ENGINEERING RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN
ANN ARBOR

THE THEORY OF SIGNAL DETECTABILITY

PART I. THE GENERAL THEORY

ISSUED SEPARATELY:

PART II. APPLICATIONS WITH GAUSSIAN NOISE

Technical Report No. 13
Electronic Defense Group
Department of Electrical Engineering

By: W. W. Peterson
T. G. Birdsall

Approved by: H. W. Welch, Jr.
H. W. Welch, Jr.

Project M970

TASK ORDER NO. EDG-3
CONTRACT NO. DA-36-039 sc-15358
SIGNAL CORPS, DEPARTMENT OF THE ARMY
DEPARTMENT OF ARMY PROJECT NO. 3-99-04-042
SIGNAL CORPS PROJECT NO. 29-194B-0

June, 1953

ERRATA FOR EDG TECHNICAL REPORT NO. 13

Part I

pg. 10 — The notation $A_2(k)$ means an optimum criterion such that $P_N(A_2(k)) = k$. This notation is defined on page 16.

pg. 11 — Line 3, The sentence should read "If at any point $(P_N(A), P_{SN}(A))$ on curve (1) a line is drawn with slope β_k given by the operating level graph, it will be tangent to the curve and will intersect the axis at the value $P_{SN}(A) - \beta_k P_{SN}(A)$."

pg. 24 — The seoond line from the bottom of the page should start "$P_{SN}(A_2 - A_1) = 0$"

pg. 35 — Omit the x_o between lines 4 and 5.

pg. 40 — Line 6, This should read "measurable set B_o contained in A_o such that $P(B_o) = \gamma$."

Part II

pg. 5 — Footnote 2 should read "If $\frac{1}{2\pi N}$ exp . . . . . . etc."

pg. 37 — Line 3, should read "times the amplitude squared of its envelope, etc."

pg. 64 — Line 1, replace "when" by "for which".

Note: An introduction to the theory of signal detectability, using as little mathematics as possible and including discussions of the applications of sequential analysis as well as the types of optimum criteria discussed in Part I, has been prepared as EDG Technical Report No. 19. Enough theoretical material will be included so that this report could be used in place of Part I as an introduction to Part II.

TABLE OF CONTENTS

PART I. THE GENERAL THEORY

TABLE OF CONTENTS (cont.)

PART II. APPLICATIONS WITH GAUSSIAN NOISE (Issued Separately)

ABSTRACT

PART I

The several statistical approaches to the problem of signal detectability which have appeared in the literature are shown to be essentially equivalent. A general theory based on likelihood ratio embraces the criterion approach, for either restricted false alarm probability or minimum weighted error type optimum, and the a posteriori probability approach. Receiver reliability is shown to be a function of the distribution functions of likelihood ratio. The existence and uniqueness of solutions for the various approaches is proved under general hypothesis.

PART II

The full power of the theory of signal detectability can be applied to detection in Gaussian noise, and several general results are given. Six special cases are considered, and the expressions for likelihood ratio are derived. The resulting optimum receivers are evaluated by the distribution functions of the likelihood ratio. In two of the special cases studied, the uncertainty of the signal ensemble can be varied, throwing some light on the effect of uncertainty on probability of detection.

ACKNOWLEDGEMENTS

In the work reported here, the authors have been influenced greatly by their association with the other members of the Electronic Defense Group. In particular, Mr. H. W. Batten contributed much to the early phases of the work on signal detectability. The authors are indebted to Mr. W. C. Fox and Mr. Paul Roth for the proofs of Lemma 1 and Lemma 2 in Appendix B and also to Mr. Fox for the proof of Lemma 4 and for the many helpful suggestions and corrections resulting from his careful reading of the text.

The authors also wish to acknowledge their indebtedness to Dr. A. B. Macnee, Dr. H. W. Welch, and Mr. C. B. Sharpe for the many suggestions resulting from their reading the report, and to Geraldine L. Preston for her assistance in the preparation of the text.

THE THEORY OF SIGNAL DETECTABILITY

PART I. THE GENERAL THEORY

ISSUED SEPARATELY:

PART II. APPLICATIONS WITH GAUSSIAN NOISE

1. Concepts and Theoretical Results

1.1 Introduction

Random interference plays the key role in the theory of signal detectability. It not only places a limit on the energy which a signal must have to be detected reliably, but it also limits the bandwidth of a receiver for strong signals, or generally the variety of signals which can be detected consistently in a given receiver. Part I of this report presents the basic theory of detecting signals in random interference and Part II applies it to some simple problems in design and evaluation of receivers.

The signal detectability problem is represented schematically in Fig. 1.1. The operator has available a voltage varying with time, which will be referred to as the receiver input. This voltage is in some way different when a signal is present from when there is noise alone.

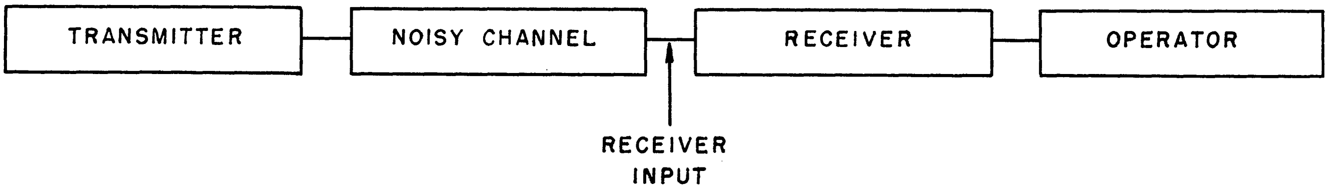

FIG. 1.1. BLOCK DIAGRAM OF SIGNAL DETECTION PROBLEM.

The receiver is the operator's tool or analyzing system; it enables him to study the input to the receiver by observing the receiver output. He can use the receiver input to his advantage only if (1) the receiver input is different when there is a signal than when there is no signal, and (2) he knows enough about the signals and the noise to analyze the input so as to recognize the difference. The operator can do better than random guessing in deciding whether or not there is a signal present only when he has information about the signals, the noise, and his receiver; this must be recognized before treating this problem. The information about the signal and about the noise is usually of a statistical nature because of the random nature of noise, and the uncertainty as to the exact signal that will be transmitted.

Signal detectability has been recognized as a statistical problem by a number of authors.[1] There have been two distinct approaches to the problem. The first, the criterion approach, is first presented in Threshold Signals by J. L. Lawson and G. I. Uhlenbeck.[2] The second, using a posteriori probability,

---

[1]Lawson and Uhlenbeck, Ref. 1; Woodward and Davies, Refs. 2, 3, 4, and 5; Reich and Swerling, Ref. 6; Middleton, Ref. 7; Slattery, Ref. 8; Hanse, Ref. 9; Schwartz, Ref. 10; North, Ref. 11; Kaplan and Fall, Ref. 12.

[2]Lawson and Uhlenbeck, Ref. 1.

was studied by P. M. Woodward and I. L. Davies.[1] The difference between the two methods lies mainly in the approach. Both are presented in this report, and the very close connection between the results of the two will be demonstrated in Section 2; namely, the basic receiver required can be the same for either case, only the final manner of analysis and presentation of the output is different. The criterion approach requires less of this analysis, and has been given more attention in this report because it is somewhat simpler.

1.2 Detectability Criteria

Suppose the operator is required to guess whether or not there is a signal present. He will, for certain receiver inputs, say that a signal is present.[2] Such receiver inputs will be said to satisfy the criterion, or to be in the criterion. Those receiver inputs which lead him to guess that there is no signal present are not in the criterion.

There are two distinct kinds of errors which the operator may make. He may say there is a signal present if there is only noise; this is a false alarm. He may say there is only noise when signal plus noise is present; he misses the signal. One of these errors may be more serious than the other, so that they must be considered separately.

It will be convenient to use the ordinary notation of probability theory. Events will be represented by letters, and in particular, the following symbols will be used for the following events:

---

[1]Davies, Ref. 2., and Woodward and Davies, Ref. 3.

[2]We shall assume the operator is scientifically logical, i.e., for the same receiver input he will always give the same response. An alternative approach is described in Appendix A.

SN	There is signal plus noise
N	There is noise alone
A	The operator says there is a signal, i.e., the receiver input is in the criterion
CA	The operator says there is only noise, i.e., the receiver input is not in the criterion.

If B and C are events, P(B) is the probability of occurrence of event B, $P(B \cdot C)$ is the probability of occurrence of events B and C together, and $P_B(C)$ is the (conditional) probability of occurrence of event C if event B is known to occur.

From the statistical information given about the signal and the interference it turns out to be convenient to calculate $P_N(A)$ and $P_{SN}(A)$, because these quantities do not depend upon the a priori probability that a signal is present. This will be done in Part II of this report for some interesting cases. If these probabilities, $P_N(A)$ and $P_{SN}(A)$, are given as well as P(SN), the a priori probability that a signal is present, then the probability of any combination of the events in this discussion can be calculated. In fact, any three (algebraically) independent probabilities can be used to calculate all the others. That there are just three (algebraically) independent probabilities can be seen by noting that all of the events discussed are combinations of the four events $SN \cdot A$, $N \cdot A$, $SN \cdot CA$, and $N \cdot CA$, and any probabilities can be calculated from the probabilities of these four. But the sum of the probabilities of these four is unity, so only three of these are independent. Thus, for example,

$$\begin{aligned}
P(SN \cdot A) &= P(SN)\, P_{SN}(A)\ , \\
P(N \cdot A) &= \left[1 - P(SN)\right] P_N(A), \\
P(SN \cdot CA) &= P(SN)\, P_{SN}(CA) = P(SN) \left[1 - P_{SN}(A)\right]\ , \qquad (1.1) \\
P(A) &= P(SN \cdot A) + P(N \cdot A), \\
P_A(SN) &= \frac{P(SN \cdot A)}{P(A)}\ , \text{etc.}
\end{aligned}$$

1.3 A Posteriori Probability and Signal Detectability

As an alternative to requiring the operator to say whether a signal is present or not, the operator might 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 operator 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.[1] The method of making the best estimate of the a posteriori probability that a signal is present will be discussed in this report. A good discussion of this approach is also found in the original papers by Woodward and Davies.[2]

It is shown in Section 2 that the a posteriori probability is given by the following equation:

$$P_x(SN) = \frac{\ell(x)\ P(SN)}{\ell(x)\ P(SN) + 1 - P(SN)} \tag{1.2}$$

where $P_x(SN)$ is the a posteriori probability for the receiver input denoted by x and $\ell(x)$ is the likelihood ratio for the same receiver input. Likelihood ratio for a particular receiver input is usually defined as the ratio of probability density for that receiver input if there is signal plus noise to the probability density if there is noise alone. It is a measure of how likely that receiver input is when there is signal plus noise as compared with when there is noise alone. It is a random variable; its value depends upon what the receiver input happens to be. If a receiver which has likelihood ratio as its output

[1]Ref. 3.
[2]Ref. 2, 3, 4, and 5.

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, making the receiver an optimum receiver for obtaining a posteriori probability.

1.4 Optimum Criteria

An important question is whether or not it is possible to find the optimum criterion for a given situation. A first step toward the answer is to define what is meant by optimum, and this definition depends upon the situation. It may be possible to put a numerical value upon the correct responses and a numerical cost on the errors. Suppose

$$\begin{aligned} V_{SN\cdot A} &= \text{Value of the correct response } SN\cdot A \\ V_{N\cdot CA} &= \text{Value of the correct response } N\cdot CA \\ K_{SN\cdot CA} &= \text{Cost of the error } SN\cdot CA \\ K_{N\cdot A} &= \text{Cost of the error } N\cdot A \end{aligned} \tag{1.3}$$

Then

$$V = V_{SN\cdot A}P(SN\cdot A) + V_{N\cdot CA}P(N\cdot CA) - K_{SN\cdot CA}P(SN\cdot CA) - K_{N\cdot A}P(N\cdot A) \tag{1.4}$$

is the expected value of the response of the equipment for a given criterion. An optimum criterion then would be one which would maximize this expression. Since the later sections will calculate $P_N(A)$ and $P_{SN}(A)$, it will be an advantage to express the expected value V of the response in terms of these quantities.

$$\begin{aligned} V = {} & V_{SN\cdot A}P(SN)\, P_{SN}(A) + V_{N\cdot CA}\left[1 - P(SN)\right]\left[1 - P_N(A)\right] \\ & - K_{SN\cdot CA}P(SN)\left[1 - P_{SN}(A)\right] - K_{N\cdot A}\left[1 - P(SN)\right] P_N(A) \end{aligned}$$

$$\begin{aligned} V = {} & P_{SN}(A)\, P(SN)\, (V_{SN\cdot A} + K_{SN\cdot CA}) - P_N(A)\left[1 - P(SN)\right](V_{N\cdot CA} + K_{N\cdot A}) \\ & + V_{N\cdot CA}\left[1 - P(SN)\right] - K_{SN\cdot CA}\, P(SN). \end{aligned} \tag{1.5}$$

Thus maximizing V is equivalent to requiring that

$$P_{SN}(A) - \beta\, P_N(A) \text{ is a maximum, where}$$

$$\beta = \frac{1 - P(SN)}{P(SN)} \; \frac{(V_{N\cdot CA} + K_{N\cdot A})}{(V_{SN\cdot A} + K_{SN\cdot CA})} \quad . \tag{1.6}$$

Note that P(SN) is the a priori probability that there is a signal present.

In another case it may be required to limit the probability of a false alarm and to minimize the probability of a missed signal with this restriction. In symbols, it is required that,

$$\begin{aligned} &P(N\cdot A) \leqq Po \\ &P(SN\cdot CA) \text{ is a minimum.} \end{aligned} \tag{1.7}$$

This also can be expressed in terms of $P_N(A)$, $P_{SN}(A)$, and the a priori probability P(SN):

$$\begin{aligned} &P(N\cdot A) = \left[1 - P(SN)\right] P_N(A) \leqq Po, \quad \text{or} \quad P_N(A) \leqq k = \frac{Po}{1 - P(SN)} \quad , \text{ and} \\ &P(SN\cdot CA) = P(SN)\left[1 - P_{SN}(A)\right] \text{ is a minimum, i.e., } P_{SN}(A) \text{ is a maximum.} \end{aligned} \tag{1.8}$$

1.5 Theoretical Results

Both of the above problems of finding an optimum criterion will be discussed in later sections, and it will be shown that under very general conditions both problems have essentially the same solution. The optimum criterion consists of all receiver inputs with likelihood greater than some number β. For the first type of optimum criterion, β is the parameter in Eq. (1.6), and for the second type of criterion, β can be determined from the value of the parameter k in Eq. (1.8). It has already been mentioned that a posteriori probability is the simple function of likelihood ratio given in Eq. (1.2). Thus a receiver which could calculate the likelihood ratio for each receiver input can be used as an a posteriori probability type receiver or as either of the criterion type receivers. Part II of this report, which treats some specific cases, deals only with the likelihood ratio.

1.6 Receiver Evaluation[1]

Usually a receiver is judged on the basis of probability of false alarm if no signal is sent, i.e., $P_N(A)$, and the probability of detection if a signal is sent, $P_{SN}(A)$. The reliability of any receiver in any given situation can be summarized in one graph, called the receiver operating characteristic, on which $P_{SN}(A)$ is plotted against $P_N(A)$. For any criterion and any fixed set of signals, there is fixed value for $P_{SN}(A)$ and a fixed value for $P_N(A)$. Thus the criterion can be represented as a point on the receiver operating characteristic graph. A criterion-type receiver may operate at any level (i.e., any value of β or any value of K), and hence is represented by a curve. Two types of optimum criteria have been discussed, and the graph points up the relation between the two. In Fig. 1.2 curve (1) is based on optimum operation for which $P_{SN}(A)$ is maximized for $P_N(A)$ fixed. Thus, no receiver can operate above the first curve. The third curve is a lower limit in operation found by rotating the optimum curve about the center point of the graph; it would result if an optimum receiver operator minimized $P_{SN}(A)$, i.e., said no whenever he should say yes, and vice versa. No receiver, no matter how poor, can be made to operate below the third curve. The diagonal could be achieved by turning the receiver off and guessing, in which case $P_{SN}(A) = P_N(A)$.

In the next section it will be shown that the derivative of curve (1) sketched in the lower plot, is the operating level β of the optimum receiver; that is, if the slope at some point is β, then the corresponding optimum criterion

---

[1]Only evaluation of criterion type receivers is discussed here. Evaluation of an a posteriori probability type receiver is considered in Section 2.5.

