.

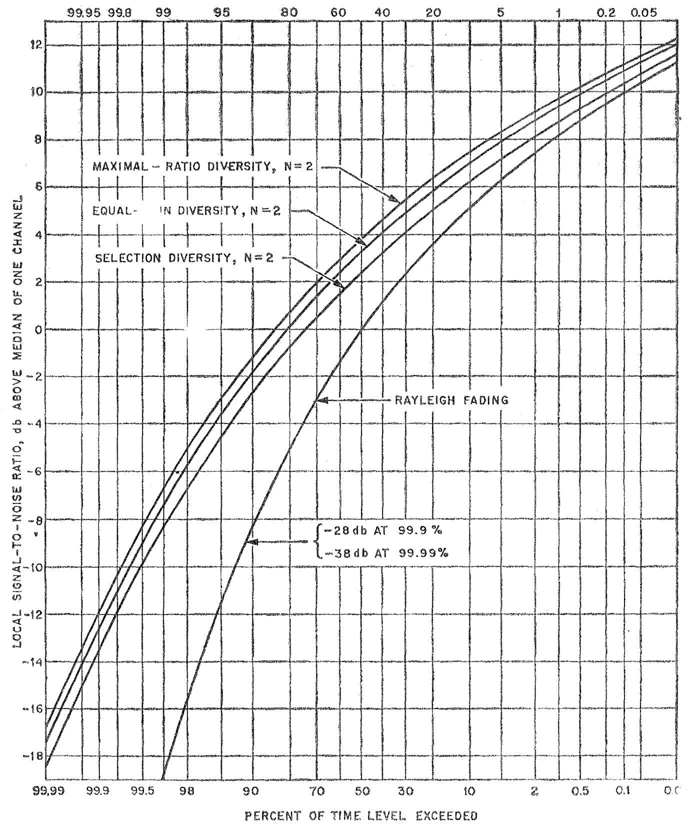

Fig. 9—Dual diversity distributions, conditions of Fig. 8.

TABLE I
COMPARATIVE AVERAGE SNR (SAME CONDITIONS AS IN FIG. 8)

Number of Channels N	Number of DB by which Maximal-Ratio Exceeds		
	Equal-Gain	Selection	One Channel
2	0.49	1.25	3.01
3	0.67	2.14	4.77
4	0.76	2.83	6.02
6	0.85	3.89	7.78
8	0.90	4.69	9.03
.	.	.	.
.	.	.	.
.	.	.	.
∞	1.05	∞	∞

the differences between the maximal-ratio values and the lower curves of Fig. 8, counting the zero axis as a curve. The last entry in the equal-gain column is essentially the assertion that no matter how far the curves of Fig. 8 were continued, the top two would never differ by more than 1.05 db, although they would get farther and farther away from the selection diversity curve and the base axis.

A brief discussion of the significance of these data is in order. These results are useful, for example, in estimating relative average system capacities, or in other circumstances where the average value alone of p is of interest. Most recent diversity systems have been designed for a specified percentage of reliability, *i.e.*, a specified percentage of time during which the system performance will exceed some given criterion. This requires information about probability distributions. This approach is appropriate whenever high reliability is a primary requisite, *e.g.*, in important military communication systems, or in relay systems carrying commercial television programs. However, it should be pointed out here that some systems do not require very high local reliability or they may effectively achieve it by other means, such as coding. In such circumstances, the data of Table I may be more meaningful than results based on the distributions to be presented.

Let us next compare the probability distributions of p realized by the three systems for different orders of diversity. The case $N=2$ (dual diversity) is illustrated in Fig. 9, together with the distribution of p for a single channel with Rayleigh fading for comparison. The term "median" in the designation of the ordinate scale of Fig. 9 refers to a value x_0 of a random variable x for which $P(x_0)=\frac{1}{2}$, *i.e.*, a value x_0 for which $x \leq x_0$ for 50 per cent of the time and $x \geq x_0$ for 50 per cent of the time. Thus, the median p_0 of the one-channel distribution (11) is obtained by setting $G(p_0)=\frac{1}{2}$ and solving for p_0, from which $p_0=\log_e 2 \cong 0.693$. The ordinate scale of Fig. 9 is expressed in decibels relative to this p_0. That is, the $N=2$ curves of Fig. 9 are plots of $10 \log_{10}(p/p_0)$ vs $100\,[1-D_2(p)]$, where $D_2(p)$ is, respectively, $G_2(p)$ of (31) (maximal-ratio), $A_2(p)$ of (40) (equal-gain) and

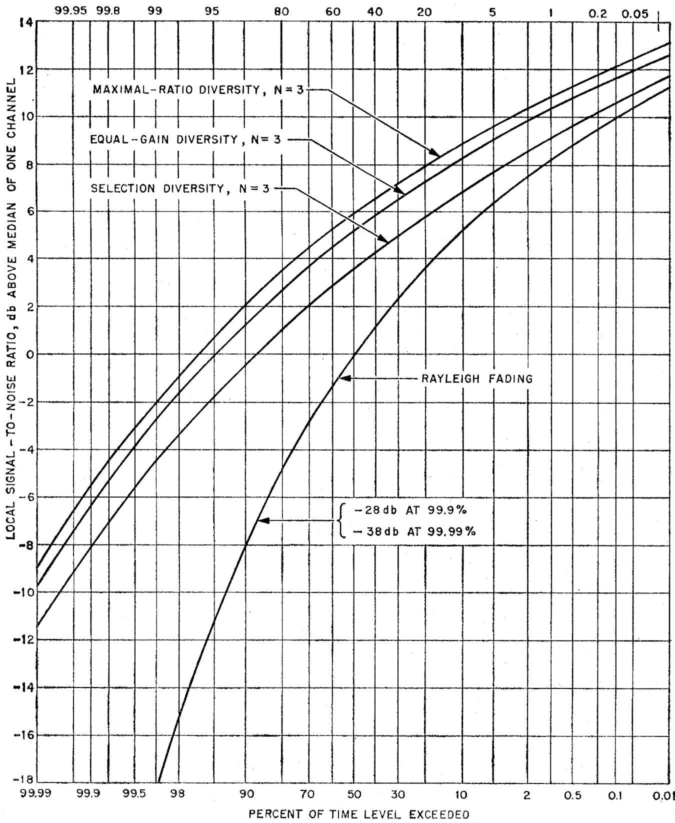

Fig. 10—$N=3$.

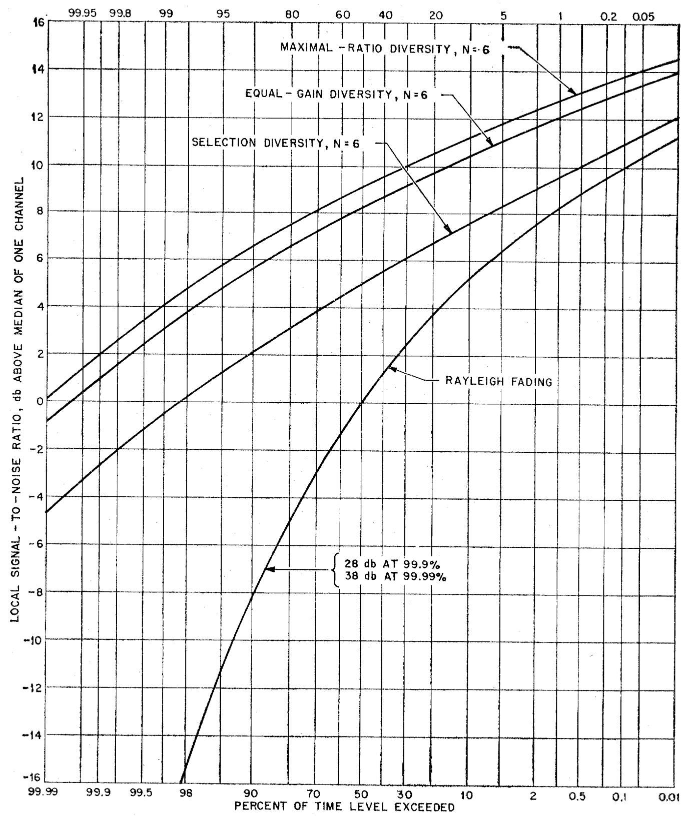

Fig. 12—$N=6$.

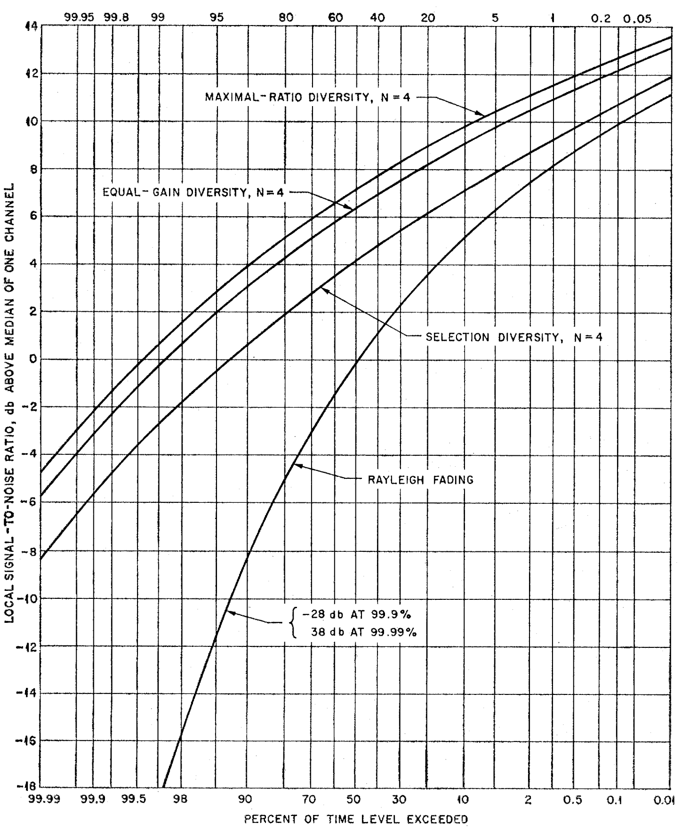

Fig. 11—$N=4$.

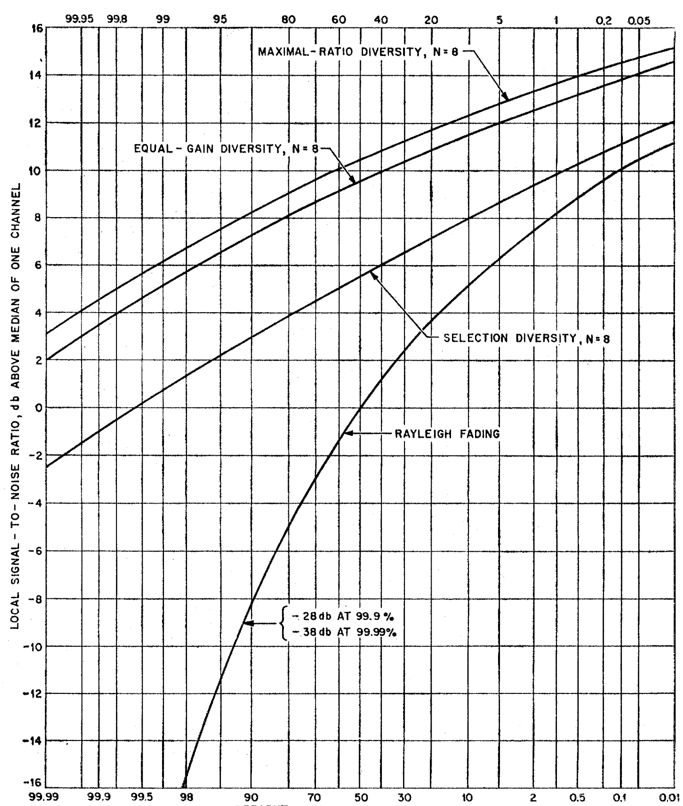

Fig. 13—$N=8$.

