THE THEORY OF SIGNAL DECTECTABILITY *

W. W. Peterson, T. G. Birdsall, and W. C. Fox
University of Michigan
Ann Arbor, Michigan

ABSTRACT

The problem of signal detectability treated in this paper is the following: Suppose an observer is given a voltage varying with time during a prescribed observation interval and is asked to decide whether its source is noise or is signal plus noise. What method should the observer use to make this decision, and what receiver is a realization of that method? After giving a discussion of theoretical aspects of this problem, the paper presents specific derivations of the optimum receiver for a number of cases of practical interest.

The receiver whose output is the value of the likelihood ratio of the input voltage over the observation interval is the answer to the second question no matter which of the various optimum methods current in the literature is employed including the Neyman - Pearson observer, Siegert's ideal observer, and Woodward and Davies' "observer." An optimum observer required to give a yes or no answer simply chooses an operating level and concludes that the receiver input arose from signal plus noise only when this level is exceeded by the output of his likelihood ratio receiver.

Associated with each such operating level are conditional probabilities that the answer is a false alarm and the conditional probability of detection. Graphs of these quantities, called receiver operating characteristic, or ROC, curves are convenient for evaluating a receiver. If the detection problem is changed by varying, for example, the signal power, then a family of ROC curves is generated. Such things as betting curves can easily be obtained from such a family. The operating level to be used in a particular situation must be chosen by the observer. His choice will depend on such factors as the permissable false alarm rate, a priori probabilities, and relative importance of errors.

With these theoretical aspects serving as an introduction, attention is devoted to the derivation of explicit formulas for likelihood ratio, and for probability of detection and probability of false alarm, for a number of particular cases. Stationary, band-limited, white Gaussian noise is assumed. The seven special cases which are presented were chosen from the simplest problems in signal detection which closely represent practical situations.

Two of the cases form a basis for the best available approximation to the important problem of finding probability of detection when the starting time of the signal, signal frequency, or both, are unknown. Furthermore, in these two cases uncertainty in the signal can be varied, and a quantitative relationship between uncertainty and ability to detect signals is presented for these two rather general cases. The variety of examples presented should serve to suggest methods for attacking other simple signal detection problems and to give insight into problems too complicated to allow a direct solution.

1. INTRODUCTION

The problem of signal detectability treated in this paper is that of determining a set of optimum instructions to be issued to an "observer" who is given a voltage varying with time during a prescribed observation interval and who must judge whether its source is "noise" or "signal plus noise." The nature of the "noise" and of the "signal plus noise" must be known to some extent by the observer.

Any equipment which the observer uses to make this judgement is called the "receiver." Therefore the voltage with which the observer is presented is called the "receiver input." The optimum instructions may consist primarily in specifying the "receiver" to be used by the observer.

The first three sections of this article survey the applications of statistical methods to this problem of signal detectability. They are intended to serve as an introduction to the subject to those who possess a minimum of mathematical training. Several definitions of "optimum" instructions have been proposed by other authors. Emphasis is placed here on the fact that these various definitions lead to essentially the same receiver. In subsequent sections the actual specification of the optimum receiver is carried out and its performance is evaluated numerically for some cases of practical interest.[17]

---

* The work reported in this paper was done under U.S. Army Signal Corps Contract No. DA - 36 - 039 sc - 15358.

1.1 Population SN and N

Either noise alone or the signal plus noise may be capable of producing many different receiver inputs. The totality of all possible receiver inputs when noise alone is present is called "Population N"; similarly, the collection of all receiver inputs when signal plus noise is present is called "Population SN." The observer is presented with a receiver input from one of the two populations, but he does not know from which population it came; indeed, he may not even know the probability that it arose from a particular population. The observer must judge from which population the receiver input came.

1.2 Sampling Plans[1]

A sampling plan is a system of making a sequence of measurements on the receiver input during the observation interval in such a way that it is possible to reconstruct the receiver input for the observation interval from the measurements. Mathematically, a sampling plan is a way of representing functions of time as sequences of numbers. The simplest way to describe this idea is to list a few examples.

A: Fourier Series on an Interval Suppose that the observation interval begins at time t_o and is T seconds long, and that each function in the population SN and N can be expanded in a Fourier series on the observation interval. The Fourier coefficients for each particular receiver input can be obtained by making measurements on that input, which can in turn be reconstructed from these measurements by the formula

$$x(t) = a_o + \sum_{n=1}^{\infty} a_n \cos \frac{2\pi nt}{T} + b_n \sin \frac{2\pi nt}{T}, \quad t_o < t < t_o + T. \tag{1}$$

Thus the process representing each function x(t) by the sequence of its Fourier coefficients (a_o, a_1, b_1, ... , a_n, b_n, ...) is a sampling plan in the sense described above.

The pair of terms in the Fourier series which involve the cosine and sine of $2\pi nt/T$ is of frequency n/T cycles per second. Suppose that for a particular population of receiver inputs the terms of frequency greater than n_o/T are zero; i.e., the population is bandlimited in the Fourier series sense or simply "series-bandlimited." For such a population the process of representing each receiver input x(t) by the finite sequence (a_o, a_1, b_1, ... , a_{n_o}, b_{n_o}) is a finite sample plan.*

B: Shannon's Sampling Plan Suppose that the observation interval includes all time and that the populations are "transform-bandlimited" to a band from 0 to W cycles per second, i.e., the Fourier transform of every receiver input is zero for frequencies greater than W. A sampling plan for this population is to represent each function x(t) by its amplitude measured at times spaced 1/2W seconds apart, (... $x(t_o-n/2W)$, ... , $x(t_o-1/2W)$, $x(t_o)$, $x(t_o+1/2W)$, ... $x(t_o+n/2W)$, ...). In this case the formula[2] for the reconstruction of the receiver input is

$$x(t) = \sum_{n=-\infty}^{\infty} x(t_o + \frac{n}{2W}) \frac{\sin \pi (2W(t-t_o) - n)}{\pi (2W(t-t_o) - n)} \cdot \tag{2}$$

The instants of time $t_o+n/2W$ are called sampling-times. Each choice of t_o between 0 and 1/2W yields a different sampling plan. If the observation interval again includes all time, but the populations are transform-bandlimited to a frequency band from $f_o-W/2$ to $f_o+W/2$ which does not contain zero frequency, then each receiver input x(t) can be considered as an amplitude and frequency modulated waveform, $x(t)=r(t) \cos (2\pi f_o t + \theta(t))$; r(t) is the amplitude of the envelope and $\theta(t)$ is the instantaneous phase of the carrier. A sampling plan employing sampling-times is obtained in this case by representing each receiver input by the sequence (... $r(t_o)$, $\theta(t_o)$, ... , $r(t_o+n/W)$, $\theta(t_o+n/W)$, ...) of envelope amplitudes and carrier phases measured at sampling-times spaced by 1/W seconds apart.[1] The reconstruction of the receiver input from this sequence is given by

$$x(t) = \sum_{n=-\infty}^{\infty} r(t_o+\frac{n}{W}) \cdot \cos\left(2\pi t_o t + \theta(t_o+\frac{n}{W})\right) \frac{\sin \pi (W(t-t_o)-n)}{\pi (W(t-t_o) - n)} \cdot \tag{3}$$

C: Sampling Plan Using Sampling-Times for a Finite Observation Interval Only functions known for all times have Fourier transforms, and therefore the hypothesis that the populations are transform-bandlimited applies only when the observation interval includes all time. If the observation interval is of finite length and if the populations are series-bandlimited, then there are sampling

---

* A sampling plan is finite if there is a finite maximum length for the sequences for all receiver inputs in the population.

plans utilizing sampling-times which are similar to those described in paragraph B for transform-band-limited populations and an infinite observation interval. Suppose that time is measured from the beginning of the observation interval, which is T seconds long, and suppose that the populations are series-bandlimited from 0 to W cycles per second. A finite sampling plan for this situation can be obtained by representing each receiver input by the sequence of its amplitudes measured 1/2W seconds apart,[1]

$$\left(x(t_o),\ x(t_o + \frac{1}{2W}),\ \ldots\ ,\ x(t_o + T - \frac{1}{2W})\right) \tag{4}$$

and the reconstruction of the receiver input from this sequence is

$$x(t) = \sum_{n=0}^{2WT-1} x(t_o + \frac{n}{2W}) \quad \frac{\sin \pi(2W(t-t_o) - n)}{2WT \sin\left(\frac{2W(t-t_o) - n}{2WT}\pi\right)}, \quad 0 < t < T\ . \tag{5}$$

Again each choice of the (initial) sampling-time t_o between 0 and 1/2W yields a different sampling plan. In a similar fashion, if the observation interval is unchanged but the populations are series-band-limited on this interval to a frequency band from f_o-W/2 to f_o+W/2 which does not include zero frequency, then each receiver input can be represented by a finite sequence $(r(t_o), \theta(t_o), r(t_o+1/W), \theta(t_o+1/W), \ldots, r(t_o+T-1/W), \theta(t_o+T-1/W))$ of envelope amplitudes and carrier phases measured at sample points 1/W seconds apart; t_o is again used to denote the initial sampling-time which may be chosen anywhere from 0 to 1/W. The reconstruction of the receiver input from this sequence of measurements is given by

$$x(t) = \sum_{n=0}^{WT-1} r(t_o+\frac{n}{W}) \cos\left(2\pi f_o t + \theta(t_o+\frac{n}{W})\right) \frac{\sin \pi\ (W(t-t_o) - n)}{WT \sin \pi\ \frac{W(t-t_o) - n}{WT}}, \quad 0 < t < T\ . \tag{6}$$

From these examples it can be seen that there are a number of important differences between various sampling plans such as i) the length of the observation interval, ii) whether sampling-times are employed, and iii) whether the measurements are all to be of the same kind, e.g., instantaneous amplitude measurements, or of different kinds, e.g., envelope amplitude and carrier phase. However, they all have in common the property that the receiver input can be reconstructed from the measurements made on it.

The role which the sampling plan plays in the theory presented in this paper is primarily one of mathematical convenience. The populations N and SN will be represented as sequences through the use of sampling plans in order to apply statistical methods. Once an answer is obtained concerning an "optimum" receiver, it is often possible to translate this answer back to the more familiar language of receiver inputs. If a finite sampling plan is not available for a particular application of the theory, then recent work by Grenander[3] shows that the desired parameters of the "optimum" receiver can be approximated by using finite sampling plans. Both for this reason and in order to simplify the exposition, the theory presented here is restricted to cases where finite sampling plans are available.

2. OPTIMUM TESTS ON FIXED OBSERVATION INTERVALS

2.1 Probability Density Functions

This part of the paper is concerned with a method of statistical analysis which requires for raw data a finite sequence of numbers $(x_1, x_2, \ldots, x_n)$, which is the result of the measurements made at the receiver input according to some particular finite sampling plan. The sequence is often called a "sample" of the population from which it arose, and is denoted by a single letter; thus, if the receiver input is x(t), and the sampling plan yields a sequence $(x_1, x_2, \ldots, x_n)$, then this sequence is called the sample X. The theory to be developed here is intended to specify an optimum receiver and is couched in the language of samples, $X = (x_1, x_2, \ldots, x_n)$. If n is very large, a receiver which had to make the measurements called for by a sampling plan would certainly be impractical. However, this practical difficulty is avoided when the specification of the receiver is translated back from the language of samples to the language of the receiver inputs; this can be done because it is possible to reconstruct the inputs from the samples.

For the purposes of the subsequent development any finite sampling plan may be considered provided

enough properties are known of the associated sample X so that certain probabilities may be calculated. Specifically, the probability density functions $f_N(X)$ and $f_{SN}(X)$ of the sample variable X for the cases when X is drawn from populations N and SN respectively must be known.* The two basic properties of density functions are

$$f_N(X) \geq 0 \qquad \int f_N(X)\, dX = 1, \tag{7}$$

and

$$f_{SN}(X) \geq 0 \qquad \int f_{SN}(X)\, dX = 1$$

where the integration symbol represents the multiple integral taken over the entire range of the sample variable $X = (x_1, x_2, \ldots, x_n)$.

2.2 The Concept of a Criterion

Consider now an observer who has as available data the sample $X = (x_1, \ldots, x_n)$. The observer's job is to judge for each sample whether or not it was taken from population SN. Although it is not possible to determine the (probably subconscious) criterion used by the observer, it is quite possible to find an external manifestation of it. Ideally all that is necessary is to submit each possible sample to the observer and to record his judgement. This will yield a tabulation of those samples which the observer decided were drawn from population SN. If any other observer is given this tabulation and instructed to base his decisions on it, he will behave exactly as did the first observer. Thus, the tabulation of these responses can be used to replace the mental criterion employed by the observer. Such a tabulation will also be called a criterion and will be denoted by the letter A, which refers to the phraseology common in statistics of "Accepting the hypothesis that a signal is present." The tabulation of the remaining samples, those which the observer concluded were drawn from population N, will be denoted by B.

2.3 Probabilities Associated with Criteria

There are of course as many different criteria as there are observers. Among all possible criteria it is necessary to select those that are best for various purposes. To do so, certain numerical quantities must be associated with each criterion. It will be necessary to know the probability that a sample from one of the populations will be listed in a particular criterion A. According to the standard definitions, these probabilities are given by

$$P_{SN}(A) = \int f_{SN}(X)\, dX$$

and (8)

$$P_N(A) = \int f_N(X)\, dX$$

where the multiple integral is taken over all samples listed in the criterion A.

For example, a particular sample plan might have a density function of the form $f_N(x_1, x_2, \ldots, x_n) = K \exp -(x_1^2 + x_2^2 + \ldots + x_n^2)$. A possible criterion would consist of those samples $X = (x_1, x_2, \ldots, x_n)$ which lie outside a sphere of radius one centered at the origin. Then the integral would be taken over the exterior of this sphere.

These probabilities have a special significance. $P_N(A)$ is the conditional probability that a sample from population N will be listed in criterion A, that is, will be judged as a sample from population SN. Thus $P_N(A) = F$ is the conditional false alarm probability. Also, $P_{SN}(A)$ is the conditional probability of a certain kind of correct response called a hit (that of judging correctly that a sample is from population SN). The conditional probability of judging falsely that a sample is from population SN is therefore given by $1 - P_{SN}(A) = M$, the conditional probability of a miss. The only errors which can occur are false alarms and misses; their conditional probabilities, F and M, are called briefly the error probabilities.

A reader familiar with the formal content of probability theory should note that these quantities

---

* In this discussion it should be kept in mind that "the event of the sample being drawn from population SN" corresponds to signal and noise being present at the receiver input. Also "the event of population SN being sampled" means the same thing.

are true conditional probabilities; the first is conditional on the sample being drawn from population SN; the second is conditional on its being drawn from population N. This is to distinguish them from a priori probabilities (the probabilities that a certain population will be sampled, for example) which are not as yet assumed known.

2.4 Likelihood Ratio and the Ratio Criteria

It is convenient to introduce a new function called the likelihood ratio, $\ell(X)$, defined as the ratio $f_{SN}(X)/f_N(X)$ for sample points $X = (x_1, \ldots, x_n)$; $\ell(X)$ represents the likelihood that the sample X was drawn from SN relative to the likelihood that it was drawn from N. Hence, if $\ell(X)$ is sufficiently large, it would be reasonable to conclude that X was in fact drawn from population SN, i.e., that X should be listed in the desired "best" criterion. Thus, for each number $\beta \geq 0$, a certain criterion A (β) will be selected; A(β) is chosen by listing each sample X for which $\ell(X) \geq \beta$. The problem then reduces to that of making a wise choice of β; that is, to determine how large "sufficiently large" is. Criteria of the form A(β) will be called ratio criteria.

A number of writers have presented varying definitions of a criterion being "optimum." It turns out that each of these optimum criteria can be expressed as a ratio criterion, so that a receiver designed to yield likelihood ratio as output could be used with any of them.

2.5 Weighted Combination Criteria

Suppose it is possible to assign a certain number w as a weighting factor representing the importance of a false alarm relative to a hit. Since $P_{SN}(A)$ is the probability of a hit, and $P_N(A)$ the probability of a false alarm, it would then be reasonable to find a criterion A which maximizes the quantity

$$P_{SN}(A) - wP_N(A). \tag{9}$$

But this quantity can be written as

$$\int_A \left[f_{SN}(X) - wf_N(X)\right] \, dX \tag{10}$$

where the integration is taken over the sample points X listed in A. To maximize this integral, one would list in A every sample for which the integrand was not negative. Solving that inequality for w, one sees that A should contain those sample points X for which

$$\ell(X) = \frac{f_{SN}(X)}{f_N(X)} \geq w. \tag{11}$$

Thus the desired criterion A is simply A(w), and so it is a ratio criterion.