M-970 A-G3-67 JEM 6-11-53

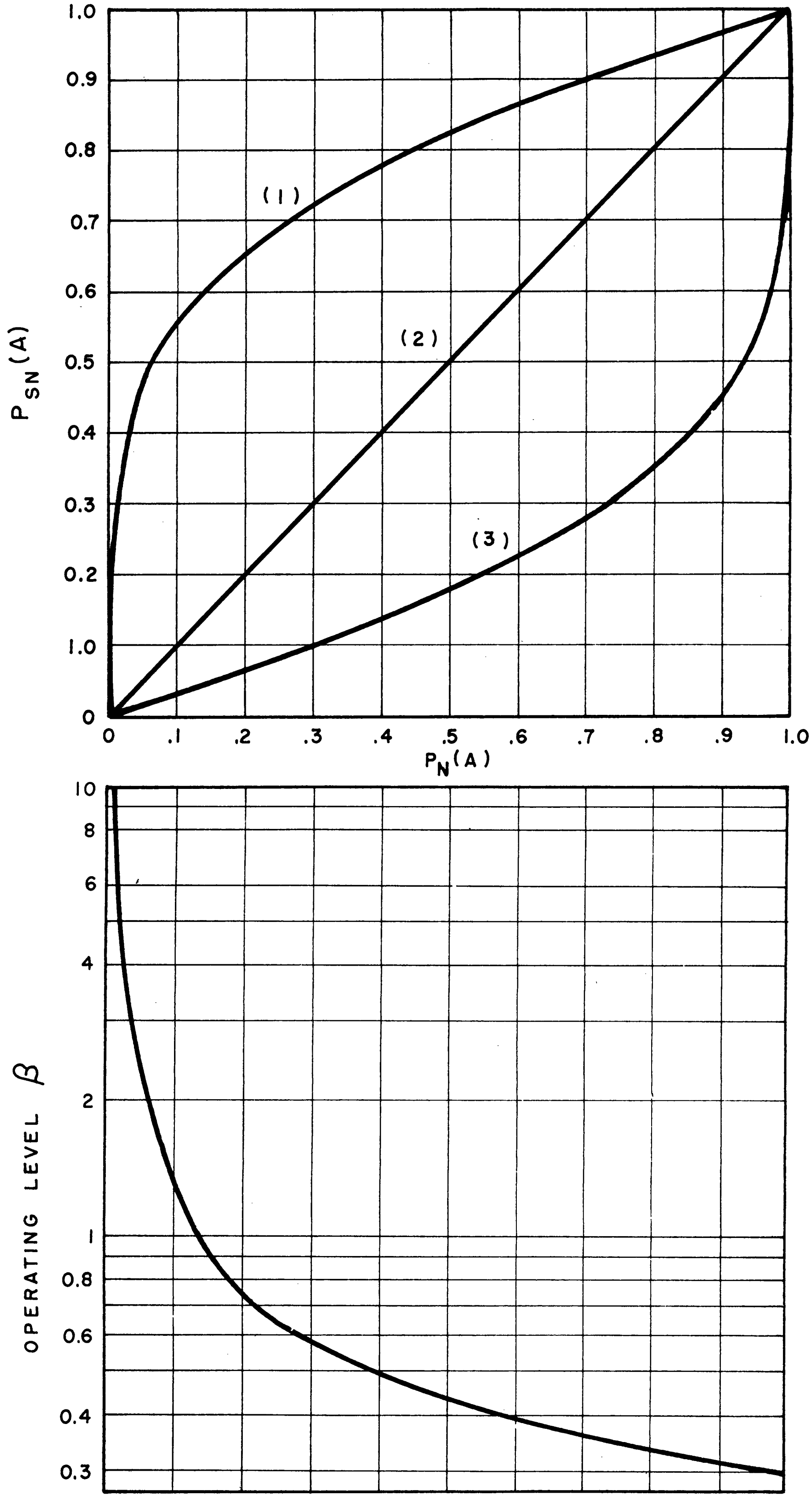

FIG. 1.2

TYPICAL RECEIVER OPERATING CHARACTERISTIC.

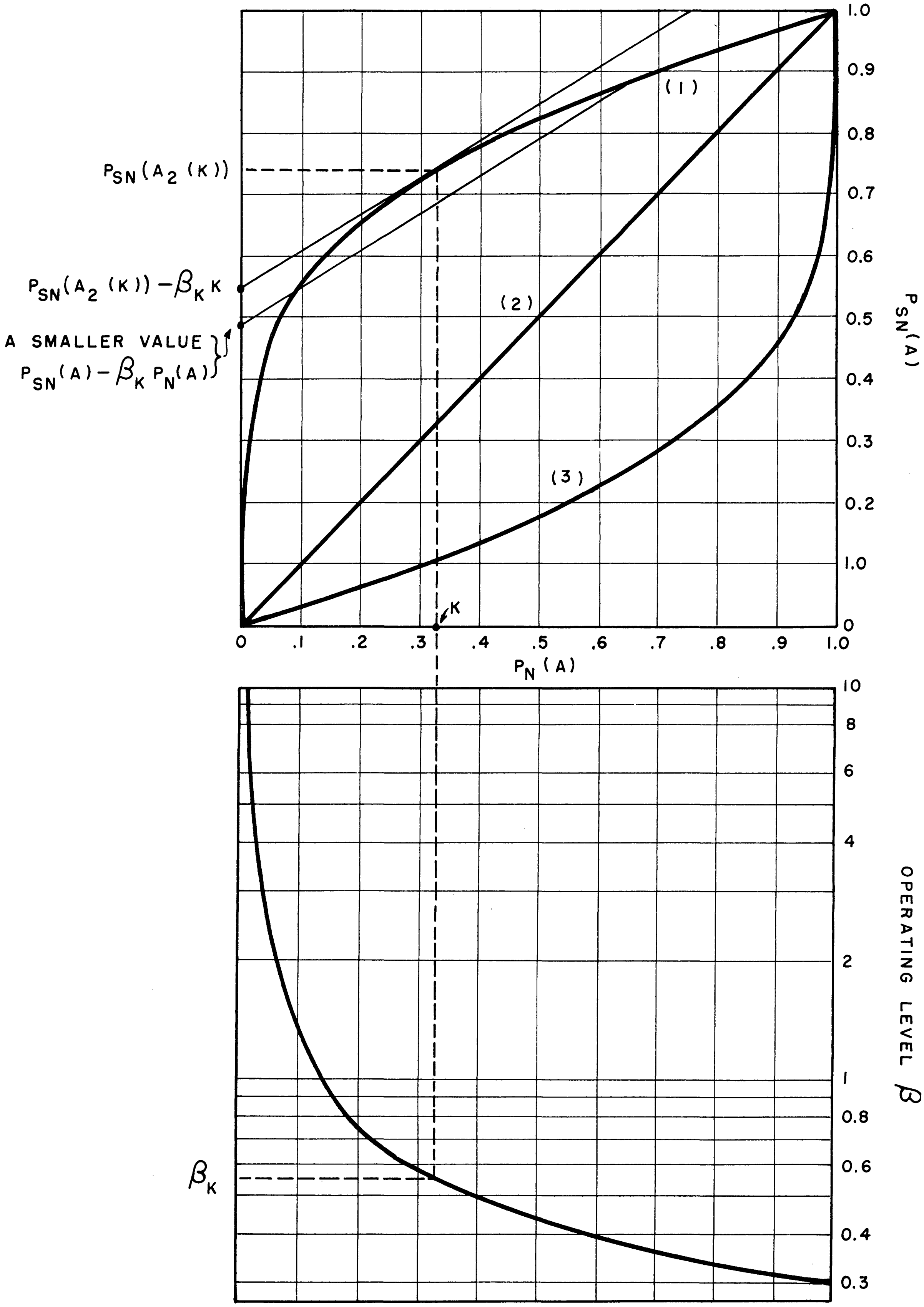

FIG. 1.3

TYPICAL RECEIVER OPERATING CHARACTERISTIC.

is made up of all inputs which have likelihood ratio greater than or equal to β. The relationship between the first and second types of optimum criteria is graphically illustrated in Fig. 1.3. If at any point $(P_N(A), P_{SN}(A))$ on curve (1) a line is drawn with slope β, it will be tangent to the curve and will intersect the axis at the value $P_{SN}(A) - \beta P_N(A)$. This is the quantity to be maximized for the first type of optimum criterion, and if a line with the same slope is drawn through any other point on or between curves (1) and (3), it will cut the axis below the point where the tangent cuts the axis. Thus, curve (1) is not only the curve for the optimum of the type when $P_N(A)$ is bounded and $P_{SN}(A)$ maximized, but also the curve for the optimum criterion when values are placed on the operator's responses.

A non-optimum receiver can be evaluated in a given situation if its receiver operating characteristic is drawn together with that of the optimum. One receiver is better than another over a range if it is closer to the optimum than the other. In some instances the optimum curve for a given situation will nearly match another receiver's operation in the same situation except that the optimum will require less signal energy. In this case, the non-optimum receiver can be given a db rating for that situation.

Each application of the theory treated in Part II of this report is accompanied by the receiver operating characteristic of the optimum receiver.

2. MATHEMATICAL THEORY

2.1 Introduction

The method for handling the signal detectability problem mathematically is described in this section. The first step is the presentation of the appropriate mathematical description of the signals and noise. In these terms the signal detectability problem is restated in several forms discussed in Section 1 of this report. It is then shown that in each case, if the likelihood ratio can be determined for each receiver input, the problem is essentially solved. Thus the conclusion is that the receiver design problem should be treated in terms of likelihood ratio; this is the approach used in Part II.

2.2 Mathematical Description of Signals and Noise

Any receiver input, noise or signal plus noise, is a voltage which is a function of time. Thus we shall be considering a set of functions. In this report it will be assumed that the receiver input is limited to bandwidth W, and that the observation is of finite duration T. By the sampling theorem,[1] any such function is completely determined when its values at "sampling" points spaced 1/2W seconds apart through the observation interval are known. There are 2WT sampling points in all. Thus a receiver input can be considered as a point in a 2WT dimensional space, the values at the sample points being taken as coordinates. Let us call the space R.

If there is noise at the receiver input, the receiver input voltage may usually be any of an infinite number of functions, i.e., any of an infinite number of points in the 2WT dimensional space R. With Gaussian noise any point is

[1] Shannon, C. E., "Communication in the Presence of Noise," Proc. IRE, Vol. 37, p. 10, January 1949; also Appendix D of Part II.

theoretically possible. It is a matter of chance which one occurs. Thus it appears that the appropriate way to describe the noise is to give the probability density for points in the space of receiver inputs.[1] The same is true when there is signal plus noise, so that we shall deal with the space R and two probability density functions, $f_N(x)$ for the case of noise alone, and $f_{SN}(x)$ for the case of signal plus noise. Here x denotes a point of the space R.

In a practical application, information will be 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. If the signals can be described by a probability density function $f_S(x)$,

$$f_{SN}(x) = \int_R f_N(x-s)\, f_S(s)\, ds \quad , \tag{2.1}$$

where the integration is over the whole space R. The receiver input x(t) could be caused by any signal s(t), and noise x(t) - s(t). The probability density for x is the probability that both s(t) and x(t) - s(t) will occur at the same time, summed over all possible s(t).

If the signals cannot be described by a probability density function, a more general form must be used, in which the signals are described by a probability measure, P_S; the formula for this case is

$$f_{SN}(x) = \int_R f_N(x-s)\, d\, P_S(s) \quad . \tag{2.2}$$

This is what is called a Lebesgue integral, and it means essentially to average

[1]We shall assume that the probability density function exists. See Appendix A.

$f_N(x-s)$ over all values of s in the whole space weighting according to the probability P_S of the points s appearing as signals.[1]

2.3 A Posteriori Probability

The approach of Woodward and Davies[2] to the signal detectability problem is to ask the operator, "What is the probability that a signal is present?" He is to give the probability, using knowledge of the receiver input, i.e., he gives the a posteriori probability.

If the probability density functions are continuous, the a posteriori probability $P_x(SN)$ can be found for any particular receiver input x. Bayes' theorem[3] is used, but not directly, since $P_{SN}(x)$ and $P_N(x)$ are both zero. Consider a small sphere U with radius r and center x. Then $P_U(SN)$ can be obtained by Bayes' theorem, and $P_x(SN)$ can be defined as the

$$P_x(SN) = \lim_{r \to 0} P_U(SN) \quad . \tag{2.3}$$

Denote by $P(SN \cdot U)$ the probability that signal plus noise will be present and the receiver output will be in U. Then

$$P(SN \cdot U) = P(SN) \cdot P_{SN}(U) = P_U(SN) \cdot P(U) \tag{2.4}$$

and

$$P(U) = P_{SN}(U)\, P(SN) + P_N(U)\,(1 - P(SN)) \tag{2.5}$$

Solving for $P_U(SN)$,

$$P_U(SN) = \frac{P(SN)\, P_{SN}(U)}{P(SN)\, P_{SN}(U) + [1 - P(SN)]\, P_N(U)}$$

$$= \frac{P(SN)\, \dfrac{P_{SN}(U)}{P_N(U)}}{P(SN)\, \dfrac{P_{SN}(U)}{P_N(U)} + (1 - P(SN))} \tag{2.6}$$

[1] Cramér, Ref.14, pp. 62, 188. [2] Woodward and Davies, Ref. 3. [3] Cramér, Ref.14, p. 507.

By the definition of probability density function,

$$P_{SN}(U) = \int_U f_{SN}(x)\, dx$$

$$P_N(U) = \int_U f_N(x)\, dx \quad , \tag{2.7}$$

where the integral is really a multiple integral over the volume of the sphere U in the n-dimensional space. Then

$$\frac{P_{SN}(U)}{P_N(U)} = \frac{\int_U f_{SN}(x)\, dx}{\int_U f_N(x)\, dx} \quad , \tag{2.8}$$

and if $f_{SN}(x)$ and $f_N(x)$ are continuous,

$$\lim_{r \to 0} \frac{P_{SN}(U)}{P_N(U)} = \frac{f_{SN}(x)}{f_N(x)} = \ell(x) \quad . \tag{2.9}$$

The ratio of probability densities $f_{SN}(x)/f_N(x) = \ell(x)$ is called the likelihood ratio. It follows that

$$P_x(SN) = \lim_{r \to 0} P_U(SN) = \frac{P(SN)\ \ell(x)}{P(SN)\ \ell(x) + [1 - P(SN)]} \tag{2.10}$$

This is the existence probability as defined by Woodward and Davies.[1] Notice that the likelihood ratio $\ell(x)$ is the all-important quantity. $P_x(SN)$ is a simple monotone increasing function of the likelihood ratio. Therefore if P(SN) is known and if the receiver produces $\ell(x)$, a calibration will convert this to $P_x(SN)$.

2.4 Criteria and the Optimum Criteria

2.4.1 Definitions. Suppose the operator is only required to guess whether or not there is a signal present. For certain receiver inputs he will guess there is a signal present. These receiver inputs form a subset of

[1] Ref. 3.

the space R of all possible receiver inputs. Let us call this subset the criterion and denote it by A. That is, a point x is in the criterion A if the operator will say there is a signal present when x occurs as receiver input.

It will be convenient to have a symbol for each of the two types of optimum criteria described in Section 1.4. The first type will be denoted by $A_1(\beta)$; that is, $A_1(\beta)$ is any subset of R such that for fixed $\beta \geq 0$,

$$P_{SN}\left[A_1(\beta)\right] - \beta\, P_N \left[A_1(\beta)\right] \text{ is maximum.} \tag{2.11}$$

The second type will be denoted by $A_2(k)$; that is, $A_2(k)$ is any subset of R such that

$$\begin{aligned} &P_N\,(A_2(k)) \leq k \quad , \quad \text{and} \\ &P_{SN}\,(A_2(k)) \quad \text{is maximum.} \end{aligned} \tag{2.12}$$

The likelihood ratio $\mathcal{L}(x)$, which is defined as ratio of the probability density functions, $f_{SN}(x)/f_N(x)$ plays an important role in the following discussion. It is a measure of how much more likely the receiver input is to be if there is signal plus noise than if there is noise alone.

2.4.2 Theorems on Optimum Criteria. The optimum criterion is closely related to the likelihood ratio. For the first type of criterion the connection is given by the following theorems.

Theorem 1: Denote by A the set of points for which the likelihood ratio $\mathcal{L}(x) \geq \beta$. Then A is an optimum criterion $A_1(\beta)$.

Proof: The condition that A be an optimum criterion $A_1(\beta)$ is that $P_{SN}(A) - \beta\, P_N(A)$ is maximum; i.e., for any other set B of receiver inputs $P_{SN}(A) - \beta\, P_N(A) \geq P_{SN}(B) - \beta\, P_N(B)$.

$$P_{SN}(A) - \beta\, P_N(A) = \int_A f_{SN}(x)\, dx - \beta \int_A f_N(x)\, dx$$
$$= \int_A \left[f_{SN}(x) - \beta\, f_N(x) \right] dx \quad , \tag{2.13}$$

where the integration is over the set A, and so is really a multiple integral over a part of the space R which has 2WT dimensions. Let B be any set different from A. Denote by A-B the set of points which are in A and not in B, by B-A the set of points which are in B but not in A, and by $B \cap A$ the set of points which belong to both A and B. Then since A is the union of A-B and $A \cap B$, and A-B and $A \cap B$ have no points in common,

$$P_{SN}(A) - \beta\, P_N(A) = \int_A \left[f_{SN}\, x) - \beta\, f_N(x) \right] dx$$
$$= \int_{A \cap B} \left[f_{SN}(x) - \beta\, f_N(x) \right] dx \tag{2.14}$$
$$+ \int_{A-B} \left[f_{SN}(x) - \beta\, f_N(x) \right] dx$$

Likewise

$$P_{SN}(B) - \beta\, P_N(B) = \int_{A \cap B} \left[f_{SN}(x) - \beta\, f_N(x) \right] dx$$
$$+ \int_{B-A} \left[f_{SN}(x) - \beta\, f_N(x) \right] dx \tag{2.15}$$

Thus

$$P_{SN}(A) - \beta\, P_N(A) - \left[P_{SN}(B) - P_N(B) \right] =$$
$$\int_{A-B} \left[f_{SN}(X) - \beta\, f_N(x) \right] dx - \int_{B-A} \left[f_{SN}(x) - \beta\, f_N(x) \right] dx \tag{2.16}$$

The points in A-B are in A, and so for them $f_{SN}(x)/f_N(x) = \mathcal{L}(x) \geq \beta$, so that $f_{SN}(x) - \beta f_N(x) \geq 0$, and the first integral in Eq. (2.16) is not less than zero. The points in the set B-A are not in A, so $f_{SN}(x)/f_N(x) < \beta$, and the second integral in Eq. (2.16) is no greater than zero. Thus

$$P_{SN}(A) - \beta P_N(A) \geq P_{SN}(B) - \beta P_N(B) \quad , \tag{2.17}$$

$P_{SN}(A) - \beta P_N(A)$ is a maximum, and A is an optimum criterion $A_1(\beta)$

There is not a unique optimum criterion $A_1(\beta)$. In the first place "optimum" was defined in terms of probability. Thus a change in $A_1(\beta)$ which would not change $P_{SN}\left[A_1(\beta)\right]$ or $P_N\left[A_1(\beta)\right]$ would result in an equally good criterion. Such a change might consist of adding or taking out a single point, a finite number of points, or generally any set of probability zero.[1] More insight into the uniqueness is given by the following theorem.

Theorem 2: If A is an optimum criterion $A_1(\beta)$, then the set of points in A for which $\mathcal{L}(x) < \beta$ has probability zero, and the set of points not in A for which $\mathcal{L}(x) > \beta$ has probability zero.