$S_2(p)$ of (16) (selection). 100 $[1-D_2(p)]$ is the per cent of time ordinate exceeded. The Rayleigh fading curve is 100 $[1-G(p)]$ of (11). (For the distributions considered here, the median values of p do not differ from the corresponding mean values by more than about 1.6 db. The reason for using the median value of the Rayleigh distribution as a reference here is that this is commonly presented as an experimental datum, since median values can be read directly from the distribution function determined by a totalizer.)

It is clear that the differences between the various dual diversity curves of Fig. 9 are quite small, especially in comparison to the difference between any one of them and the Rayleigh fading curve. For example, the 99.99 per cent exceeded level of the selection diversity curve is almost 20 db above the Rayleigh curve, while the maximal-ratio curve is only 1.4 db above the selection curve at the 99.99 per cent point. Evidently one would choose among the three types of two-channel systems on the basis of Fig. 9 only if one were fighting for the last decibel. Even then, one would wish to make very sure that that last decibel could actually be realized, the selection diversity curve does not depend on the important assumptions (**B**) and (**C**), which must be satisfied for the equal-gain and maximal-ratio systems to work properly, as will be seen.

However, the differences in the performance of the various combining techniques become more important as the number N of channels is increased. The maximal-ratio and equal-gain systems improve much more rapidly than selection diversity does, as can be seen in Figs. 10–13, which give the distributions for $N=3$, 4, 6, and 8, respectively.[20] (Note that the ordinate scales of Figs. 9–13 cover different ranges.) However, the maximal-ratio and equal-gain curves remain quite close together; indeed, the difference between them is hardly significant even for $N=8$. This is one of the facts that makes equal-gain diversity quite attractive and suggests that there are many applications where it should be exploited. It can be seen that the maximal-ratio and equal-gain curves differ approximately by the constants in the equal-gain column of Table I; that is, the equal-gain diversity distributions can be approximated quite well by translating the maximal-ratio distributions downward by the values in the second column of Table I.

The data of Figs. 9–13 are useful in the design of radio communication systems and radar and navigation systems of the type discussed in the Introduction. One such application is as follows. Suppose a high-reliability communication system is to be designed for a fixed information rate, which cannot be maintained whenever the received local power ratio p drops below a certain value. That is, it is desired to maintain the local power ratio above a certain value for, say, 99 per cent of the time during an interval of length T_1, for which the curves of Figs. 9–13 are applicable if the relevant conditions are satisfied. Referring to the 99 per cent exceeded values of Fig. 9, it can be seen that the difference between the Rayleigh fading curve and the dual selection diversity curve is about 10 db at the 99 per cent point. But this implies that whatever transmitter power was required for a single-channel system, a transmitter of 10 db less power would be adequate if dual selection diversity were employed at the receiving terminal. Of course, part of the reduction could be applied to the antenna gains, etc. Similarly, reference to the 99 per cent values of Fig. 11 shows that the use of fourth-order maximal-ratio diversity would enable a reduction in transmitter power of 19 db relative to that required for a single-channel system.

This reduction in transmitter power required for a given grade of local reliability has been called "diversity gain." The term was apparently introduced by Jelonek, *et al.* [3]. Here the term "local reliability diversity gain," or simply "local reliability gain," is used to emphasize the fact that it is not a gain in the usual sense and that it depends very heavily on the local reliability percentage chosen. The dependence of the local reliability gains on the percentage selected can be seen in Table II, which gives the values realized by the three types of systems for $N=2$, 3, 4, 6, and 8 corresponding to local reliability percentages of 99 per cent and 99.9 per cent.

TABLE II

LOCAL RELIABILITY GAINS (IN DB), CONDITIONS OF FIG. 8, FOR 99 PER CENT AND 99.9 PER CENT LOCAL RELIABILITY

N	Selection		Equal-Gain		Maximal-Ratio	
	99 per cent	99.9 per cent	99 per cent	99.9 per cent	99 per cent	99.9 per cent
2	10	14.5	11	15.5	12	16
3	14	20	16	21.5	16.5	22.5
4	16	22.5	18.5	25	19.5	26
6	18	25.5	21.5	29	22.5	30
8	19	27	23.5	31.5	24.5	32.5

It can be seen that the values of Table II are much larger than those of Fig. 8. They can be made to appear even larger by computing the local reliability gains corresponding to 99.99 per cent or higher percentages; however, considering the present or immediately foreseeable state of the art, such values would not be meaningful. Among other things, the Rayleigh distribution does not provide an accurate model for actual fading distributions outside of the 0.1 per cent to 99.9 per cent range.

Various modifications and extensions of these considerations as they occur in practice will be considered next. However, it should be noted that there are many practical situations in which the conditions assumed above are realistic approximations, and for which the results can be used without significant modifications.

[20] High-resolution graphs of the curves of Figs. 9–13 are available from the author to those having serious need of such graphs. Letters requesting the same should describe the nature of said need. Requests on postal cards or form letters will not be honored. This offer may be withdrawn at any time.

VII. Non-Rayleigh Fading Distributions

Only in the case of long-range UHF and SHF tropospheric transmission does it appear that observed fading distributions are most often Rayleigh, when observed in intervals of length T_1. For short-range UHF circuits and normal or scatter ionospheric transmission at VHF and below, other distributions are often observed. Indeed, at frequencies of a few megacycles and below, an accurate fit to the Rayleigh distribution is more nearly the exception than the rule. It is therefore of interest to discuss the way in which these results are modified by other distributions, assuming that the other conditions still hold.

Certain of these results are easily discussed. The maximal-ratio curve of Fig. 8 and the last column of Table I do not in any way depend on the fading distribution and are not modified at all. In order to discuss the effect on the distributions of Figs. 9–13, it will be convenient to return to the geometric approach mentioned in Appendix I and consider the case $N=2$ channels. Let $x=x_1$ and $y=x_2$ be the local amplitude ratios of the two channels. The probability that a maximal-ratio system has a local power ratio $\leq p$, *i.e.*, the maximal-ratio distribution function $G_2(p)$, is obtained by integrating the joint density function of x and y over the interior of the quarter circle $x^2+y^2=p$ in the $x-y$ plane. Similarly, the equal-gain distribution function $A_2(p)$ is obtained by integrating the same density function over the triangle bounded by the line $x+y=\sqrt{2p}$. The corresponding region for the selection diversity distribution $S_2(p)$ is the square bounded by $x=\sqrt{p}$, $y=\sqrt{p}$. These three regions are shown together in Fig. 14. Now, the fact that the maximal-ratio system outperforms the other two is intimately connected with the fact that, for any fixed p, the probability that the maximal-ratio output ratio is $\leq p$ is smaller than it is for the others, that is, $G_2(p)<A_2(p)$ and $G_2(p)<S_2(p)$. This is reflected in Fig. 9 in the fact that the maximal-ratio curve is strictly above the others. The reason for this can be seen at once in Fig. 14; the region of integration for the maximal-ratio system is smaller than it is for the others and interior to both of the others. Hence, no matter what fading distribution is involved, the joint density function will still be non-negative and, therefore, its integral over the maximal-ratio region of Fig. 14, *i.e.*, $G_2(p)$, will still be smaller than for the others. Thus, the maximal-ratio curve of Fig. 9 would be above the others for any fading distribution.

Of course, this result could also be seen from the fact that the maximal-ratio system yields an output power ratio that is indeed maximal. But there is no similar fact to use as a guide in comparing the other two, for which we must rely on Fig. 14. It can be seen there that the areas of the selection and equal-gain regions are identical and that neither region includes the other. This would lead one to suspect that the relative performance of selection diversity and equal-gain diversity depends on the form of the fading distribution. In order to discuss this, consider the nature of the possible departures from the Rayleigh distribution.

For purposes here, two cases may be distinguished: fading distributions more disperse (broader or more smeared-out) than the Rayleigh distribution, which are associated with frequent or persistent deep fades, and distributions less disperse (narrower or shallower) than the Rayleigh distribution, which are associated with shallow fading. These cases are illustrated in Fig. 15, together with the Rayleigh distribution. Curve (b), one of a family of distributions given by Rice, illustrates the less disperse or shallow fading often encountered at frequencies below UHF.[21] Curve (c) illustrates the more disperse case sometimes found in short-range UHF circuits and in high- and medium-frequency systems.

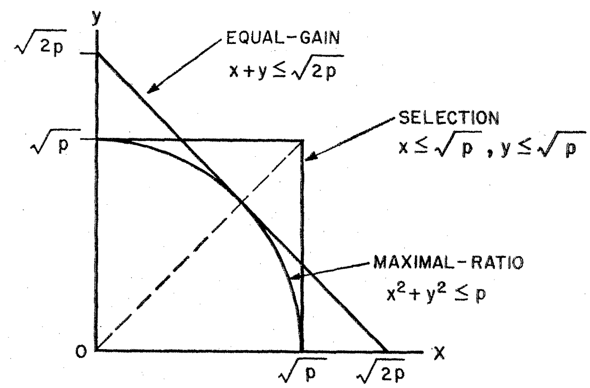

Fig. 14—Regions of integration for three types of dual diversity systems, after Altman and Sichak [8].

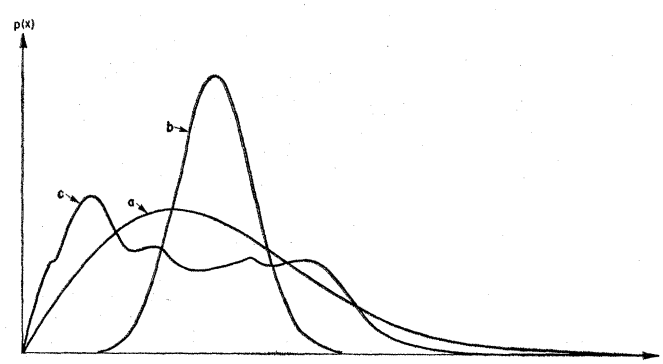

Fig. 15—Representative fading distributions. (a) Rayleigh density function. (b) Representative Rice distribution. (c) Typical distribution of the unpleasant sort often observed at frequencies below UHF.

Returning now to Fig. 14, it is not difficult to see that independent shallow fading will tend to improve the performance of an equal-gain system. This is because the height of the joint density function will be small in the region near the origin common to both the equal-gain and selection regions, and the bulk of the density function will be "pushed out" along the diagonal where

[21] R. W. E. McNicol, "The fading of radio waves of medium and high frequencies," *Proc. IEE*, vol. 96, pp. 517–524; October, 1949.

it will contribute more to the integral over the selection region than to the integral over the equal-gain region. Thus, an equal-gain combiner will continue to outperform selection diversity in the presence of shallow fading; indeed, its performance will more nearly approximate a maximal-ratio system. This can also be seen directly by considering the basic operation of a two-channel equal-gain system, and visualizing the case where the two signals are approximately constant.