2.6 Neyman-Pearson Criteria

If it is critically important to keep the probability of a false alarm $P_N(A)$ below a certain level k, then it would be reasonable to choose from among such criteria that one which maximizes the probability of a hit. Thus Neyman and Pearson proposed[4] as a type of optimum criterion any criterion A_k for which

(1) $P_N(A_k) \leq k$, and
(2) $P_{SN}(A_k)$ is a maximum for all the criteria A with the property $P_N(A) \leq k$.

The A_k type criterion can also be expressed as a ratio criterion. This can be made plausible as follows. To begin with, it is necessary to consider only those criteria A for which $P_N(A) = k$, because A will be taken as large as possible in order to meet condition (2). Now consider the curve given parametrically by the equations

$$X = X(\beta) = P_N\,(A(\beta))$$

and

$$Y = Y(\beta) = P_{SN}\,(A(\beta)). \tag{12}$$

This curve will be called the Receiver Operating Characteristic (briefly, ROC) curve, for a receiver whose output is likelihood ratio and with which ratio criteria are being used.

The ROC curve passes through the points (0, 0) and (1, 1), the first at $\beta = \infty$, the second at $\beta = 0$. At $\beta = 0$, $\ell(X) \geq \beta = 0$ for all X, so A(0) consists of all possible samples, Thus the observer will report that every sample is drawn from SN, so he will be certain to make a false alarm and to make a hit. (This assumes that the samples will not be drawn exclusively from one of the populations.)

This can be verified, using the basic property of the density functions expressed by the following equations:

$$P_{SN}(A(0)) = \int f_{SN}(X)\, dX = 1$$

and (13)

$$P_N(A(0)) = \int f_N(X)\, dX = 1$$

where the integration is taken over all possible samples X. These equations mean that $X(0) = Y(0) = 1$. Moreover, $X(\infty) = Y(\infty) = 0$, because for $\beta = \infty$ there are no samples X with $\ell(X) \geq \infty$; i.e., $A(\infty)$ contains no samples at all and the operator will never report a signal is present. Therefore the operator cannot possibly make a false alarm nor can he make a hit. Thus $P_{SN}(A(\infty)) = 0$ and $P_N(A(\infty)) = 0$.

These considerations, together with those of the next section, show that the ROC curve can be sketched somewhat as in Fig. 1.

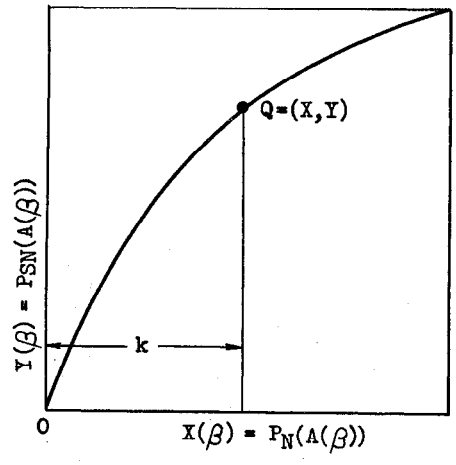

FIG. 1. TYPICAL ROC CURVE

To determine the desired A_k, recall that all probabilities lie between zero and one, so that $P_N(A_k) = k$ is between zero and one. Then there is a point Q of the ROC curve which lies vertically above the point (k,0). The coordinates (X,Y) of Q are $X = P_N(A(\beta)) = k$ and $Y = P_{SN}(A(\beta))$, for some β, which will be written β_k. Now $A(\beta_k)$ satisfies condition (1) because $P_N(A(\beta_k)) = k$, and therefore $A(\beta_k)$ will be the desired A_k if $P_{SN}(A) \leq P_{SN}(A(\beta_k))$ for any criterion with the property that $P_N(A) = k$. From paragraph 2.5, it is clear that the ratio criterion $A(\beta_k)$ is an optimum weighted-combination criterion with the weighting factor $w = \beta_k$. Therefore, if $w = \beta_k$, the weighted-combination using the criterion $A(\beta_k)$ is greater than or equal to the same weighted-combination using any other criterion A, i.e.,

$$P_{SN}(A(\beta_k)) - \beta_k P_N(A(\beta_k)) \geq P_{SN}(A) - \beta_k P_N(A) \tag{14}$$

In this case both $P_N(A(\beta_k))$ and $P_N(A)$ are equal to k. If this value is substituted into the inequality above, one obtains

$$P_{SN}(A(\beta_k)) \geq P_{SN}(A). \tag{15}$$

Therefore, the desired Neyman-Pearson criterion A_k should be chosen to be this particular ratio criterion, $A(\beta_k)$.

2.7 ROC Curve

It is desirable to digress for a moment to study the ROC curve more closely. Its value lies

in the fact that if the type of criterion chosen for a particular application is a ratio criterion, $A(\beta)$, then a complete description of the detection system's performance can be read off the ROC curve. By the very definition of the ROC curve, the X coordinate is the conditional probability, F, of false alarm, and the Y coordinate is the conditional probability of a hit. Similarly (1-X) is the conditional probability of being correct when noise alone is present, and (1-Y) = M is the conditional probability of a miss. It will be shown in a moment that the operating level β for the ratio criterion $A(\ )$ can also be determined from the ROC curve as the slope at the point

$$(P_N(A(\beta)),\ P_{SN}(A(\beta)))\ .$$

Since most proposed kinds of optimum criteria can be reduced to ratio criteria, the ROC curve assumes considerable importance.

In order to determine some of its geometric properties, it will be assumed that the parametric functions

$$X = X(\beta) = P_N(A(\beta))$$

and (16)

$$Y = Y(\beta) = P_{SN}(A(\beta))$$

are differentiable functions of β. The slope of the tangent to the ROC curve is given by the quotient $(dY/d\beta)/dX/d\beta$. To calculate the slope at the point $(X(\beta_o),\ Y(\beta_o))$, notice that among all criteria A, the quantity $P_{SN}(A) - \beta_o P_N(A)$ is maximized by $A = A(\beta_o)$. Therefore, in particular, the function

$$Y(\beta) - \beta_o X(\beta) = P_{SN}(A(\beta)) - \beta_o P_N(A(\beta)) \tag{17}$$

has a maximum at $\beta = \beta_o$, so that its derivative must vanish there. Thus differentiating,

$$\frac{dY}{d\beta} - \beta_o \frac{dX}{d\beta} = 0 \quad \text{at } \beta = \beta_o\ . \tag{18}$$

Solving for β_o, one obtains

$$\beta_o = \frac{\left(\frac{dY}{d\beta}\right)_{\beta = \beta_o}}{\left(\frac{dX}{d\beta}\right)_{\beta = \beta_o}} = \text{the slope of the tangent to the ROC curve at the point } (X(\beta_o),\ Y(\beta_o))\ . \tag{19}$$

This shows that the slope of the ROC curve is given by its parameter β, and so is always positive. Hence the curve rises steadily. In addition, this means that $Y(\beta)$ can be written as a single valued function of $X(\beta)$, $Y = Y(X)$, which is monotone increasing, and where $Y(0) = 0$ and $Y(1) = 1$. These remarks make fully warranted the sketch of the ROC curve given in Fig. 1. The next two sections are concerned with determining the best value to use for the weighting factor w when a priori probabilities are known.

2.8 Siegert's "Ideal Observer's" Criteria

Here it is necessary to know beforehand the a priori probabilities that population SN and that population N will be sampled. This is an additional assumption. These probabilities are denoted respectively by P(SN) and P(N). Moreover, P(SN) + P(N) = 1 because at least one of the populations must be sampled. The criterion associated with Siegert's Ideal Observer is usually defined as a criterion for which a priori probability of error is minimized (or, equivalently, the a priori probability of a correct response is maximized).[5] Frequently the only case considered is that where P(SN) and P(N) are equal, but this restriction is not necessary.

Since the conditional probability F of a false alarm is known as well as the a priori probability of the event (that population N was sampled) upon which F is conditional, then the probability of a false alarm is given by the product

$$P(N)F\ . \tag{20}$$

In the same way the probability of a miss is given by

$$P(SN)M\ . \tag{21}$$

Because an error E can occur in exactly these two ways, the probability of error is the sum of these quantities

$$P(E) = P(N)F + P(SN)M \quad . \tag{22}$$

It has already been pointed out that $F = P_N(A)$ and $M = 1-P_{SN}(A)$. If these are substituted into the expression for P(E) a simple algebraic manipulation gives

$$P(E) = P(SN) - P(SN)\left[P_{SN}(A) - \frac{P(N)}{P(SN)} \cdot P_N(A)\right] \quad . \tag{23}$$

It is desired to minimize P(E). But from the last equation this is equivalent to maximizing the quantity

$$P_{SN}(A) - \frac{P(N)}{P(SN)} \cdot P_N(A) \tag{24}$$

and, of course, this will yield a weighted combination criterion with $w = P(N)/P(SN)$, which is known to be simply a ratio criterion A(w).

2.9 Maximum Expected-Value Criteria

Another way to assign a weighting factor w depends on knowing the "expected value" of each criterion. This can be determined if the a priori probabilities P(SN) and P(N) are known, and if numerical values can be assigned to the four alternatives. Let V_D be the value of detection and V_Q the value of being "quiet", that is, of correctly deciding that noise alone is present. The other two alternatives are also assigned values, V_M, the value of a miss, and V_F, the value of a false alarm. The expected value associated with a criterion can now be determined. In this case it is natural to define an optimum criterion as one which maximizes the expected value. It can be shown that such a criterion maximizes

$$P_{SN}(A) - \left[\frac{P(N)}{P(SN)} \cdot \frac{V_Q - V_F}{V_D - V_M}\right] P_N(A) \tag{25}$$

By definition (see paragraph 2.5), this criterion is a weighted combination criterion with weighting factor

$$w = \frac{P(N)}{P(SN)} \cdot \frac{V_Q - V_F}{V_D - V_M} \tag{26}$$

and hence a likelihood ratio criterion. Seigert's "Ideal Observer" criterion is the special case for which $V_Q - V_F = V_D - V_M$.

2.10 A Posteriori Probability and Signal Detectability

Heretofore the observer has been limited to two possible answers, "signal plus noise is present" or "noise alone is present". Instead he may be asked what, to the best of his knowledge, is the probability that a signal is present. This approach has the advantage of getting more information from the receiving equipment. In fact, Woodward and Davies point out that if the observer makes the best possible estimate of this probability for each possible transmitted message, he is supplying all the information which his equipment can give him.[6] A good discussion of this approach is found in the original papers by Woodward and Davies.[6,7] Their formula for the a posteriori probability, $P_X(SN)$, becomes, in the notation of this paper,

$$P_X(SN) = \frac{f_{SN}(X)\ P(SN)}{f_{SN}(X)\ P(SN) + (1 - P(SN))\ f_N(X)} \quad , \quad \text{or} \tag{27}$$

$$P_X(SN) = \frac{\ell\ (X)\ P(SN)}{\ell\ (X)\ P(SN) + 1 - P(SN)} \tag{28}$$

If a receiver which has likelihood ratio as its output can be built, and if the a priori probability P(SN) is known, a posteriori probability can be calculated easily. The calculation could be built into the receiver calibration, since (28) is a monotonic function of $\ell(X)$; this would make the receiver an optimum receiver for obtaining a posteriori probability.

3. SEQUENTIAL TESTS WITH MINIMUM AVERAGE DURATION

3.1 Sequential Testing

The idea of sequential testing is this: make one measurement x_1 on the receiver input; if the evidence x_1 is sufficiently persuading, decide as to whether the receiver input was drawn from population SN or from population N. If the evidence is not so strong, make a second measurement x_2 and consider the evidence (x_1, x_2). Continue to make measurements until the resulting sequence of measurements is sufficiently pursuading in favor or one population or the other. Obviously this involves the theoretical possibility of making arbitrarily many measurements before a final decision is made. This does not mean that infinitely many measurements must be made in an actual application, nor does it necessarily mean that the operation might entail an arbitrarily long interval of time. If in a particular application measurements are taken at evenly spaced times then the "time base" of such a measurement plan is infinite. However, another plan might call for measurements to be made at the instants $t = 0$, $t = 1/2$, ... , $t = (n-1)/n$, ... and as these times all lie in the time interval from zero to one, such a measurement plan would have a time base of only one unit of time.

If the measurement plan has been carried out to the stage where n measurements $x_1, x_2, \dots, x_n$ have been made, the variable $X_n = (x_1, x_2, \dots, x_n)$ is called the n^{th} stage sample variable. A specific plan for measurements will be considered only if for each possible stage n, the two density functions $f_{SN}(X_n)$ and $f_N(X_n)$ of the n^{th} stage sample variable X_n are known; the first of these density functions is applicable when population SN is being sampled and the second is applicable when population N is being sampled. These density functions may very well differ at different stages, so that they should be written $f^n_N(X_n)$ and $f^n_{SN}(X_n)$; however, the n appearing in the argument X_n should always make the situation clear, and the superscript on the density functions themselves will be omitted.

3.2 Sequential Tests

A sequential test will consist of two things:

1) An (infinite) measurement plan with density functions $f_N(X_n)$ and $f_{SN}(X_n)$
2) An assignment of three criteria to each stage of the measurement plan.

These three criteria represent the three possible conclusions:

A) Signal plus noise is present, i.e. the sample comes from population SN
B) Noise alone is present, i.e. the sample comes from population N
C) Another measurement should be made.

At the first stage of the measurement plan, any (real) number at all could theoretically result from the first measurement. This means that the first stage sample variable $X_1 = (x_1)$ ranges through the entire number system, which will be written S_1 to stand for the first stage sample space. Suppose the three first-stage criteria A_1, B_1, and C_1, have been chosen. If the sample X_1 is listed in A_1, the conclusion that a signal is present is drawn and the test terminated. If it is listed in B_1 the conclusion is that noise alone is present, and again the test is terminated. If X_1 should be listed in C_1, another measurement will be made, and the test moves on to the second stage instead of terminating.

When the first stage criteria have been chosen, a limitation is placed on S_2, the space through which the second stage sample variable $X_2 = (x_1, x_2)$ ranges. The only way the test can proceed to the second stage is for $X_1 = (x_1)$ to be listed in C_1. Therefore, S_2 does not contain all possible second stage samples $X_2 = (x_1, x_2)$ but only those for which (x_1) is listed in C_1. Three second stage criteria, A_2, B_2, and C_2, must now be chosen from those samples X_2 listed in S_2. They must be chosen in such a way that there are no duplications in the listings and no sample in S_2 is omitted. These criteria carry exactly the same significance as those chosen in the first stage. That is, the three conclusions that a signal is or is not present, or that the test should be continued, are drawn when the sample X_2 is listed in A_2, B_2, or C_2 respectively.

The selection of criteria proceeds in the same way. If the n^{th} stage criteria A_n, B_n, and C_n, have been chosen, then the next stage's sample space S_{n+1} consists of those samples $X_{n+1} = (x_1, x_2, \dots, x_n, x_{n+1})$ for which $X_n = (x_1, x_2, \dots, x_n)$ was listed in C_n. Then from S_{n+1} are drawn the three (n+1) stage criteria A_{n+1}, B_{n+1}, and C_{n+1}.

When an entire sequence

$$\begin{aligned}&(A_1, B_1, C_1) \ ,\\ &(A_2, B_2, C_2) \ ,\\ &\quad\vdots\\ &(A_n, B_n, C_n) \ ,\\ &\quad\vdots\end{aligned}$$

of criteria is selected, a "sequential test" has been determined. This does not mean of course that the test will necessarily be particularly useful. However, among all the possible ways of selecting a sequence of criteria and hence a sequential test, there may be particular ones which are very useful.

3.3 Probabilities Associated with Sequential Tests

If Q_n is any n^{th} stage criterion, then the quantities*

and

$$\begin{aligned}P_N(Q_n) &= \int_{Q_n} f_N(X_n)\, dX_n\\ P_{SN}(Q_n) &= \int_{Q_n} f_{SN}(X_n)\, dX_n\end{aligned} \tag{29}$$

represent the (N or SN) conditional probabilities that an n^{th} stage sample X_n will be listed in the criterion Q_n. Conditional probabilities of particular interest are:

1) The n^{th} stage conditional error probabilities:
If population N is sampled, then the probability that the sample variable X_n will be listed in A_n is $P_N(A_n)$. This is the N-conditional probability of a false alarm.
If population SN is sampled, then the probability that the sample variable X_n will be listed in B_n is $P_{SN}(B_n)$. This is the SN-conditional probability of a miss.