Proof: We will show that any criterion which does not have these two properties is not an optimum criterion. Consider any criterion B with a subset C, of non-zero probability, such that the likelihood ratio of each point in C is less than β. There is a positive number ϵ and a subset C_ϵ of C, having non-zero probability, such that $\mathcal{L}(x) \leq \beta - \epsilon$ for the points in C_ϵ . If this were not true, then for any positive small number ϵ , the subset C_ϵ would have probability zero. These subsets C_ϵ are monotone, that is,

[1] A set E will be said to have probability zero if both $P_{SN}(E)$ and $P_N(E)$ are zero.

if $\epsilon_2 < \epsilon_1$, then C_{ϵ_2} contains C_{ϵ_1}, and, since C contains no points with likelihood ratio equal to β, the union of all C_ϵ is C itself, and would have probability zero.[1]

As in Eq. (2.14),

$$P_{SN}(C_\epsilon) - \beta P_N(C_\epsilon) = \int_{C_\epsilon} \left[f_{SN}(x) - \beta f_N(x)\right] dx = \int_{C_\epsilon} f_N(x)\left[\ell(x) - \beta\right] dx$$

and since $\ell(x) \leq \beta - \epsilon$ or $\ell(x) - \beta \leq -\epsilon$,

$$P_{SN}(C_\epsilon) - \beta P_N(C_\epsilon) \leq -\epsilon \int_{C_\epsilon} f_N(x)\, dx = -\epsilon P_N(C_\epsilon) \quad . \tag{2.19}$$

Therefore, if $P_N(C_\epsilon) > 0$,

$$P_{SN}(C_\epsilon) - \beta P_N(C_\epsilon) < 0 \quad . \tag{2.20}$$

But C_ϵ is a subset of A, and therefore

$$P_{SN}(B - C_\epsilon) - \beta P_N(B - C_\epsilon) > P_{SN}(B) - \beta P_N(B) \quad , \tag{2.21}$$

and B is not an $A_1(\beta)$. It can be shown in an analogous manner that if there is a set D of non-zero measure outside of criterion B such that $\ell(x) > \beta$ in D, then there is a subset D_ϵ of D such that

$$P_{SN}(D_\epsilon) - \beta P_N(D_\epsilon) > 0 \tag{2.22}$$

and therefore

$$P_{SN}(B \cup D_\epsilon) - \beta P_N(B \cup D_\epsilon) > P_{SN}(B) - \beta P_N(B) \quad , \tag{2.23}$$

and B is not an $A_1(B)$.

[1] Cramér, Ref. 14, p. 50, Eq. 6.2.3; and p. 77, paragraph 8.2.

This theorem says nothing about the points for which $\ell(x) = \beta$. It is not hard to show that $P_{SN}(A) - \beta\, P_N(A)$ is not affected by including or excluding points where $\ell(x) = \beta$. Thus a criterion $A_1(\beta)$ must include all points for which $\ell(x) > \beta$ (except perhaps a set of probability zero), none of the points where $\ell(x) < \beta$ (except perhaps a set of probability zero), and it may or may not include a point for which $\ell(x) = \beta$.

In the most general case, when the noise is Gaussian, the following two theorems show the uniqueness of $A_1(\beta)$.

Theorem 3: If the probability density function for noise alone, $f_N(x)$, is an analytic function, then the set of points for which $\ell(x) = \beta$ has probability zero.[1]

A function is said to be analytic if it is analytic in the ordinary sense when considered as a function of each single coordinate. The proof of the theorem is quite involved, and so it is given in Appendix B.

Theorem 4 follows immediately from Theorem 2 and Theorem 3.

Theorem 4: If the probability density function for noise alone $f_N(x)$ is analytic, any two optimum criteria $A_1(\beta)$ can differ only by a set of probability zero.

Now let us turn to the second type of optimum criterion.

Theorem 5: Let A be a set such that if x is in A, the likelihood ratio $\ell(x) \geq \beta$, while if x is not in A, $\ell(x) \leq \beta$. Then if $P_N(A) = k$, A is an optimum criterion $A_2(k)$.

Proof: An optimum criterion $A_2(k)$ must satisfy the conditions $P_N(A) \leq k$, and $P_{SN}(A)$ is maximum. The first is satisfied by hypothesis. Suppose B is any other set such that $P_N(B) \leq k$. Denote by A-B the set of points in A which are not in B, by B-A

---

[1]A little more is needed in the hypothesis for Theorem 3 than that $f_N(x)$ is analytic. See Appendix B.

the set of points in B which are not in A, and by $B \cap A$ the set of points common to B and A. Since A is the union of A-B and $A \cap B$, and since A-B and $A \cap B$ have no points in common,

$$P_N(A) = \int_A f_N(x)\,dx = \int_{A-B} f_N(x)\,dx + \int_{A\cap B} f_N(x)\,dx$$
$$= P_N(A-B) + P_N(A \cap B) = k \quad . \tag{2.24}$$

Likewise

$$P_N(B) = P_N(B-A) + P_N(A \cap B) \leq k, \tag{2.25}$$

and thus

$$P_N(A-B) \geq P_N(B-A) \quad . \tag{2.26}$$

Also,

$$P_{SN}(B-A) = \int_{B-A} f_{SN}(x)\,dx \quad , \tag{2.27}$$

and since any point x in B-A is not in A, $\ell(x) = \dfrac{f_{SN}(x)}{f_N(x)} \leq \beta$ and

$$P_{SN}(B-A) = \int_{B-A} \frac{f_{SN}(x)}{f_N(x)}\, f_N(x)\,dx \quad \beta \int_{B-A} f_N(x)\,dx \quad ,$$

or

$$P_{SN}(B-A) \leq \beta\, P_N(B-A) \quad . \tag{2.28}$$

Likewise

$$P_{SN}(A-B) \geq \beta\, P_N(A-B) \quad . \tag{2.29}$$

Collecting Eqs. (2.26), (2.28), and (2.29),

$$P_{SN}(B-A) \leq \beta\, P_N(B-A) \leq \beta\, P_N(A-B) \leq P_{SN}(A-B) \quad . \tag{2.30}$$

As in Eq. (2.24),

$$P_{SN}(A) = \int_A f_{SN}(x)\,dx = \int_{A-B} f_{SN}(x)\,dx + \int_{A\cap B} f_{SN}(x)\,dx$$
$$= P_{SN}(A-B) + P_{SN}(A\cap B) \quad , \tag{2.31}$$

and

$$P_{SN}(B) = P_{SN}(B-A) + P_{SN}(A\cap B) \quad . \tag{2.32}$$

Therefore

$$P_{SN}(A) - P_{SN}(B) = P_{SN}(A-B) - P_{SN}(B-A). \tag{2.33}$$

From Eqs. (2.30) and (2.33) it follows that

$$P_{SN}(A) \geq P_{SN}(B) \quad , \tag{2.34}$$

and $P_{SN}(A)$ is a maximum.

It follows from Theorem 5 that every optimum of the first type, $A_1(\beta)$, is an optimum of the second type. More precisely, if set A is an optimum of the first type it is associated with the fixed β for which it is an $A_1(\beta)$. By Theorem 2, the likelihood ratio in A is not less than β, and outside A the likelihood ratio is not greater than β, except on a set of probability zero. But the introduction or omission of such a set has no effect on $P_{SN}(A)$ or $P_N(A)$. Since $P_N(A)$ has some value, call it a; A will be an $A_2(a)$ by Theorem 5.

Theorem 6: For every k between 0 and 1 there is an optimum criterion of the first type A_k, such that $P_N(A_k) = k$.

Proof: For each value β we consider the maximal $A_1(\beta)$; by Theorem 2 this is the set consisting of all points of likelihood ratio not less than β:

$$M_\beta = \left\{ x \;\middle|\; \ell(x) \geq \beta \right\} \quad . \tag{2.35}$$

Now if for k there is a β such that $P_N(M_\beta) = k$, then because M_β is an $A_1(\beta)$ the proof is complete.

Next we point out that M_o is the whole space R and M_∞ is the empty set, and therefore $P_N(M_o) = 1$ and $P_N(M_\infty) = 0$. For any value of k, if there is no M_β such that $P_N(M_\beta) = k$, let $\beta^* = \min \left\{ \beta \,\middle|\, P_N(M_\beta) \geq k \right\} = g\ell b \left\{ \beta \,\middle|\, P_N(M_\beta) < k \right\}$ that is, $P_N(M_\beta{}^*) > k$ and if $\beta > \beta^*$, $P_N(M_\beta) < k$. Thus the jump in P_N is due to those points in $M_\beta{}^*$ for which $\ell(\mathbf{x}) = \beta^*$.

Because the probability density functions exist, every point has probability zero and therefore there is a subset S of these points with $\ell(x) = \beta^*$ for which $P_N = P_N(M_\beta{}^*) - k$. This is shown in Appendix B (Lemma 4).

Removing this subset from $M_\beta{}^*$, $$P_N(M_\beta{}^* - S) = k \quad . \tag{2.36}$$

Because $M_\beta{}^* - S$ satisfies Theorem 1, it is an $A_1(\beta^*)$. Of course, by Theorem 5, it is an $A_2(k)$ also.

The following theorem completes this circle of proof.

Theorem 7: For any k there is a β_k such that every $A_2(k)$ is an $A_1(\beta_k)$.

Proof: Let A be any $A_2(k)$.

By Theorem 6 there exists a β_k and an $A_1(\beta_k)$, which we will denote by A^*, such that $P_N(A^*) = k$. Then by Theorem 5, A^* is also an $A_2(k)$, and hence for both A and A^*, P_{SN} is maximum and $P_N \leqq k$. Therefore

$$P_{SN}(A^*) = P_{SN}(A) \tag{2.37}$$

$$P_N(A^*) = k \geqq P_N(A) \quad , \tag{2.38}$$

Multiplying Eq. (2.38) by $-\beta_k$ and adding gives

$$P_{SN}(A^*) - \beta_k P_N(A^*) \leqq P_{SN}(A) - \beta_k P_N(A) \quad . \tag{2.39}$$

Since A* maximizes this expression, the equality must hold, and A is also an $A_1(\beta_k)$.

In summary, these theorems show that β can be written as a multivalued function of k and that k can be written as a multivalued function of β. These relations can be sharpened somewhat.

Theorem 8: Let $a < b$ be two values taken on by $\ell(x)$. If no set of the form $\left\{x \mid \ell_1 < \ell(x) < \ell_2\right\}$ for $a \le \ell_1 < \ell_2 \le b$ has probability zero, then β_k is a single valued function of k on some interval I, with $a \le \beta_k \le b$, and $d\, P_{SN}(A_1(\beta_k))/dk$ exists and equals β_k for every k in I.

Proof: 1) In general, if a function is monotone on an interval and its range of values is also an interval, then it is continuous. If it were not, then at some point the left and right hand limits would be unequal, which would introduce a gap in the range of values, contradicting the hypotheses.

2) If $\beta_{k1} > \beta_{k_2}$ and if the interval from β_{k_1} to β_{k_2} contains a subinterval of $\left[a, b\right]$ of length greater than zero, then $k_2 > k_1$. There are, by Theorem 6, criteria of the first type A_i (for i = 1, 2), which, by Theorem 2, may be chosen so that A_i contains all points for which $\ell(x) > \beta_{k_i}$ and no points for which $\ell(x) < \beta_{k_i}$. Also $P_N(A_i) = k_i$, by Theorem 5. By applying P_N to the equation $A_2 = A_1 \cup (A_2 - A_1)$, one obtains $k_2 = k_1 + P_N(A_2 - A_1)$. If $P_N(A_2 - A_1) = 0$, then from Eqs. 2.7 and the fact that $\ell(x)$ is bounded on $A_2 - A_1$, it follows that $P_{SN}(A_2 - A_1) - 0$ also. But, by hypotheses, $A_1 - A_2$ cannot have probability zero. Hence $k_2 > k_1$.

3) Let I be the set of points k for which at least one β_k is in the open interval from a to b, and let β_k denote the possibly multivalued function defined on I. Then 2) says that β_k is both single valued and monotone, and Theorems 1 and 6 imply that the range of values of β_k is the interval from a to b. Hence I is an interval, for if it were not, there would exist three values $k_1 < k_2 < k_3$ with only the middle one not in I. Then $\beta_{k_1} < \beta_{k_2} < \beta_{k_3}$ and β_{k_2} would not be in the interval from a to b, yet the other two would be--a contradiction. Thus 1) can be applied to β_k and β_k is therefore continuous on I.

4) To form the derivative, let

$$\begin{aligned} D &= A_1(\beta_k) - A_1(\beta_{k_o}) \text{ if } \beta_k \leq \beta_{k_o} \\ &= A_1(\beta_{k_o}) - A_1(\beta_k) \text{ if } \beta_k \geq \beta_{k_o} \quad . \end{aligned} \tag{2.42}$$

Then

$$\lim_{k \to k_o^+} \frac{P_{SN}(A_1(\beta_k)) - P_{SN}(A_1(\beta_{k_o}))}{k - k_o} = \lim_{k \to k_o +} \frac{P_{SN}(D)}{k - k_o} \tag{2.43}$$

Since $k \geq k_o$, $\beta_k \leq \beta_{k_o}$, and in D, $\beta_k \leq \ell(x) \leq \beta_{k_o}$, $\beta_k f_N(x) \leq f_{SN}(x) \leq \beta_{k_o} f_N(x)$. But

$$P_{SN}(D) = \int_D f_{SN}(x)\, dx = \int_D \ell(x)\, f_N(x)\, dx \tag{2.44}$$

and

$$P_N(D) = k - k_o = \int_D f_N(x)\, dx \tag{2.45}$$

and therefore $\beta_{k} P_N(D) \leq P_{SN}(D) \leq \beta_{k_o} P_N(D)$. Similarly if $k \leq k_o$ $\beta_{k_o} P_N(D) \leq P_{SN}(D) \leq \beta_k P_N(D)$. Thus

$$\lim_{k \to k_o} \frac{P_{SN}(D)}{k - k_o} = \beta_{k_o} , \tag{2.46}$$

by virtue of the result that β_k is a continuous function of k.

2.5 Evaluation of Optimum Receivers

2.5.1 Introduction. This section treats the problem of determining how well a given receiver will perform its task of detecting signals. For the criterion type receiver, the probability of false alarm if no signal is sent, $P_N(A)$, and the probability of detection if a signal is sent, $P_{SN}(A)$, give a good measure of receiver performance. For the a posteriori probability type receivers, the average or mean a posteriori probability with signal plus noise and with noise alone describe the receiver's ability to discriminate between signal plus noise and noise alone.

2.5.2 Evaluation of Criterion Type Receivers. For simplicity, let us restrict this discussion to the case in which the probability density function for noise alone, $f_N(x)$ is analytic.

Denote by $F_{SN}(\beta)$ the probability that the likelihood ratio $\ell(x)$ is equal to or greater than β if there is signal plus noise, and similarly, let $F_N(\beta)$ be the probability that $\ell(x)$ is equal to or greater than β if there is noise alone. These are the complimentary distribution functions for $\ell(x)$. Then for any $A_1(\beta)$,

$$P_{SN}(A_1(\beta)) = F_{SN}(\beta), \text{ and} \tag{2.47}$$

$$P_N(A_1(\beta)) = F_N(\beta), \tag{2.48}$$

because the set of points for which $\ell(x) \geqq \beta$, and differs from any $A_1(\beta)$ only by a set of probability zero (Theorem 4). By Theorem 7, every $A_2(k)$ is an $A_1(\beta)$. The β_k corresponding to k can be found from Eq. (2.48)

$$P_N(A_1(\beta_k)) = F_N(\beta_k) = k \quad . \tag{2.49}$$

Then

$$P_{SN}(A_2(k)) = F_{SN}(\beta_k) \quad . \tag{2.50}$$

Thus, if the distribution functions $F_{SN}(\beta)$ and $F_N(\beta)$ are known, any criterion type receiver can be evaluated.

It turns out that not both $F_{SN}(\beta)$ and $F_N(\beta)$ are necessary. Theorem 8 states that

$$\frac{d\, F_{SN}(\beta)}{d\, F_N(\beta)} = \beta \quad , \tag{2.51}$$

since $P_{SN}(A_1(\beta_k)) = F_{SN}(\beta_k)$, and $k = F_N(\beta_k)$. Thus, if $F_N(\beta)$ is known, $F_{SN}(\beta)$ can be found by integrating Eq. (2.51).[1]

$$F_{SN}(\beta) = - \int_{\beta}^{\infty} y \, d\, F_N(y) \quad . \tag{2.52}$$

As an alternative, $F_{SN}(\beta)$ might be given as a function of $F_N(\beta)$; this is the receiver operating characteristic graph. Then β can be found from Eq. (2.51); i.e., β is the slope of the graph.

---

[1]The change in sign is because the functions $F_{SN}(\beta)$ and $F_N(\beta)$ are complimentary distribution functions. If the density function associated with $F_N(\beta)$ is $g(\beta)$,

then $\frac{d\, F_N(\beta)}{d\beta} = - g(\beta)$ and $F_{SN}(\beta) = \int_{\beta}^{\infty} g(\beta)\, d\,\beta$.