However, this is not true for the more disperse distributions. Consider first the case where the individual amplitude ratios x and y are rectangularly distributed,[2] say on $0 \leq x \leq A$ and $0 \leq y \leq A$, with a joint density function $p(x, y) = 1/A^2$ on the square $x \leq A$, $y \leq A$. It is then easy to see that for values of p for which both the equal-gain and selection regions fit inside this square (*i.e.*, for $\sqrt{2p} < A$, or $p < A^2/2$), their distribution functions are identical. That is, with respect to the smaller values of p, the equal-gain and selection systems yield identical performance. Next, suppose the independent amplitude ratios are exponentially distributed, say e^{-x} and e^{-y}, so that their joint density function is $e^{-(x+y)}$. Noting that the contours of constant height of this density function are the lines $x+y=$constant, parallel to the boundary of the equal-gain region, it is easy to see that the integral of this density function over the equal-gain region is strictly larger than it is over the selection region. This can also be verified by direct computation, as the relevant distribution functions are easily evaluated. Hence, for exponential amplitude fading, the local reliability gain of dual equal-gain diversity is, for any percentage, strictly less than it is for dual selection diversity.

It is thus seen that the relative performance of selection diversity and equal-gain diversity depends to some extent on the fading distribution involved. Consequently, the application of equal-gain diversity should be viewed with a modicum of caution in cases where very disperse fading distributions might be encountered. However, the exponential distribution used above is probably extreme in this respect,[22] and even for this case, the equal-gain system is not significantly poorer than selection diversity. For high reliability percentages, the local reliability gain of the dual maximal-ratio system over either the selection system or the equal-gain system is exactly (approximately) $10 \log_{10} (4/\pi) = 1.05$ db for rectangular (exponential) fading.

It was noted above that the maximal-ratio data of Fig. 8 were independent of the fading distribution. However, the mean power ratios of equal-gain systems do depend on the distribution, but only to the extent of the parameter $r^2 = (\bar{x})^2/\overline{x^2}$ of (44). For the rectangular and exponential distributions considered above, $r^2 = \frac{3}{4}$ and $r^2 = \frac{1}{2}$, respectively, indicating that the average local power ratio of an equal-gain system is not substantially degraded by even very disperse fading distributions. Unfortunately, no such simple and clear dependence of the selection diversity mean values on the form of the distribution exists. The result $\bar{p} = \sum_{k=1}^{N} (1/k)$ of (18) is intimately wrapped up with the Rayleigh distribution, not merely the first two moments. But it is certainly clear that moderate changes in the form of the fading distribution could not lead to substantial changes in the selection diversity values of Fig. 8.

VIII. Relative Effects of Correlated Fading

Two smoothly varying random variables such as the x_j cannot, in general, be strictly independent. Of course, they may fail to be even approximately independent. It is therefore of interest to estimate the effect of dependent fading.

It is convenient to estimate this in terms of a parameter called the correlation coefficient. For two random variables x and y with positive variances[2] σ_x^2 and σ_y^2, this is defined by

$$\rho_{xy} = \frac{\langle (x - \bar{x})(y - \bar{y}) \rangle}{\sigma_x \sigma_y}, \tag{45}$$

which reduces readily to

$$\rho_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y} \cdot \tag{46}$$

If x and y are independent, then $\overline{xy} = \bar{x}\bar{y}$.[2] Hence, if x and y are independent, then $\rho_{xy} = 0$. It is known[23] that $-1 \leq \rho_{xy} \leq 1$, and $\rho_{xy} = \pm 1$ if and only if $y = \pm ax + b$ $(a > 0)$. x and y are said to be correlated if $\rho \neq 0$, uncorrelated if $\rho = 0$, and partially correlated if $0 < |\rho| < 1$.

The problem of correlated fading in selection diversity systems has been studied by Staras [7] and others [3], [11], [13]. (See Appendix V for certain questions related to this subject.) Packard [14] and Bolgiano, *et al.*,[8] have studied this problem for two-channel maximal-ratio systems. Quite recently, Pierce[24] and Stein[25] independently studied correlated fading in N-channel maximal-ratio systems, and their results will be published in the near future. Some of Staras' results will simply be reproduced here in Fig. 16, in a form suitable for direct comparison with Fig. 9. The curve $\rho = 1$ is the Rayleigh fading curve of Fig. 9, while $\rho = 0$ denotes the dual selection diversity curve of that figure. It can be seen that approximately half of the uncorrelated local reliability gain is realized even for $\rho = 0.8$, and that the effect is negligible for $0 < \rho < 0.3$.

To consider the relative effect of correlated fading on

[22] So far as conventional applications are concerned. It should be noted that it is not extreme, or even sufficient, for postdetection distributions in FM systems, or special applications, such as that of Price and Green, *op. cit.*

[23] Cramér, *op. cit.*, p. 265, or other standard sources. It is also known that the vanishing of ρ_{xy} does not necessarily imply that x and y are independent.

[24] J. N. Pierce, Air Force Cambridge Res. Center, Bedford, Mass., private communication.

[25] S. Stein, Hycon Eastern, Inc., Cambridge, Mass., private communication.

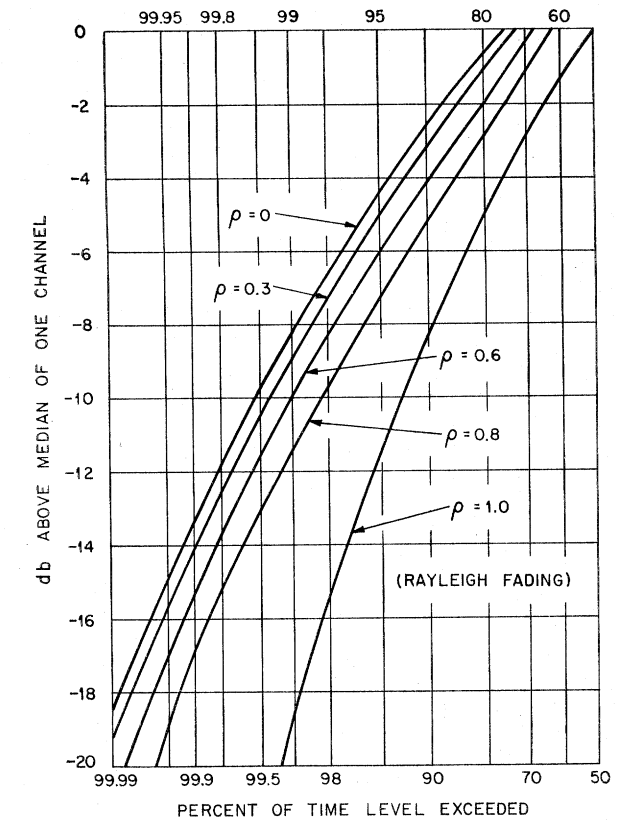

Fig. 16—Dual selection diversity distributions, Rayleigh fading, for various degrees of correlation.

the other systems, refer once again to Fig. 14. Now, in terms of the joint density function, there are two major effects of correlation: first, the mass of the density function tends to concentrate around the diagonal line $y=x$; second, the mass tends to be pulled back nearer the origin. The first effect is simply an expression of the fact that as the correlation increases, the probability that y can differ appreciably from x necessarily decreases. (Variables with the same distribution are being considered here.) The second fact can be inferred from the behavior of Fig. 16. Given these facts, it is not difficult to see from Fig. 14 that appreciable correlation will, if anything, tend to improve the performance of equal-gain diversity, relative to selection diversity. (Of course, all three systems degrade in an absolute sense with increasing correlation.) Indeed, as ρ approaches 1, when the density function approaches zero except on the line $y=x$, it is clear that the equal-gain system approaches the maximal-ratio system in performance. This can also be seen by considering the basic operation of the two systems. From these considerations, it is not difficult to visualize the way in which the maximal-ratio and equal-gain curves of Fig. 9 follow the selection diversity curves of Fig. 16. At $\rho=1$, the dual maximal-ratio and equal-gain curves coincide and are uniformly 3 db above the selection curve for $\rho=1$.

With respect to the average values of Fig. 8, the maximal-ratio data are unaltered by correlated fading. The equal-gain values actually increase toward the maximal-ratio values with increasing correlation. This can be seen either from physical considerations, or by noting that the terms $\overline{x_i}\,\overline{x_j}$ of (42) are replaced by $\overline{x_i x_j}$; for $\rho>0$, $\overline{x_i x_j}>\overline{x_i}\,\overline{x_j}$, and in fact $\overline{x_i x_j}$ approaches 1 as ρ_{ij} approaches 1. In contrast, the selection diversity values of Fig. 8 approach zero as the ρ_{ij} approach 1, as is easily seen.

In space-diversity communication systems, an antenna separation of 30 to 50 wavelengths is typically required to obtain correlation coefficients consistently less than 0.3. However, 10 to 15 wavelengths will often yield coefficients less than 0.6. Van Wambeck and Ross[26] measured the performance of certain HF selection diversity systems directly, without measuring correlation coefficients, and apparently found that even shorter spacings led to useful results. More recently, Grisdale, *et al.*,[5] have obtained numerous data bearing on this question in the 6- to 18-mc region.

IX. Variable Local Noise Powers

Many of the data above were obtained on the assumption $\overline{n_j^2}=$constant. This will not be usually strictly true and in many cases will not even be approximately true. If the noises are principally due to interference from remote sources, the $\sqrt{\langle n_j^2\rangle}$ themselves may well follow the Rayleigh distribution, a case that has recently been studied by Bond and Meyer [12] for dual selection diversity. Related material has also been given by Clarke and Cohn.[27] If the n_j are principally due to receiver front end noise, then the $\overline{n_j^2}$ may be approximately constant; the actual amount of fluctuation to be expected is a function of the noise bandwidth and the duration T of the local averages. This fluctuation has been studied by Rice,[28] whose results are quite useful in determining a suitable value of T.

In terms of the analysis above, the principal effect of variable $\overline{n_j^2}$ is to modify the distribution of the $p_j=x_j^2/\overline{n_j^2}$ with results as discussed in Section VII above. (The distribution of the p_j becomes more disperse as the noise power fluctuations increase.) It is not difficult to obtain quantitative estimates of the degradation in particular cases. It should be pointed out that extreme fluctuations in noise power can lead to very poor performance of an equal-gain system, which has no provision for cutting off a very noisy channel, in contrast to the maximal-ratio and selection systems.

[26] S. H. Van Wambeck and A. H. Ross, "Performance of diversity receiving systems," Proc. IRE, vol. 39, pp. 256–264; March, 1951.
[27] K. K. Clark and J. Cohn, "Carrier-to-noise statistics for various carrier and interference characteristics," Proc. IRE, vol. 46, pp. 889–895; May, 1958.
[28] S. O. Rice, "Filtered thermal noise—fluctuation of energy as a function of interval length," *J. Acoust. Soc. Amer.*, vol. 14, pp. 216–227; April, 1943. For additional results, see also, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 46–156; January, 1945. (Section 3.9, p. 87 ff.) These results have recently been extended by D. Slepian, "Fluctuations of random noise power," *Bell Sys. Tech. J.*, vol. 37, pp. 163–184; January, 1958.

X. Failure of the Noises to Be Locally Incoherent

The failure of assumption (C) would have no effect on selection diversity systems, for which the assumption $\overline{n_i n_j}=0$ if $i \neq j$ has no relevance whatever. However, this assumption is of vital importance for the maximal-ratio and equal-gain systems, as will be seen.