2) The conditional error probabilities of the entire test:

$F = \sum_{n=1}^{\infty} P_N(A_n)$, the N-conditional probability of a false alarm, and (30)

$M = \sum_{n=1}^{\infty} P_{SN}(B_n)$, the SN-conditional probability of a miss, (31)

are merely the sums of the same error probabilities over all stages.

3) The conditional probabilities of terminating at stage n are

$$T_N^n = P_N(A_n) + P_N(B_n) \ , \text{ and} \tag{32}$$

$$T_{SN}^n = P_{SN}(A_n) + P_{SN}(B_n) \ . \tag{33}$$

These equations can be justified by a simple argument. The only way the test can terminate at stage n is for the sample variable X_n to be listed in either A_n or B_n. The probability of this event is the sum of the probabilities of the component events which are mutually exclusive since X_n can be listed in at most one of A_n and B_n.

---

* The notation $\int_{Q_n}$ indicates that the integration is to be carried out over all sample points listed in Q_n.

4) The conditional probabilities that the entire test will terminate are

$$T_N = \sum_{n=1}^{\infty} T_N^n \quad , \text{ and} \tag{34}$$

$$T_{SN} = \sum_{n=1}^{\infty} T_{SN}^n \quad . \tag{35}$$

3.4. Average Sample Numbers

There are two other quantities which must be introduced. One feature of the sequential test is that it affords an opportunity of arriving at a decision early in the sampling process when the data happens to be unusually convincing. Thus one might expect that, on the average, the stage of termination of a well-constructed sequential test would be lower than could be achieved by an otherwise equal, good standard test. It is therefore important to obtain expressions for the average or expected value of the stage of termination. As with other probabilities, there will be two of these quantities: one conditional on population N being sampled; the other conditional on population SN being sampled. They are given by

$$E_N = \sum_{n=1}^{\infty} n\, T_N^n \tag{36}$$

and

$$E_{SN} = \sum_{n=1}^{\infty} n\, T_{SN}^n \tag{37}$$

The letter E is used to refer to the term "expected value." The quantities E_N and E_{SN} are called the average sample numbers. The form these formulas take can be justified (somewhat freely) on the grounds that each value, n, which the variable "stage of termination" may take on must be weighted by the (conditional) probability that the variable will in fact take on that value.

It should be heavily emphasized that the average sample numbers are strictly average figures. In actual runs of a sequential test, the stages of termination will sometimes be less than the average sample numbers but will also be upon occasion much larger. Any sequential test whose average sample numbers are not finite would be useless for applications. Therefore the only ones to be considered are those with finite average sample numbers. Under this assumption,* it can be shown that $T_N = T_{SN} = 1$ so that the test is certain to terminate (in the sense of probability). On the other hand, if it is known that $T_N = T_{SN} = 1$ it does not always follow that the average sample numbers are finite. Such a situation would mean only that if a sequence of runs of the test were made, each run would probably terminate, but the average stage of termination would become arbitrarily large as more runs were made.

3.5 Sequential Ratio Tests

In studying non-sequential tests using finite samples it was found that the best criterion could always be expressed in terms of likelihood ratio. Therefore, it may be useful to introduce likelihood ratios at each stage of an infinite sample plan. The n^{th} stage likelihood ratio function $\ell(X_n)$ is defined as the ratio $f_{SN}(X_n)/f_N(X_n)$. Optimum criteria in the finite sample tests turned out to be criteria listing all samples X for which $\ell(X)$ is greater than or equal to a certain number. It should be possible to choose sequential criteria (A_n, B_n, C_n) in the same way. For each stage two numbers a_n and b_n with $b_n \leq a_n$ could be chosen. Then the criteria (A_n, B_n, C_n) determined by the numbers a_n and b_n would be

A_n lists all samples X_n of the sample space S_n for which $\ell(X_n) \geq a_n$
B_n lists all samples X_n of the sample space S_n for which $\ell(X_n) \leq b_n$
C_n lists all samples X_n of the sample space S_n for which $b_n < \ell(X_n) < a_n$.

If criteria selected in this way meet the requirements that the average sample numbers be finite, then the resulting sequential test is called a "sequential ratio test."

3.6 Optimum Sequential Tests

---

* Remember that the sampling process is not assumed to yield independence among the X_i.

It is customary[8] to define an optimum sequential test as that one for which the average sample numbers E_N and E_{SN} are minimum among all sequential tests with fixed error probabilities F and M.

In addition to the formulas given in Section 3.4, alternative formulas[9] for the average sample numbers are

$$E_N = 1 + \sum_{i=1}^{\infty} P_N(C_i) \tag{38}$$

and

$$E_{SN} = 1 + \sum_{i=1}^{\infty} P_{SN}(C_i) \, . \tag{39}$$

Thus, if a set of sequential criteria (A_n^*, B_n^*, C_n^*) is presented as a possible optimum test, then its optimum character is decided by ascertaining whether the inequalities

$$\sum P_N(C_i^*) \leq \sum P_N(C_i) \tag{40}$$

and

$$\sum P_{SN}(C_i^*) \leq \sum P_{SN}(C_i) \tag{41}$$

hold for every other set of sequential criteria $\{(A_n, B_n, C_n)\}$ with the same error probabilities, i.e., with

$$\sum P_N(A_i^*) = \sum P_N(A_i) \tag{42}$$

and

$$\sum P_{SN}(B_i^*) = \sum P_{SN}(B_i) \tag{43}$$

The problem of constructing an optimum sequential test is difficult because the equalities (42) and (43) can be satisfied even when there is no apparent term-by-term relation between the sequences $\{P_N(C_i^*)\}$ and $\{P_N(C_i)\}$. Wald has proposed as optimum the tests in which each of the sequences $\{a_n\}$ and $\{b_n\}$ is constant, that is, $b_1 = b_n$ and $a_1 = a_n$ for all n. Moreover Wald and Wolfowitz[10] proved that these tests are optimum whenever the density functions at successive stages are independent, as can be the case for example when both noise and signal plus noise consist of "random noise." However, this "randomness" is not met with in most applications of the theory of signal detectability at least not in the sense that the hypotheses of Wald and Wolfowitz are satisfied.

Consider a test of fixed length as described in Section 2, with error probabilities F and M. Although the optimum sequential test with these same error probabilities generally requires less time on the average, it has the disadvantage that it will sometimes use much more time than the fixed length test requires. In a conversation with the authors, Professor Mark Kac of Cornell University suggested that the dispersion, or variance, of the sample numbers may be so large as seriously to affect the usefulness of the sequential tests in applications to signal detectability. Certainly this matter should be investigated before a final decision is reached concerning the merits of sequential tests relative to tests on a fixed observation interval. However it is a difficult matter to calculate the variance of the sample numbers. Therefore an electronic simulator is being built at the University of Michigan which will simulate both types of tests and will provide data for ROC curves of both types as well as the distribution of the (sequential) sample numbers.

4. OPTIMUM DETECTION FOR SPECIFIC CASES

4.1 Introduction

The chief conclusion obtained from the general theory of signal detectability presented in Section 2 of this paper is that a receiver which calculates the likelihood ratio for each receiver input is the optimum receiver for detecting signals in noise.

It is the purpose of Section 4 to consider a number of different ensembles of signals with band-limited white Gaussian noise. For each case, a possible receiver design is discussed. The primary emphasis, however, is on obtaining the probability of detection and probability of false alarm, and hence on estimates of optimum receiver performance for the various cases.

The cases which are presented were chosen from the simplest problems in signal detection which closely represent practical situations. They are listed in Table I along with examples of engineering problems in which they find application. In the last two cases the uncertainty in the signal can be varied, and some light is thrown on the relationship between uncertainty and the ability to detect

signals. The variety of examples presented should serve to suggest methods for attacking other simple signal detection problems and to give insight into problems too complicated to allow a direct solution.

The reader will find the discussion of likelihood ratio and its distribution easier to follow if he keeps in mind the connection between a criterion type receiver and likelihood ratio. In an optimum criterion type system, the operator will say that a signal is present whenever the likelihood ratio is above a certain level β. He will say that only noise is present when the likelihood ratio is below β. For each operating level β, there is a false alarm probability and a probability of detection. The false alarm probability is the probability that the likelihood ratio $l(X)$ will be greater than β if no signal is sent; this is by definition the complementary distribution function $F_N(\beta)$. Likewise, the complementary distribution $F_{SN}(\beta)$ is the probability that $l(X)$ will be greater than β if there is signal plus noise, and hence $F_{SN}(\beta)$ is the probability of detection if a signal is sent.

TABLE I

Section	Description of Signal Ensemble	Application
4.4	Signal Known Exactly*	Coherent radar with a target of known range and character
4.5	Signal Known Except for Phase*	Ordinary pulse radar with no integration and with a target of known range and character.
4.6	Signal a Sample of White Gaussian Noise	Detection of noise-like signals; detection of speech sounds in Gaussian noise.
4.7	Detector Output of a Broad Band Receiver	Detecting a pulse of known starting time (such as a pulse from a radar beacon) with a crystal-video or other type broad band receiver.
4.8	A Radar Case (A train of pulses with incoherent phase)	Ordinary pulse radar with integration and with a target of known range and character.
4.10	Signal One of M Orthogonal Signals	Coherent radar where the target is at one of a finite number of non-overlapping positions.
4.11	Signal One of M Orthogonal Signals Known Except for Phase	Ordinary pulse radar with no integration and with a target which may appear at one of a finite number of non-overlapping positions.

4.2 Gaussian Noise

In the remainder of this paper the receiver inputs will be assumed to be defined on a finite observation interval, $0 < t < T$. It will further be assumed that the receiver inputs are series-bandlimited. By the sampling plan C (Section 1.2) any such receiver input $x(t)$ can be reconstructed from sample values of the function taken at points $1/2W$ apart throughout the observation interval, i.e.,

---

* Our treatment of these two fundamental cases is based upon Woodward and Davies' work, but here they are treated in terms of likelihood ratio, and hence apply to criterion type receivers as well as to a posteriori probability type receivers. These first two cases have been solved for the more general problem in which the noise is Gaussian but has an arbitrary spectrum.[11, 12] Those solutions require the use of an infinite sampling plan and are consideredly more involved than the corresponding derivations in this report.

$$x(t) = \sum_{k=1}^{2WT} x_k \psi_k(t), \tag{44}$$

where

$$\psi_k(t) = \frac{\sin \pi 2WT(\frac{t}{T} - \frac{k}{2WT})}{2WT \sin \pi (\frac{t}{T} - \frac{k}{2WT})} \quad \text{and} \quad x_k = x(\tfrac{k}{2W}). \tag{45}$$

Therefore the receiver inputs can be represented by the sample $(x_1, x_2, \ldots, x_{2WT})$. In Section 4 the notation x will be used to denote either the receiver input function x(t) or the sample $(x_1, x_2, \ldots, x_{2WT})$. Similarly the signal s(t), or simply s, can be represented by the sample $(s_1, \ldots, s_{2WT})$ where $s_k = s(k/2W)$.

Only the probability distributions for receiver inputs x(t) can be specified. The distribution must be given for the receiver inputs both with noise alone and with signal plus noise. The probabili distributions are described by giving the probability density functions $f_{SN}(x)$ and $f_N(x)$ for the receiver inputs x.

The probability density function for the receiver inputs with noise alone are assumed to be

$$f_N(x) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi N}} \exp \left[- \frac{x_i^2}{2N} \right] \right\}, \tag{46}$$

or

$$f_N(x) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp \left[- \frac{1}{2N} \sum_{i=1}^{n} x_i^2 \right]$$

where n is 2WT and N is the noise power. It can be verified easily that this probability density function is the description of noise which has a Gaussian distribution of amplitude at every time, is stationary, and has the same average power in each of its Fourier components. Thus we shall refer to it as "stationary band-limited white Gaussian noise."

The functions $\psi_k(t)$ are orthogonal and have energy 1/2W, and therefore

$$\sum x_i^2 = 2W \int_0^T \left[x(t) \right]^2 dt \quad , \tag{47}$$

so that

$$f_N(x) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp \left[- \frac{1}{N_o} \int_0^T x(t)^2 \, dt \right] , \tag{48}$$

where $N_o = N/W$ is the noise power per unit bandwidth.

In a practical application, information is given about the signals as they would appear without noise at the receiver input, rather than about the signal plus noise probability density. Then $f_{SN}(x)$ must be calculated from this information and the probability density function $f_N(x)$ for the noise. The noise and the signals will be assumed independent of each other.

If the input to the receiver is the sum of the signal and the noise, then the receiver input x(t) could have been caused by any signal s(t) and noise n(t) = x(t) - s(t). The probability density for the input x in signal plus noise is thus the probability (density) that s(t) and x(t) - s(t) will occ together, averaged over all possible s(t). If the probability of the signals is described by a densit function $f_S(s)$, then

$$f_{SN}(x) = \int f_N(x-s) f_S(s) ds \tag{49}$$

where the integration is over the entire range of the sample variable s. A more general form is used when the probability of the signals is described by a probability measure P_S; the formula in this case is

$$f_{SN}(x) = \int f_N(x-s)\, dP_S(s). \tag{50}$$

This integral is a Lebesgue integral, and is essentially an "average" of $f_N(x-s)$ over all values of s weighted by the probability P_S. If $f_N(x)$ is taken from Eq. (46), this becomes

$$\begin{aligned} f_{SN}(x) &= \int f_N(x-s) dP_S(s) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \int \exp\left[-\frac{1}{2N}\sum_{i=1}^{n}(x_i-s_i)^2\right] dP_S(s) \\ &= \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2N}\sum_{i=1}^{n} x_i^2\right] \int \exp\left[-\frac{1}{2N}\sum_{i=1}^{n} s_i^2\right] \exp\left[\frac{1}{N}\sum_{i=1}^{n} x_i s_i\right] dP_S(s) \end{aligned} \tag{51}$$

$$\begin{aligned} f_{SN}(x) &= \int f_N(x-s)\, dP_S(s) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \int \exp\left[-\frac{1}{N_o}\int_0^T \left[x(t)-s(t)\right]^2 dt\right] dP_S(s) \\ &= \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{N_o}\int_0^T x^2\, dt\right] \int \exp\left[-\frac{1}{N_o}\int_0^T s^2 dt\right] \exp\left[\frac{2}{N_o}\int_0^T xs\, dt\right] dP_S(s) \end{aligned} \tag{52}$$

The factor $\exp\left[-(1/N_o)\int_0^T x^2(t)\, dt\right] = \exp\left[-(1/2N)\sum x_i^2\right]$ can be brought out of the integral since it does not depend on s, the variable of integration. Note that the integral

$$\int_0^T s(t)^2 dt = \frac{1}{2W}\sum s_i^2 = E(s) \tag{53}$$

is the energy* of the expected signal, while

$$\int_0^T x(t)\, s(t)\, dt = \frac{1}{2W}\sum x_i s_i \tag{54}$$

is the cross correlation between the expected signal and the receiver input.

4.3 Likelihood Ratio with Gaussian Noise

Likelihood ratio is defined as the ratio of the probability density functions $f_{SN}(x)$ and $f_N(x)$. With white Gaussian noise it is obtained by dividing Eq. (51) and (52) by (46) and (48) respectively.

---

* This assumes that the circuit impedance is normalized to one ohm.

$$\ell(x) = \int \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{1}{N}\sum_{i=1}^{n} x_i s_i\right] dP_S(s), \quad \text{or} \tag{55}$$

$$\ell(x) = \int \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{2}{N_o}\int_0^T x(t)\, s(t)\, dt\right] dP_S(s) \,. \tag{56}$$

If the signal is known exactly or completely specified, the probability for that signal is unity, and the probability for any set of possible signals not containing s is zero. Then the likelihood ratio becomes

$$\ell_s(x) = \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{1}{N}\sum_{i=1}^{n} x_i s_i\right], \quad \text{or} \tag{57}$$

$$\ell_s(x) = \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{2}{N_o}\int_0^T x(t)\, s(t)\, dt\right] \,. \tag{58}$$

Thus the general formulas (55) and (56) for likelihood ratio state that $\ell(x)$ is the weighted average of $\ell_s(x)$ over the set of all signals, i.e.,

$$\ell(x) = \int \ell_s(x)\, d\, P_S(s) \,. \tag{59}$$

An equipment which calculates the likelihood ratio $\ell(x)$ for each receiver input x is the optimum receiver. The form of equation (58) suggests one form which this equipment might take. First, for each possible expected signal s, the individual likelihood ratio $\ell_s(x)$ is calculated. Then these numbers are averaged. Since the set of expected signals is often infinite, this direct method is usually impractical. It is frequently possible in particular cases to obtain by mathematical operations on equation (58) a different form for $\ell(x)$ which can be recognized as the response of a realizable electronic equipment, simpler than the equipment specified by the direct method. It is essentially this which is done in the following paragraphs.