A corollary of Theorem 8 is the following: The nth moment of the distribution for noise alone is the (n-1)st moment of the signal plus noise distribution.

$$\int_{-\infty}^{\infty} y^n \, d\,F_N(y) = \int_{-\infty}^{\infty} y^{n-1} \, (y \, d\,F_N(y)) = \int_{-\infty}^{\infty} y^{n-1} \, d\,F_{SN}(y) \qquad (2.53)$$

As an example of the application of this corollary, note that the mean value of likelihood ratio with noise alone is always unity. If the variance with noise alone is σ_N^2, the second moment of $F_N(\beta)$ is $1 + \sigma_N^2$; then the mean of the signal plus noise distribution is $1 + \sigma_N^2$, and the difference of the means is σ_N^2. For detection corresponding roughly to Fig. 2.1, the difference of the means of the two distributions must be of the order of the standard deviation of the distributions, so that

$$\sigma_N^2 \approx \sigma_N \quad , \qquad (2.54)$$

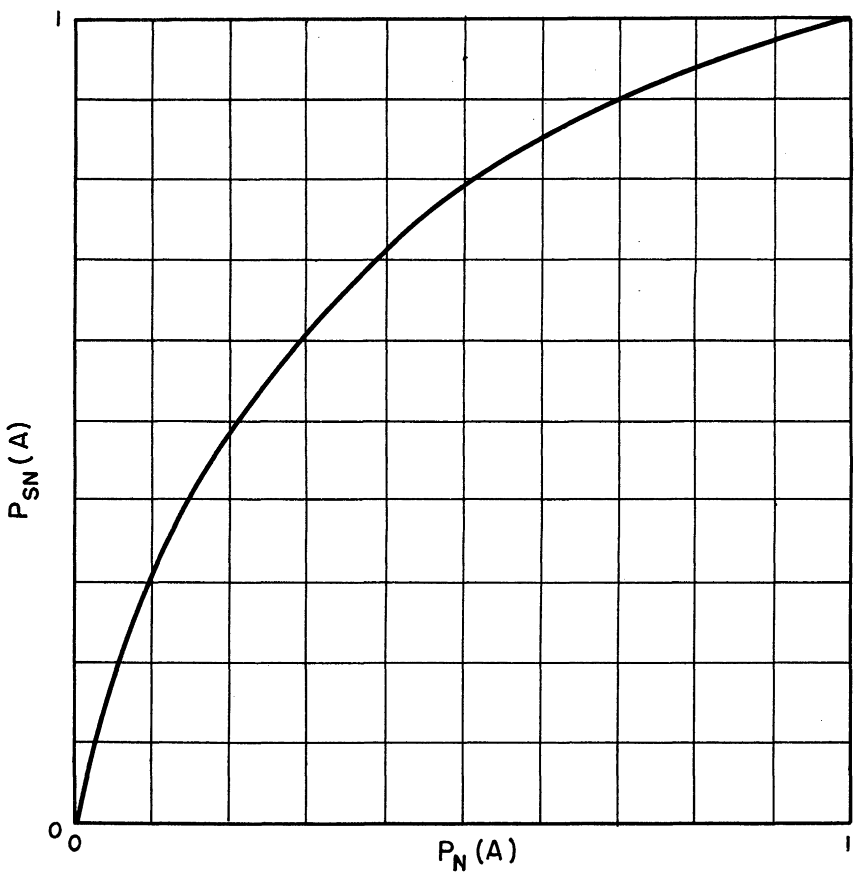

FIG. 2.1

RECEIVER OPERATING CHARACTERISTIC

For $\sigma_N^2 = 1$.

or the variance of the distribution with noise alone must be of the order of unity. For better detection, σ_N^2 must be greater.

2.5.3 Evaluation of A Posteriori Probability Woodward and Davies Type Receivers. Davies proposes the mean a posteriori probability as a measure of the efficiency of a receiver. The mean a posteriori probability is defined as:

$$\mu_{SN}\,(P_x(SN)) = \int_R P_x(SN)\, f_{SN}(x)\, dx \tag{2.55}$$

$$\mu_N(P_x(SN)) = \int_R P_x(SN)\, f_N(x)\, dx \tag{2.56}$$

These can be evaluated if the distribution functions $F_{SN}(\beta)$ and $F_N(\beta)$ for likelihood ratio are known. Since

$$P_x(SN) = \frac{P(SN)\ \ell(x)}{P(SN)\ \ell(x) + 1 - P(SN)} , \tag{2.57}$$

the mean a posteriori probabilities are

$$\mu_{SN}(P_x(SN)) = \int \frac{y\ P(SN)}{y\ P(SN) + 1 - P(SN)}\, d\, F_{SN}(y), \text{ and} \tag{2.58}$$

$$\mu_N(P_x(SN)) = \int \frac{y\ P(SN)}{y\ P(SN) + 1 - P(SN)}\, d\, F_N(y) \quad . \tag{2.59}$$

Davies presents the formula

$$\mu_{SN}\left[P_x(SN)\right] + \frac{1 - P(SN)}{P(SN)}\, \mu_N\left[P_x(SN)\right] = 1 \quad , \tag{2.60}$$

which enables one to calculate easily either one of the mean a posteriori probabilities once the other has been calculated.

2.6 Conclusions

It is possible to combine the most common statistical approaches to the theory of signal detectability into one general theory. In this theory likelihood ratio plays the central role: the result of the theory is that a receiver built so that its output is likelihood ratio can be adapted easily to accomplish the task specified in any of the well-known approaches to signal detectability. If the probability distribution of likelihood ratio is known, then the receiver reliability can be evaluated.

In Part II of this report, likelihood ratio and its distribution functions are calculated for a number of specific cases, and the problems of receiver design are discussed.

APPENDIX A

It was assumed throughout the discussion of the criterion approach to signal detectability that for any given receiver input, the operator would always give the same response. This is certainly not the case with threshold signals and a human operator. A more realistic approach might be to assume that for any receiver input x, the operator would say with probability $\rho(x)$ that there is signal plus noise. Finding the optimum receiver would then consist of finding the optimum $\rho(x)$. This approach does not lead to any interesting new results; if $\rho(x) = 1$ on an optimum criterion and zero on its compliment, then $\rho(x)$ is optimum.

The theorems on signal detectability are proved in Section II in more general form than has yet been found necessary in an application. However, they can be generalized somewhat, and this appendix discusses some of the possibilities.

It is certainly possible to consider more general spaces of signals. Any space on which a probability measure can be defined might be used. In order to prove the theorems on optimum criteria, however, some sort of likelihood ratio seems necessary. One possibility is to assume the measure $P_N(A)$ and the random variable $\ell(x)$ are given and to define $P_{SN}(A)$ through the integral

$$P_{SN}(A) = \int_A \ell(x)\, d\, P_N(A) \quad . \tag{A.1}$$

The mean value of $\ell(x)$ must be unity, of course.

If the space is a Euclidean space of finite dimension, then it is possible to define an arbitrary measure through distribution functions. These functions, being monotone, have a derivative almost everywhere, and thus afford a means of defining likelihood ratio. For any point which has measure zero, the likelihood is the ratio of the derivatives of the distribution function for signal plus noise and for noise alone. Points which do not have measure zero can always be treated separately. There can be only a countable number of these and likelihood ratio for such a point x can be defined as

$$\ell(x) = \frac{P_{SN}(x)}{P_N(x)} \tag{A.2}$$

Any point with infinite likelihood ratio belongs in the criterion, of course, and such a point has a posteriori probability unity. Then likelihood ratio is defined except for a set of points of measure zero.

In any case where likelihood ratio is defined and satisfies Eq. (A.1), Theorems 1 and 2 can be proved. The lemma (Appendix B, Lemma 1) which is needed for the proof of Theorem 5 can be proved for any space and measure for which sets of arbitrarily small measure can be found containing each point. If this holds and likelihood ratio is defined, then Theorems 5, 6, 7, and 8 can be proved.

APPENDIX B

This appendix contains the proof of Theorem 3 and the lemma required to complete the proof of Theorem 6. It is convenient to prove three lemmas from which Theorem 3 will follow directly.

Lemma 1: Let S be a sphere (i.e., the set of all points whose distance to a fixed point is less than or equal to a fixed positive number) in n-dimensional Euclidean space E^n. Let $f(x)$ be a continuous real function defined on S. Then the graph $G = \{[x, f(x)]\}$ of $f(x)$ in E^{n+1} has (n+1)-measure zero.

Proof: Let the volume (the n-measure) of S be V. Since $f(x)$ is uniformly continuous on S, for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever the distance between x_1 and x_2 is less than δ it follows that $|f(x_1) - f(x_2)| < \epsilon/4V$.

Moreover, for each $\delta > 0$ there is a decomposition of E^n into pairwise disjoint congruent n-dimensional cubes each with its greatest diagonal of length less than $\delta/2$. This decomposition may be chosen so that, if $\{C_i\}$ $i = 1, 2, \ldots, k$ are the cubes that touch S, then

$$\sum_i \text{(volume } C_i) < 2V \quad . \tag{B.1}$$

Thus $I_i = f(C_i)$ is an interval of length less than $2(\epsilon/4V) = \epsilon/2V$.

Now, let C_i^* be the (n+1)-cube formed by the Cartesian product $C_i \times I_i$; by construction, the graph G is covered by the (n+1)-cubes C_i^*. Also

$$\sum_i \left[\text{(n+1)-volume } C_i^*\right] \leq \sum_i \left[\text{(n)-volume } C_i\right] \epsilon/2V \leq 2V \cdot \epsilon/2V = \epsilon \quad . \tag{B.2}$$

Thus for each $\epsilon > 0$ there is a covering of G by (n+1)-cubes whose total (n+1)-volume is less than ϵ. This means (n+1)-measure of G is zero.

Lemma 2: Let D be an open set in Euclidean n-dimensional space E^n and $f(x)$ a real function defined for all points x in D which has continuous partial derivatives of all orders such that at each point x in D at least one partial derivative (of any order) does not vanish. Then, if b is some value taken on by f, the set $f^{-1}(b)$ of all points x such that $f(x) = b$ has n-measure zero.

Proof: A point x in D is said to have "order zero" if some first order derivative of f does not vanish at x; x has "order r" (r a positive integer) if all partial derivatives of f of order $\leq r$ vanish at x, but at least one partial derivative of f of order r+1 does not vanish at x. By the hypotheses, every point of D has finite order.

For each integer $r \geq 0$ let C_r be the set of points in $f^{-1}(b)$ of order r; then $f^{-1}(b) = \bigcup_{r=0}^{\infty} C_r$. The theorem is proved if it is shown that the n-measure of C_r is zero for each r. This will be done in two steps.

I. At each point x° in C_r, there is a sphere $S(x^\circ)$ centered at x° such that $S(x^\circ) \cap C_r$ has n-measure zero.

II. There is a countable collection $\left\{S(x^i)\right\}$, $i = 1, 2, \ldots,$ of such spheres such that C_r is contained in the union $\bigcup_{i=1}^{\infty} S(x^i)$.

Steps I and II together show that n-measure of C_r is zero because

$$0 \leq \text{n-measure } C_r \leq \sum_{i=1}^{\infty} \text{n-measure } \left[S(x^i) \cap C_r\right] = 0 \quad . \qquad \text{(B.3)}$$

Step II is an application of the Lindelöff theorem which asserts that every collection of spheres contains a countable subcollection whose union is equal to the union of all the original spheres.

The proof of I follows:

Since $x°$ is of order r, one of the derivatives of order r of f(x), say ω (x), has a first order derivative which does not vanish at $x°$. By a change in notation, this can be written as: $\frac{\partial \omega}{\partial x_n} = \omega_n$ does not vanish at

$x° = (x°_1, \ldots, x°_n)$. The implicit function theorem can then be applied to ω, yielding these results:

1) there is a sphere $S(x°)$ centered at $x°$ and contained in D.

2) writing π for the projection of $S(x°)$ onto the $x_1, \ldots, x_{n-1}$ "coordinate plane," π is an (n-1) sphere. There is a real valued continuous function $X(x_1, \ldots, x_{n-1})$ defined on π whose graph $G = \left\{\left[x_1, \ldots, x_{n-1}, X(x_1, \ldots, x_{n-1})\right]\right\}$ is the set of all points x in $S(x°)$ such that $\omega(x°) = \omega(x)$; that is $G = S(x°) \cap \omega^{-1}\left[\omega(x°)\right]$.

Note: 2) says that, in particular, $\omega\left[x_1, \ldots, x_{n-1}, X(x_1, \ldots, x_{n-1})\right] = \omega(x°)$. This is the usual way of stating the theorem.

By Lemma 1, the n-measure of G is zero. Thus step I is proved if $S(x°) \cap C_r \subset G$.

Case 1: $r = 0$. If x is in $S(x°) \cap C_o$, then x is of order $r = 0$ and $f(x) = f(x°)$. But in this case ω must have been chosen to be f, so $\omega(x) = \omega(x°)$, which implies that x is in G.

Case 2: $r > 0$. If x is in $S(x°) \cap C_r$, then x is of order r, which means that in particular all r-order partials of f vanish at x. Hence $\omega(x) = 0$. Also, by the same argument $\omega(x°) = 0$, and $\omega(x) = \omega(x°)$ implies that x is in G. This completes the proof of Lemma 2.

<u>Lemma 3</u>: If $f_N(x_1, x_2, \ldots, x_n)$ is an analytic function defined on n-dimensional Euclidean space E^n, and if $P(S_1, S_2, \ldots, S_n)$ is a probability measure on E^n such that there exists a bounded set in E^n whose probability is unity, then

$$f_{SN}(x_1, \ldots, x_n) = \int_{E^n} f_N(x_1-s_1, \ldots, x_n-s_n)\, dP(s_1, \ldots, s_n) \qquad \text{(B.4)}$$

exists and is analytic.

Proof: Let B be a bounded set such that $P(B) = 1$. Then $\bar{B}$, the closure of B, is such a set also; it is certainly bounded, and it can be assigned the measure unity, since

$$B \subset \bar{B} \subset E^n \text{ and } 1 = P(B) \leqq P(\bar{B}) \leqq P(E^n) = 1 \ . \qquad \text{(B.5)}$$

The probability of the complement of $\bar{B}$ is zero, and hence the integration can be restricted to the set $\bar{B}$ rather than to the whole of E^n.

For a fixed $(x_1, \ldots, x_n)$ and for $(s_1, \ldots, s_n)$ in $\bar{B}$, $f_N(x_1-s_1, \ldots, x_n-s_n)$ is bounded, since f_N is continuous and $\bar{B}$ is closed and bounded. The function f_N is also measurable, since it is continuous. (This assumes open sets are measurable.) Then the integral exists.[1]

The function $f_N(x_1, \ldots, x_n)$ being analytic means that $f_N(x_1, \ldots, x_n)$ is an analytic function in the ordinary sense when considered as a function of any single coordinate x_i. Let us forget about the other coordinates for the present. Then $f_N(x_i)$ has a power series expansion at each point $x°_i$, which converges in a neighborhood of the point $(x°_i, 0)$ in the complex plane. Thus $f_N(x_i)$ can be extended for complex values of x_i in a region containing the real axis.

Formally,

$$\frac{\partial}{\partial x_i} f_{SN}(x_i) = \lim_{h \to 0} \frac{f_{SN}(x_i+h)-f_{SN}(x_i)}{h} \qquad \text{(B.6)}$$

[1]Cramér, Ref. 14, Section 5.2, p. 37.

$$= \lim_{h \to 0} \frac{1}{h} \left[\int_B f_N(x_1-s_1, \ldots, x_i+h-s_i, \ldots, x_n-s_n)\, dP(s_1, \ldots, s_1) \right.$$
$$\left. - \int_B f_N(x_1-s_1, \ldots, x_i s_i, \ldots, x_n-s_n)\, dP(s_1, \ldots, s_n) \right] \tag{B.7}$$

$$= \lim_{h \to 0} \int_B \frac{1}{h} \left[f_N(x_1-s_1, \ldots, x_i+h-s_i, \ldots, x_n-s_n) \right.$$
$$\left. - f_N(x_1-s_1, \ldots, x_i-s_i, \ldots, x_n-s_n) \quad dP(s_1, \ldots, s_n) \right] \tag{B.8}$$

$$= \int_B \lim_{h \to 0} \frac{1}{h} \left[f_N(x_1-s_1, \ldots, x_i+h-s_n, \ldots, x_n-s_n) \right.$$
$$\left. - f_N(x_1-s_1, \ldots, x_i, \ldots, s_i-x_n-s_n) \quad dP(\ _1, \ldots, s_n) \right] \tag{B.9}$$

$$= \int_B \frac{\partial f_N}{\partial x_i}\, dP(s_1, \ldots, s_n) \quad . \tag{B.10}$$

The only question now is whether or not it is permissible to interchange the order of integration and taking the limit of the difference quotient at step (B.9). This is permissible if the difference quotient converges uniformly, which turns out to be the case.

The function $f_N(x_i)$ is analytic in a domain which extends to complex values of x_i near the real axis. The function $f_N(x_i + h - s_i)$ can be considered as a function of $h - s_i$, and is analytic for complex values of $h - s_i$ in a domain containing the real axis. Since the values of $s = (s_1, \ldots, s_n)$ in B are a closed bounded set, and the values of h can certainly be bounded, the set V of values $h - s_i$ is bounded. V can also be taken as closed, and it can be chosen so that no point s_i is on its boundary. Then there will be a minimum distance $h_o > 0$ from points s_i to the boundary of V. Consider the function

$$\psi(s_1, \ldots, s_n, h) = \frac{1}{h}\Big[f_N(x_1-s_i, x_i+h-s_i, \ldots, x_n-s_n)$$

$$-f_N(x_1-s_1, \ldots, x_1-s, \ldots, x_n-s_n)\Big]$$

if $h \neq 0$, and

$$= \frac{\partial f_N}{\partial s_i}, \qquad \text{if } h = 0 \quad ,$$

defined for $|h| \leqq h_o$, and s in $\bar{B}$. ψ is continuous at every point, and it is defined for all points (h, s) with $h = u+iv$ and $s = (s_1, \ldots, s_n)$ of a compact subset of E^{n+2}. ψ is therefore uniformly continuous, and its convergence to $\frac{\partial f_N}{\partial x_i}$ as h approaches zero along <u>any</u> complex valued path is uniform in s. Thus the difference quotient converges uniformly.

<u>Lemma 3'</u>: Let $f_N(x_1, \ldots, x_n)$ be a function of n complex variables, and suppose that for each i, there is a domain D_i in the complex plane and a number h_o such that the domain D_i contains all points within a distance of h_o of the real axis, and $f_N(x_1, \ldots, x_i, \ldots, x_n)$ is an analytic function of x_i in D_i for all real values of the other coordinates. Then, if $P(s_1, \ldots, s_n)$ is a probability measure on the n-dimensional Euclidean space E^n,

$$f_{SN}(x_1, \ldots, x_n) = \int_{E^n} f_N(x_1-s_1, \ldots, x_n-s_n)\, dP(s_1, \ldots, s_n) \qquad (B.11)$$

is analytic if it exists.[1]

[1] If f_N is bounded, the integral must exist, as in the previous case.

The proof will be omitted. The idea of the proof is as follows: one must form the difference quotient for $f_{SN}(x_1, \ldots, x_n)$ for each coordinate x_i

$$\frac{1}{h}\left[f_{SN}(x_1, \ldots, x_i+h, \ldots, x_n) - f_{SN}(x_1, \ldots, x_i, \ldots, x_n)\right]$$

and show that the limit as $h \rightarrow 0$ exists, and is equal to what is obtained by differentiating under the integral sign. The space can be divided into two parts such that one will have arbitrarily small measure and contribute an arbitrarily small amount to the integrals, while the other will be closed and bounded and hence on it the order of integration and taking the limit as $h \rightarrow 0$ can be interchanged, as in Lemma 3. The domain D_i is required so that differentiation in the complex plane will be possible.