There are essentially two ways in which $\overline{n_i n_j}$ may fail to be identically zero, the first of which is simply due to the fact that the average $\overline{n_i n_j}$ is over a short interval of duration T and the local average $\overline{n_i n_j}$ will fluctuate about zero if the noises are basically unrelated. The amount of fluctuation will decrease as T is increased and will be small if the lowest frequency of the noise is large in comparison to $1/T$. In this case, the $\overline{n_i n_j}$ terms will be negative as often as positive and will simply contribute a small perturbation to the output noise power $\overline{n^2}$ of a maximal-ratio or equal-gain system. This case is not troublesome.

The troublesome case arises when the noises have a definite positive correlation, as can happen, for example, in a postdetection combiner when the noises stem largely from sources of external interference. To consider this, let $\overline{n_j^2}=1$ and $\overline{n_j}=0$; then $\rho_{ij}=\overline{n_i n_j}$ is the correlation of n_i and n_j. (Note that the local correlation over intervals of length T, in contrast to the correlation over length T_1 of the x_j discussed in Section VIII above, is considered here.) Let $\rho_{ij}=\rho$ if $i \neq j$. Then the output local noise power of an equal-gain system becomes

$$\begin{aligned}\overline{n^2} &= \left\langle \left[\sum_{j=1}^{N} n_j \right]^2 \right\rangle \\ &= \sum_{j=1}^{N} \overline{n_j^2} + \sum_{i \neq j} \overline{n_i n_j} \\ &= N[1 + (N-1)\rho], \end{aligned} \tag{47}$$

instead of $\overline{n^2}=N$. Hence, the local noise power is increased by the factor $[1+(N-1)\rho]$, which is to say the output power ratio p of (42) is decreased by this factor, which may not be trivial. To see how untrivial it can be, consider $\bar{p}$ for an equal-gain system. Eqs. (45) and (46) become

$$\begin{aligned}\bar{p} &= \frac{\sum \overline{x_j^2} + \sum_{i \neq j} \overline{x_i x_j}}{N[1 + (N-1)\rho]} \\ &= \frac{N + N(N-1)r^2}{N[1 + (N-1)\rho]} \\ &= \frac{1 + (N-1)r^2}{1 + (N-1)\rho}, \end{aligned} \tag{48}$$

which reduces to (46) when $\rho=0$. Hence , if $\rho > r^2$—a situation by no means impossible—it would follow that $\bar{p}(1) > \bar{p}(2)$; *i.e.*, the average local power ratio of a two-channel equal-gain system would be less than for a single channel, and the performance gets worse as the number of channels is increased. It is probably gratuitous to point out explicitly that, in such a case, it would be much better to use a selection diversity system, for which (18) would still hold. Similar considerations show that the average local noise power of a maximal-ratio system—by which is meant one for which the coefficients are given by (14), though this is no longer "maximal"—is increased by the factor $[1+(N-1)\rho r^2]$.

It follows that the use of maximal-ratio or equal-gain diversity in circumstances where the noise voltages may be highly correlated is hazardous.

XI. Predetection vs Postdetection Combining

In systems where the power ratio at the output of the detector is essentially the same as at the input, there is no fundamental change required in the conclusions developed above. Of course, there are always practical differences between predetection and postdetection combining; *e.g.*, a predetection maximal-ratio or equal-gain combiner will require the addition of phase-control circuitry in order to satisfy the local-coherence assumption (B). On the other hand, predetection selection diversity will sometimes produce smaller switching transients than postdetection selection. Once phase control is established, it is easier to satisfy the conditions (B) and (C) required for maximal-ratio and equal-gain combiners in the case of a predetection system.

However, substantial changes are required in the case of FM systems with a large deviation ratio, or in other bandwidth-exchange systems with a pronounced threshold effect. In such systems, an SNR at the detector input that is more than a few db above threshold yields a large output ratio, while an input ratio that is more than a few db below threshold yields a very small output ratio. That is, the output ratio changes from "completely useful" to "completely useless" with a few db change of input ratio. This fact has important consequences.

To begin with, a Rayleigh distribution of input signal strength for an FM system will emphatically not lead to a Rayleigh distribution of the postdetection amplitude ratio. Hence, the distribution-sensitive results of Figs. 8–13 and Tables I and II are not realistic for postdetection combining in FM systems. Furthermore, equal-gain combiners are not even suitable for postdetection combining in conventional FM systems; this may be regarded as a consequence of the fact that the detection gain of such systems is not constant. An alternative point of view would be that the distribution of the amplitude ratios at the input of the combiner would be such as to eliminate the equal-gain combiner from consideration; cf. (36), and note the unfortunate effect if any one of the $\overline{n_j^2}$ becomes large.

Of course, a maximal-ratio system can be used for postdetection combining in an FM system. The requirement $a_j = x_j/\overline{n_j^2}$ for the coefficients insures that any channel with large $\overline{n_j^2}$ contributes very little to the output. However, a maximal-ratio system will not yield much improvement over selection diversity in such circumstances. It will eliminate switching tra[nsients] otherwise will not usually make a significant difference in the operation of the system. This can be seen on the basis of various qualitative considerations. When dealing with postdetection combination in sharp-threshold FM systems, at least for $N \leq 8$, it would probably be best to use only the selection diversity values of Table II, whether selection or maximal-ratio diversity is actually used. In any event, the actual local reliability gains of such maximal-ratio systems—which could be computed from specific detector characteristics, such as those given by Middleton[29] or obtained by measurement—would certainly be less than the maximal-ratio values in Table II. A specific distribution computed on the basis of a highly simplified detector characteristic has been given. [16]

If the local reliability gain is defined in terms of the transmitter power required to maintain the input level of the detector above a certain value for more than a specified percentage of time, then the selection diversity values of Table II are applicable whether the selection is predetection or postdetection. It is clear that the operation is identical in either case. Furthermore, the maximal-ratio and equal-gain data of Table II are completely applicable to predetection combining, as is easily seen. Accordingly, the full advantages of maximal-ratio and equal-gain combiners can be realized in FM systems when and only when they are employed before detection. Taking the selection values of Table II as being the gains obtained by a postdetection maximal-ratio combiner, the differences between the maximal-ratio and selection values of Table II then illustrate the added advantage of predetection maximal-ratio or equal-gain combining. This may be regarded as due to an effective lowering of the detector threshold resulting from these techniques.

An additional advantage of predetection combining in FM systems is that FM multipath distortion can be reduced by this method. It has been shown by Adams and Mindes [16], both theoretically and experimentally, that a predetection equal-gain combiner yields substantially less multipath distortion than is obtained with a postdetection maximal-ratio combiner, when both are operated under the same circumstances.

Instrumentation for postdetection maximal-ratio combining has been discussed by Kahn [6] and by Morrow, *et al.*,[11] for what amount to AM systems, and by Mack[30] for FM systems. An ingenious predetection maximal-ratio combiner has been devised by Granlund.[31] A particularly elegant predetection equal-gain combiner has been developed by the Federal Telecommunication Laboratories (now the ITT Laboratories), indicated in Fig. 17. This combiner, called simply a phase combiner in FTL literature, is the same one used in the experiments reported by Adams and Mindes [16]. The phase control and adder circuits require only two semiconductor diodes and 16 passive linear elements. Phase control is established via a phase discriminator, the output of which is applied as a bias voltage to one of the local oscillators. This corrects the phase of the local oscillator via Miller-effect changes in the oscillator tube capacity.

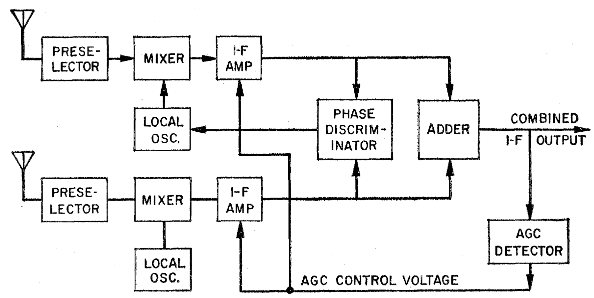

Fig. 17—FTL predetection equal-gain combiner. This can be used with any type of modulation.

The problem of adequate phase control for predetection maximal-ratio or equal-gain combining leads naturally to the next topic, namely

XII. Failure of the Local-Coherence Condition (B)

It is obviously of interest to estimate the possible degradation in performance of maximal-ratio and equal-gain combiners when the local-coherence condition (**B**) is not satisfied. The following treatment is due to Stein.[32]

Recall that (**B**) was the assumption $s_j(t) = x_j m(t)$ where x_j was the slowly varying local rms value of s_j. If the s_j are not all in phase, we must write $s_j(t) = x_j m_j(t)$, where the m_j have different phases. Consider the case $m_j(t) = \sqrt{2} \cos(\omega t - \phi_j)$ where the ϕ_j are locally constant in the sense that the x_j are. Then $\langle m_j^2 \rangle = 1$ and $\langle s_j^2 \rangle = x_j^2$, as before, averaging over a few cycles (or more) of ωt. Then, for any locally linear combiner of the type (6),

$$
\begin{aligned}
\overline{s^2} &= \left\langle \left[\sum_{j=1}^{N} a_j s_j \right]^2 \right\rangle \\
&= \Big\langle \sum a_j^2 x_j^2 2 \cos^2(\omega t - \phi_j) \\
&\qquad + \sum_{i \neq j} a_i a_j x_i x_j \cdot 2 \cos(\omega t - \phi_i) \cos(\omega t - \phi_j) \Big\rangle \\
&= \sum a_j^2 x_j^2 + \sum_{i \neq j} a_i a_j x_i x_j \langle 2 \cos(\omega t - \phi_i) \cos(\omega t - \phi_j) \rangle \\
&= \sum a_j^2 x_j^2 + \sum_{i \neq j} a_i a_j x_i x_j \cos(\phi_i - \phi_j), \qquad (49)
\end{aligned}
$$

where the last step used $2 \cos A \cos B = \cos(A+B)$

[29] D. Middleton, "On theoretical signal-to-noise ratios in FM receivers: a comparison with amplitude modulation," *J. Appl. Phys.*, vol. 20, pp. 334–351; April, 1949.

[30] C. L. Mack, "Diversity reception in UHF long-range communications," Proc. IRE, vol. 43, pp. 1281–1289; October, 1955.

[31] J. Granlund, "Topics in the Design of Antennas for Scatter," Lincoln Lab., M.I.T., Lexington, Mass., Tech. Rep. No. 135, pp. 105–113; November, 1956. See also recent Quarterly Progress Reports of the M.I.T. Res. Lab. of Electronics.

[32] S. Stein, private communication; August, 1957.

$+\cos(A-B)$ and the fact that $\langle\cos(2\omega t-\phi_i-\phi_j)\rangle=0$. Eq. (49) may also be written

$$\overline{s^2} = \sum_{i,j} a_i a_j x_i x_j \cos(\phi_i - \phi_j) \tag{50}$$

since $\cos(\phi_j-\phi_j)=1$, and this reduces to

$$\overline{s^2} = [\sum a_j x_j]^2 = \sum_{i,j} a_i a_j x_i x_j \tag{23'}$$

when $(\phi_i-\phi_j)=0$. Let p denote the output power ratio of the general combiner (6) when (23′) holds, and p' denotes the same for the phase-degraded case (50). Then (assuming (**C**) still holds, so that $\overline{n^2}=\sum a_j^2\overline{n_j^2}$)

$$\begin{aligned} p' &= \frac{\sum_{i,j} a_i a_j x_i x_j \cos(\phi_i - \phi_j)}{\sum a_j^2 \overline{n_j^2}} \\ &= \frac{\sum_{i,j} a_i a_j x_i x_j \cos(\phi_i - \phi_j)}{[\sum a_j x_j]^2} \frac{[\sum a_j x_j]^2}{\sum a_j^2 \overline{n_j^2}} \\ &= kp, \end{aligned} \tag{51}$$

where

$$k = \frac{\sum_{i,j} a_i a_j x_i x_j \cos(\phi_i - \phi_j)}{[\sum a_j x_j]^2} \tag{52}$$

is the "phase degradation ratio" p'/p.