If the distribution function $P_S(s)$ depends on various parameters such as carrier phase, signal energy, or carrier frequency, and if the distributions in these parameters are independent, the expression for likelihood ratio can be simplified somewhat. If these parameters are indicated by r_1, r_2, ... , r_n, and the associated probability density functions are denoted by $f_1(r_1)$, $f_2(r_2)$, ..., $f_n(r_n)$, then

$$d\, P_S(s) = f_1(r_1) \cdots f_n(r_n)\, dr_1 \cdots dr_n \,.$$

The likelihood ratio becomes

$$\begin{aligned} \ell(x) &= \int \cdots \int \ell_s(x)\, f_1(r_1) \cdots f_n(r_n)\, dr_1 \cdots dr_n \\ &= \int \left[f_n(r_n) \cdots \left[\int f_1(r_1)\, \ell_s(x)\, dr_1 \right] \cdots \right] dr_n \,. \end{aligned} \tag{60}$$

Thus the likelihood ratio can be found by averaging $\ell_s(x)$ with respect to the parameters.

4.4 The Case of a Signal Known Exactly

The likelihood ratio for the case when the signal is known exactly has already been presented in Section 4.3.

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \exp\left[\frac{1}{N}\sum_{i=1}^{n} x_i s_i\right] , \tag{61}$$

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \exp\left[\frac{2}{N_o}\int_0^T x(t)\, s(t)\, dt\right] \tag{62}$$

As the first step in finding the distribution functions for $\ell(x)$, it is convenient to find the distribution for $(1/N)\sum x_i s_i$ when there is noise alone. Then the input $x = (x_1, x_2, \ldots, x_n)$ is due to white Gaussian noise. It can be seen from Eq. (46) that each x_i has a normal distribution with zero mean and variance $N = WN_o$ and that the x_i are independent. Because the s_i are constants depending on the signal to be detected, $s = (s_1, s_2, \ldots, s_n)$, each summand $(x_i s_i)/N$ has a normal distribution with mean s_i/N times the mean of x_i, and with variance $(s_i/N)^2$ times the variance of x_i, which are zero and s_i^2/N respectively. Because the x_i are independent, the summands $(s_i x_i)/N$ are independent, each with normal distribution, and therefore their sum has a normal distribution with mean the sum of the means -- i.e., zero -- and variance the sum of the variances.

$$\sum \frac{s_i^2}{N} = \frac{2WE(s)}{N} = \frac{2E}{N_o} = 2 \times \frac{\text{Signal Energy}}{\text{Noise Power Per Unit Bandwidth}} . \tag{63}$$

The distribution for $(1/N)\sum x_i s_i$ with noise alone is thus normal with zero mean and variance $2E/N_o$. Recalling from Eq. (61)

$$\ell(x) = \exp\left[-\frac{E}{N_o} + \frac{1}{N}\sum x_i s_i\right] , \tag{64}$$

one sees that the distribution for $(1/N)\sum x_i s_i$ can be used directly by introducing α defined by

$$\beta = \exp\left[-\frac{E}{N_o} + \alpha\right] , \quad \text{or } \alpha = \frac{E}{N_o} + \ell n\, \beta . \tag{65}$$

The inequality $\ell(x) \geq \beta$ is equivalent to $(1/N)\sum x_i s_i \geq \alpha$, and therefore

$$F_N(\beta) = \sqrt{\frac{N_o}{4\pi E}} \int_\alpha^\infty \exp\left[-\frac{1}{2}\frac{N_o}{2E} y^2\right] dy . \tag{66}$$

The distribution for the case of signal plus noise can be found by using Eq. (19), which states that

$$\left(\frac{d\, P_{SN}(A(\beta))}{d\, P_N\,(A(\beta))}\right)_{\text{at } \beta=\beta_o} = \beta_o . \tag{67}$$

Because these probabilities are equal to the complimentary distribution functions for likelihood ratio, this can be written as

$$d\, F_{SN}\,(\beta) = \beta\, d\, F_N(\beta). \tag{68}$$

Differentiating Eq. (66),

$$dF_N(\beta) = -\sqrt{\frac{N_o}{4\pi E}} \exp\left(-\frac{N_o \alpha^2}{4E}\right) d\alpha , \tag{69}$$

and combining (65), (68), and (69), one obtains

$$dF_{SN}(\beta) = -\sqrt{\frac{N_o}{4\pi E}} \exp\left[-\frac{E}{N_o} + \alpha - \frac{N_o\alpha^2}{4E}\right] d\alpha . \tag{70}$$

Thus

$$F_{SN}(\beta) = \sqrt{\frac{N_o}{4\pi E}} \int_{\alpha}^{\infty} \exp\left[-\frac{N_o}{4E}\left(y - \frac{2E}{N_o}\right)^2\right] dy . \tag{71}$$

In summary, α and therefore $\ell n\ \beta$, have normal distributions with signal plus noise as well as with noise alone; the variance of each distribution is $2E/N_o$, and the difference of the means is $2E/N_o$.

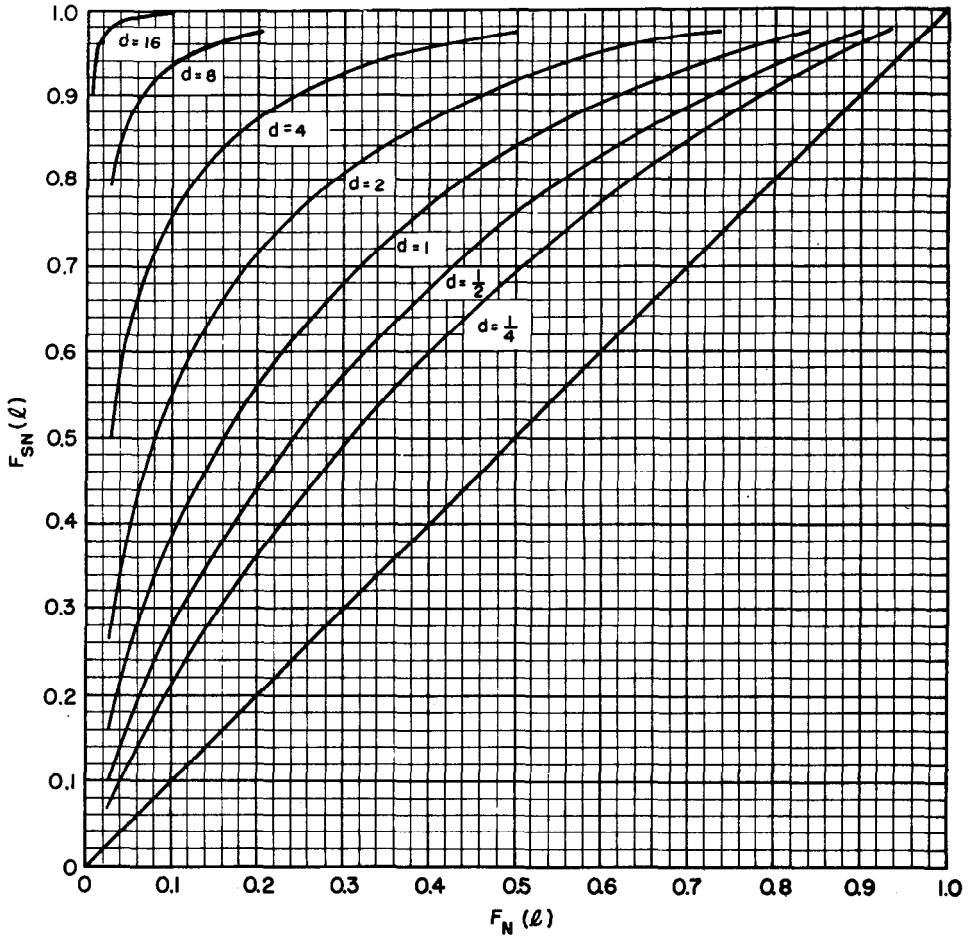

FIG. 2

RECEIVER OPERATING CHARACTERISTIC

$\ell n\ \ell$ IS A NORMAL DEVIATE WITH $\sigma_N^2 = \sigma_{SN}^2$, $(M_{SN}-M_N)^2 = d\cdot\sigma_N^2$

The receiver operating characteristic curves in Figs. 2 and 3* are plotted for any case in which

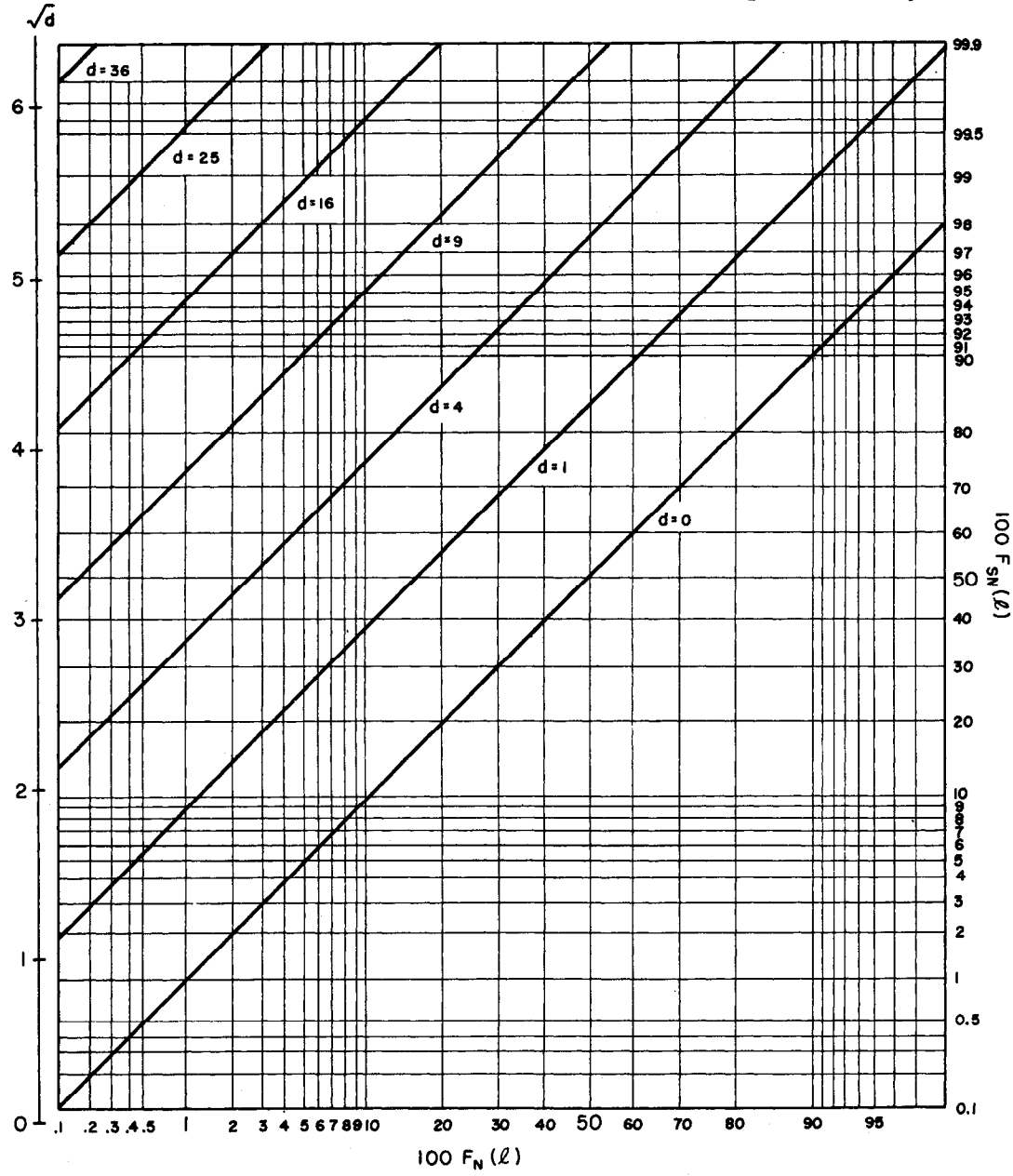

FIG. 3
RECEIVER OPERATING CHARACTERISTIC.

$\ell n\ \ell$ IS A NORMAL DEVIATE, $\sigma_{SN}^2 = \sigma_N^2$, $(M_{SN} - M_N)^2 = d\ \sigma_N^2$

* In Fig. 3, the receiver operating characteristic curves are plotted on "double probability" paper. On this paper both axes are linear in the error function erf $(x) = (1/\sqrt{2\pi}) \cdot \int_{-\infty}^{x} \exp\left[-t^2/2\right] dt$; this makes the receiver operating characteristic straight lines.

$\ell n\ \ell$ has a normal distribution with the same variance both with noise alone and with signal plus noise. The parameter d in this figure is equal to the square of the difference of the means, divided by the variance. These receiver operating characteristic curves apply to the case of the signal known exactly, with $d = 2E/N_o$.

Eq. (62) describes what the ideal receiver should do for this case. The essential operation in the receiver is obtaining the correlation, $\int_0^T s(t)x(t)dt$. The other operations, multiplying by a constant, adding a constant, and taking the exponential function, can be taken care of simply in the calibration of the receiver output. Electronic means of obtaining cross correlation have been developed recently.[13]

If the form of the signal is simple, there is a simple way to obtain this cross correlation.[6,7] Suppose h(t) is the impulse response of a filter. The response $e_o(t)$ of the filter to a voltage x(t) is

$$e_o(t) = \int_{-\infty}^{t} x(\tau)\, h(t-\tau)\, d\tau . \tag{72}$$

If a filter can be synthesized so that

$$\begin{aligned} h(t) &= s(T-t) && 0 \leqq t \leqq T \\ h(t) &= 0 && \text{otherwise,} \end{aligned} \tag{73}$$

then

$$e_o(T) = \int_0^T x(\tau)\, s(\tau)\, d\tau , \tag{74}$$

so that the response of this filter at time T is the cross correlation required. Thus, the ideal receiver consists simply of a filter and amplifiers.

It should be noted that this filter is the same, except for a constant factor, as that specified when one asks for the filter which maximizes peak signal to average noise power ratio.[14]

4.5 Signal Known Except for Carrier Phase

The signal ensemble considered in this section consists of all signals which differ from a given amplitude and frequency modulated signal only in their carrier phase, and all carrier phases are assumed equally likely.

$$s(t) = f(t) \cos\left(\omega t + \phi(t) - \theta\right) . \tag{75}$$

Since the unknown phase angle θ has a uniform distribution,

$$dP_S(\theta) = \frac{1}{2\pi}\, d\theta . \tag{76}$$

The likelihood ratio can be found by applying Eq.(56), and since the signal energy E(s) is the same for all values of the carrier phase θ,

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \int \exp\left[\frac{1}{N}\sum x_i s_i\right] dP_S(s) . \tag{77}$$

Expanding s into the coefficients of $\cos\theta$ and $\sin\theta$ will be helpful:

$$s(t) = f(t)\cos\left(\omega t+\phi(t)\right)\cos\theta + f(t)\sin\left(\omega t+\phi(t)\right)\sin\theta , \tag{78}$$

and

$$\frac{1}{N}\sum x_i s_i = \cos\theta \frac{1}{N}\sum x_i f(t_i)\cos\left(\omega t_i + \phi(t_i)\right) + \sin\theta \frac{1}{N}\sum x_i f(t_i)\sin\left(\omega t_i + \phi(t_i)\right) \quad * \tag{79}$$

Because we wish to integrate with respect to θ to find the likelihood ratio, it is easiest to introduce parameters similar to polar coordinates (r, θ_o) such that

$$\frac{1}{N} r\cos\theta_o = \frac{1}{N}\sum x_i f(t_i)\cos\left(\omega t_i + \phi(t_i)\right)$$
$$\frac{1}{N} r\sin\theta_o = \frac{1}{N}\sum x_i f(t_i)\sin\left(\omega t_i + \phi(t_i)\right) \tag{80}$$

and therefore

$$\frac{1}{N}\sum x_i s_i = \frac{r}{N}\cos(\theta - \theta_o) \tag{81}$$

Using this form the likelihood ratio becomes

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right]\int_0^{2\pi} \exp\left[\frac{r}{N}\cos(\theta - \theta_o)\right]\frac{d\theta}{2\pi}$$
$$= \exp\left[-\frac{E}{N_o}\right] I_o\left(\frac{r}{N}\right) \tag{82}$$

where I_o is the Bessel function of zero order and pure imaginary argument.