Now let us discuss Theorem 3. Suppose $f_N(x)$ is analytic, and suppose either Lemma 3 or Lemma 3' holds. Then $f_{SN}(x)$ is analytic, and their ratio

$$\ell(x) = \frac{f_{SN}(x)}{f_N(x)} \quad ,$$

is analytic except where $f_N(x) = 0$. This is a set of measure zero, by Lemma 2. Since $\ell(x)$ is analytic, the points where $\ell(x) = \beta$ form a set of measure zero, by Lemma 2.[1] This proves Theorem 3.

<u>Theorem 3</u>: If the probability density function for noise alone, $f_N(x)$, is an analytic function, (and if either Lemma 3 or Lemma 3' holds,) then the set of points for which $\ell(x) = \beta$ has measure zero.

The restriction that Lemma 3 or Lemma 3' holds is not at all serious. If the signals have bounded energy, Lemma 3 holds. Lemma 3' would be expected to hold for most analytic probability density functions, and in particular it does hold if the noise is Gaussian.

[1]Note that Lebesgue measure zero implies probability zero, since the probability is defined through density functions.

The following lemma is needed to complete the proof of Theorem 6.

Lemma 4: Let $f(x)$ be a probability density function defined on the n-dimensional Euclidean space E^n. Denote by $P(A)$ the value of the integral $\int_A f(x)\,dx$ for all subsets A of E^n for which the integral exists. If A_o is any P-measurable set whose measure $P(A_o)$ is finite, and if $0 < \gamma < P(A_o)$, then there is a P-measurable set B_o such that $P(B_o) = \gamma$.

The following proof makes the theorem valid for any measure on any space M with the property "C" defined below.

Proof: Under the hypotheses above, the measure P has a special property relative to the space E^n.

Property "C": There is a countable class $\left[C_i\right]$, $i = 1, 2, \ldots,$ of P-measurable sets such that if x is a point and $\epsilon > 0$ then there is a C_i containing x such that $P(C_i) < \epsilon$.

One can obtain such a class by choosing all (n-dimensional) spheres of rational radius centered at points whose coordinates are rational. This class is countable because the rational numbers are countable. Its members are P-measurable because $\int_A f(x)\,dx$ exists for any sphere A. That it has property "C" is a way of stating a fundamental property of integrals.

The desired set B_o will be constructed as the union of a special sequence $\left[D_i\right]$ of P-measurable sets. Define D_1 to be $C_1 \cap A_o$ if $P(C_1 \cap A_o) \leq \gamma$; otherwise define D_1 to be empty. If D_n has been defined, define $D_{n+1} = D_n \cup \left[C_{n+1} \cap A_o\right]$ if $P\left\{D_n \cup \left[C_{n+1} \cap A_o\right]\right\} \leqq \gamma$; otherwise define $D_{n+1} = D_n$. Since $D_n \subset D_{n+1}$, $P(D_n) \leq P(D_{n+1}) \leq \gamma$. Hence the sequence $\left[P(D_n)\right]$ of real numbers converges. A general property of measures yields the result that $P\left[\bigcup_{n=1}^{\infty} D_n\right] = \lim_{n \to \infty} P(D_n)$. Write $B_o = \bigcup_{n=1}^{\infty} D_n$; then $P(B_o) = \lim_{n \to \infty} P(D_n) \leq \gamma$.

It remains to be shown that $P(B_o) = \gamma$. Suppose $P(B_o) < \gamma$; then writing $\epsilon = \gamma - P(B_o) > 0$, one has $P(B_o) = \gamma - \epsilon$. Since $P(B_o) < P(A_o)$, there is a point x in A_o but not in B_o. By property "C", there is some C_k containing x such that $P(C_k) < \epsilon$. Return to the definition of D_k. If $P\left\{D_{k-1} \cup \left[C_k \cap A_o\right]\right\} \leqq \gamma$, then D_k was defined to be $D_{k-1} \cup \left[C_k \cap A_o\right]$. Here

$$P\left\{D_{k-1} \cup \left[C_k \cap A_o\right]\right\} \leq P(D_{k-1}) + P(C_k) \leq P(B_o) + P(C_k) \leq (\gamma - \epsilon) + \epsilon = \gamma .$$

Thus it was the case that $C_k \cap A_o \subset D_k \subset B_o$. But $C_k \cap A_o$ contains a point x not in B_o. This contradiction shows that $P(B_o)$ is actually equal to γ and not less than as was supposed.

APPENDIX C

The following theory was developed as the preparation of the text of this report neared completion. The subject matter is appropriate to this report, and so it is included.

The purpose of this material is to characterize uniformly best tests, or criteria. If there are a family of signal distributions (or hypotheses, in statistical terms), and if a criterion A is an $A_2(k)$ for each of them, then A is a uniformly best test.[1] Theorem C1 states that if all distributions in a family of signal distributions are k-equivalent, all optimum criteria are uniform best tests, and Theorem C2 states the converse.

In the first three cases considered in Part II of The Theory of Signal Detectability, the signal known exactly, the signal known except for carrier phase, and the signal a sample of white Gaussian noise, two signal distributions differing only in signal energy are k-equivalent. Thus, by Theorem C4, a signal distribution with fixed signal energy and one with the signal energy having an arbitrary distribution are k-equivalent in these three cases. These three cases have for the boundaries of their optimum criteria, planes, cylinders, and spheres, respectively. For the other cases, with more complicated criterion boundaries, k-equivalence cannot be expected when energy is changed.

Definition: If $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ and $f_N(x)$ are defined on E^n, and if there exists a set X of probability zero such that for any two points x and y in in E^n, but not in X,

$$\ell_1(x) \geqq \ell_1(y) \text{ if and only if } \ell_2(x) \geqq \ell_2(y) \quad ,$$

then $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ are said to yield k-equivalent distributions.

[1] Neyman and Pearson, Ref. 13.

Theorem C1: If $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ give k-equivalent distributions, then a criterion is an $A_2(k)$ for the first if and only if it is an $A_2(k)$ for the second.

Proof: Suppose A is an $A_2(k)$ for the first distribution. Then by Theorem 7, there is a β such that A is a $A_1(\beta)$. By Theorem 2, A contains all points for which $\ell(x) > \beta$ and none for which $\ell(x) < \beta$, except for a set of probability zero. Except for a set of probability zero, if x and y are any two points such that x is in A and y is not in A, then $\ell_1(x) \geqq \ell_1(y)$. By definition of k-equivalence, there is a set X of probability zero, such that if x and y are also not in X, $\ell_2(x) \geqq \ell_2(y)$. Then there must exist a number β_2 such that for any x except a set of probability zero, $\ell_2(x) \geqq \beta_2$ if x is in A and $\ell_2(x) \leqq \beta_2$ if x is not in β_2. If follows that A is an $A_1(\beta_2)$ with respect to the second distribution. Furthermore, $P_N(A) = k$, for either distribution since the probability density with noise alone is the same for both distributions. It follows by Theorem 5 that A is an $A_2(k)$ for the second distribution.

Theorem C2: If $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ lead to two distributions such that for every k, any criterion A is an $A_2(k)$ for one if and only if it is for the other also, then $f_{SN}^{(1)}(x)$ and $f_{SN}^{(2)}(x)$ lead to k-equivalent distributions.

Proof: Consider the family of sets A_α where $A_\alpha = \left\{x \mid \ell_1(x) \geqq \alpha\right\}$, and α takes on all rational number values greater than zero. Each A_α is an $A_2(k)$ for some k with respect to the first distribution, by Theorem 5. Then it is for the second also, by hypothesis. Each A_α is an $A_1[\beta(\alpha)]$ for some $\beta(\alpha)$, by Theorem 7. For each A_α, the set of points C_α such that x is in A_α and $\ell(x) < \beta(\alpha)$ or x is not in A_α and $\ell(x) > \beta(\alpha)$ has probability zero, by Theorem 2. Let X_1 be the union of all the sets C_α, and since each C_α has probability zero, and the rational numbers and hence the family C_α is countable, it follows the the set X_1 has probability zero.

Now consider the family of sets

$$A_r = \bigcap_{\text{all } \alpha < r} A_\alpha = \left\{x \mid \ell_1(x) \geqq r\right\} \qquad (C.1)$$

defined for every positive real number r. Also define

$$g(r) = \underset{\text{all } \alpha < r}{\ell.\ u.\ b.}\ \beta(\alpha) \qquad (C.2)$$

Then for any point x not in X_1, if x is in A_r, $\ell_2(x) \geqq g(r)$. Also consider the family of sets

$$A^*_r = \bigcup_{\text{all } s > r} A_s = \left\{x \mid \ell_1(x) > r\right\} \qquad (C.3)$$

defined for every positive real number r. If x is a point not in X_1, and if x is not in A^*_r,

$$\ell_2(x) \leqq \underset{\text{all } r^* > r}{g.\ \ell.\ b.}\ g(r^*) \quad . \qquad (C.4)$$

For any value of r at which g(r) is continuous,

$$g(r) = \underset{\text{all } r^* > r}{g.\ \ell.\ b.}\ g(r^*) \quad . \qquad (C.5)$$

Any point x which is not in X_1 and for which $\ell_1(x) = r$ is in A_r but not in A^*r, and therefore

$$g(r) \leqq \ell_2(x) \leqq g(r), \text{ i.e., } \ell_2(x) = g(r) \quad . \qquad (C.6)$$

Clearly g(r) is a monotone increasing function of r. It can therefore have at most a countable number of discontinuities. Let r_o denote a discontinuity in g(r) and suppose that the set of points $B = \left\{x \mid \ell_1(x) = r_o\right\}$ has probability greater than zero. Define

$$h(r_o) = \ell.\ u.\ b.\ \left\{\beta \mid P\left\{x \mid x \epsilon\ B \text{ and } \ell_2(x) < \beta\right\} = 0\right\}$$

$$h^*(r_o) = g.\ \ell.\ b.\left\{\beta \mid P\left(\left\{x \mid x \epsilon\ B \text{ and } \ell_2(x) > \beta\right\}\right) = 0\right\} . \qquad (C.7)$$

The claim is made that $h(r_o) = h^*(r_o)$. Suppose $h(r_o) \neq h^*(r_o)$. Then there exists a number γ such that $h(r_o) < \gamma < h^*(r_o)$. Define

$$C_1 = \left\{ x \mid h(r_o) \leqq \ell_2(x) \leqq \gamma \right\}$$
$$C_2 = \left\{ x \mid \gamma < \ell_2(x) \leqq h^*(r_o) \right\} \quad \cdot \quad , \tag{C.8}$$

both C_1 and C_2 have probability greater than zero, by Eq. (C.7). Now consider the set $A_r - C_2$. It is an $A_2(k)$ for the first distribution, by Theorem 5. Clearly, by Theorems 7 and 2, it cannot be an $A_2(k)$ for the second distribution. The contradiction leads us to conclude that $h(r_o) = h^*(r_o)$. Then for each discontinuity r_o there exists a set of probability zero, say $S(r_o)$, such that if $\ell_1(x) = r_o$ and x is not in $S(r_o)$, $\ell_2(x) = h(r_o)$. Let $X_2 = \bigcup_{\text{all } r_o} S(r_o)$. Then X_2 has probability zero, since there are only a countable number of points of discontinuity r_o. Now define $X = X_1 \cup X_2$, X also has probability zero. Let the function h(r) be defined as follows:

$$h(r) = g(r) \quad \text{if } g(r) \text{ is continuous at } r$$
$$h(r) = h(r_o) \text{ at } r = r_o, \text{ a discontinuity of } g(r). \tag{C.9}$$

The function h(r) has the following properties: (1) h(r) is a monotone increasing function of r, and (2) if $\ell_1(x) = r$, and x is not in X, then $\ell_2(x) = h(r)$. The first assertion is an obvious consequence of the way in which h(r) is defined, and the fact that g(r) is monotone. The second assertion has been shown separately first for points where g, and hence h, is continuous, Eq. (C.6), secondly for the points of discontinuity of h, in the preceding paragraph.

Now suppose x and y are not elements of X, and $\ell_1(x) \geqq \ell_1(y)$. If $\ell_1(x) = r_x$ and $\ell_1(y) = r_y$, then $r_x \geqq r_y$. It follows from the fact that h(r) is monotone increasing that $h(r_x) \geqq h(r_y)$, and since $\ell_2(x) = h(r_x)$ and

$\ell_2(y) = h(r_y)$, $\ell_2(x) \geqq \ell_2(y)$. Since X has probability zero, this completes the proof.

Theorem C3. If $f_{SN}^{(1)}(x)$ is k-equivalent to $f_{SN}^{(i)}(x)$ for each value of i between 2 and n, (or between 2 and ∞), and a_i are positive real numbers such that $\sum_1^n a_i = 1$, (or $\sum_1^\infty a_i = 1$), then $f_{SN}^{(1)}(x)$ and $\sum_1^n a_i f_{SN}^{(i)}(x)$, (or $\sum_1^\infty a_i f_{SN}^{(i)}(x)$) yield k-equivalent distributions.

The set X (in the definition of k-equivalence) for the distribution given by the sum is taken as the union of the sets X for the individual distributions. Then the proof is obvious.

Theorem C4: If $f_{SN}^{(\alpha)}(x)$ is a continuous function of α in an interval $[a, b]$, if for any two numbers α_1 and α_2, $f_{SN}^{(\alpha_1)}(x)$ is k-equivalent to $f_{SN}^{(\alpha_2)}(x)$, and if $F(\alpha)$ is a monotone function which is zero at the left end of the interval and 1 at the right end of the interval, then

$$\int_a^b f_{SN}^{(\alpha)}(x)\, dF(\alpha)$$

is k-equivalent to any $f_{SN}^{(\alpha)}(x)$.

Proof: Choose any α_o in the interval $[a, b]$. Then for each rational value of α in the interval $[a, b]$, $f_{SN}^{(\alpha)}(x)$ and $f_{SN}^{(\alpha_o)}(x)$ are k-equivalent. There is a set X_α, which has probability zero, such that if x, y are not in X_α, $\ell_\alpha(x) \geqq \ell_\alpha(y)$ if and only if $\ell_{\alpha_o}(x) \geqq \ell_{\alpha_o}(y)$. The union X of all X_α with rational α also has probability zero, since the rational numbers are countable. Furthermore, if x and y are not in X, then $\ell_\alpha(x) \geqq \ell_\alpha(y)$ for any rational value of α implies $\ell_{\alpha_o}(x) \geqq \ell_{\alpha_o}(y)$, and $\ell_{\alpha_o}(x) \geqq \ell_{\alpha_o}(y)$ implies $\ell_\alpha(x) \geqq \ell_\alpha(y)$ for all rational values of α. Since $f_{SN}^{(\alpha)}(x)$ is continuous in α, $\ell_\alpha(x)$ must be continuous in α also, and it must follow that for any real α in $[a, b]$ and for any x, y not in X, $\ell_\alpha(x) \geqq \ell_\alpha(y)$ if and only if $\ell_{\alpha_o}(x) \geqq \ell_{\alpha_o}(y)$. Then it is easy to show that if x and y are not in X,

$$\int_a^b \left[\ell_\alpha(x) - \ell_\alpha(y)\right] dF(\alpha) \geqq 0$$

if and only if $\ell_{\alpha_o}(x) \geqq \ell_{\alpha_o}(y)$, and hence $\int_a^b f_{SN}^{(\alpha)}(x)\, dF(\alpha)$ is equivalent to $f_{SN}^{(\alpha_o)}(x)$.

BIBLIOGRAPHY

On Statistical Approaches to the Signal Detectability Problem:

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

   This book is certainly the outstanding reference on threshold signals. It presents a great variety of both theoretical and experimental work. Chapter 7 presents a statistical approach of the criterion type for the signal detection problem, and the idea of a criterion which minimizes the probability of an error is introduced. (This is a special case of an optimum criterion of the first type.)

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

3. Woodward, P. M., and Davies, I. L., "Information Theory and Inverse Probability in Telecommunication," Proc. I.E.E. (London), Vol. 99, Part III, p. 37, March, 1952.

4. Woodward, P. M., and Davies, I. L., "A Theory of Radar Information," Phil. Mag., Vol. 41, p. 1001, 1950.

5. Woodward, P. M., "Information Theory and the Design of Radar Receivers," Proc. I.R.E., Vol. 39, p. 1521.

   Woodward and Davies have introduced the idea of a receiver having a posteriori probability as its output, and they point out that such a receiver gives a maximum amount of information. They have handled the case of an arbitrary signal function known exactly or known except for phase with no more difficulty than other authors have had with a sine wave signal. Their methods serve as a basis for the second part of this report.

6. Reich, E., and Swerling, P., "The Detection of a Sine Wave in Gaussian Noise," Jour. App. Phys., Vol. 24, p. 289, March, 1953.

   This paper considers the problem of finding an optimum criterion of the second type presented in this report for the case of a sine wave of limited duration, known amplitude and frequency, but unknown phase in the presence of Gaussian noise of arbitrary autocorrelation. The method probably could be extended to more general problems. On the other hand, the methods of this report can be applied if the signals are band limited even in the case of non-uniform noise by putting the signals and noise through an imaginary filter to make the noise uniform before applying the theory. See The Theory of Signal Detectability, Part II, Section 3.

7. Middleton, D., "Statistical Criteria for the Detection of Pulsed Carriers in Noise, *Jour. Appl. Phys.*, Vol. 24, p. 371, April, 1953.

A thorough discussion is given of the problem of detecting pulses (of unknown phase) in Gaussian noise. Both types of optimum criteria are discussed, but not in their full generality. The sequential type of test is discussed also.

8. Slattery, T. G., "The Detection of a Sine Wave in Noise by the Use of a Non-Linear Filter," *Proc. I.R.E.*, Vol. 40, p. 1232, October, 1952.

This article considers the problem of detecting a sine wave of known duration, amplitude, and frequency, but unknown phase in uniform Gaussian noise. The article contains several errors, and although the results appear to be correct, they are not clearly presented.

9. Hanse, H., "The Optimization and Analysis of Systems for the Detection of Pulsed Signals in Random Noise," Doctoral Dissertation (MIT), January, 1951.

10. Schwartz, M., "A Statistical Approach to the Automatic Search Problem," Doctoral Dissertation (Harvard), June, 1951.

These dissertations both consider the problem of finding the optimum receiver of the criterion type for radar type signals.

11. North, D. O., "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems," RCA Laboratory Report PTR-6C, 1943.

The ideas of false alarm probability and probability of detection are introduced. North argues that these probabilities will be most favorable when peak signal to average noise ratio is largest. The ideal filter, which maximizes this ratio, is derived. (This commentary is based on second-hand knowledge of the report.)