Apart from the fact that $0\leq k\leq 1$, not much can be said about k in the general case (52) in the absence of additional information about the ϕ_j. It is easy to see that k may actually vanish; *e.g.*, $N=2$, $a_1=a_2=x_1=x_2=1$, $\phi_1-\phi_2=180°$. Then $p'=k=0$, which is entirely to be expected when adding two signals of equal magnitudes and opposite phases. This illustrates the fact that the condition (**B**) cannot be ignored. On the other hand, it is not necessary that it should be satisfied with great precision. Suppose that the magnitudes of the phase differences, $|\phi_i-\phi_j|$, do not exceed 90°, and let $\Delta=$ maximum of $|\phi_i-\phi_j|$, $i, j=1, 2, \cdots, N$; $0\leq\Delta\leq 90°$. Then $0\leq\cos\Delta\leq\cos(\phi_i-\phi_j)$, so

$$k \geq \frac{\sum_{i,j} a_i a_j x_i x_j \cos\Delta}{[\sum a_j x_j]^2} = \cos\Delta, \tag{53}$$

or

$$p' \geq p\cos\Delta. \tag{54}$$

That is, the local power ratio is not reduced by more than $\cos\Delta$, or $-10\log_{10}\cos\Delta$ db, in any combiner whatever of the general type (6), provided $\Delta\leq 90°$. In particular, this conclusion holds for equal-gain and maximal-ratio combiners. Thus, to restrict the reduction in p due to imperfect phase control to 1 db or less, it is only necessary to maintain the phases within 37.5° of each other, while 51° is sufficient to guarantee a maximum loss of 2 db. Furthermore, it is clear that the estimate $p\cos\Delta$ is actually conservative.

XIII. Long-Term Variability

Recall that the distributions of the x_j and p, and mean values of these quantities, were to be determined in intervals of length T_1, relative to which the x_j were assumed to be approximately Rayleigh-distributed and approximately independent. It is important to understand the nature of this situation.

It is an experimental fact that, for a suitable choice of T_1, both of these assumptions are often satisfied. It is also an experimental fact that if T_1 is made too long or too short, neither assumption is satisfied. Hence, the approach used above and all of the results developed above depend entirely on the use of finite intervals of observation that are neither too long nor too short.

Specific suitable values of T_1 depend on the circumstances, primarily the carrier frequency and transmission distance. Of course, it is necessary to understand that the results of Figs. 8–13, etc., refer only to intervals of length T_1, whatever this may be. Specific values are roughly as follows for long-range transmission. At frequencies below VHF, intervals of 30 minutes to an hour are usually suitable. In VHF ionospheric scatter systems, values of one or two minutes usually seem to be appropriate; in UHF and SHF tropospheric systems, intervals of five to 30 minutes are often used.

It is manifestly necessary to consider the behavior of diversity systems over longer intervals than those of length T_1. This may be done as follows. The previous results were obtained using a Rayleigh distribution (9) of unit mean square, with a definite median value $M_0=\sqrt{\log_e 2}$. Now, the experimental fact is that the fading distributions observed over different intervals of length T_1 will not usually have the same median values. However, the medians obtained in two adjacent or overlapping intervals of length T_1 will not usually differ by very much. One way to represent this fact is to let $M=M(t)$ denote the median of the distribution obtained in the interval from $t-T_1$ to the present time t. Then this median function is a continuous function of time and the experimental fact is that $M(t)$ is usually approximately constant over intervals of length T_1. It should be clear that this does not depend on having the distributions, for which the values of $M(t)$ are the medians, all be of the same form, Rayleigh or other.

This may be used as follows. If x is any random variable with a nonzero median M_0 and M is any nonzero constant, then it is easy to see that $y=(M/M_0)x$ has a distribution of the same form as that of x, differing only in the scale factor M/M_0, and that the median value of y is M. (The only reason for writing this scale factor in terms of the medians is that these are easily determined experimental quantities.) Then, instead of taking the local rms signals to be the x_j with a fixed median M_0, the actual local rms signals may be written as $y_j=(M_j/M_0)x_j$, with median values M_j, relative to a period of length T_1. Here, however, another experimental fact enters. The medians M_j are usually approximately the same for different channels, and we may write $M_j = M$, $j = 1, 2, \cdots, N$. If the median function $M(t)$, above, is approximately constant over intervals of length T_1, we may take $M = M(t)$.

To apply this to our previous results, note that the local linearity of (6) implies that any common scale factor multiplying the signal components s_j may be taken outside the sum $s = \sum s_j$. Hence, the combined output signal $s(t)$ is simply multiplied by M/M_0 and the local power ratio becomes $(M/M_0)^2 p$ wherever p was before. This becomes even simpler when expressed in decibels. Let

$$\begin{aligned} w &= 10 \log_{10} \left[(M/M_0)^2 p\right] \\ &= 20 \log_{10} M + 10 \log_{10} (p/M_0^2) \\ &= u + v \end{aligned} \tag{55}$$

be the local power ratio in db, where $u = 20 \log_{10} M$ and $v = 10 \log_{10} (p/M_0^2)$. Then this expresses the actual local power ratio delivered by any combiner of the type (6) as the sum of a variable v whose T_1 median does not depend on time and a variable u that is approximately constant over every interval of length T_1. Now, the distributions plotted in Figs. 9–13 are precisely the distributions of the variable v for the conditions of Fig. 8, for different combiners and orders of diversity. Hence, to apply the results of Figs. 9–13 to any particular interval of length T_1, it is only necessary to translate their ordinate scales by $u = 20 \log_{10} M$ where M is the median of the single-channel fading distribution for the interval concerned.

In order to describe the long-term variability of the actual local power ratio w, it is necessary to have information about the long-term variability of the (T_1-) median u. Distributions of u are usually studied in intervals of length T_2 = one month to one year; such distributions are often called distributions of hourly medians, though they should properly be distributions of T_1-medians. Several such distributions for frequencies at VHF and above have been given.[33] Unfortunately, no comparable single source of information for MF and HF systems presently exists; the relevant data are largely scattered in (generally unobtainable) Signal Corps reports, FCC hearing transcripts, and URSI and CCIR documents, though a few such data have been published. Observed distributions of u are sometimes approximately Gaussian (normal) in form, especially at the higher frequencies, which is why the distributions of M are often said to be log-normal.

Once a distribution of u pertinent to the proposed circuit is available, there are two ways it may be used. The first method, which is applicable to high-reliability systems at VHF and above, is to estimate the lowest T_1-median likely to be encountered on the circuit. (The great virtue of these systems is that this minimum value of u is not $-\infty$.) The ordinate scales of Figs. 9–13 are then translated to this value, after which a rational choice among the various possibilities of transmitter power, order and type of diversity system, etc., may be made on the basis of economic and other factors, and on the basis of the local reliability percentage it is desired to maintain during such worst hours. Of course, the data of Figs. 9–13 must be modified in accordance with the discussion of Sections VII to XII if the circumstances so dictate.

The second method is applicable in circumstances where all distributions of v in all intervals of length T_1 are approximately the same.[34] It is then easy to see that the distribution of v in an interval of length T_2 = one year would also be the same; furthermore, the variables u and v of (55) would then be independent, relative to T_2, to a very high degree of accuracy. Then the T_2-distribution of the local power ratio w would be the distribution of the sum of two independent variables with known distributions, and could be computed. It would usually be found that the "exact" determination of the T_2-distribution of w would require numerical methods of integration in (67). Such distributions have been computed by Shimony[35] and Sichak[36] among others. However, Staras [9] has observed that, for the larger values of N, the relevant distributions of v are approximately normal (cf. the $N = 6$ and $N = 8$ curves of Figs. 12 and 13, on which figures a normal distribution would be a straight line) and since the distribution of u is approximately normal, the T_2-distribution of w would therefore[2] approximate a normal distribution with a median and variance respectively equal to the sums of the medians and variances of the u and v distributions. But this approximation is not very accurate for $N \leq 4$.

It should be added, however, that this second method has not been universally accepted by designers of high-reliability systems, for the following reasons. Computing the long-term distribution of w serves to obscure the question of whether the periods of very low signal are a few long periods or many short ones. This question can be important; *e.g.*, there are systems in operation in which the loss of two hours in a year would not be troublesome if split into a number of separated intervals of a minute or two each, but which could be disastrous if concentrated in a single interval of two hours. Since two hours in a year corresponds to the 99.98 per cent exceeded level, it would be necessary to compute the T_2-distribution of w down to something like the 99.999 per cent exceeded level in order to insure approximately the reliability obtained by the first method. However, the empirical distributions on which this computation must ultimately rest are not known to anything approaching this degree of accuracy, and the validity of such a computation would seem to be open to question. In addition, the problem mentioned in footnote 34

[33] Scatter Propagation Issue, Proc. IRE, vol. 43; October, 1955.

[34] It should be noted that such circumstances, however, are not too common in practice.

[35] A. Shimony, Final Rep. MU-156, Evans Signal Lab., Fort Monmouth, N. J., (cited in footnote 36).

[36] W. Sichak, Fed. Telecommun. Labs., Nutley, N. J., Tech. Memo. 619; December, 1956. See Appendix D, "Diversity Theory."

would often infect such a computation.

Two additional points should also be noted in connection with long-term distributions of w. The first is that the variability or dispersion of w will increase as the dispersion of the T_1-median u increases. In other words, the variability of v will tend to be obscured by that of u. However, it is precisely the variability of v that can be reduced by diversity techniques, while that of u cannot. It has been noted[36] that w distributions for different orders of diversity show less difference than would be indicated by Table II, but this is simply a reflection of the dispersion contribution by u. Thus, computing long-term distributions of w tends to obscure the gains (Table II) that actually *are* realized by diversity techniques.

The second point to be noted is that any long-term distribution of w is highly specific to the circuit for which it was computed, because u distributions are highly specific. This is indicated in Fig. 18,[37] which shows percentage points of u distributions as a function of distance at 400 mc. (Note that Fig. 18 gives only the distributions of the T_1 median "scatter" loss; the free-space loss has been subtracted out.) It can be seen that the dispersion of u decreases with distance; *e.g.*, the interdecile range is about 12.5 db at 200 miles, but only 5 db at 600 miles. This indicates that a long-term distribution of w would only be valid for the distance on which the distribution of u was based.

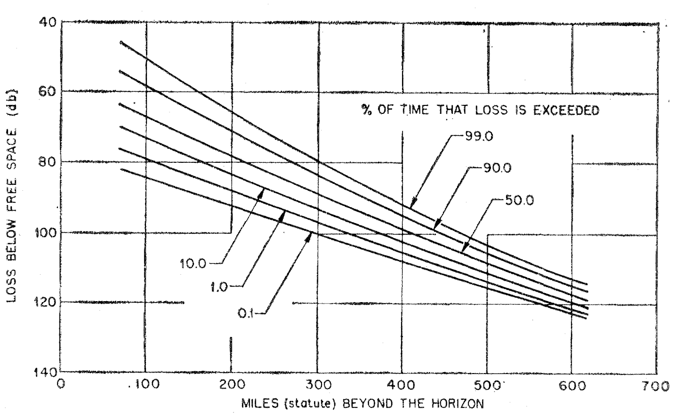

Fig. 18—Distributions of hourly medians as a function of distance, wintertime propagation at 400 mc, experimental data from several Lincoln Laboratory circuits. (After Morrow[37].)