I_o is a strictly monotone increasing function, and therefore the likelihood ratio will be greater than a value β if and only if r/N is greater than some value corresponding to β.

In the previous section it was shown that the sum $(1/N)\sum x_i s_i$ has a normal distribution with zero mean and variance $2E/N_o$ if the receiver input $x(t)$ is due to noise alone; E is the energy of the signal known exactly, $s(t)$, and N_o is the noise power per cycle. Since $f(t)\cos(\omega t + \phi(t))$ and $f(t)\sin(\omega t + \phi(t))$ are signals known exactly, both $(r/N)\cos\phi_o$ and $(r/N)\sin\phi_o$ have normal distributions with zero mean and variance $2E/N_o$. The probability that due to noise alone $r/N = \sqrt{(r/N\cos\theta_o)^2 + (r/N\sin\theta_o)^2}$ will exceed any fixed value, is given by the well known chi-square distribution for two degrees of freedom, $K_2(\alpha^2)$. The proper normalization yielding zero mean and unit variance requires that the variable be $(r/N)\sqrt{N_o/2E(s)}$, that is

$$P_N\left(\frac{r}{N}\sqrt{\frac{N_o}{2E}} \geq \alpha\right) = K_2(\alpha^2) = \exp\left[-\frac{\alpha^2}{2}\right]. \quad ** \tag{83}$$

---

* t_i denotes the i^{th} sampling time, i.e., $t_i = i/2W$.

** The symbol $P(x \geq \alpha)$ denotes the probability that the variable x is not less than the constant α.

If α is defined by the equation

$$\beta = \exp\left[-\frac{E}{N_o}\right] I_o\left(\sqrt{\frac{2E}{N_o}}\ \alpha\right), \tag{84}$$

the distribution for $\ell(x)$ in the presence of noise alone is in the simple form

$$F_N(\beta) = \exp\left[-\frac{\alpha^2}{2}\right] . \tag{85}$$

It follows from (85) that

$$dF_N(\beta) = -\alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha . \tag{86}$$

If in equation (68), namely

$$\beta\ dF_N(\beta) = dF_{SN}(\beta) , \tag{87}$$

β is replaced by the expression given in (84) and $dF_N(\beta)$ is replaced by that given in (86), then

$$dF_{SN}(\beta) = -\exp\left[-\frac{E}{N_o}\right]\alpha \exp\left[-\frac{\alpha^2}{2}\right] I_o\left(\sqrt{\frac{2E}{N_o}}\,\alpha\right) d\alpha \tag{88}$$

is obtained. Integration of (88) yields

$$F_{SN}(\beta) = \exp\left[-\frac{E}{N_o}\right]\int_{\alpha}^{\infty} \alpha \exp\left[-\frac{\alpha^2}{2}\right] I_o\left(\sqrt{\frac{2E}{N_o}}\,\alpha\right) d\alpha . \tag{89}$$

Eqs. (85) and (89) yield the receiver operating characteristic in parametric form, and Eq. (84) gives the associated operating levels.[15] These are graphed in Fig. 4 for some of the same values of signal energy to noise power per unit bandwidth as were used when the phase angle was known exactly, Figs. 2 and 3, so that the effect of knowing the phase can be easily seen.

If the signal is sufficiently simple so that a filter could be synthesized to match the expected signal for a given carrier phase Θ as in the case of a signal known exactly, then there is a simple way to design a receiver to obtain likelihood ratio. For simplicity let us consider only amplitude modulated signals ($\phi(t)=0$) in Eq. (75). Let us also choose $\Theta = 0$. (Any phase could have been chosen.) Then the filter has impulse response

$$\begin{aligned} h(t) &= f(T-t)\cos\left[\omega(T-t)\right] && 0 \le t \le T \\ &= 0 && \text{otherwise.} \end{aligned} \tag{90}$$

The output of the filter in response to x(t) is then

$$\begin{aligned} e_o(t) &= \int_{-\infty}^{t} x(\tau)\, h(t-\tau)\, d\tau = \int_{t-T}^{t} x(\tau)\, f(\tau+T-t)\cos\omega(\tau+T-t)\, d\tau \\ &= \cos\omega(T-t)\int_{t-T}^{t} x(\tau)\, f(\tau+T-t)\cos\omega\tau\ d\tau \\ &\quad - \sin\omega(T-t)\int_{t-T}^{t} x(\tau)\, f(\tau+T-t)\sin\omega\tau\ d\tau . \end{aligned} \tag{91}$$

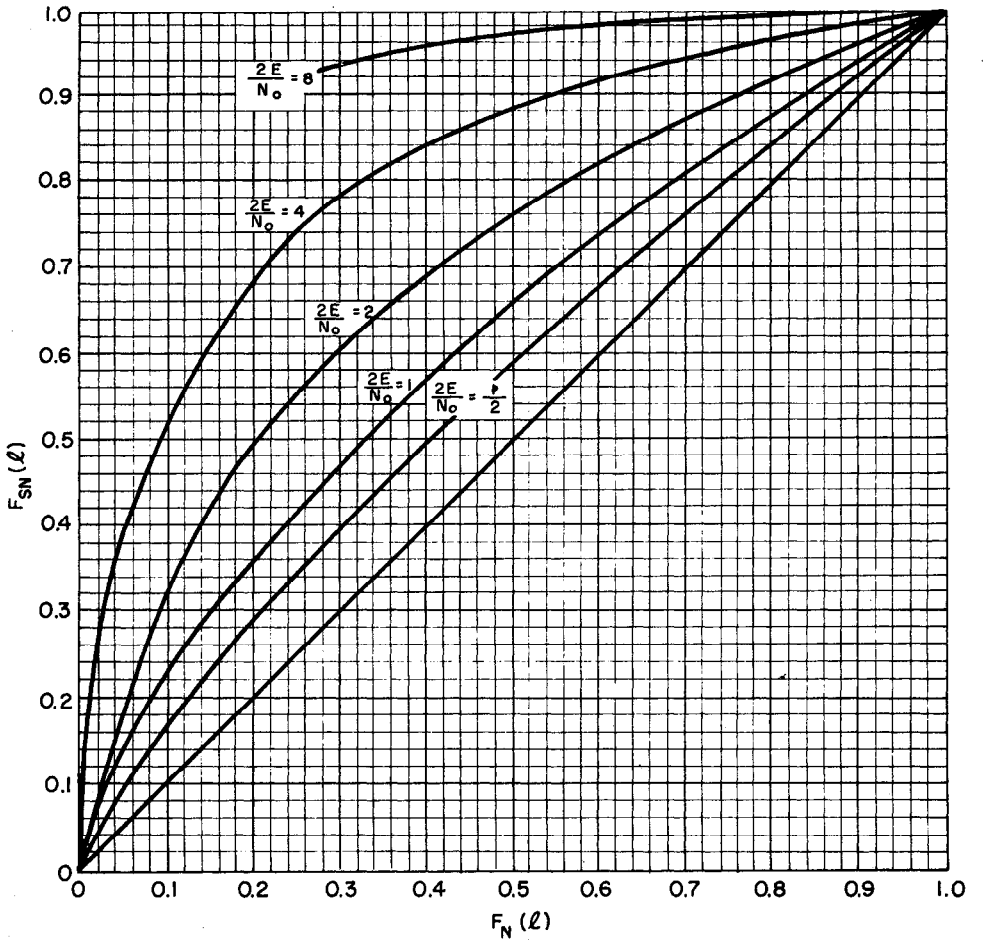

Fig. 4

RECEIVER OPERATING CHARACTERISTIC.

SIGNAL KNOWN EXCEPT FOR PHASE.

The envelope of the filter output will be the square root of the sum of the squares of the integrals;* and the envelope at time T will be proportional to r/N, since

$$\left(\frac{r}{2W}\right)^2 = \left[\int_0^T x(\tau)\, f(\tau) \cos \omega\tau \, d\tau\right]^2 + \left[\int_0^T x(\tau)\, f(\tau) \sin \omega\tau \, d\tau\right]^2, \qquad (92)$$

which can be identified as the square of the envelope of $e_o(t)$ at time T. If the input x(t) passes through the filter with an impulse response given by Eq. (90), then through a linear detector, the output will be $(N_o/2)r/N$ at time T. Because the likelihood ratio, Eq. (82), is a known monotone function of r/N, the output can be calibrated to read the likelihood ratio of the input.

---

* If the line spectrum of s(t) is zero at zero frequency and at all frequencies equal to or greater than $2w/2\pi$, then it can be shown that these integrals contain no frequencies as high as $\omega/2\pi$.

4.6 Signal Consisting of a Sample of White Gaussian Noise

Suppose the values of the signal voltage at the sample points are independent Gaussian random variables with zero mean and variance S, the signal power. The probability density due to signal plus noise is also Gaussian, since signal plus noise is the sum of two Gaussian random variables:

$$f_{SN}(x) = \left(\frac{1}{2\pi(N+S)}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\frac{1}{N+S}\sum x_i^2\right], \tag{93}$$

where n = 2WT.

The likelihood ratio is

$$\ell(x) = \left(\frac{N}{N+S}\right)^{\frac{n}{2}} \exp\left[\frac{1}{2}\frac{1}{N}\sum x_i^2 - \frac{1}{2}\frac{1}{N+S}\sum x_i^2\right] . \tag{94}$$

In determining the distribution functions for ℓ, it is convenient to introduce the parameter α, defined by the equation

$$\beta = \left(\frac{N}{N+S}\right)^{\frac{n}{2}} \exp\left(\frac{S}{N+S}\frac{\alpha^2}{2}\right) . \tag{95}$$

Then the condition $\ell(x) \geq \beta$ is equivalent to the condition that $(1/N)\sum x_i^2 \geq \alpha^2$. In the presence of noise alone the random variables $x_i/\sqrt{N}$ have zero mean and unit variance, and they are independent. Therefore, the probability that the sum of the squares of these variables will exceed α^2 is the chi-square distribution with n degrees of freedom, i.e.,

$$F_N(\beta) = K_n(\alpha^2) . \tag{96}$$

Similarly, in the presence of signal plus noise the random variables $x_i/\sqrt{N+S}$ have zero mean and unit variance. The condition $(1/N)\sum x_i^2 \geq \alpha^2$ is the same as requiring that $(1/N+S)\sum x_i^2 \geq (N/N+S)\alpha^2$, and again making use of the chi-square distribution,

$$F_{SN}(\beta) = K_n\left(\frac{N}{N+S}\alpha^2\right) . \tag{97}$$

For large values of n, the chi-square distribution is approximately normal over the center portion; more precisely,[16] for $\alpha^2 >> 0$,

$$F_N(\beta) = K_n(\alpha^2) \approx \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2\alpha^2}-\sqrt{2n-1}}^{\infty} \exp\left[-\frac{1}{2}y^2\right] dy \tag{98}$$

and

$$F_{SN}(\beta) = K_n\left(\frac{N}{N+S}\alpha^2\right) \approx \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\frac{2N\alpha^2}{N+S}}-\sqrt{2n-1}}^{\infty} \exp\left[-\frac{1}{2}y^2\right] dy . \tag{99}$$

If the signal energy is small compared to that of the noise, $\sqrt{N/N+S}$ is nearly unity and both distribu-

tions have nearly the same variance. Then Figs. 2 and 3 apply to this case too, with the value of d given by

$$d = (2n-1)\left(1 - \sqrt{\frac{N}{N+S}}\right)^2 \tag{100}$$

For these small signal to noise ratios and large samples, there is a simple relation between signal to noise ratio, the number of samples, and the detection index d.

$$1 - \sqrt{\frac{N}{N+S}} \approx \frac{1}{2}\frac{S}{N} \qquad \text{for } \frac{S}{N} << 1 \quad , \quad \text{and}$$
$$d \approx \frac{nS^2}{2N^2} \tag{101}$$

Two signal to noise ratios, $(S/N)_1$ and $(S/N)_2$, will give approximately the same operating characteristic if the corresponding numbers of sample points, n_1 and n_2, satisfy

$$\frac{n_1}{n_2} = \frac{\left(\frac{S}{N}\right)_1^2}{\left(\frac{S}{N}\right)_2^2} \tag{102}$$

By Eq. (94), the likelihood is a monotone function of $\sum x_i^2$. But the output of an energy detector,

$$e_o(t) = \int_0^T \left[x(t)\right]^2 dt = \frac{1}{2W}\sum x_i^2 \tag{103}$$

is proportional to $\sum x_i^2$. Therefore an energy detector can be calibrated to read likelihood ratio, and hence can be used as an optimum receiver in this case.

4.7 Video Design of a Broad Band Receiver

The problem considered in this section is represented schematically in Fig. 5. The signals

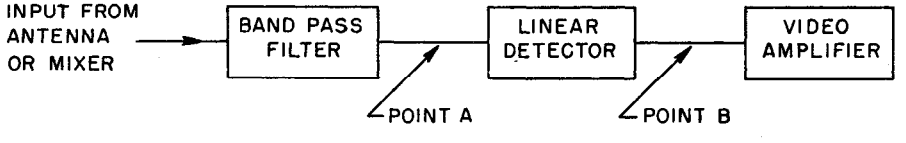

Fig. 5

BLOCK DIAGRAM OF A BROAD BAND RECEIVER

and noise are assumed to have passed through a band pass filter, and at the output of the filter, point A on the diagram, they are assumed to be limited in spectrum to a band of width W and center frequency $\omega/2\pi > W/2$. The noise is assumed to be Gaussian noise with a uniform spectrum over the band. The signals and noise then pass through a linear detector. The output of the detector is the envelope of the signals and noise as they appeared at point A; all knowledge of the phase of the receiver input is lost at point B. The signals and noise as they appear at point B are considered receiver inputs,

and the theory of signal detectability is applied to these video inputs to ascertain the best video design and the performance of such a system. The mathematical description of the signals and noise will be given for the signals and noise as they appear at point A. The envelope functions, which appear at point B, will be derived, and the likelihood ratio and its distribution will be found for these envelope functions.

The only case which will be considered here is the case in which the amplitude of the signal as it would appear at point A is a known function of time.

Any function at point A will be band limited to a band of width W and center frequency $\omega/2\pi > W/2$. Any such function f(t) can be expanded as follows:

$$f(t) = x(t) \cos \omega t + y(t) \sin \omega t \tag{105}$$

where x(t) and y(t) are band limited to frequencies no higher than W/2, and hence can themselves* be expanded by sampling plan C, yielding

$$f(t) = \sum_i \left[x\left(\frac{i}{W}\right) \psi_i(t) \cos \omega t + y\left(\frac{i}{W}\right) \psi_i(t) \sin \omega t \right]. \tag{106}$$

The amplitude of the function f(t) is

$$r(t) = \sqrt{\left[x(t)\right]^2 + \left[y(t)\right]^2} \tag{107}$$

and thus the amplitude at the i^{th} sampling point is

$$r\left(\frac{i}{W}\right) = r_i = \sqrt{x_i^2 + y_i^2} \quad . \tag{108}$$

The angle

$$\theta_i = \arctan \frac{y_i}{x_i} = \arccos \frac{x_i}{r_i} \tag{109}$$

might be considered the phase of f(t) at the i^{th} sampling point. The function f(t) then might be described by giving the r_i and θ_i rather than the x_i and y_i.