12. Kaplan, S. M., and Fall, R. W., "The Statistical Properties of Noise Applied to Radar Range Performance," *Proc. I.R.E.*, Vol. 39, p. 56, January, 1951.

The ideas of false alarm probability and probability of detection are introduced and an example of their application to a radar receiver is given.

On Statistics:

13. Neyman, J., and Pearson, E. S., "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Phil. Trans. Roy. Soc.*, Vol. 231, Series A, p. 289, 1933.

14. Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, 1951.

On Related Topics:

15. Dwork, B. M., "Detection of a Pulse Superimposed on Fluctuation Noise," Proc. I.R.E., Vol. 38, p. 771, July, 1950.

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

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

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

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

20. Rice, S. O., "Mathematical Analysis of Random Noise," B.S.T.J., Vol. 23, p. 282-332 and Vol. 24, p. 46-156, 1945-6.

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

LIST OF SYMBOLS

A	The event "The operator says there is signal plus noise present," or a criterion, i.e., the set of receiver inputs for which the operator says there is a signal present.
$A_1(\beta)$	Any criterion A which maximizes $P_{SN}(A) - \beta P_N(A)$, i.e., an optimum criterion of the first type.
$A_2(k)$	Any criterion A for which $P_N(A) \leqq k$, and $P_{SN}(A)$ is maximum, i.e., an optimum criterion of the second type.
CA	The event "The operator says there is noise alone."
d	A parameter describing the ability of a receiver to detect signals. (See Section 5.1 and Fig. 5.1.)
E, $E(s)$	The signal energy.
E^n	The n-dimensional Euclidean space.
$f_N(x)$	The probability density for points x in R if there is noise alone.
$f_{SN}(x)$	The probability density for points x in R if there is signal plus noise.
$F_N(\beta)$, $F_N(\ell)$	The complementary distribution function for likelihood ratio if there is noise alone, i.e., $F_N(\beta)$ is the probability that the likelihood ratio will be greater than β if there is noise alone.
$F_{SN}(\beta)$, $F_{SN}(\ell)$	The complementary distribution function for likelihood ratio if there is signal plus noise.
k	A symbol used primarily for the upper bound placed on false alarm probability $P_N(A)$ in the definition of the second kind of optimum criterion.
$\ell(x)$	The likelihood ratio for the receiver input x. $\ell(x) = \dfrac{f_{SN}(x)}{f_N(x)}$.
n	The dimension of the space of receiver inputs. $n = 2WT$.
N	The event "There is noise alone," or the noise power.
N_o	The noise power per unit bandwidth. $N_o = N/W$.
$P_N(A)$	The probability that the operator will say there is signal plus noise if there is noise alone, i.e., the false alarm probability.

$P_{SN}(A)$	The probability that the operator will say there is signal plus noise if there is signal plus noise, i.e., the probability of detection.
$P_x(SN)$	The a posteriori probability that there is signal plus noise present. (See Sections 1.3 and 2.3.)
$P_S(s)$	The probability measure defined on R for the set of expected signals.
R	The space of all receiver inputs. (The set of all possible signals is the same space.)
s	A signal s(t), which may also be considered as a point s in R with coordinates $(s_1, s_2, \ldots, s_n)$.
SN	The event "There is signal plus noise."
t	Time.
T	The duration of the observation.
W	The bandwidth of the receiver inputs.
x	A receiver input x(t), which may also be considered as a point x in R with coordinates $(x_1, x_2, \ldots, x_n)$
β	A symbol usually used for the likelihood ratio level of an optimum criterion.
$\mu_{SN}(z)$	The mean of the random variable z if there is signal plus noise.
$\mu_N(z)$	The mean of the random variable z if there is noise alone.
$\sigma_N^2(z)$	The variance of the random variable z if there is noise alone.
σ_N^2	The variance of likelihood ratio if there is noise alone.

DISTRIBUTION LIST

1 copy	Director, Electronic Research Laboratory Stanford University Stanford, California Attn: Dean Fred Terman
1 copy	Commanding Officer Signal Corps Electronic Warfare Center Fort Monmouth, New Jersey
1 copy	Chief, Engineering and Technical Division Office of the Chief Signal Officer Department of the Army Washington 25, D. C. Attn: SIGGE-C
1 copy	Chief, Plans and Operations Division Office of the Chief Signal Officer Washington 25, D. C. Attn: SIGOP-5
1 copy	Countermeasures Laboratory Gilfillan Brothers, Inc. 1815 Venice Blvd. Los Angeles 6, California
1 copy	Commanding Officer White Sands Signal Corps Agency White Sands Proving Ground Las Cruces, New Mexico Attn: SIGWS-CM
1 copy	Signal Corps Resident Engineer Electronic Defense Laboratory P. O. Box 205 Mountain View, California Attn: F. W. Morris, Jr.
75 copies	Transportation Officer, SCEL Evans Signal Laboratory Building No. 42, Belmar, New Jersey For - Signal Property Officer Inspect at Destination File No. 25052-PH-51-91(1443)

1 copy	W. G. Dow, Professor Dept. of Electrical Engineering University of Michigan Ann Arbor, Michigan
1 copy	H. W. Welch, Jr. Engineering Research Institute University of Michigan Ann Arbor, Michigan
1 copy	Document Room Willow Run Research Center University of Michigan Willow Run, Michigan
10 copies	Electronic Defense Group Project File University of Michigan Ann Arbor, Michigan
1 copy	Engineering Research Institute Project File University of Michigan Ann Arbor, Michigan

ENGINEERING RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN
ANN ARBOR

THE THEORY OF SIGNAL DETECTABILITY

PART II. APPLICATIONS WITH GAUSSIAN NOISE

ISSUED SEPARATELY:

PART I. THE GENERAL THEORY

Technical Report No. 13
Electronic Defense Group
Department of Electrical Engineering

By: W. W. Peterson
T. G. Birdsall

Approved by: H. W. Welch, Jr.
H. W. Welch, Jr.
Project Engineer

Project M970

TASK ORDER NO. EDG-3
CONTRACT NO. DA-36-039 sc-15358
SIGNAL CORPS, DEPARTMENT OF THE ARMY
DEPARTMENT OF ARMY PROJECT NO. 3-99-04-042
SIGNAL CORPS PROJECT NO. 29-194B-0

July, 1953

TABLE OF CONTENTS

PART II. APPLICATIONS WITH GAUSSIAN NOISE

TABLE OF CONTENTS (Cont.)

PART I. THE GENERAL THEORY (Issued Separately)

LIST OF ILLUSTRATIONS

ABSTRACT

PART I

The several statistical approaches to the problem of signal detectability which have appeared in the literature are shown to be essentially equivalent. A general theory based on likelihood ratio embraces the criterion approach, for either restricted false alarm probability or minimum weighted error type optimum, and the a posteriori probability approach. Receiver reliability is shown to be a function of the distribution functions of likelihood ratio. The existence and uniqueness of solutions for the various approaches is proved under general hypothesis.

PART II

The full power of the theory of signal detectability can be applied to detection in Gaussian noise, and several general results are given. Six special cases are considered, and the expressions for likelihood ratio are derived. The resulting optimum receivers are evaluated by the distribution functions of the likelihood ratio. In two of the special cases studied, the uncertainty of the signal ensemble can be varied, throwing some light on the effect of uncertainty on probability of detection.

ACKNOWLEDGEMENTS

In the work reported here, the authors have been influenced greatly by their association with the other members of the Electronic Defense Group. In particular, Mr. H. W. Batten contributed much to the early phases of the work on signal detectability. Mr. W. C. Fox assisted in the calculations. The authors are indebted to Dr. A. B. Macnee and Dr. J. L. Stewart for the many suggestions resulting from their reading the report.

The authors also wish to acknowledge their indebtedness to Geraldine L. Preston and Jenny-Lea E. Mesler for their assistance in the preparation of the text.

THE THEORY OF SIGNAL DETECTABILITY

Part II. APPLICATIONS WITH GAUSSIAN NOISE

ISSUED SEPARATELY:

Part I. THE GENERAL THEORY

3. INTRODUCTION AND GAUSSIAN NOISE

3.1 Introduction

The chief conclusion obtained from the general theory of signal detectability presented in Part I is that a receiver which calculates the likelihood ratio for each receiver input is the optimum receiver. The receiver can be evaluated (e.g., false alarm probability and probability of detection can be found) if the distribution functions for likelihood ratio are known. It is the purpose of Part II to consider a number of different ensembles of signals with Gaussian noise. For each case, a possible receiver design is discussed. The primary emphasis, however, is on obtaining the distribution functions for likelihood ratio, and hence on estimates of receiver performance for the various cases.

The special 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.

TABLE I

Section	Description of Signal Ensemble	Application
4.2	Signal Known Exactly[1]	Coherent radar with a target of known range and character
4.3	Signal Known Except for Phase[1]	Ordinary pulse radar with no integration and with a target of known range and character.
4.4	Signal a Sample of White Gaussian Noise	Detection of noise-like signals; detection of speech sounds in Gaussian noise.
4.5	Video Design 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.
4.6	A Radar Case (A train of pulses with incoherent phase)[2]	Ordinary pulse radar with integration and with a target of known range and character.
4.8	Signal One of M Orthogonal Signals	Coherent radar where the target is at one of a finite number of non-overlapping positions.
4.9	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.

---

[1] 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.

[2] This is essentially the case treated by Middleton in Ref. 7.

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.[1] 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.

It should be borne in mind that this report discusses the detection of signals in noise; the problem of obtaining information from signals or about signals, except as to whether or not they are present, is not discussed. Furthermore, in treating the special cases, the noise was assumed to be Gaussian.[2]

The reader will probably 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 $\ell(x)$ 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 $\ell(x)$ 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.

---

[1] The only discussion in the literature on the effect of uncertainty on signal detectability which has come to our attention is in Davies, Ref. 2, where the effect upon signal detectability of not knowing carrier phase is shown quantitatively.

[2] See the footnote on page 4 with reference to the spectrum of the assumed noise.

3.2 Gaussian Noise

Throughout Part II, receiver input voltages, which are functions of time, are assumed to be defined for all times t in an observation interval, $0 \leqq t \leqq T$. They are also assumed to be limited to a band of frequencies of width W. By the sampling theorem,[1] each receiver input can be thought of as a point in a 2WT dimensional space, the coordinates of the point being the value of the function at the "sample points" $t = \frac{i}{2W}$, for $1 \leqq i \leqq 2WT$. The notation x(t), or simply x, denotes a receiver input, and x_i denotes the i^{th} sample value, or coordinate. The signal as it would appear at the receiver input in the absence of noise is denoted by s(t), or simply s, and the coordinates, or sample values, of s are denoted by s_i. The receiver input, which may be due to noise alone or to signal plus noise, is random because of the presence of noise. Therefore, only the probability distribution for the receiver inputs x(t) can be specified. The distribution must be given for the receiver inputs both when there is noise alone and when there is signal plus noise. The probability distributions are described in this report by giving the probability density function $f_{SN}(x)$ and $f_N(x)$ for the receiver inputs x in the 2WT dimensional space.

The noise considered in Part II is always Gaussian noise limited to the bandwidth W, and having a uniform spectrum over the band.[2] This is ordinarily called white Gaussian noise. The probability density function for white Gaussian noise, and hence for the receiver inputs when there is noise alone, is:

$$f_N(x) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi N}} \exp \left[- \frac{x_i^2}{2N} \right] \right\}, \text{ or} \tag{3.1}$$

[1] See Appendix D.

[2] If the noise spectrum is band limited, but not uniform, the noise and signals can be put through a filter which makes the noise uniform, and then the theory can be applied. See H. W. Bode and C. E. Shannon, "A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory," Proc. I.R.E., Vol. 38, p. 417, April 1950.

$$f_N(x) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2N}\sum x_i^2\right]^1 \qquad (3.2a)$$

where n is the dimension of the space, i.e., 2WT,[2] and N is the noise power.[3] It can be shown that this ensemble of noise functions has a Gaussian distribution at every time and that its spectrum is uniform.

By the sampling theorem,[4]

$$\sum x_i^2 = 2W \int_0^T \left[x(t)\right]^2 dt \quad . \qquad (3.3)$$

Therefore

$$f_N(x) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{N_o}\int_0^T x(t)^2\, dt\right] , \qquad (3.2b)$$

where $N_o = \frac{N}{W}$ is the noise power per unit bandwidth.[5]

If the signals and their probabilities are known, then the signal plus noise probability density function, $f_{SN}(x)$, can be found by the convolution integral, as described in Section 2.[6]

---

[1]Unless otherwise indicated, the limits on the sum are i = 1 to i = n = 2WT.

[2]If $\frac{1}{2\pi N} \exp\left[-\frac{x_i^2}{2N}\right]$ is called $f_N(x_i)$, then $f_N(x) = \prod_{i=1}^{n} f_N(x_i)$, i.e., the x_i are independent and each has $f_N(x_i)$ for its probability density function. For a discussion of "independent," see Cramér, Ref. 14, p. 159.

[3]This assumes the circuit impedance is normalized to one ohm.

[4]See Appendix D.

[5]This form of the expression for $f_N(x)$, and the corresponding forms of the equations for $f_{SN}(x)$ and $\mathcal{L}(x)$ were first derived by Woodward. See Woodward and Davies, Refs. 2 and 3.

[6]See page 13 of Part I.

$$f_{SN}(x) = \int_R f_N(x-s)\,dP_S(s) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \int_R \exp\left[-\frac{1}{2N}\sum_{i=1}^{n}(x_i-s_i)^2\right] dP_S(s)$$

$$= \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2N}\sum_{i=1}^{n}x_i^2\right]\int_R \exp\left[-\frac{1}{2N}\sum_{i=1}^{n}s_i^2\right]\exp\left[\frac{1}{N}\sum_{i=1}^{n}x_i s_i\right] dP_S(s) \tag{3.4a}$$

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

$$= \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{N_o}\int_0^T x^2\,dt\right]\int_R \exp\left[-\frac{1}{N_o}\int_0^T s^2 dt\right]\exp\left[\frac{2}{N_o}\int_0^T xs\,dt\right] dP_S(s) \tag{3.4b}$$

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

$$\int_0^T s^2 dt = \frac{1}{2W}\sum s_i^2 = E(s) \quad ^{1} \tag{3.5}$$

is the energy of the expected signal, while

$$\int_0^T x(t)\,s(t)\,dt = \frac{1}{2W}\sum x_i s_i \tag{3.6}$$

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

---

[1]See footnote 3 on page 5.

3.3 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 (3.4) by Eq (3.2).

$$\ell(x) = \int_R \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{1}{N} \sum_{i=1}^{n} x_i s_i\right] dP_S(s), \quad \text{or} \tag{3.7a}$$

$$\ell(x) = \int_R \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{2}{N_o} \int_0^T x(t)\, s(t)\, dt\right] dP_S(s) \ . \tag{3.7b}$$

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

$$\ell_s(x) = \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{1}{N} \sum_{i=1}^{n} x_i s_i\right], \quad \text{or} \tag{3.8a}$$

$$= \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{2}{N_o} \int_0^T x(t)\, s(t)\, dt\right] \ . \tag{3.8b}$$

Thus the general formulas (3.7a) and (3.7b) for likelihood ratio state that $\ell(x)$ is the weighted average of $\ell_s(x)$ over the set of all signals, i.e.,

$$\ell(x) = \int_R \ell_s(x)\, d\, P_S(s) \ . \tag{3.9}$$

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,[1] the expression for likelihood ratio can be simplified somewhat. If these parameters are indicated by $r_1, r_2, \ldots, r_n$, and the associated probability density functions are denoted by $f_1(r_1), f_2(r_2), \ldots, f_n(r_n)$, then

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

The likelihood ratio becomes

$$\ell(x) = \int \cdots \int \ell_s(x)\, f_1(r_1) \cdots f_n(r_n)\, dr_1 \cdots dr_n$$

$$= \int \left[f_n(r_n) \cdots \left[\int f_1(r_1)\ \ell_s(x)\, dr_1 \right] \cdots \right] dr_n \quad . \tag{3.10}$$

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

---

[1]Cramér, Ref. 14, p. 159.

4. LIKELIHOOD RATIO AND ITS DISTRIBUTION FOR SPECIAL CASES

4.1 Introduction

The purpose of this section is to derive expressions or approximate expressions for likelihood ratio and its distribution functions for a number of special signals in the presence of Gaussian noise. The results obtained in this section are summarized and discussed in Section 5.

4.2 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 3.3, Eq (3.8).

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \exp\left[\frac{1}{N} \sum_{i=1}^{n} x_i s_i\right] , \tag{4.1a}$$

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \exp\left[\frac{2}{N_o} \int_0^T x(t)\, s(t)\, dt\right] \tag{4.1b}$$

As the first step in finding the distribution functions for $\ell(x)$, it is convenient to find the distribution for $\frac{1}{N} \sum x_i s_i$ when there is noise alone. Then the input $x = (x_i, x_2, \ldots, x_i)$ is due to white Gaussian noise. It can be seen from Eq (3.1) 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 $\frac{1}{N}(x_i s_i)$ has a normal distribution with mean $\frac{s_i}{N}$ times the mean of x_i, and variance $\frac{s_i^2}{N^2}$ times the variance of x_i — zero and $\frac{s_i^2}{N^2} N = \frac{s_i^2}{N}$ respectively. Because the x_i are independent, the summands $\frac{1}{N} s_i x_i$ are independent, each with normal distributions, 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,[1]

$$\sum \frac{s_i^2}{N} = \frac{2WE(s)}{N} = \frac{2E}{N_o} = 2\times\frac{\text{Signal Energy}}{\text{Noise Power Per Unit Bandwidth}} \quad . \tag{4.2}$$

The distribution for $\frac{1}{N} \sum x_i s_i$ with noise alone is thus normal with zero mean and variance $\frac{2E}{N_o}$. Recalling (4.1a)

$$\ell(x) = \exp\left[-\frac{E}{N_o} + \frac{1}{N}\sum x_i s_i\right] \tag{4.1a}$$

it is seen that the distribution for $\frac{1}{N}\sum x_i s_i$ can be used directly by introducing α defined by

$$\beta = \exp\left[-\frac{E}{N_o} + \alpha\right] \quad , \text{ or } \alpha = \frac{E}{N_o} + \ell n\, \beta \tag{4.3}$$

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

$$F_N(\beta) = \sqrt{\frac{N_o}{4\pi E}} \int_{\alpha}^{\infty} \exp\left[-\frac{1}{2}\frac{N_o}{2E}y^2\right] dy \tag{4.4}$$

The distribution for the case of signal plus noise can be found by using Theorem 8, which states that[2]

$$dF_{SN}(\beta) = \beta\, dF_N(\beta) \quad . \tag{4.5}$$

---

[1] Cramér, Ref. 14, p. 212.