XIV. Case of Unequal Median Signals

Most of the material above presupposed that the T_1-medians for the several channels were all the same. Experimentally, this is found to be a reasonable approximation in most cases, provided that the interval length T_1 is not made too short. However, there are cases, most especially angle diversity,[8–10] in which it is not a reasonable approximation. The first treatment of this problem apparently was given in [3]. It is quite simple to plot selection diversity distributions for unequal medians or even for dissimilar distributions; the identical factors $G(p)$ of (16) are simply replaced by the proper distribution functions. Maximal-ratio distributions for unequal-median Rayleigh signals can also be expressed explicitly. Unfortunately, this is not true of complete equal-gain distributions (*cf.* Appendix IV). However, it would be possible to obtain the low-signal ends of such distributions by taking the first few terms of a power series, but a detailed analysis of this problem would appear to be premature at the present time.

XV. Experimental Results

Experimental data relating to diversity systems have been given by several workers, including Glaser and Van Wambeck,[38] Van Wambeck and Ross,[26] Glaser and Faber,[4] and Grisdale, *et al.*[5] It is unfortunate for our present purposes that most of these data relate to selection diversity only; clear-cut and unambiguous experimental data bearing on the comparative performance of the three combining techniques studied above are so rare to be as essentially nonexistent. It is hoped that some comparative experimental results will be available within the next year.

Perhaps the best single datum presently available was obtained in unpublished experiments conducted by the Signal Corps a few years ago. Two high-frequency systems were compared, one of which used dual maximal-ratio diversity and the other used something approximating selection diversity. It appeared that the maximal-ratio system yielded an average power ratio of 1.0 to 1.5 db above the selection system when averaged over periods of about 30 minutes.[39] This compares very favorably with the value 1.25 db entered in the "selection" column of Table I for $N=2$. However, there were many periods during which the performance of the maximal-ratio system was inferior to the other. This could probably be traced to the failure of one or both of the conditions (**B**) and (**C**) during such periods, which would not affect a selection system.

XVI. Conclusions

Perhaps the most important conclusion to be drawn is that, all things considered, no one of the diversity combining techniques studied deserves to be called "the optimum system." All three have their merits and defects, and the one to be used will depend on the circumstances. However, the simplicity and efficacy of the equal-gain system suggest that this may well become the principal standard of the art. In addition to the data set forth above, it should be especially noted that the instrumentation for an equal-gain combiner is completely independent of what one chooses to think of as a SNR. The importance of this fact is considerable.

[37] W. E. Morrow, Jr., "Etude de systemes de radiocommunication troposphérique UHF a longue distance," *Onde Elect.*, vol. 37, pp. 444–449; May, 1957.

[38] J. L. Glaser and S. H. Van Wambeck, "Experimental evaluation of diversity receiving systems," Proc. IRE, vol. 39, pp. 252–255; March, 1951.

[39] F. E. Bond and H. F. Meyer, Signal Corps Eng. Labs., Fort Monmouth, N. J., private communication; June, 1957.

Appendix I

Certain Facts About Probability Theory

It will be recalled[2] that, in general, a distribution function $P(x)$ is the probability that (some random variable) is less than or equal to x. A particularly simple special case of this arises when the random variable in question is some voltage or current waveform given as a function of time, say $f(t)$, and the probability that $f \leq x$ is simply the fraction of some interval $t_1 - T \leq t \leq t_1$ in which $f \leq x$. In this case, $P(x)$ is determined, for any given value of x, simply by adding up the lengths of the t intervals for which $f(t) \leq x$ and dividing their sum by the total duration of the observation, as indicated in Fig. 19. Instruments for measuring $P(x)$ at selected values of x are known as "totalizers" or "level distribution recorders" and exist in various forms. Another method of obtaining the distribution function of some random variable is to sample it at discrete intervals and count the fraction of sampled values that are $\leq x$; however, it is not difficult to see that this will lead to the same $P(x)$ as defined above. The associated density function[2] $p(x) = dP(x)/dx$, so that $p(x)dx = dP(x)$.

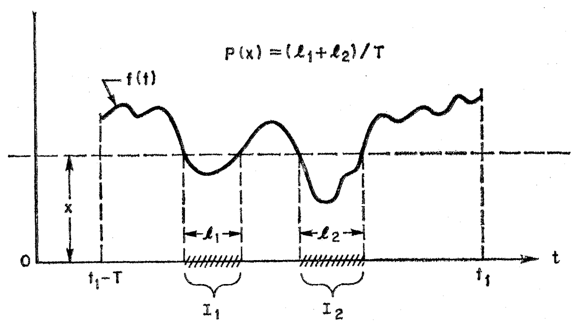

Fig. 19—Definition of $P(x)$ = fraction of the time of observation that $f(t) \leq x$.

The central purpose in using such distribution functions in radio engineering stems from the fact that many different individual waveforms have approximately the same distribution function, or at least have distribution functions that differ in describable ways, as in Section XIII. This is an experimental fact, no more, but, what is important, no less. Thus, in circumstances where the distribution function of some waveform can be approximately predicted from either theoretical or empirical grounds, one has available a method of predicting many important facts about the situation.

One such fact, of the first importance, is that all time averages of f in the interval $t_1 - T \leq t \leq t_1$ are given by the moments[2] of the distribution function $P(x)$. For example, suppose one is interested in the average value of $f(t)$. Then

$$\frac{1}{T}\int_{t_1-T}^{t_1} f(t)dt = \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\infty} xdP(x); \qquad (56)$$

i.e., $\bar{f}$ is the first moment of the distribution. More generally, for any value of n, the time average of $[f(t)]^n$ is given by the nth moment of the distribution:

$$\frac{1}{T}\int_{t_1-T}^{t_1} [f(t)]^n dt = \int_{-\infty}^{\infty} x^n dP(x), \qquad (57)$$

a result that is especially useful in computing the average power when f is a voltage or current and $n = 2$. In the light of (57), $\overline{f^n}$ or $\langle f^n \rangle$ or $\overline{x^n}$ or $\langle x^n \rangle$ can be and is written interchangeably for such averages, using whichever notation seems most convenient for the expression involved.

It is important to understand the sense in which (57) is applicable. No theory or representation or mathematical model whatever can predict the particular distribution function of a particular waveform in a particular interval exactly, but to the extent it can be predicted by whatever means, (57) is applicable. In many applications, one uses a specific mathematical model distribution (*e.g.*, the Rayleigh distribution, much used in the body of this paper) for predicting facts such as averages of the form (57), but with a clear understanding that any realized distribution function can only approximate the Rayleigh distribution, however long or short the interval of observation. But the approximation may be very close. In cases where the nature of the possible departures from the model distribution can be estimated, and there are many such cases, the possible departures in the corresponding time averages can similarly be estimated via (57). [Some statisticians and noise theorists may be bothered by the absence from this discussion of any reference to the classical notions of sample and population. The reason for this is that a parent population in the classical sense does not usually exist in this environment. (See Section XIII.) No fixed distribution can serve as a population distribution for any non-stationary process. In the notation of Section XIII, it would be possible, but not necessarily desirable, to discuss parent populations for T_2-distributions, but certainly not T_1-distributions. One may, however, discuss a "distribution" of distributions, as is done in engineering language in that section. It would be a simple matter to provide a more formal framework for this material by defining suitable classes of functions f; *e.g.*, all those f whose half-hour local distribution functions (in the sense of Fig. 19) were all within a specified distance (in the sense of Lévy's metric) of some Rayleigh distribution function, and whose half-hour medians (*ipso facto* unique) had yearly distribution functions within a specified distance of a fixed log-normal distribution function. This class is non-empty for positive distances, and would suffice for most of the purposes of Sections II through VI. One could similarly replace our heuristic language about approximate constants with a more formal treatment that was liberally seasoned with epsilons and deltas and rigorous inequalities. However, there is probably little to be gained by this formalism in the present context.]

These considerations above extend directly to several random variables given as in Fig. 1 and their corresponding multidimensional distribution functions. Thus, all of the multidimensional probability theory given by Bennett[2] can be directly applied to our present circumstances. Of course, such distributions will generally depend on the present time t_1 and the duration T, but for suitable (not necessarily long) values of the duration T, this dependence may be considered to be negligible for certain engineering purposes.

It is quite well known that averaging is a linear operation; *i.e.*, if x and y are random variables and a and b are constants, then $\langle ax+by\rangle = a\bar{x}+b\bar{y}$. This is clear when considered as time averages and, as an immediate consequence of (56), also holds for the corresponding distribution averages. (Let P_1, P_2, and P_3, denote the distribution functions of f, g, and $f+g$, respectively, and write (56) three times. This does not require independence of f and g.)

Although this fact is well known, such extensive use of it is made in the body of the paper that it is advisable to mention a few consequences here. First, if x_1, $x_2, \cdots, x_n$ are random variables, not necessarily independent, and $a_1, a_2, \cdots, a_n$ are constants and

$$y = \sum_{k=1}^{n} a_k x_k, \tag{58}$$

then

$$\bar{y} = \sum_{k=1}^{n} a_k \overline{x_k}. \tag{59}$$

In order to consider higher moments than the first, note the simple algebraic fact that $(a_1x_1+a_2x_2)^2 = a_1^2x_1^2 + a_2^2x_2^2 + 2a_1a_2x_1x_2$ can be written $a_1^2x_1^2+a_2^2x_2^2+a_1a_2x_1x_2 + a_2a_1x_2x_1$. More generally, one can write

$$\begin{aligned} y^2 &= \left[\sum_{k=1}^{n} a_k x_k\right]^2 \\ &= \sum_{k=1}^{n} a_k^2 x_k^2 + \sum_{i\neq j} a_i a_j x_i x_j. \end{aligned} \tag{60}$$

Hence, by (59),

$$\overline{y^2} = \sum_{k=1}^{n} a_k^2 \overline{x_k^2} + \sum_{i\neq j} a_i a_j \langle x_i x_j\rangle \cdot \tag{61}$$

If the x_i are independent, then $\langle x_i x_j\rangle = \bar{x}_i \bar{x}_j$ if $i \neq j$.[40] Then (61) becomes

$$\overline{y^2} = \sum_{k=1}^{n} a_k^2 \overline{x_k^2} + \sum_{i\neq j} a_i a_j \bar{x}_i \bar{x}_j. \tag{62}$$

Hence, if the first two moments $\bar{x}_i$ and $\overline{x_i^2}$ of the x_i are known, the average square of y can be computed without even knowing the distribution of the x_i, much less the explicit distribution of y. It will be seen in several sections of this paper that these simple facts can lead to interesting and important results, some of which are by no means obvious.