Let us denote by x_i, y_i, or r_i, θ_i, the sample values for a receiver input after the filter (i.e., at the point A in Fig. 5). Let a_i, b_i, or f_i, ϕ_i, denote the sample values for the signal as it would appear at point A if there were no noise. The envelope of the signal, hence the amplitude sample values f_i, are assumed known. Let us denote by $F_S(\phi_1, \phi_2, \ldots, \phi_{n/2})$ the distribution function of the phase sample values ϕ_i. The probability density function for the input at A when there is white Gaussian noise and no signal, with n = 2WT, is

$$f_N(x, y) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2N} \sum_{i=1}^{n/2} x_i^2 + \sum_{i=1}^{n/2} y_i^2 \right] \tag{110}$$

and for signal plus noise, it is

$$f_{SN}(x, y) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \int_R \exp\left[-\frac{1}{2N} \left(\sum_{i=1}^{n/2} (x_i - a_i)^2 + \sum_{i=1}^{n/2} (y_i - b_i)^2 \right) \right] dP_S(a_i b_i) \quad . \tag{111}$$

---

* Because any function f(t) at A has no frequency greater than $(\omega/2\pi) + (W/2)$, the usual sampling plan C might have been used on f(t). However, the distribution in noise alone, $f_N(x_i)$, would probably not be applicable.

Expressed in terms of the (r, θ) sample values, Eq. (110) and Eq. (111) become

$$f_N(r,\theta) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \prod_{i=1}^{n/2} r_i \exp\left[-\frac{1}{2N}\sum_{i=1}^{n/2} r_i^2\right], \tag{112}$$

and

$$f_{SN}(r,\theta) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \prod_{i=1}^{n/2} r_i \int_R \exp\left[-\frac{1}{2N}\sum_{i=1}^{n/2}\left\{r_i^2+f_i^2-2r_i f_i\cos(\theta_i-\phi_i)\right\}\right] dF_S\left(\phi_1,\ldots,\phi_{\frac{n}{2}}\right). \tag{113}$$

The factors $\prod r_i$ are introduced because they are the Jacobian of the transformation from the x, y sampling plan to the r, θ sampling plan.[16], *

The probability density function for r alone, i.e., the density function for the output of the detector, is obtained simply by integrating the density functions for r and θ with respect to θ.

$$f_N(r) = \int_0^{2\pi}\int_0^{2\pi}\cdots\int_0^{2\pi} f_N(r_i,\theta_i)\, d\theta_1\, d\theta_2 \cdots d\theta_{\frac{n}{2}}, \tag{114}$$

or

$$f_N(r) = \left(\frac{1}{N}\right)^{\frac{n}{2}} \prod_{i=1}^{n/2} r_i \exp\left[-\frac{1}{2N}\sum_{i=1}^{n/2} r_i^2\right],$$

and

$$f_{SN}(r) = \int_0^{2\pi}\int_0^{2\pi}\cdots\int_0^{2\pi} f_{SN}(r_i,\theta_i)\, d\theta_1\, d\theta_2\cdots d\theta_{\frac{n}{2}};$$

or

$$f_{SN}(r) = \left(\frac{1}{N}\right)^{\frac{n}{2}} \int_R \prod_{i=1}^{n/2} r_i \exp\left[-\frac{1}{2N}\sum_{i=1}^{n/2}(r_i^2+f_i^2)\right] \prod_{i=1}^{n/2} I_o\left(\frac{r_i f_i}{N}\right) dF(\phi_1,\phi_2\cdots\phi_{\frac{n}{2}}), \tag{115}$$

or

$$f_{SN}(r) = \left(\frac{1}{N}\right)^{\frac{n}{2}} \prod_{i=1}^{n/2} r_i\, I_o\left(\frac{r_i f_i}{N}\right) \exp\left[-\frac{1}{2N}\sum_{i=1}^{n/2}(r_i^2+f_i^2)\right].$$

Notice that the probability density for r is completely independent of the distribution which the ϕ_i had; all information about the phase of the signals has been lost.

The likelihood ratio for a video input r(t), is

$$\ell(r) = \frac{f_{SN}(r)}{f_N(r)} = \exp\left[-\frac{1}{2N}\sum_{i=1}^{n/2} f_i^2\right] \prod_{i=1}^{n/2} I_o\left(\frac{r_i f_i}{N}\right). \tag{116}$$

---

* For example, in two dimensions, $f_N(x,y)dx\,dy = f_N(r,\theta)\, r\, dr\, d\theta$.

Again it is more convenient to work with the logarithm of the likelihood ratio. Thus

$$\frac{1}{2N}\sum_{i=1}^{n/2} f_i^2 = \frac{W}{2N}\int \left[f(t)\right]^2 dt = \frac{E}{N_o}, \text{ and} \tag{117}$$

$$\ell n\ \ell\ (r) = -\frac{E}{N_o} + \sum_{i=1}^{n/2} \ell n\ I_o\left(\frac{r_i f_i}{N}\right) , \tag{118}$$

which is approximately

$$\ell n\ \ell\big(r(t)\big) = -\frac{E}{N_o} + W \int_0^T \ell n\ I_o\left(\frac{r(t)\ f(t)}{N}\right)dt . \tag{119}$$

The function $\ell n\ I_o(x)$ is approximately the parabola $x^2/4$ for small values of x and is nearly linear for large values of x. Thus, the expression for likelihood ratio might be approximated by

$$\ell n\,\ell\big(r(t)\big) = -\frac{E}{N_o} + \frac{W}{4N^2}\int_0^T \left[r(t)\right]^2\left[f(t)\right]^2 dt \tag{120}$$

for small signals, and by

$$\ell n\,\ell\big(r(t)\big) = C_1 + C_2 \int_0^T r(t)\ f(t)\ dt \tag{121}$$

for large signals, where C_1 and C_2 are chosen to approximate $\ell n\ I_o$ best in the desired range.

The integrals in Eqs. (120) and (121) can be interpreted as cross correlations. Thus the optimum receiver for weak signals is a square law detector, followed by a correlator which finds the cross correlation between the detector output and $(f(t))^2$, the square of the envelope of the expected signal. For the case of large signal to noise ratio, the optimum receiver is a linear detector, followed by a correlator which has for its output the cross correlation of the detector output and f(t), the amplitude of the expected signal.

The distribution function for $\ell(r)$ cannot be found easily in this case. The approximation developed here will apply to the receiver designed for low signal to noise ratio, since this is the case of most interest in detection studies. An analogous approximation for the large signal to noise ratios would be even easier to derive.

First we shall find the mean and standard deviation for the distribution of the logarithm of the likelihood ratio as shown above,

$$\ell n\ \ell(r) \approx -\frac{1}{2N}\sum f_i^2 + \frac{1}{4N^2}\sum_{i=1}^{n/2} r_i^2 f_i^2 , \tag{122}$$

for the case of small signal to noise ratio. The probability density functions for each r_i are

$$g_{SN}(r_i) = \frac{r_i}{N}\ \exp\left[-\frac{r_i^2+f_i^2}{2N}\right] I_o\left[\frac{r_i f_i}{N}\right], \text{ and} \tag{123}$$

$$g_N(r_i) = \frac{r_i}{N}\ \exp\left[-\frac{r_i^2}{2N}\right] .$$

The notation $g_N(r_i)$ and $g_{SN}(r_i)$ is used to distinguish these from the joint distributions of all the r_i which were previously called $f_N(r)$ and $f_{SN}(r)$. The mean of each term $r_i^2 f_i^2/4N^2$ in the sum in Eq. (122) is

$$\mu_{SN}\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{f_i^2}{4N}\int_0^\infty \frac{r_i^2}{N}\ g_{SN}(r_i)\ dr_i , \text{ or} \tag{124}$$

$$\mu_{SN}\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{f_i^2}{4N}\int_0^\infty \frac{r_i^3}{N^2}\exp\left[-\frac{(r_i^2+f_i^2)}{2N}\right] I_o\left(\frac{r_i f_i}{N}\right) dr_i \ .$$

Similarly,

$$\mu_N\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{f_i^2}{4N}\int_0^\infty \frac{f_i^2}{N} g_N(r_i)\, dr_i = \frac{f_i^2}{4N}\int_0^\infty \frac{r_i^3}{N^2}\exp\left[-\frac{r_i^2}{2N}\right] dr_i \qquad (124)$$

The second moment of each term $r_i^2 f_i^2/4N^2$ is

$$\mu_{SN}\left(\frac{r_i^4 f_i^4}{16N^4}\right) = \frac{f_i^4}{16N^2}\int_0^\infty \frac{r_i^4}{N^2} g_{SN}(r_i)\, dr_i \ , \text{ or}$$

$$\mu_{SN}\left(\frac{r_i^4 f_i^4}{16N^4}\right) = \frac{f_i^4}{16N^2}\int_0^\infty \frac{r_i^5}{N^3}\exp\left[-\frac{(r_i^2+f_i^2)}{2N}\right] I_o\left(\frac{r_i f_i}{N}\right) dr_i \ . \qquad (125)$$

Similarly,

$$\mu_N\left(\frac{r_i^4 f_i^4}{16N^4}\right) = \frac{f_i^4}{16N^2}\int_0^\infty \frac{r_i^4}{N^2} g_N(r_i)\, dr_i \ , \text{ or}$$

$$\mu_N\left(\frac{r_i^4 f_i^4}{16N^4}\right) = \frac{f_i^4}{16N^2}\int_0^\infty \frac{r_i^5}{N^3}\exp\left[-\frac{r_i^2}{2N}\right] dr_i \ .$$

The integrals for the case of noise alone can be evaluated easily:

$$\mu_N\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{f_i^2}{2N} \ , \qquad (126)$$

and

$$\mu_N\left(\frac{r_i^4 f_i^4}{16N^4}\right) = \frac{f_i^4}{2N^2} \ .$$

The integrals for the case of signal plus noise can be evaluated in terms of the confluent hypergeometric function, which turns out for the cases above to reduce to a simple polynomial. The required formulas are collected in convenient form in Threshold Signals[5] on page 174. The results are

$$\mu_{SN}\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{1}{2}\frac{f_i^2}{N}\left(1+\frac{f_i^2}{2N}\right) \ ,$$

and

$$\mu_{SN}\left(\frac{r_i^4 f_i^4}{16N^4}\right) = \frac{1}{2}\frac{f_i^4}{N^2}\left(1+\frac{f_i^2}{N}+\frac{f_i^4}{8N^2}\right) . \qquad (127)$$

Since

$$\sigma^2(Z) = \mu(Z^2) - \left[\mu(Z)\right]^2, \tag{128}$$

the variances of $r_i^2 f_i^2/4N^2$ are

$$\sigma^2_{SN}\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{1}{4}\,\frac{f_i^4}{N^2}\left(1 + \frac{f_i^2}{N}\right)$$

and (129)

$$\sigma^2_{N}\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{f_i^4}{4N^2} \quad .$$

For the sum of independent random variables, the mean is the sum of the means of the terms and the variance is the sum of the variances. Therefore the means of $\ell n\, \ell(r)$ are

$$\mu_{SN}\left(\ell n\, \ell(r)\right) = -\frac{1}{2N}\sum_{i=1}^{n/2} f_i^2 + \sum_{i=1}^{n/2}\left[\frac{1}{2}\,\frac{f_i^2}{N} + \frac{1}{4}\,\frac{f_i^4}{N^2}\right] = \sum_{i=1}^{n/2}\frac{f_i^4}{4N^2}$$

and (130)

$$\mu_{N}\left(\ell n\, \ell(r)\right) = -\sum_{i=1}^{n/2}\frac{f_i^2}{2N} + \frac{1}{2}\sum_{i=1}^{n/2}\frac{f_i^2}{N} = 0 \,,$$

and the variances of $\ell n\, \ell(r)$ are

$$\sigma^2_{SN}\left(\ell n\, \ell(r)\right) = \sum_{i=1}^{n/2}\left(\frac{1}{4}\,\frac{f_i^4}{N^2} + \frac{1}{4}\,\frac{f_i^6}{N^3}\right)$$

and (131)

$$\sigma^2_{N}\left(\ell n\, \ell(r)\right) = \sum_{i=1}^{n/2}\frac{f_i^4}{4N^2} \quad .$$

If the distribution functions of $\ell n\, \ell(r)$ can be assumed to be normal, they can be obtained immediately from the mean and standard deviation of the logarithm of likelihood ratio.

Let us consider the case in which the incoming signal is a rectangular pulse which is M/W seconds long.* The energy of the pulse is half its duration times the amplitude squared of its envelope, for a normalized circuit impedance of one ohm.

---

* The problem of finding the distribution for the sum of M independent random variables, each with a probability density function $f(x) = x \exp\left[-(1/2)(x^2 + a^2)\right] I_o(ax)$ arises in the unpublished report by J. I. Marcum, A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix, Project Rand Report R - 113. Marcum gives an exact expression for this distribution which is useful only for small values of M, and an approximation in Gram-Charlier series which is more accurate than the normal approximation given here. Marcum's expressions could be used in this case, and in the case presented in Section 4.6.

Thus of the WT numbers $\{f_i\}$, there are M consecutive ones which are not zero. These are given by

$$f_i = \sqrt{\frac{2EW}{M}} \quad , \tag{132}$$

where E is the pulse energy at point A in Fig. 5 in the absence of noise. For this case, Eq. (130) and Eq. (131) become

$$\begin{aligned} \mu_{SN}\left(\ell n\ \ell(r)\right) &= \frac{1}{M}\,\frac{E^2}{N_o^2} \quad , \\ \mu_{N}\left(\ell n\ \ell(r)\right) &= 0 \quad , \\ \sigma_{SN}^2\left(\ell n\ \ell(r)\right) &= \frac{E^2}{MN_o^2}\left(1+\frac{2}{M}\frac{E}{N_o}\right) , \end{aligned} \tag{133}$$

and

$$\sigma_{N}^2\left(\ell n\ \ell(r)\right) = \frac{E^2}{MN_o^2} \quad .$$

The distribution of $\ell n\ \ell(r)$ is approximately normal if M is much larger than one, for, by the central limit theorem, the distribution of a sum of M independent random variables with a common distribution must approach the normal distribution as M becomes large. The actual distribution for the case of noise alone can be calculated in this case, since the convolution integral for the $g_N(r_i)$ with itself any number of times can be expressed in closed form. The distribution of $\ell n\ \ell(r)$ for signal plus noise is more nearly normal than its distribution with noise alone, since the distributions $g_{SN}(r_i)$ are more nearly normal than $g_N(r_i)$.

The receiver operating characteristic for the case M = 16 is plotted in Fig. 6 using the normal distribution as approximation to the true distribution. In many cases it will be found that

$$\frac{1}{M}\cdot\frac{2E}{N_o} << 1 \quad . \tag{134}$$

In such a case the distributions have approximately the same variance. Assuming normal distribution then leads to the curves of Figs. 2 and 3, with

$$d = \frac{1}{4M}\left(\frac{2E}{N_o}\right)^2 \quad . \tag{135}$$

4.8 A Radar Case

This section deals with detecting a radar target at a given range. That is, we shall assume that the signal, if it occurs, consists of a train of M pulses whose time of occurrence and envelope shape are known. The carrier phase will be assumed to have a uniform distribution for each pulse independent of all others, i.e., the pulses are incoherent.