[2] See Part I, pp. 24 and 27.

Differentiating Eq (4.4),

$$dF_N(\beta) = -\sqrt{\frac{N_o}{4\pi E}} \exp\left(-\frac{N_o\alpha^2}{4E}\right) d\alpha \quad ,^1 \qquad (4.6)$$

and combining (4.3), (4.5), and (4.6),

$$dF_{SN}(\beta) = -\sqrt{\frac{N_o}{4\pi E}} \exp\left[-\frac{E}{N_o} + \alpha - \frac{N_o\alpha^2}{4E}\right] d\alpha \qquad (4.7)$$

Thus

$$F_{SN}(\beta) = \sqrt{\frac{N_o}{4\pi E}} \int_{\alpha}^{\infty} \exp\left[-\frac{N_o}{4E}\left(y - \frac{2E}{N_o}\right)^2\right] dy \qquad (4.8)$$

In summary, α, and therefore $\ell n\ \beta$, has a normal distribution with signal plus noise as well as with noise alone; the variance of both distributions is $\frac{2E}{N_o}$, and the difference of the means is $\frac{2E}{N_o}$.

The receiver operating characteristic curves in Fig. 4.1 are plotted for any case in which $\ell n\ \ell$ 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 $d = \frac{2E}{N_o}$.

Eq (4.1b) describes what the ideal receiver should do for this case. The essential operation in the receiver is obtaining the correlation, $\int_0^T s(t)x(t)dt$.

[1]The change in sign appears because the distribution functions $F_{SN}(\beta)$ and $F_N(\beta)$ are probabilities that $\ell(x)$ will lie between β and ∞, not $-\infty$ and β as is usually the case. If the density function for $F_{SN}(\beta)$ is called $g(\beta)$, then

$$\frac{dF_{SN}(\beta)}{d\beta} = -g(\beta), \text{ and } F_{SN}(\beta) = \int_{\beta}^{\infty} g(\beta)\, d\beta.$$

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.[1]

If the form of the signal is simple, there is a simple way to obtain this cross correlation.[2] Suppose h(t) is the impulse response of a filter. The response $e_o(t)$ of the filter to a voltage x(t) is[3]

$$e_o(t) = \int_{-\infty}^{t} x(\tau)\, h(t-\tau)\, d\tau \tag{4.9}$$

If a filter can be synthesized so that

$$\begin{aligned} h(t) &= s(T-t) && 0 \leqq t \leqq T \\ h(t) &= 0 && \text{otherwise,} \end{aligned} \tag{4.10}$$

then

$$e_o(T) = \int_{0}^{T} x(\tau)\, s(\tau)\, d\tau \quad , \tag{4.11}$$

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.[4]

---

[1] Harrington and Rogers, Ref. 16; Harting and Meade, Ref. 17; Lee, Cheatham, and Wiesner, Ref. 18; Levin and Reintzes, Ref. 19.

[2] This appears to be due to Woodward. See Woodward, Ref. 5, and Woodward and Davies, Ref. 3.

[3] S. Goldman, Transformation Calculus and Electrical Transients, Prentice Hall, New York, 1949, p. 112.

[4] Lawson and Uhlenbeck, Ref. 1, p. 206; North, Ref. 11.

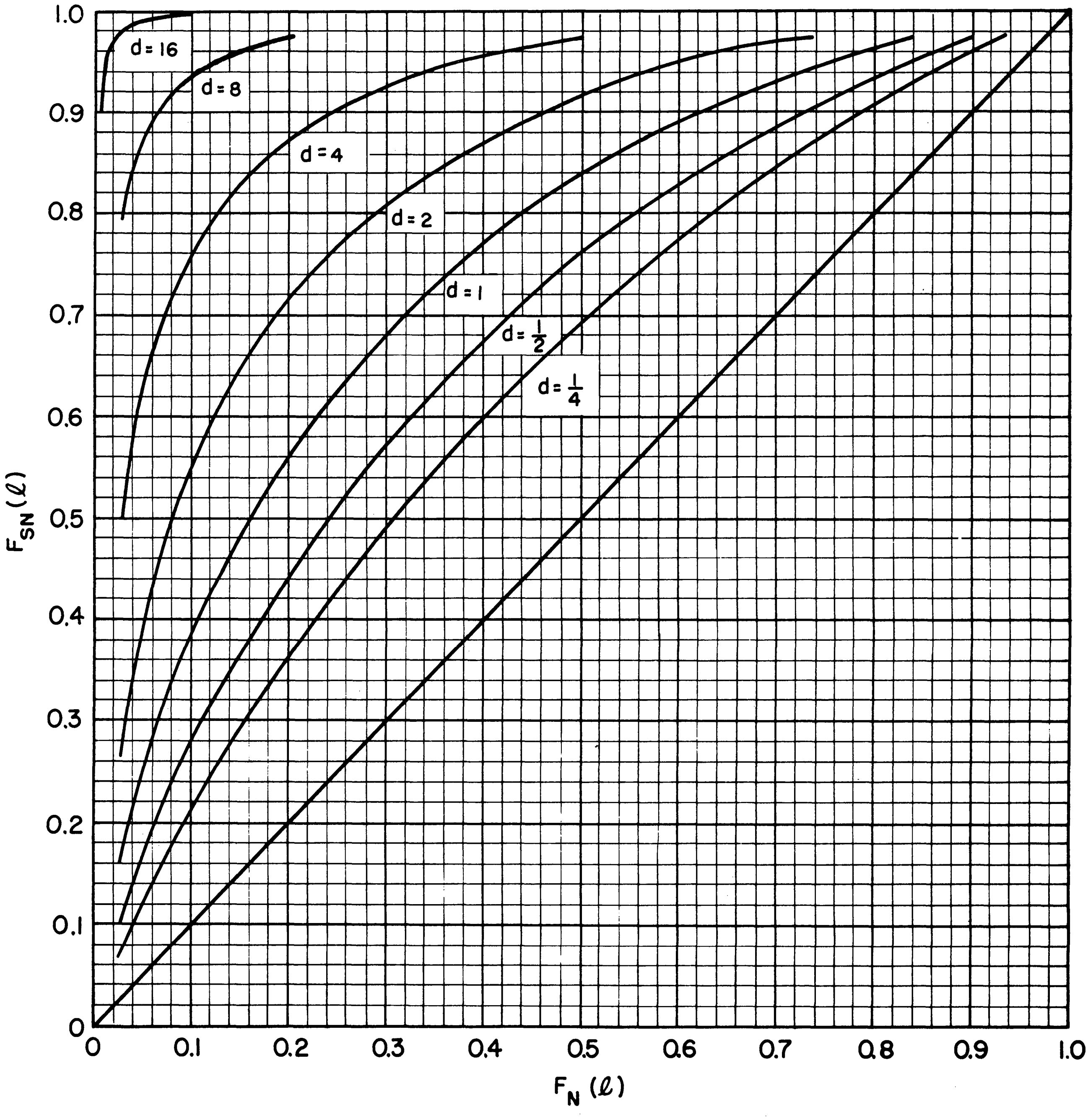

FIG. 4.1

RECEIVER OPERATING CHARACTERISTIC.

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

M-970 A-63-73 JEM 7-17-53

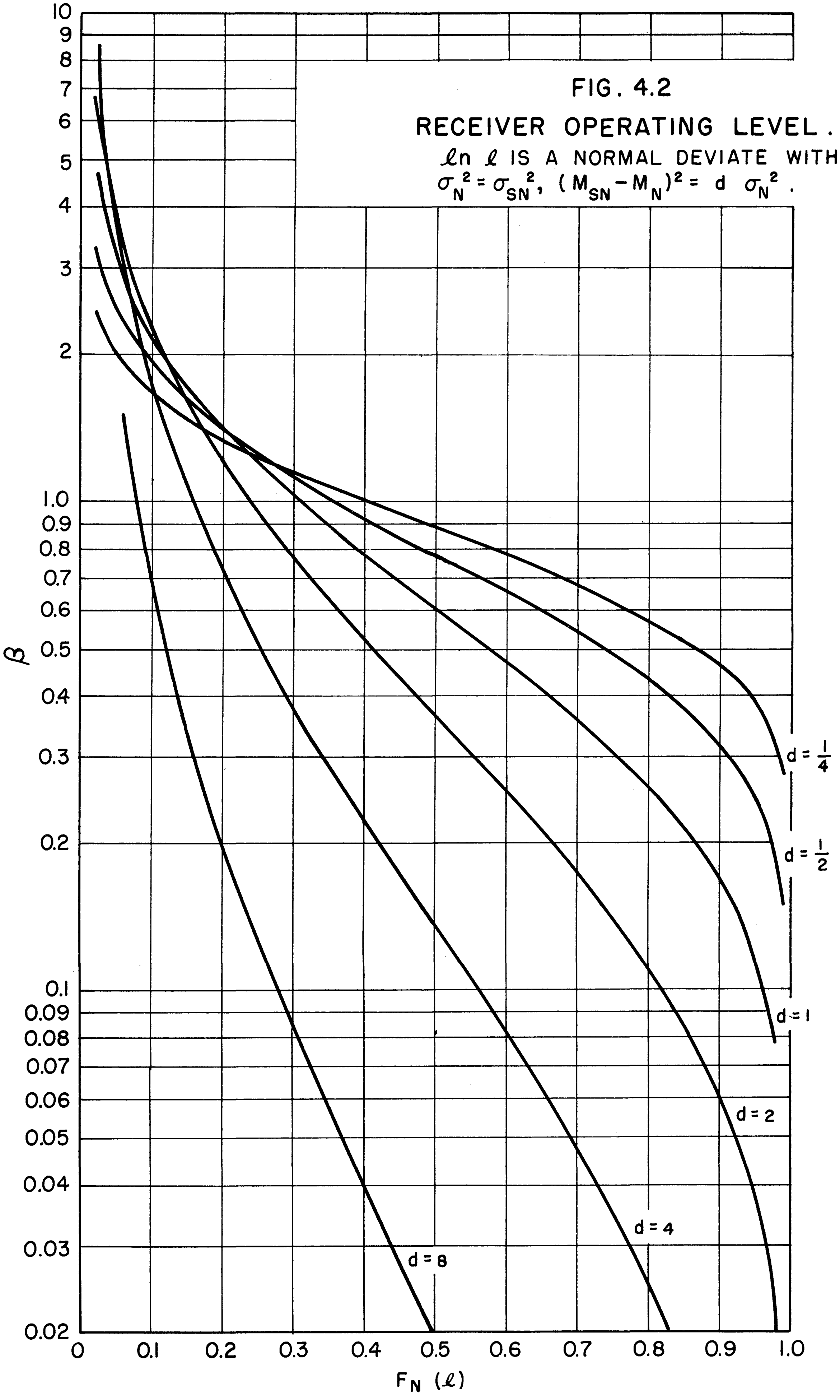

FIG. 4.2

RECEIVER OPERATING LEVEL.

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

M-970 A-63-71 JEM 7-13-53

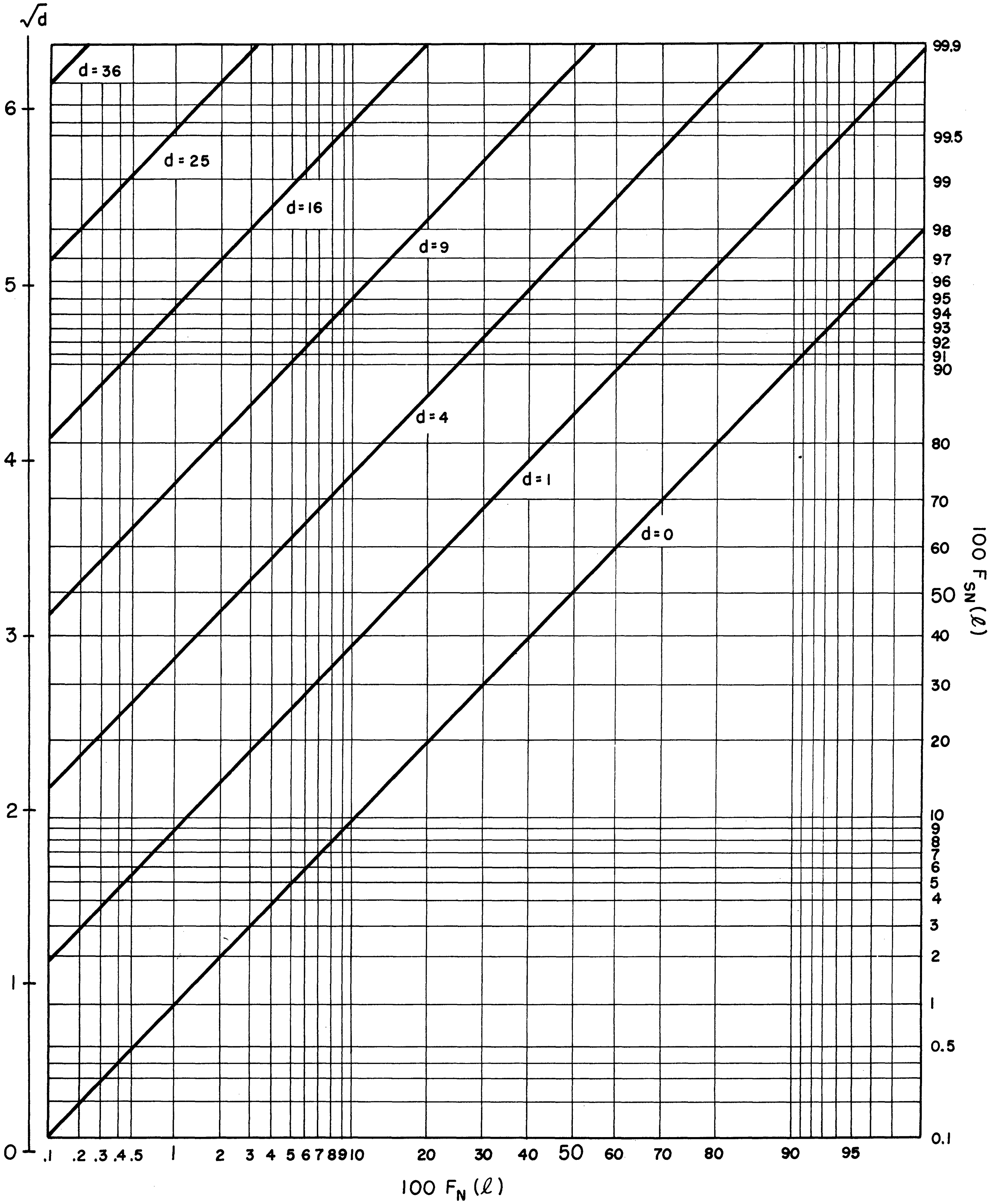

FIG. 4.3

RECEIVER OPERATING CHARACTERISTIC.

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

M-970 A-G3-74 JEM 7-17-53

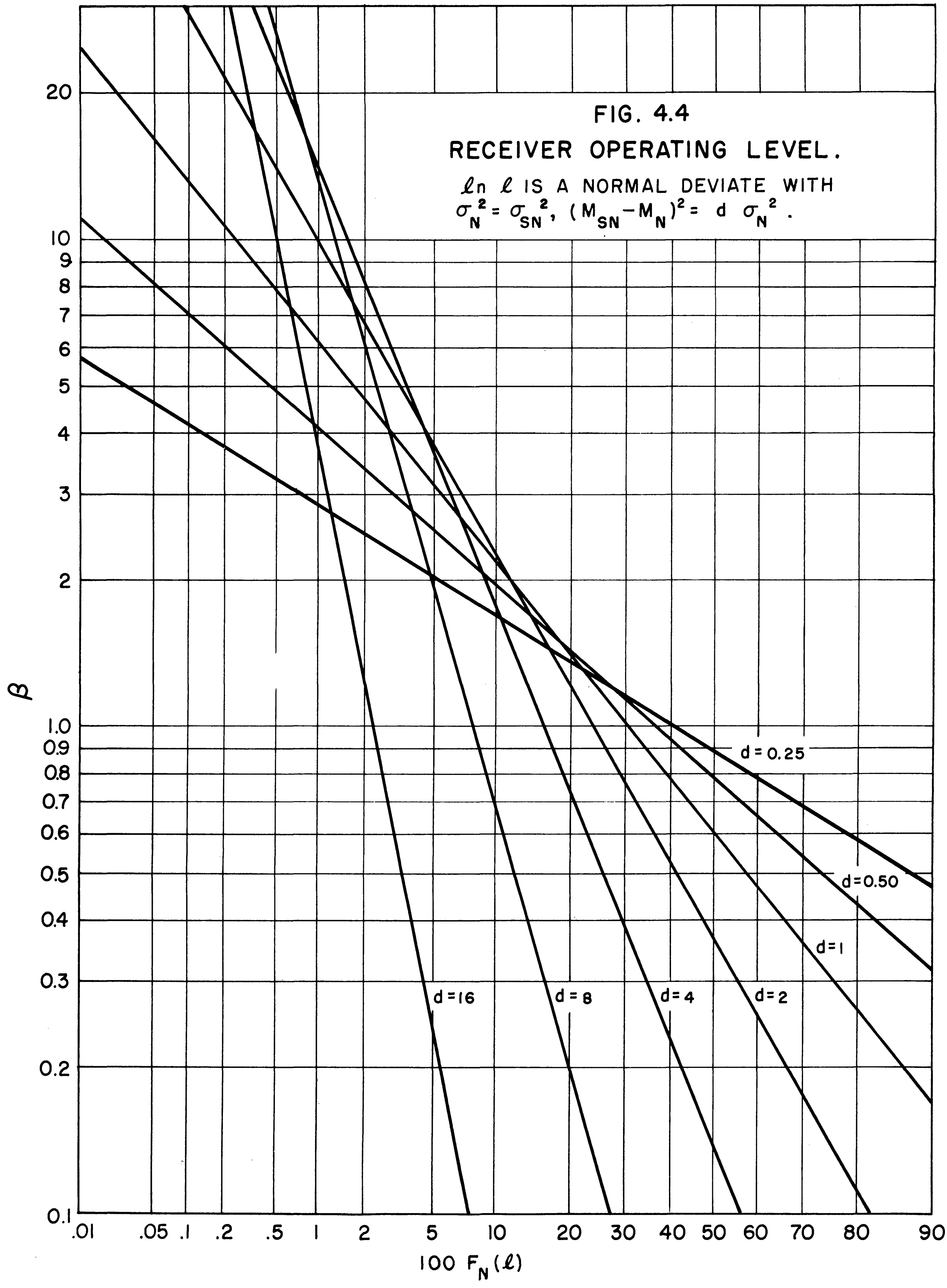

FIG. 4.4
RECEIVER OPERATING LEVEL.
$\ell n\ \ell$ IS A NORMAL DEVIATE WITH $\sigma_N^2 = \sigma_{SN}^2$, $(M_{SN} - M_N)^2 = d\ \sigma_N^2$.