Another well-known fact is that a joint density function[2] $p(x, y)$ of two random variables can be integrated over a region in the plane to obtain the probability of the region. Thus, the probability that $x_1 < x \leq x_2$, $y_1 < y \leq y_2$ is given by

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y)\,dy\,dx, \tag{63}$$

and the joint distribution function $P(x, y)$ is simply

$$P(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(s, t)\,dt\,ds. \tag{64}$$

Notice that both (63) and (64) can be written in the form

$$\iint_E p(x, y)\,dx\,dy \tag{65}$$

of an integral of $p(x, y)$ over a certain region E in the x-y plane. In the case of (63), the region is an ordinary rectangle, while in the case of (64), it is a semi-infinite rectangle. The virtue of this geometric approach is that it often enables the expression of an event of practical interest in terms of such a region, not necessarily a "rectangle," after which the probability of the event in question can be computed by integrating the joint density function over the region. Joint density functions of three or more variables can similarly be integrated over regions in space of three or more dimensions. This is used at several places in the body of the paper.

Finally, a few words on computing the distribution functions of sums of independent random variables may be useful. It was pointed out by Bennett[2] that if x and y are independent random variables with density functions p_1 and p_2, respectively, the density function p_3 of the sum $z = x+y$ is given by

$$p_3(z) = \int_{-\infty}^{\infty} p_1(z-y)p_2(y)\,dy, \tag{66}$$

which is called the "convolution" or "composition" of the density functions p_1 and p_2. (This has sometimes been referred to as "combining" the x and y distributions.) However, in many—possibly most—practical applications, the distribution function of z, *i.e.*, the probability that $z \leq u$, is of more interest than the density function. This can be expressed in terms of the component distribution functions as

$$P_3(u) = \int_{-\infty}^{\infty} P_1(u-y)\,dP_2(y), \tag{67}$$

which can be seen from (66) by writing $p_2(y)dy = dP_2(y)$ and integrating (66) on z from $-\infty$ to u. The integral (67), which is well known among mathematicians,[41] may be defined as the limit of approximating sums of the form

$$\sum_k P_1(u-y_k)\left[P_2(y_k) - P_2(y_{k-1})\right], \tag{68}$$

[40] Bennett, footnote 2, p. 619.

[41] Cramer, *op. cit.*, p. 190, (15.12.2), and other sources.

where the y_k form a suitably fine partition of the range of interest. This and other numerical integration techniques may be used to evaluate (67) numerically, at least as easily as a numerical evaluation of (66). (See Appendix IV.) Hence, if the distribution functions of two variables are given numerically and the distribution function of the sum is required, there is no need to transform the given distribution functions into approximate density functions, evaluate (66) numerically, and then sum the resulting approximate density function, a procedure unnecessarily involved, but fairly commonly used.

Appendix II

The Schwarz Inequality

There are several ways of proving the inequality

$$\left[\sum_{j=1}^{N} u_j v_j\right]^2 \leq \left[\sum_{j=1}^{N} u_j^2\right]\left[\sum_{j=1}^{N} v_j^2\right], \tag{19}$$

of which the simplest is perhaps to notice the algebraic identity

$$\left[\sum_{k=1}^{N} u_k v_k\right]^2 = \left[\sum_{k=1}^{N} u_k^2\right]\left[\sum_{k=1}^{N} v_k^2\right] - \frac{1}{2}\sum_{i,j=1}^{N}(u_i v_j - v_i u_j)^2, \tag{69}$$

which can be verified at once simply by expanding both sides. It is clear that (19) follows immediately from (69), for the term

$$\sum_{i,j=1}^{N}(u_i v_j - v_i u_j)^2$$

is obviously non-negative. However, more precise information can be extracted from this. Evidently there is equality holding in (19) if and only if every term in the double sum on the right in (69) vanishes, *i.e.*, if and only if

$$u_i v_j - v_i u_j = 0,\ i, j = 1, 2, \cdots, N, \tag{70}$$

and it is not difficult to see that this will happen if and only if there are constants a and b, not both zero, such that $au_i = bv_i$, $i = 1, 2, \cdots, N$. That is, equality holds in (19) if and only if the u's and v's are proportional. It is this fact that accounts for the "and only if" assertion following (14); the material of Section IV does not justify this assertion.

Appendix III

The Maximal-Ratio Distribution

The characteristic function[2] of the p_j is

$$\phi(t) = \int_{-\infty}^{\infty} e^{itp_j} g(p_j)\,dp_j = \int_{0}^{\infty} e^{-p_j + itp_j}\,dp_j = \frac{1}{1 - it}, \tag{71}$$

where $i = \sqrt{-1}$ is not an index. The characteristic function of p is then simply

$$\phi_N(t) = [\phi(t)]^N = \frac{1}{(1 - it)^N}, \tag{72}$$

so that the density function of

$$p = \sum_{j=1}^{N} p_j$$

is

$$g_N(p) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ipt}\phi_N(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-ipt}}{(1 - it)^N}\,dt. \tag{73}$$

The integral (73) is easily evaluated by contour integration and the residue theorem. The result is

$$g_N(p) = \frac{1}{(N - 1)!}\,p^{N-1}e^{-p} \tag{74}$$

for $p > 0$, while $g_N(p) = 0$ for $p < 0$. This is precisely the result (29).

Appendix IV

Computation of the Equal-Gain Distribution

The function $B_N(u)$ of (41) is given recursively by

$$B_N(u) = \int_0^u B_{N-1}(u - t)\,dB_1(t) = \int_0^u B_{N-1}(u - t)B_1'(t)\,dt, \tag{75}$$

as can be seen from (67), where $B_1'(t) = 2te^{-t^2}$ is the Rayleigh density function. Tables of $B_N(u)$ have been constructed[19] from (75) for $N = 2, 3, \cdots, 8$, using an IBM 704 computer. Tables were constructed for various increments Δ of u, ranging from $\Delta = 0.2$ down to $\Delta = 0.02$, and for the range $0 \leq u \leq 17$. The Rayleigh density and distribution functions were generated in the computation program by rational approximations accurate to six decimals. Each value $B_N(u_k)$, where $u_k = k\Delta$, was then computed from

$$B_N(u_k) = \sum_{l=1}^{k}\int_{(l-1)\Delta}^{l\Delta} B_{N-1}(t)B_1'(u_k - t)\,dt, \tag{76}$$

where each integral over the range of length Δ was computed by a 16-point Gaussian quadrature formula. The values of $B_{N-1}(t)$ for this integration were obtained from the previously constructed table of B_{N-1} by a modified Tchebycheff-Everett interpolation formula. Tables of $B_2(u)$ constructed by this method agreed with (39) (separately tabulated) to six decimals. For all $N \leq 8$, tables for the smaller values of Δ were consistent to four decimals for the entire range of u.

An additional check was provided. The function defined by (41) for all complex values of u is an entire function, which therefore admits a power series,

$$B_N(u) = \sum_{k=0}^{\infty} b_k^N u^k, \quad (77)$$

valid for all values of u, and which coincides with the desired distribution function for positive real values of u. It is not especially difficult to show that $b_k^N = 0$ for all odd k and for $k < 2N$. It can be shown on the basis of extensive computations from (41) that for even $k \geq 2N$,

$$b_k^N = (-1)^{k/2-N} \frac{2^{k/2}}{k!} \sum_{l=1}^{l=(k/2)+1-N} p_l \times \cdot \sum_{j_1+\cdots+j_{N-1}=k/2-l}^{j_i \geq 1} p_{j_1} p_{j_2} \cdots p_{j_{N-1}}, \quad (78)$$

where

$$p_l = (2l-1)(2l-3) \cdots 5 \cdot 3 \cdot 1. \quad (79)$$

In particular, the coefficient b_{2N}^N of the leading term is $b_{2N}^N = 2^N/(2N)!$, *i.e.*,

$$B_N(u) = \frac{2^N}{(2N)!} u^{2N} + \cdots, \quad (80)$$

but this term alone is not sufficiently accurate for useful values of u. For larger values of k, (78), which has resisted strenuous attempts at simplification, is not as useful for the explicit computation of coefficients as the recursion relation

$$b_k^N = (-1)^{k/2-N} \frac{1}{k!} \sum_{j=1}^{j=(k/2)-1} 2^j p_j (k-2j)! \left| b_{k-2j}^{N-1} \right|, \quad (81)$$

which can also be established from (41). In connection with (81), one uses $b_k^1 = (-1)^{(k/2)-1}/(k/2)!$.

It can be shown from (78) that for $|u| \leq u_0$ and for

$$k \geq 2\left(N - 1 + \frac{1}{N}\right) u_0^2 - 2, \quad (82)$$

the terms of (77) are monotonically decreasing in magnitude. Since the terms alternate in sign, this means that the error in terminating (77) at the kth term is less than the magnitude of the kth term, provided (82) is satisfied. This was used to construct a table of $B_N(u)$ for $N = 2, 3, \cdots, 8$ for $0 \leq u \leq 1.5$ with a guaranteed accuracy of six decimals. This table agreed in this range with that constructed from (76) to six decimals or better.

The results (78), (81), and (82) are principally due to Michael Ginsburg.

Appendix V

Certain Questions Related to the Problem of Correlated Fading

All of the presently published treatments of correlated fading known to the present writer rely on a result due to Uhlenbeck,[42] which was reproduced in a paper by Booker, Ratcliffe, and Shinn.[43] Uhlenbeck's result rests in turn on the joint distribution of two Rayleigh variables given by Rice.[44] However, meaningful sufficient conditions under which this distribution is applicable to correlated Rayleigh fading do not appear to be known. It is essentially certain that it is applicable to narrow-band random noise of the type originally studied by Rice and Uhlenbeck, but it is far from clear that it is equally applicable to fading radio waves in general. For example, if Uhlenbeck's result always held, then the correlation of two Rayleigh variables could not be negative; however, several investigators, including Grisdale, *et al.*,[5] and McNicol[45] have found such negative correlation. It seems very probable that Uhlenbeck's result is satisfactory as a first approximation for engineering purposes. This is why Fig. 16 was unhesitatingly included in Section VII; however, such results should be understood as representative, rather than absolute. In other words, the correlation coefficient does not uniquely determine diversity performance, even when ρ and the separate input distributions are known.

A closely related problem that sometimes arises in this connection is the assumption, which has not always been recognized as such, that two random variables that are individually Gaussian or normal have a joint distribution that is a two-dimensional normal distribution. This assumption is the basis of the common statement that "uncorrelated normal variables are independent." As a mathematical matter, this need not be true; it is not difficult to give counter-examples. The prevalence of the quoted statement probably stems in part from some insufficiently explicit language of Cramér.[46] The two-dimensional form of the central limit theorem suggests that this assumption would often be very reasonable, but should be recognized as an assumption.

Appendix VI

Selection Diversity Mean Power Ratios

In the integral (17) for $\bar{p}(N)$,

$$\bar{p}(N) = N \int_0^{\infty} p(1 - e^{-p})^{N-1} e^{-p} dp, \quad (17)$$

we make the change of variable $y = 1 - e^{-p}$, obtaining

$$\bar{p}(N) = N \int_0^1 [-\log(1-y)] y^{N-1} dy. \quad (83)$$

[42] G. E. Uhlenbeck, Rad. Lab., M.I.T., Cambridge, Mass., Rep. No. 454; October 15, 1943.

[43] H. G. Booker, J. A. Ratcliffe, and D. H. Shinn, "Diffraction from an irregular screen with applications to ionospheric problems," *Phil. Trans. Royal Soc. (London) A*, vol. 242, pp. 579–607; September, 1950.

[44] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 46–156; January, 1945. See (3.7–13).

[45] McNicol, *op. cit.*, Fig. 4.

[46] Cramér, *op. cit.*, p. 289.