The set of signals can be described as follows:

$$s(t) = \sum_{m=0}^{M-1} f(t+m\tau)\cos\left(\omega t+\theta_i\right) , \tag{136}$$

where the M angles θ_i have independent uniform distributions, and the function f, which is the envelope of a single pulse, has the property that

$$\int_0^T f(t+i\tau)\, f(t+j\tau)\, dt = \frac{2E}{M}\,\delta_{ij} \quad , \tag{137}$$

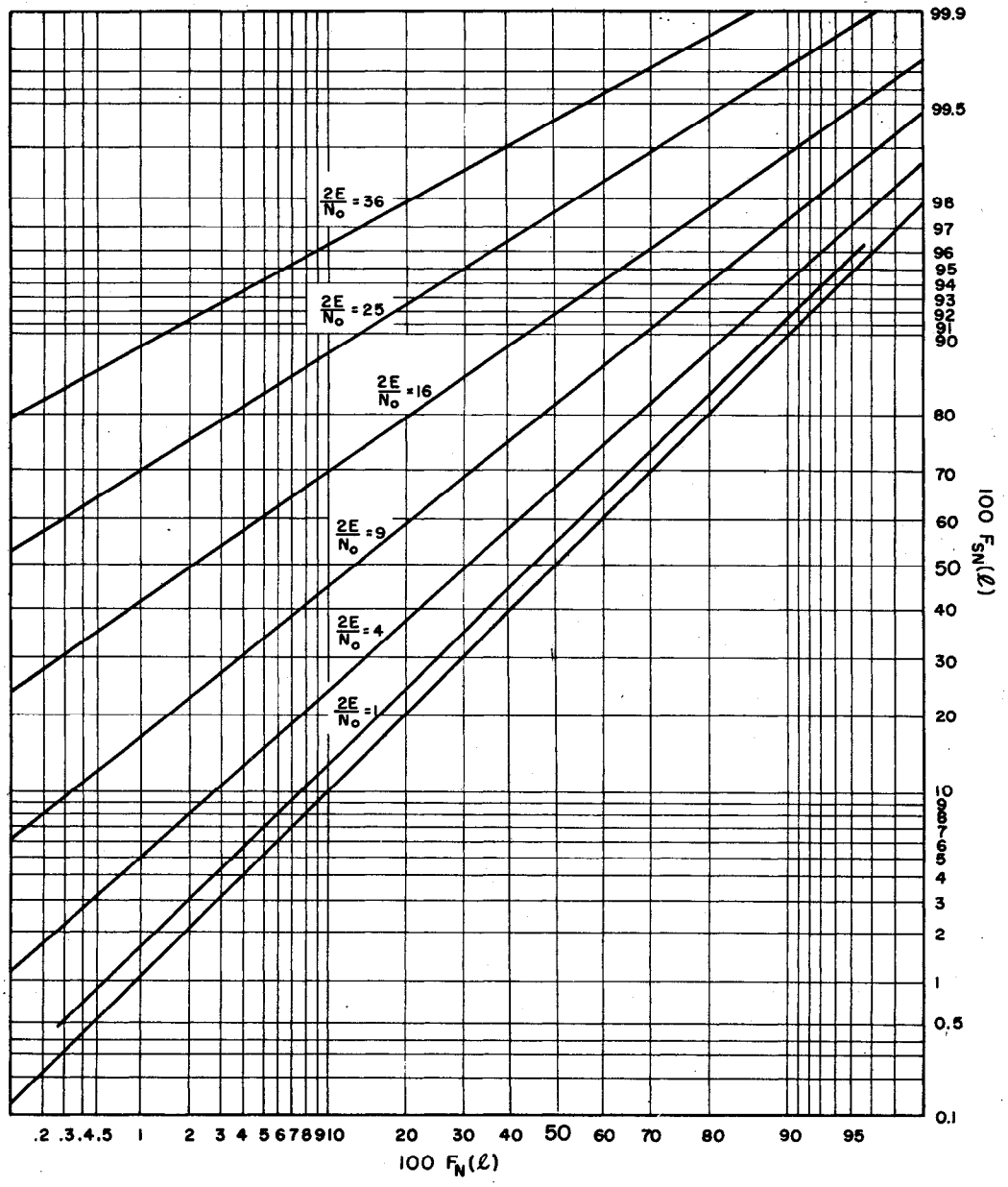

Fig. 6

RECEIVER OPERATING CHARACTERISTIC

BROAD BAND RECEIVER WITH
OPTIMUM VIDEO DESIGN, M = 16

where δ_{ij} is the Kronecker delta function, which is zero if $i \neq j$, and unity if $i = j$. The time τ is the interval between pulses. Eq. (137) states that the pulses are spaced far enough so that they are orthogonal, and that the total signal energy is E.* The function f(t) is also assumed to have no frequency components as high as $\omega/2\pi$.

The likelihood ratio can be obtained by applying Eq. (56). Then

$$\ell(x) = \int_R \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{2}{N_o}\int_0^T s(t)\, x(t)\, dT\right] dP_S(s) \tag{138}$$

or

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \int_0^{2\pi} \cdots \int_0^{2\pi} \exp\left[\frac{2}{N_o}\int_0^T \sum_{m=0}^{M-1} f(t+m\tau)x(t)\cos(\omega t+\theta_m)dt\right] d\theta_o \cdots d\theta_{M-1} \quad . \tag{139}$$

The integral can be evaluated, as in Section 4.5, yielding

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \prod_{m=0}^{M-1} I_o\left(\frac{r_m}{N}\right) \quad , \tag{140}$$

where

$$\left(\frac{r_m}{N}\right)^2 = \left[\frac{2}{N_o}\int_0^T f(t+m\tau)x(t)\cos\omega t\, dt\right]^2 + \left[\frac{2}{N_o}\int_0^T f(t+m\tau)\, x(t)\sin\omega t\, dt\right]^2 . \tag{141}$$

This quantity r_m is almost identical with the quantity r which appeared in the discussion of the case of the signal known except for carrier phase, Section 4.5. In fact, each r_m could be obtained in a receiver in the manner described in that section. The quantity r_o is connected with the first pulse; it could be obtained by designing an ideal filter for the signal

$$s_o(t) = f(t)\cos(\omega t+\theta) \tag{142}$$

for any value of the phase angle θ, and putting the output through a linear detector. The output will be $(N_o/2)r_o/N$ at some instant of time t_o which is determined by the time delay of the filter. The other quantities r_m differ only in that they are associated with the pulses which come later. The output of the filter at time $t_o + m\tau$ will be $(N_o/2)r_m/N$.

It is convenient to have the receiver calculate the logarithm of the likelihood ratio,

$$\ln \ell(x) = -\frac{E}{N_o} + \sum_{m=0}^{M-1} \ln I_o\left(\frac{r_m}{N}\right) . \tag{143}$$

Thus the $\ln I_o(r_m/N)$ must be found for each r_m, and these M quantities must be added. As in the previous section, r_m/N will usually be small enough so that $\ln I_o(x)$ can be approximated by $x^2/4$. The quantities $1/4\,(r_m/N)^2$ can be found by using a square law detector rather than a linear detector, and the outputs of the square law detector at times t_o, $t_o+\tau$, ... , $t_o + (M-1)\tau$ then must be added. The ideal system thus consists of an IF amplifier with its passband matched to a single pulse,** a

---

* The factor 2 appears in (137) because f(t) is the pulse envelope; the factor M appears because the total energy E is M times the energy of a single pulse.

** It is usually most convenient to made the ideal filter (or an approximation to it) a part of the IF amplifier

square law detector (for the threshold signal case), and an integrating device.

We shall find normal approximations for the distribution functions of the logarithm of the likelihood ratio using the approximation

$$\ell n\, I_o\left(\frac{r_m}{N}\right) \approx \frac{r_m^2}{4N^2} \tag{144}$$

which is valid for small values of r_m/N.* Substitution of (144) into (143) yields

$$\ell n\, \ell \approx -\frac{E}{N_o} + \sum_{n=o}^{M-1} \frac{1}{4}\left(\frac{r_m}{N}\right)^2 . \tag{145}$$

The distributions for the quantities r_m are independent; this follows from the fact that the individual pulse functions $f(t+m\tau)\cos(\omega t+\theta_m)$ are orthogonal. The distribution for each is the same as the distribution for the quantity r which appears in the discussion of the signal known except for phase; the same analysis applies to both cases. Thus, by Eq. (83)**

$$P_N\left(\frac{r_m}{N}\sqrt{\frac{N_o M}{2E}} \geq \alpha\right) = \exp\left[-\frac{\alpha^2}{2}\right]$$

$$P_N\left(\frac{r_m}{N} \geq a\right) = \exp\left[-\frac{a^2 N_o M}{2E}\right] , \tag{146}$$

and by (89),

$$P_{SN}\left(\sqrt{\frac{N_o M}{2E}}\,\frac{r_m}{N} \geq \alpha\right) = \exp\left[-\frac{E}{N_o}\right]\int_{\alpha}^{\infty} \alpha \exp\left[-\frac{\alpha^2}{2}\right] I_o\left(\alpha\sqrt{\frac{2E}{N_o M}}\right) d\alpha \tag{147}$$

or

$$P_{SN}\left(\frac{r_m}{N} \geq a\right) = \frac{N_o M}{2E}\exp\left[-\frac{E}{N_o M}\right]\int_{a}^{\infty} a \exp\left(-\frac{a^2 N_o M}{4E}\right) I_o(a)\, da .$$

The density functions can be obtained by differentiating (146) and (147):

$$g_N\left(\frac{r_m}{N}\right) = \frac{MN_o}{2E}\left(\frac{r_m}{N}\right)\exp\left[-\left(\frac{r_m}{N}\right)^2\left(\frac{N_o M}{4E}\right)\right] ;$$

$$g_{SN}\left(\frac{r_m}{N}\right) = \frac{MN_o}{2E}\left(\frac{r_m}{N}\right)\exp\left[-\frac{E}{MN_o}\right]\exp\left[-\left(\frac{r_m}{N}\right)^2\left(\frac{N_o M}{4E}\right)\right] I_o\left(\frac{r_m}{N}\right). \tag{148}$$

---

* See the footnote below equation (131).

** The M appears in the following equations because the energy of a single pulse is E/M rather than E.

This is the same situation, mathematically, as appeared in the previous section. The standard deviation and the mean for the logarithm of the likelihood ratio can be found in the same manner, and they are

$$\mu_{SN}(\ell n\, \ell) = \frac{E^2}{MN_o^2} ,$$

$$\mu_{N}(\ell n\, \ell) = 0 ,$$

$$\sigma_{SN}^2(\ell n\, \ell) = \frac{E^2}{MN_o^2}\left(1 + \frac{2E}{MN_o}\right),$$

and

$$\sigma_{N}^2(\ell n\, \ell) = \frac{E^2}{MN_o^2} . \qquad (1$$

If the distributions can be assumed normal, they are completely determined by their means and variances. These formulas are identical with the formulas (133) of the previous section. The proble is the same, mathematically, and the discussion and receiver operating characteristic curves at the end of Section 4.7 apply to both cases.

4.9 Approximate Evaluation of an Optimum Receiver

In order to obtain approximate results for the remaining two cases, the assumption is made that in these cases the receiver operating characteristic can be approximated by the curves of Figs. 2 and 3, i.e., that the logarithm of the likelihood ratio is approximately normal. This section discusses the approximation and a method for fitting the receiver operating characteristic to the curves of Figs. 2 and 3.

By (68), $F_{SN}(\ell)$ can be calculated if $F_N(\ell)$ is known. Furthermore, it can be seen that the n^{th} moment of the distribution $F_N(\ell)$ is the $(n-1)^{th}$ moment of the distribution $F_{SN}(\ell)$. Hence, the mean of the likelihood ratio with noise alone is unity, and if the variance of the likelihood ratio with noise alone is σ_N^2, the second moment with noise alone, and hence the mean with signal plus noise, i $1 + \sigma_N^2$. Thus the difference between the means is equal to σ_N^2, which is the variance of the likelihood ratio with noise alone. Probably this number characterizes ability to detect signals bett than any other single number.

Suppose the logarithm of the likelihood ratio has a normal distribution with noise alone, i.e.,

$$F_N(\ell) = \frac{1}{\sqrt{2\pi d}} \int_{\ell n\, \ell}^{\infty} \exp\left[-\frac{(x-m)^2}{2d}\right] dx, \qquad (1$$

where m is the mean and d the variance of the logarithm of the likelihood ratio. The n^{th} moment of the likelihood ratio can be found as follows:

$$\mu_N(\ell^n) = \int_0^{\infty} \ell^n\, dF_N(\ell) = \frac{1}{\sqrt{2\pi d}} \int_{-\infty}^{\infty} \exp\left[nx\right]\exp\left[-\frac{(x-m)^2}{2d}\right] dx , \qquad (1$$

where the substitution $\ell = \exp x$ has been made. The integral can be evaluated by completing the squ in the exponent and using the fact that

$$\int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2d}\right] dx = \sqrt{2\pi d} .$$

Thus

$$\mu_N(\ell^n) = \exp\left[\frac{n^2 d}{2} + nm\right] . \qquad (1$$

In particular, the mean of $\ell(x)$, which must be unity, is

$$\mu_N(\ell) = 1 = \exp\left[\frac{d}{2} + m\right] , \qquad (1$$

and therefore

$$m = -\frac{d}{2} \quad . \tag{154}$$

The variance of $\ell(x)$ with noise alone is σ_N^2, and therefore the second moment of $\ell(x)$ is

$$\mu_N(\ell^2) = \left[\mu_N(\ell)\right]^2 + \sigma_N^2(\ell) = 1 + \sigma_N^2(\ell) \quad , \tag{155}$$

and this must agree with (152). It follows that

$$\mu_N(\ell^2) = 1 + \sigma_N^2 = \exp\left[2d + 2m\right] = \exp\left[d\right] \tag{156}$$

and therefore

$$d = \ell n\ (1 + \sigma_N^2)$$

The distribution of likelihood ratio with signal plus noise can be found by applying Eq. (68). Thus

$$dF_{SN}(\ell) = \ell\, dF_N(\ell) \quad ,$$
$$F_{SN}(\ell) = -\int_{\ell}^{\infty} \ell\, dF_N(\ell) \tag{158}$$

If $dF_N(\ell)$ is obtained from Eq. (150) and ℓ is replaced by exp x, then

$$F_{SN}(\ell) = \frac{1}{\sqrt{2\pi d}} \int_{\ell n \ell}^{\infty} \exp\left[x\right] \exp\left[-\frac{(x + \frac{d}{2})^2}{2d}\right] dx$$

or

$$F_{SN}(\ell) = \frac{1}{\sqrt{2\pi d}} \int_{\ell n \ell}^{\infty} \exp\left[-\frac{(x - \frac{d}{2})^2}{2d}\right] dx \quad . \tag{159}$$

Thus the distribution of $\ell n \ell$ is normal also when there is signal plus noise, in this case with mean d/2 and variance d.

In summary, it is probable that the variance σ_N^2 of the likelihood ratio measures ability to detect signals better than any other single number. If the logarithm of likelihood ratio has a normal distribution with noise alone, then this distribution and that with signal plus noise are completely determined if σ_N^2 is given. The distribution of $\ell n\ \ell(x)$ is normal in both cases. Its variance in both cases is d, which is also the difference of the means. The receiver operating characteristic curves are those plotted in Fig. 2, with the parameter d related to σ_N^2 by the equation

$$d = \ell n\ (1 + \sigma_N^2) \tag{160}$$

In the case of a signal known exactly, this is the distribution which occurs. In the cases of Section 4.6, Section 4.7, and Section 4.8 this distribution is found to be the limiting distribution when the number of sample points is large. Certainly in most cases the distribution has this general form. Thus it seems reasonable that useful approximate results could be obtained by calculating only σ_N^2 for a given case and assuming that the ability to detect signals is approximately the same as if the logarithm of the likelihood ratio had a normal distribution. On this basis, $\sigma_N^2(\ell)$ is calculated in the following sections for two cases, and the assertion is made that the receiver operating characteristic curves are approximated by those of Fig. 2 with $d = \ell n\ (1 + \sigma_N^2)$.

4.10 Signal Which is One of M Orthogonal Signals

Suppose that the set of expected signals includes just M functions $s_k(t)$, all of which have the same probability, the same energy E, and are orthogonal. That is,

$$\int_0^T s_k(t)\, s_q(t)\, dt = E\,\delta_{kq}. \tag{161}$$

Then the likelihood ratio can be found from Eq. (56) to be

$$\ell(x) = \sum_{k=1}^{M} \frac{1}{M} \exp\left[-\frac{E}{N_o}\right] \exp\left[\frac{1}{N}\sum_{i=1}^{n} x_i s_{ki}\right],$$

or

$$\ell(x) = \frac{1}{M}\sum_{k=1}^{M} \exp\left[\frac{1}{N}\sum_{i=1}^{n} x_i s_{ki} - \frac{E}{N_o}\right], \tag{162}$$

where s_{ki} are the sample values of the function $s_k(t)$.

With noise alone, each term of the form $(1/N)\sum_{i=1}^{n} x_i s_{ki}$ has a normal distribution with mean zero and variance $\sum_{i=1}^{n} s_{ki}^2/N = 2E/N_o$.* Furthermore, the M different quantities $(1/N)\sum_{i=1}^{n} x_i s_{ki}$ are independent, since the functions $s_k(t)$ are orthogonal. It follows that the terms $\exp\left[(1/N)\sum_{i=1}^{n} x_i s_{ki} - E/N_o\right]$ are independent.