Using the series

$$-\log(1-y) = \sum_{k=1}^{\infty} y^k/k \text{ for } |y| < 1,$$

$$\begin{aligned}\bar{p}(N) &= N\int_0^1 \left[\sum_{k=1}^{\infty} \frac{y^{k+N-1}}{k}\right] dy \\ &= \sum_{k=1}^{\infty} \frac{N}{k(k+N)} \\ &= \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{N+k}\right) \\ &= \lim_{m\to\infty}\left[\sum_{k=1}^{m}\left(\frac{1}{k} - \frac{1}{N+k}\right)\right] \\ &= \lim_{m\to\infty}\left[\sum_{k=1}^{N}\frac{1}{k} - \sum_{k=1}^{N}\frac{1}{m+k}\right] \\ &= \sum_{k=1}^{N}\frac{1}{k},\end{aligned} \tag{84}$$

which is the result (18). The termwise integration from the first to second line of (84) is easily justified.

The result $\bar{p}(N) = \sum_{k=1}^{N} 1/k$ was originally found essentially by accident and verified by induction on N.[47] Another direct approach was also suggested by Stein, who pointed out that (17) could be written

$$\bar{p}(N) = -\frac{\partial}{\partial x}\int_0^{\infty} N(1-e^{-p})^{N-1}e^{-xp}dp\,\Big|_{x=1}, \tag{85}$$

which, integrating by parts $(N-1)$ times, becomes

$$\bar{p}(N) = -\frac{\partial}{\partial x}\left[\frac{(x-1)!N!}{(x+N-1)!}\right]\Bigg|_{x=1}. \tag{86}$$

Stein remarked that this is

$$\bar{p}(N) = -N\frac{\partial}{\partial x}\beta(N,x)\,\Big|_{x=1} \tag{87}$$

where $\beta(N, x)$ is the Beta function [substitute $t=e^{-p}$ in (85)], and that higher moments of p are given by successive derivatives of the same function. The differentiation indicated in (86) is straightforward.

Acknowledgment

I am indebted to the several individuals who either provided helpful comments on footnote 1, furnished data or results for inclusion in the present paper, carefully read and criticized the manuscript, or did one or more of the foregoing. These included W. R. Bennett of the Bell Telephone Laboratories, E. J. Baghdady and John Granlund of the Department of Electrical Engineering, M.I.T., F. E. Bond of the Signal Corps Engineering Laboratories, H. F. Meyer of SCEL, William Sichak of the ITT Laboratories, Harold Staras of the RCA Laboratories, Princeton, Seymour Stein of Hycon Eastern, Inc., and several colleagues at Lincoln Laboratory. Among these, I should especially like to thank William Sichak, who contributed very generously to this paper under each of the indicated headings.

I am also indebted to my colleagues W. C. Mason and Michael Ginsburg for the use of the equal-gain distributions in Figs. 9–13 before the complete tables of Mason, *et al.*, have been published. Phyllis Bloom and Marguerite Glynn of the Division 3 Computing Section of Lincoln Laboratory provided substantial assistance with the material in Appendix IV and prepared several of the curves.

Bibliography

The following list includes most of the theoretical papers related to diversity techniques published since 1947, together with two papers of historical interest. The comments following the references from [3] onwards are those of the present author and are not intended to be complete summaries of the papers involved. (Additional references dealing primarily with experimental results or instrumentation techniques are included among the footnotes.)

[1] H. H. Beverage and H. O. Peterson, "Diversity receiving system of RCA Communications, Inc., for radiotelegraphy," Proc. IRE, vol. 19, pp. 531–561; April, 1931.

[2] H. O. Peterson, H. H. Beverage, and J. B. Moore, "Diversity telephone receiving system of RCA Communications, Inc.," Proc. IRE, vol. 19, pp. 562–584; April, 1931.

[3] Z. Jelonek, E. Fitch, and J. H. H. Chalk, "Diversity reception, statistical evaluation of possible gain," *Wireless Engr.*, vol. 24, pp. 54–62; February, 1947. Selection diversity for the cases 1) uncorrelated Rayleigh fading and 2) partially correlated shallow Gaussian fading for two channels is analyzed. The first case is also considered at the low end of the distribution for unequal mean signals.

[4] A. H. Hausman, "An analysis of dual diversity receiving systems," Proc. IRE, vol. 42, pp. 944–947; June, 1954. A comparison of two-channel selection diversity and scanning diversity in terms of unspecified general fading distributions is presented but no concrete results.

[5] L. R. Kahn, "Ratio squarer," Proc. IRE, vol. 42, p. 1704; November, 1954. This is a note pointing out that two-channel maximal-ratio diversity could outperform two-channel selection diversity. Instrumentation and comparative tests are discussed.

[6] D. G. Brennan, "On the maximal signal-to-noise ratio realizable from several noisy signals," Proc. IRE, vol. 43, p. 1530; October, 1955. Statement and proof of the theorem are embodied in (13) and (14) of the present paper.

[7] H. Staras, "Diversity reception with correlated signals," *J. Appl. Phys.*, vol. 27, pp. 93–94; January, 1956. An ingenious method of evaluating the effect of partially correlated Rayleigh fading in two-channel selection diversity systems is presented with results that are also useful in measuring such correlation.

[8] F. J. Altman and W. Sichak, "A simplified diversity communication system for beyond-the-horizon links," IRE Trans. on Communications Systems, vol. CS-4, pp. 50–55; March, 1956. Several comparative results on selection, maximal-ratio and equal-gain diversity systems, including a demonstration that equal-gain systems perform almost as well as maximal-ratio systems in the presence of Rayleigh fading, are presented. The importance of this paper has not yet been widely recognized, perhaps because it is in a journal with a very limited circulation and is so concise that even experts have failed to appreciate its contents. Apart from the papers by Adams and Mindes, and Morrow, *et al.*, and a passing reference by Staras, it is not even mentioned in any of the following references of this paper.

[9] H. Staras, "The statistics of combiner diversity," Proc. IRE, vol. 44, pp. 1057–1058; August, 1956. Curves of the distribution functions for maximal-ratio diversity and a method of approximating very-long-term distributions of the local SNR are presented. This method would be most accurate for systems with a large number of channels. Combiner diversity here means maximal-ratio diversity. See the end of Section IV following (33) in this paper.

[10] J. N. Pierce, "Diversity Improvement in Frequency-Shift Keying for Rayleigh Fading Conditions," Air Force Cambridge Res. Center, Bedford, Mass., Tech. Rep. No. 56-117; September, 1956. Results applicable to certain types of binary signals and four combining methods, including maximal-ratio and selection diversity, are considered. See [15].

[47] The procedure used here was suggested by one of the IRE reviewers.

[11] K. H. Schmelovsky, "Einfluss der Korrelation zwischen Empfangsfeldstarken bei Diversity-Empfang," *Hochfrequenz. Elektr.*, vol. 65, pp. 74–76; November, 1956. (Translations of Friedman available at the John Crerar Library.) A method of evaluating the effect of partially correlated Rayleigh fading in two-channel selection diversity systems is discussed.

[12] F. E. Bond and H. F. Meyer, "The effect of fading on communication circuits subject to interference," PROC. IRE, vol. 45, pp. 636–642; May, 1957. An analysis of two-channel selection diversity for the case of Rayleigh-variable signals in Rayleigh-variable noise is presented. This case often arises in practice in communication systems operating at frequencies below 30 mc. This case was also considered for a selection diversity system in which the selection is of the channel with the greatest signal-plus-noise, in contrast to the greatest SNR.

[13] E. Henze, "Theoretische Untersuchungen über einige Diversity-Verahren," *Arch. Elekt. Übertragung*, vol. 11, pp. 183–194; May, 1957. (Translations by M. D. Friedman available from the John Crerar Library, 86 E. Randolph St., Chicago 1, Ill.) A consideration of two-channel selection diversity and scanning diversity, including both uncorrelated and partially correlated Rayleigh fading cases, is presented. Several of the derivations are extraordinarily complicated and can be greatly simplified.

[14] K. S. Packard, "Effect of correlation on combiner diversity," PROC. IRE, vol. 46, pp. 362–363; January, 1958. Combiner diversity here means maximal-ratio diversity. A method of estimating the effect of partially correlated Rayleigh fading in two-channel maximal-ratio diversity systems is discussed.

[15] J. N. Pierce, "Theoretical diversity improvement in frequency-shift keying," PROC. IRE, vol. 46, pp. 903–910; May, 1958. Error rates for various combining techniques are derived and the effects of correlated fading in dual-diversity binary systems are considered. See also additional references pertinent to binary signal systems given in this paper.

[16] R. T. Adams and B. M. Mindes, "Evaluation of IF and base-band diversity combining receivers," IRE TRANS. ON COMMUNICATION SYSTEMS, vol. CS-6, pp. 8–13; June, 1958. A comparison of predetection equal-gain combining with postdetection maximal-ratio combining for a two-channel FM system is presented, with a theoretical and experimental demonstration of the reduction in FM multipath distortion achieved by using predetection equal-gain combining.

Physical Principles of Avalanche Transistor Pulse Circuits*

D. J. HAMILTON†, ASSOCIATE MEMBER, IRE, J. F. GIBBONS†, AND W. SHOCKLEY†, FELLOW, IRE

Summary—**A simple physical theory is developed which permits a calculation of the significant points of avalanche transistor transient behavior.**

A model for the transistor is defined in terms of charge variables and the physical parameters of the device. The transient performance of the model is calculated by focusing attention on the minority carrier charge stored in the base region and the influence of base-width modulation upon this stored charge. In the charge formulation of the problem, the physical details of the avalanche multiplication process need not be considered; multiplication is accounted for by the boundary conditions which it imposes upon the stored charge.

Good agreement has been obtained between calculated and experimentally observed data for a simple avalanche transistor relaxation oscillator.

I. INTRODUCTION

TRANSISTORS exhibiting avalanche multiplication have recently been shown[1] to be useful for the generation of millimicrosecond pulses. These devices thus provide a new and simple solution to a problem which previously taxed the ingenuity of both circuit and device designers.

As is frequently the case, however, a simple empirical solution poses difficult analytical problems. These analytical difficulties arise primarily from a failure to recognize the important physical principles which govern the terminal behavior of the device. When the problem is properly formulated, many of the analytical complications are removed, and a simple unified theory is obtained. It is the purpose of this paper to present such a theory.

The most significant aspect of the theory is the concept of minority carrier charge stored in the base region during the transient period, a concept which results in a considerable simplification of the problem by permitting time to be eliminated in several of the calculations.[2] A relaxation oscillator (Fig. 1) is used to illustrate the theory.

Section II of the paper deals with the circuit model for the relaxation oscillator, together with a physical model for the avalanche transistor; the concept of stored minority carrier charge is also introduced.

In Section III, the stored charge concept is used to determine the two critical values of external capacitance: the capacitance required to start regeneration and the capacitance required to forward bias the collector junction.

* Original manuscript received by the IRE, December 28, 1959; revised manuscript received, March 6, 1959. This work was supported in part by the U. S. Army Signal Corps, the U. S. Air Force, and the U. S. Navy, through ONR Contract Nonr 225(24) at Stanford Electronics Laboratories.

† Stanford Electronics Labs., Stanford University, Stanford, Calif.

[1] G. B. B. Chaplin, "A method of designing transistor avalanche circuits with application to a sensitive transistor oscilloscope," paper presented at 1958 Solid-State Circuits Conference, Philadelphia, Pa.; February, 1958.

[2] W. Shockley and J. Gibbons, "Theory of transient build-up in avalanche transistors," *Trans. AIEE (Commun. and Electronics)*, no. 40, pp. 993–998; January, 1959.