Since the logarithm of each term $Z = \exp\left[(1/N)\sum_{i=1}^{n} x_i s_{ki} - E/N_o\right]$ has a normal distribution with mean $(-E/N_o)$ and variance $2E/N_o$, the moments of the distribution can be found from Eq. (152). The n^{th} moment is

$$\mu_N(Z^n) = \exp\left[n(n-1)\frac{E}{N_o}\right]. \tag{163}$$

It follows that the mean of each term is unity, and the variance is

$$\sigma_N^2(Z) = \mu(Z^2) - \left[\mu(Z)^2\right] = \exp\left[\frac{2E}{N_o}\right] - 1. \tag{164}$$

The variance of a sum of independent random variables is the sum of the variances of the terms. Therefore

$$\sigma_N^2(M\ell) = M\left[\exp\left(\frac{2E}{N_o}\right) - 1\right], \tag{165}$$

and it follows that the variance of the likelihood ratio is

$$\sigma_N^2(\ell) = \frac{1}{M}\left[\exp\left(\frac{2E}{N_o}\right) - 1\right]. \tag{166}$$

It was pointed out in Section 4.9, that the receiver operating characteristic curves are approximately those of Fig. 2, with

$$d = \ell n\,(1 + \sigma_N^2) = \ell n\left(1 - \frac{1}{M} + \frac{1}{M}\exp\left(\frac{2E}{N_o}\right)\right). \tag{167}$$

---

* The reasoning is the same as that in Section 4.4.

This equation can be solved for $2E/N_o$:

$$\frac{2E}{N_o} = \ell n \left[1 + M (e^d - 1) \right] \quad . \tag{168}$$

Suppose it is desired to keep the false alarm probability and probability of detection constant. This requires that d be kept constant. Then from Eq. (168) it can be seen that if the number of possible signals M is increased, the signal energy E must also be increased.

4.11 Signal Which is One of M Orthogonal Signals with Unknown Carrier Phase

Consider the case in which the set of expected signals includes just M different amplitude-modulated signals which are known except for carrier phase. Denote the signals by

$$s_k(t) = f_k(t) \cos (\omega t + \theta) \quad . \tag{169}$$

It will be assumed further that the functions $f_k(t)$ all have the same energy E and are orthogonal, i.e.,

$$\int_0^T f_k(t) f_q(t) \, dt = 2E \delta_{kq} \quad , \tag{170}$$

where the 2 is introduced because the f's are the signal amplitudes, not the actual signal functions. Also, let the $f_k(t)$ be band-limited to contain no frequencies as high as . Then it follows that any two signal functions with different envelope functions will be orthogonal. Let us assume also that the distribution of phase, θ, is uniform, and that the probability for each envelope function is 1/M.

With these assumptions, the likelihood ratio can be obtained from Eq. (66), and it is given by

$$\ell(x) = \frac{1}{M} \sum_{k=1}^{M} \frac{1}{2\pi} \int_0^{2\pi} \exp \left[\frac{1}{N} \sum_{i=1}^{n} x_i s_{ki} - \frac{E}{N_o} \right] d\theta$$

where s_{ki} are the sample values of $s_k(t)$, and hence depend upon the phase θ. The integration is the same as in the case of the signal known except for phase, and the result, obtained from Eq. (82), is

$$\ell(x) = \frac{1}{M} \sum_{k=1}^{M} \exp \left[- \frac{E}{N_o} \right] I_o \left(\frac{r_k}{N} \right) \quad , \tag{172}$$

where

$$r_k = \sqrt{\left(\sum_i x_i f_k(t_i) \cos \omega t_i \right)^2 + \left(\sum_i x_i f_k(t_i) \sin \omega t_i \right)^2} \quad . \tag{173}$$

Now the problem is to find $\sigma_N^2(\ell)$. The variance of each term in the sum in Eq. (172) can be found since the distribution function with noise alone can be found in Section 4.5. Since the $f_k(t)$ are orthogonal, the distributions of the r_k are independent, and the terms in the sum in Eq. (172) are independent. Then the variance of the likelihood ratio, $\sigma_N^2(\ell)$, is the sum of the variances of the terms, divided by M^2.

The distribution function for each term $\exp(-E/N_o) \; I_o(r_k/N)$ is given in Section 4.5 by Eqs. (84) and (85). If α is defined by the equation

$$\beta = \exp \left[- \frac{E}{N_o} \right] I_o \left(\alpha \sqrt{\frac{2E}{N_o}} \right) \quad , \tag{174}$$

then the distribution function in the presence of noise for each term in Eq. (172) is

$$F_N^{(k)}(\beta) = \exp\left[-\frac{\alpha^2}{2}\right] . \tag{175}$$

The mean value of each term is

$$\mu_N^{(k)}(\beta) = \int_0^\infty \beta dF_N^{(k)}(\beta) = \int_0^\infty \exp\left[-\frac{E}{N_o}\right] I_o\left(\sqrt{\frac{2E}{N_o}}\,\alpha\right)\alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha . \tag{176}$$

This can be evaluated as on page 174 of Threshold Signals[5], and the result is that $\mu^{(k)}(\beta) = 1$.

The second moment of each term is

$$\mu_N^{(k)}(\beta^2) = \int_0^\infty \beta^2\, dF_N^{(k)}(\beta) ,$$

or

$$\mu_N^{(k)}(\beta^2) = \int_0^\infty \exp\left[-\frac{2E}{N_o}\right]\left[I_o\left(\alpha\sqrt{\frac{2E}{N_o}}\right)\right]^2 \alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha . \tag{177}$$

The integral can be evaluated as in Appendix E of Part II of reference 17, and the result is

$$\mu_N^{(k)}(\beta^2) = I_o\left(\frac{2E}{N_o}\right) . \tag{178}$$

The variance of each term in Eq. (172) is

$$\left[\sigma_N^{(k)}(\beta)\right]^2 = \mu^{(k)}(\beta^2) - \left[\mu^{(k)}(\beta)\right]^2 = I_o\left(\frac{2E}{N_o}\right) - 1 . \tag{179}$$

It follows that the variance of $M\ell$ is

$$\sigma_N^2(M\ell) = M\left[I_o\left(\frac{2E}{N_o}\right) - 1\right] , \text{ and therefore} \tag{180}$$

$$\sigma_N^2(\ell) = \frac{1}{M}\left[I_o\left(\frac{2E}{N_o}\right) - 1\right] , \tag{181}$$

since the variance for the sum of independent random variables is the sum of the variances.

If the approximation described in Section 4.9 is used, the receiver operating characteristic curves are approximately those of Fig. 2, with

$$d = \ell n\,(1 + \sigma_N^2) = \ell n\left(1 - \frac{1}{M} + \frac{1}{M} I_o\left(\frac{2E}{N_o}\right)\right) . \tag{182}$$

4.12 The Broad Band Receiver and the Optimum Receiver

A few applications of the results of Section 4 are suggested in Table I, Section 4.1. Two further examples of practical knowledge obtainable from the theory are presented in this section and in the next.

One common method of detecting pulse signals in a frequency band of width B is to build a receiv which covers this entire frequency band. Such a receiver with a pulse signal of known starting time is studied in Section 4.7. This is not a truly optimum receiver; it would be interesting to compare with an optimum receiver. We have been unable to find the distribution of likelihood ratio for the case of a signal which is a pulse of unknown carrier phase if the frequency is distributed evenly over a band. However, if the problem is changed slightly, so that the frequency is restricted to points spaced approximately the reciprocal of the pulse width apart, then pulses at different frequer cies are approximately orthogonal, and the case of the signal which is one of M orthogonal signals known except for phase can be applied. Eq. (182) should be used with M equal to the ratio of the frequency band width B to the pulse band width. Since the band width of a pulse is approximately the reciprocal of its pulse width, the parameter M used in Section 4.7 also has this value. Curves showi $2E/N_o$ as a function of d are given in Fig. 7 for both the approximate optimum receiver and the broad band receiver for several values of M. In the figure, d is calculated from Eq. (135) and Eq. (182), which hold for large values of M.

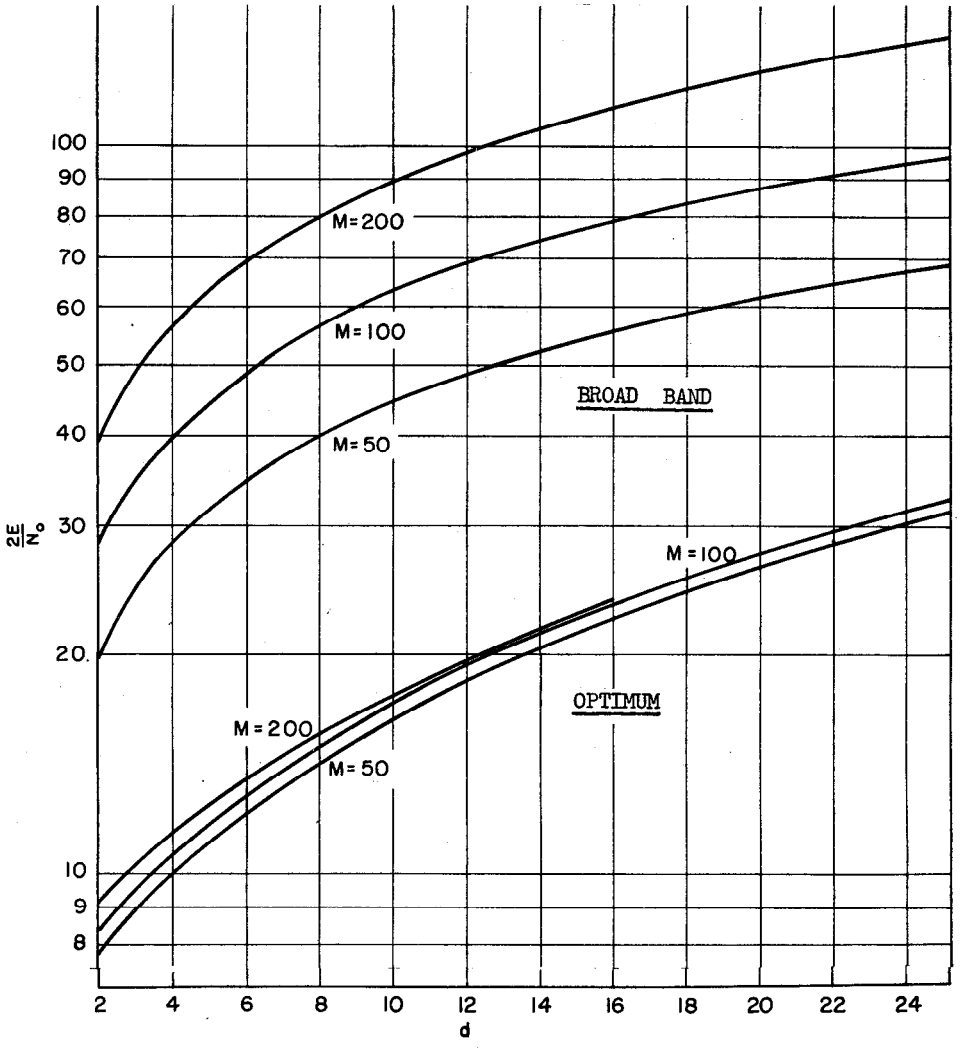

FIG. 7 COMPARISON OF OPTIMUM AND BROAD BAND RECEIVERS

4.13 Uncertainty and Signal Detectability

In the two cases where the signal considered is one of M orthogonal signals, the uncertainty of the signal is a function of M. This provides an opportunity to study the effect of uncertainty on signal detectability. In the approximate evaluation of the optimum receiver when the signal is one of M orthogonal functions, the ROC curves of Figs. 2 and 3 are used with the detection index d given by

$$d = \ell n \left[1 - \frac{1}{M} + \frac{1}{M} \exp\left(\frac{2E}{N_o}\right) \right]. \tag{167}$$

This equation can be solved for the signal energy, yielding

$$\frac{2E}{N_o} = \ell n \left[1 - M + Me^d \right] \approx \ell n \; M + \ell n \; (e^d - 1) \quad , \tag{175}$$

the approximation holding for large $2E/N_o$.* From this equation it can be seen that the signal energy is approximately a linear function of ℓn M when the detection index d, and hence the ability to detect signals, is kept constant. It might be suspected that $2E/N_o$ is a linear function of the entropy, $-\sum p_i \ell n p_i$, where p_i is the probability of the i^{th} orthogonal signal. The linear relation holds only when all the p_i are equal. The expression which occurs in this more general case is:

$$\frac{2E}{N_o} \approx - \ell n \left[\sum p_i^2 \right] + \ell n \; (e^d - 1) \; . \tag{176}$$

LIST OF REFERENCES

1. S. Goldman, Information Theory, Prentice-Hall, New York, 1953. Chapter II, pp. 65-84, is devoted to sampling plans.

2. C. E. Shannon, "Communication in the Presence of Noise," Proc. I.R.E., Vol. 37, pp. 10-21, January, 1949.

3. U. Grenander, "Stochastic Processes and Statistical Inference," Arkiv För Mathematik, Bd 1 nr 17, p. 195, 1950.

4. J. Neyman, and E. S. Pearson, "On the Problems of the Most Efficient Tests of Statistical Hypotheses," Philosophical Transactions of the Royal Society of London, Vol. 231, Series A, p. 289, 1933.

5. J. L. Lawson, and G. E. Uhlenbeck, Threshold Signals, McGraw-Hill, New York, 1950.

6. P. M. Woodward and I. L. Davies, "Information Theory and Inverse Probability in Telecommunications," Proc. I.E.E. (London), Vol. 99, Part III, pp. 37-44, March, 1952.

7. I. L. Davies, "On Determining the Presence of Signals in Noise," Proc. I.E.E. (London), Vol. 99, Part III, pp. 45-51, March, 1952.

8. A. Wald, Sequential Analysis, John Wiley and Sons, 1947.

9. W. C. Fox, "Signal Detectability: A Unified Description of Statistical Methods Employing Fixed and Sequential Observation Processes," Electronic Defense Group, University of Michigan, Technical Report No. 19 (unclassified).

* If $2E/N_o > 3$, the error is less than 10%.

10. A. Wald and J. Wolfowitz, "Optimum Character of the Sequential Probability Ratio Test," Ann. Math. Stat., Vol. 19, p. 326, September, 1948.

11. E. Reich, and P. Swerling, "The Detection of a Sine Wave in Gaussian Noise," Journal Applied Physics, Vol. 24, p. 289, March, 1953.

12. R. C. Davis, "On the Detection of Sure Signals in Noise," Journal Applied Physics, Vol. 25, pp. 76-82, January, 1954.

13. J. V. Harrington, and T. F. Rogers, "Signal-to-Noise Improvement Through Integration in a Storage Tube," Proc. I.R.E., Vol. 38, p. 1197, October, 1950.

A. E. Harting, and J. E. Meade, "A Device for Computing Correlation Functions," Rev. Sci. Instr., Vol. 23, 347, 1952.

Y. W. Lee, T. P. Cheatham, Jr., and J. B. Wiesner, "Applications of Correlation Analysis to the Detection of Periodic Signals in Noise," Proc. I.R.E., Vol. 38, p. 1165, October, 1950.

M. J. Levin, and J. F. Reintjes, "A Five Channel Electronic Analog Correlator," Proc. Nat. El. Conf., Vol. 8, 1952.

14. D. O. North, "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems," RCA Laboratory Rpt PTR-6C, 1943.
See also Reference 5, p. 206.

15. Graphs of values of the integral (89) along with approximate expressions for small and for large values of appear in Rice, S. O., "Mathematical Analysis of Random Noise," B.S.T.J., Vol. 23, p. 282-332 and Vol. 24, p. 46-156, 1944-5. Tables of this function have been compiled by J. I. Marcum in an unpublished report of the Rand Corporation, "Table of Q-Functions," Project Rand Report RM-399.

16. P. G. Hoel, Introduction to Mathematical Statistics, New York: Wiley, 1947, p. 246.

17. The material of Sections 2 and 3 of this paper is drawn from reference 9 above and from Part I of W. W. Peterson, and T. G. Birdsall, "The Theory of Signal Detectability," Electronic Defense Group, University of Michigan, Technical Report No. 13 (Unclassified), July, 1953. Part II of that report contains the material in Section 4 of this paper. Other work in this field may be found in D. Middleton, "Statistical Criteria for the Detection of Pulsed Carriers in Noise," Jour. App. Phys., Vol. 24, p. 371, April, 1953; D. Middleton, "The Statistical Theory of Detection. I: Optimum Detection of Signals in Noise," M.I.T. Lincoln Laboratory, Technical Report No. 35, November 2, 1953; D. Middleton, "Statistical Theory of Signal Detection," Trans. I.R.E., PGIT-3, p. 26, March, 1954; D. Middleton, W. W. Peterson, and T. G. Birdsall, "Discussion of 'Statistical Criteria for the Detection of Pulsed Carriers in Noise. I, II'", Journal Applied Physics, Vol. 25, pp. 128-130, January, 1954.