4.3 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.

$$s(t) = f(t) \cos\left(\omega t + \phi(t) - \theta\right) . \tag{4.12}$$

Since the unknown phase angle θ has a uniform distribution,

$$dP_S(\theta) = \frac{1}{2\pi} d\theta. \tag{4.13}$$

The likelihood ratio can be found by applying Eq (3.7), and since the signal energy $E(s)$ is the same for all values of carrier phase θ,[1]

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \int_R \exp\left[\frac{1}{N}\sum x_i s_i\right] dP_S(s) \tag{4.14}$$

Expanding s into the coefficients of cos θ and sin θ will be helpful:[2]

$$s(t) = f(t) \cos\left(\omega t + \phi(t)\right) \cos\theta + f(t) \sin\left(\omega t + \phi(t)\right) \sin\theta , \tag{4.15}$$

and

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

$$+ \sin\theta \frac{1}{N}\sum x_i f(t_i) \sin\left(\omega t_i + \phi(t_i)\right) \tag{4.16}$$

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

---

[1] For this to be rigorously true, it is sufficient that the signal be time limited and have its line spectrum zero at zero frequency and at all frequencies equal to or greater than $\frac{2\omega}{2\pi}$.

[2] t_i denotes the i^{th} sample point, i.e., $t_i = \frac{i}{2W}$.

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

$$\frac{1}{N} r \sin \theta_o = \frac{1}{N} \sum_o x_i f(t_i) \sin \left(\omega t_i + \phi(t_i) \right) \qquad (4.17)$$

and therefore

$$\frac{1}{N} \sum x_i s_i = \frac{r}{N} \cos (\theta - \theta_o) \qquad (4.18)$$

Using this form the likelihood ratio becomes

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

$$= \exp \left[- \frac{E}{N_o} \right] I_o \left(\frac{r}{N} \right) \qquad (4.19)$$

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 $\frac{r}{N}$ is greater than some value corresponding to β. The quantity r is defined by the Eq (4.17); $\frac{r}{N}$ is the square root of the sums of the squares of the right-hand sides. The probability that $\frac{r}{N}$ will exceed any certain value can be computed by observing that each of the right-hand sides is $\frac{N_o}{2}$ times the cross correlation of x(t) with a fixed signal, either $f(t) \cos \left[\omega t + \phi(t) \right]$ or $f(t) \sin \left[\omega t + \phi(t) \right]$. Therefore, the distribution of each can be found in the same manner as the distribution of $\frac{1}{N} \sum x_i s_i$ was found for the case of the signal known exactly,[1] and both $\frac{r}{N} \cos \theta_o$ and $\frac{r}{N} \sin \theta_o$ have normal distributions with zero mean and variance $\frac{2E}{N_o}$. Furthermore, $f(t) \cos \left(\omega t + \phi(t) \right)$ and $f(t) \sin \left(\omega t + \phi(t) \right)$ are out of phase, or orthogonal, and therefore $r \cos \theta_o$ and $r \sin \theta_o$ have independent distributions.[2]

[1]See page 9. [2]See footnote 1, p. 17.

Because $\frac{r}{N} = \sqrt{\left(\frac{r}{N} \cos \theta_o\right)^2 + \left(\frac{r}{N} \sin \theta_o\right)^2}$, the probability that $\frac{r}{N}$ will exceed any fixed value is given by the well-known chi-square distribution for two degrees of freedom, $K_2(\alpha^2)$.[1] The proper normalization yielding zero mean and unit variance requires that the variable be $\frac{r}{N}\sqrt{\frac{N_o}{2E(s)}}$, that is

$$P_N\left(\frac{r}{N}\sqrt{\frac{N_o}{2E}} \geq \alpha\right) = K_2(\alpha^2) = \exp\left[-\frac{\alpha^2}{2}\right]. \tag{4.20}$$

If α is defined by the equation

$$\beta = \exp\left[-\frac{E}{N_o}\right] I_o\left(\sqrt{\frac{2E}{N_o}}\ \alpha\right), \tag{4.21}$$

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

$$F_N(\beta) = \exp\left[-\frac{\alpha^2}{2}\right]. \tag{4.22}$$

Using Theorem 8 of Section 2, namely

$$\beta\ dF_N(\beta) = dF_{SN}(\beta) \tag{4.23}$$

but making use of the parameter α, we form first

$$dF_N(\beta) = -\alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha, \tag{4.24}$$

and hence

$$dF_{SN}(\beta) = -\exp\left[-\frac{E}{N_o}\right]\alpha \exp\left[-\frac{\alpha^2}{2}\right] I_o\left(\sqrt{\frac{2E}{N_o}}\ \alpha\right) d\alpha \tag{4.25}$$

Integrate from α to infinity.

$$F_{SN}(\beta) = \exp\left[-\frac{E}{N_o}\right]\int_\alpha^\infty \alpha \exp\left[-\frac{\alpha^2}{2}\right] I_o\left(\sqrt{\frac{2E}{N_o}}\ \alpha\right) d\alpha . \tag{4.26}$$

---

[1] Cramér, Ref. 14, p. 233, or Hoel, P. G., Introduction to Mathematical Statistics, Wiley, 1947, p. 134.

Eqs (4.22) and (4.26) yield the receiver operating characteristic in parametric form, and Eq (4.21) gives the associated operating levels.[1] These are graphed in Fig. 4.5 for the same values of signal energy to noise per unit bandwidth ratio as were used when the phase angle was known exactly, Fig. 4.1, 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 $\left(\phi(t) = 0\right)$ in Eq. (4.12)). Let us also choose $\theta = 0$. (Any phase could have been chosen.) Then the filter has impulse response

$$\begin{aligned} h(t) &= f(T-t) \cos\left[\omega(T-t)\right] \qquad 0 \le t \le T \\ &= 0 \qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned} \tag{4.27}$$

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

$$\begin{aligned} e_o(t) &= \int_{-\infty}^{t} x(\tau)\, h(t-\tau)\, d\tau = \int_{t-T}^{t} x(\tau)\, f(\tau+T-t) \cos\omega(\tau+T-t)\, d\tau \\ &= \cos\omega(T-t) \int_{t-T}^{t} x(\tau)\, f(\tau+T-t) \cos\omega\tau\, d\tau \\ &\qquad - \sin\omega(T-t) \int_{t-T}^{t} x(\tau)\, f(\tau+T-t) \sin\omega\tau\, d\tau \quad . \end{aligned} \tag{4.28}$$

[1] Graphs of values of the integral (4.26) along with approximate expressions for small and for large values of α appear in Rice, Ref. 20. 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.

M-970 A-63-75 JEM 7-17-53

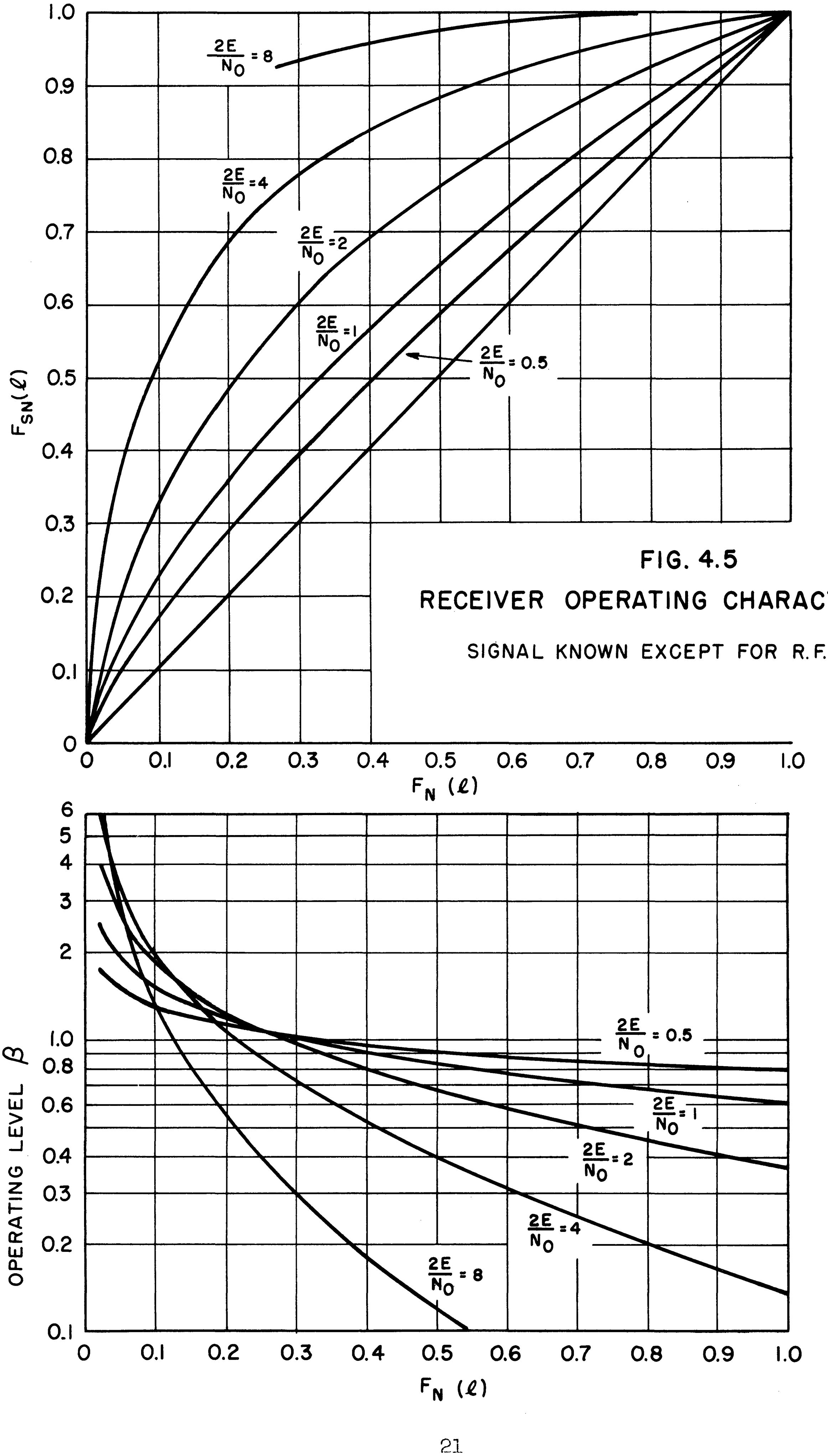

FIG. 4.5

RECEIVER OPERATING CHARACTERISTIC.

SIGNAL KNOWN EXCEPT FOR R.F. PHASE.

The envelope of the filter output will be the square root of the sum of the squares of the integrals,[1] and the envelope at time T will be proportional to $\frac{r}{N}$, since

$$\left(\frac{r}{2W}\right)^2 = \left[\int_0^T x(\tau)\, f(\tau) \cos \omega\tau\, d\tau\right]^2 + \left[\int_0^T x(\tau)\, f(\tau) \sin \omega\tau\, d\tau\right]^2. \quad (4.29)$$

$$= \text{Square of the envelope, at time T, of } e_o(t).$$

If the input x (t) passes through the filter with an impulse response given by Eq (4.27), then through a linear detector, the output will be $\frac{N_o}{2}\,\frac{r}{N}$ at time T. Because the likelihood ratio, Eq (4.19), is a known monotone function of $\frac{r}{N}$, the output can be calibrated to read the likelihood ratio of the input.

4.4 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:[2]

$$f_{SN}(x) = \left(\frac{1}{2\pi(N+S)}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\,\frac{1}{N+S} \sum x_i^2\right]. \quad (4.30)$$

The likelihood ratio is

$$\ell(x) = \left(\frac{N}{N+S}\right)^{\frac{n}{2}} \exp\left[\frac{1}{2}\,\frac{1}{N} \sum x_i^2 - \frac{1}{2}\,\frac{1}{N+S} \sum x_i^2\right] \quad (4.31)$$

---

[1] If the line spectrum of x(t) is zero at zero frequency and at all frequencies equal to or greater than $\frac{2\omega}{2\pi}$, then it can be shown that these integrals contain no frequencies as high as $\frac{\omega}{2\pi}$.

[2] Cramér, Ref. 14, p. 212.

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

$$\beta = \left(\frac{N}{N+S}\right)^{\frac{n}{2}} \exp\left(\frac{S}{N+S}\,\frac{\alpha^2}{2}\right) \quad . \tag{4.32}$$

Then the condition $\ell(x) \geq \beta$ is equivalent to the condition that $\frac{1}{N}\sum x_i \geq \alpha^2$. In the presence of noise alone the random variables $\left(\frac{x_i}{\sqrt{N}}\right)$ 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,[1] i.e.,

$$F_N(\beta) = K_n(\alpha^2) \quad . \tag{4.33}$$

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

$$F_{SN}(\beta) = K_n\left(\frac{N}{N+S}\,\alpha^2\right) \quad . \tag{4.34}$$

Receiver operating characteristic curves are presented in Figs. 4.6 and 4.7 for four possible choices of n (10^2, 10^3, 10^4, 10^5), and in each case for three values of signal to noise ratio three db apart.

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

---

[1]Cramér, Ref. 14, p. 233. Tables of $K_n(\alpha^2)$ can be found in most books on statistics. Extensive tables are listed in the bibliography of Ref. 14, p. 570.

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

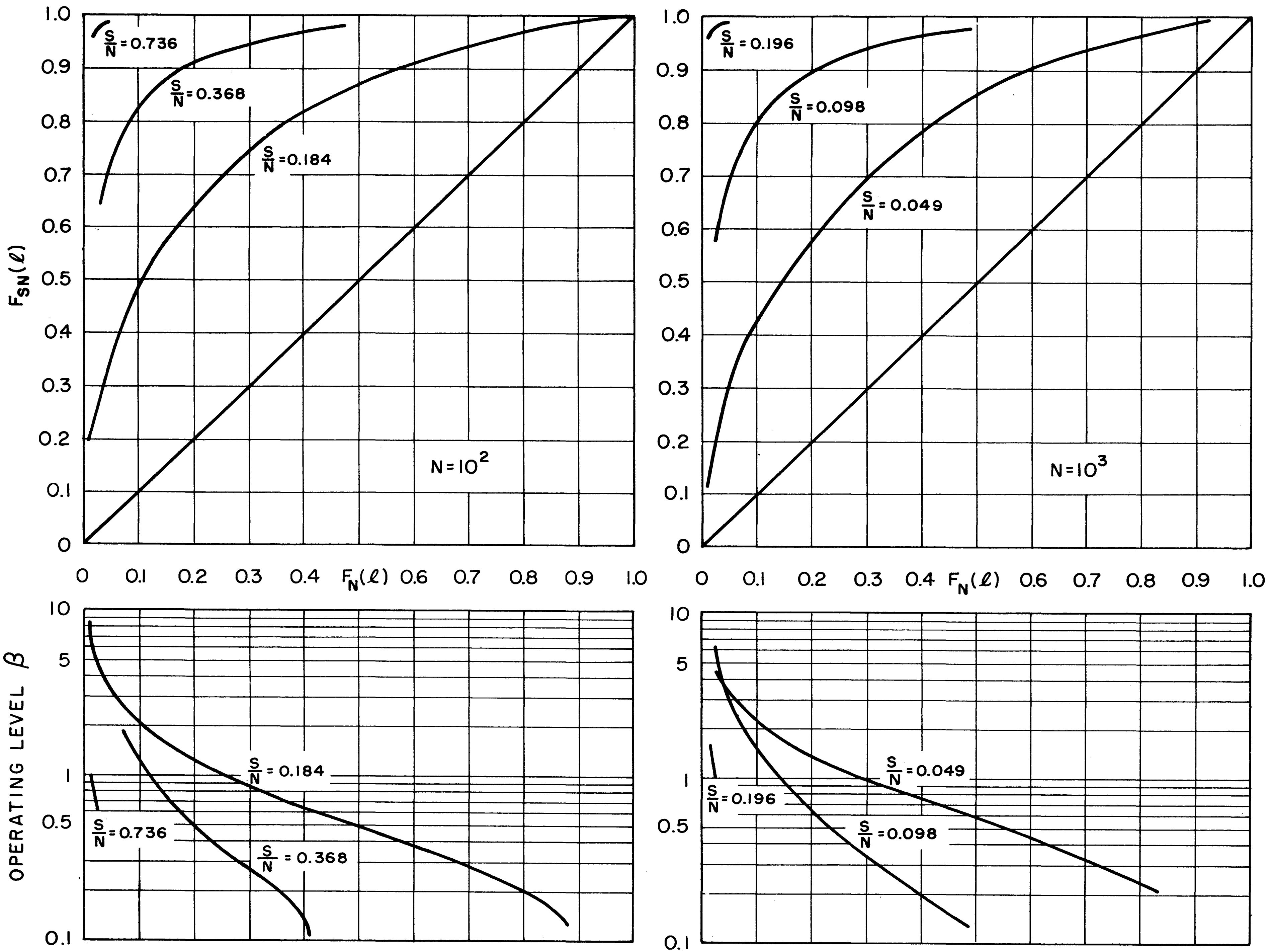

FIG. 4.6. RECEIVER OPERATING CHARACTERISTIC.
SIGNAL A SAMPLE OF WHITE GAUSSIAN NOISE.

M-970 A-63-76 JEM 7-20-53

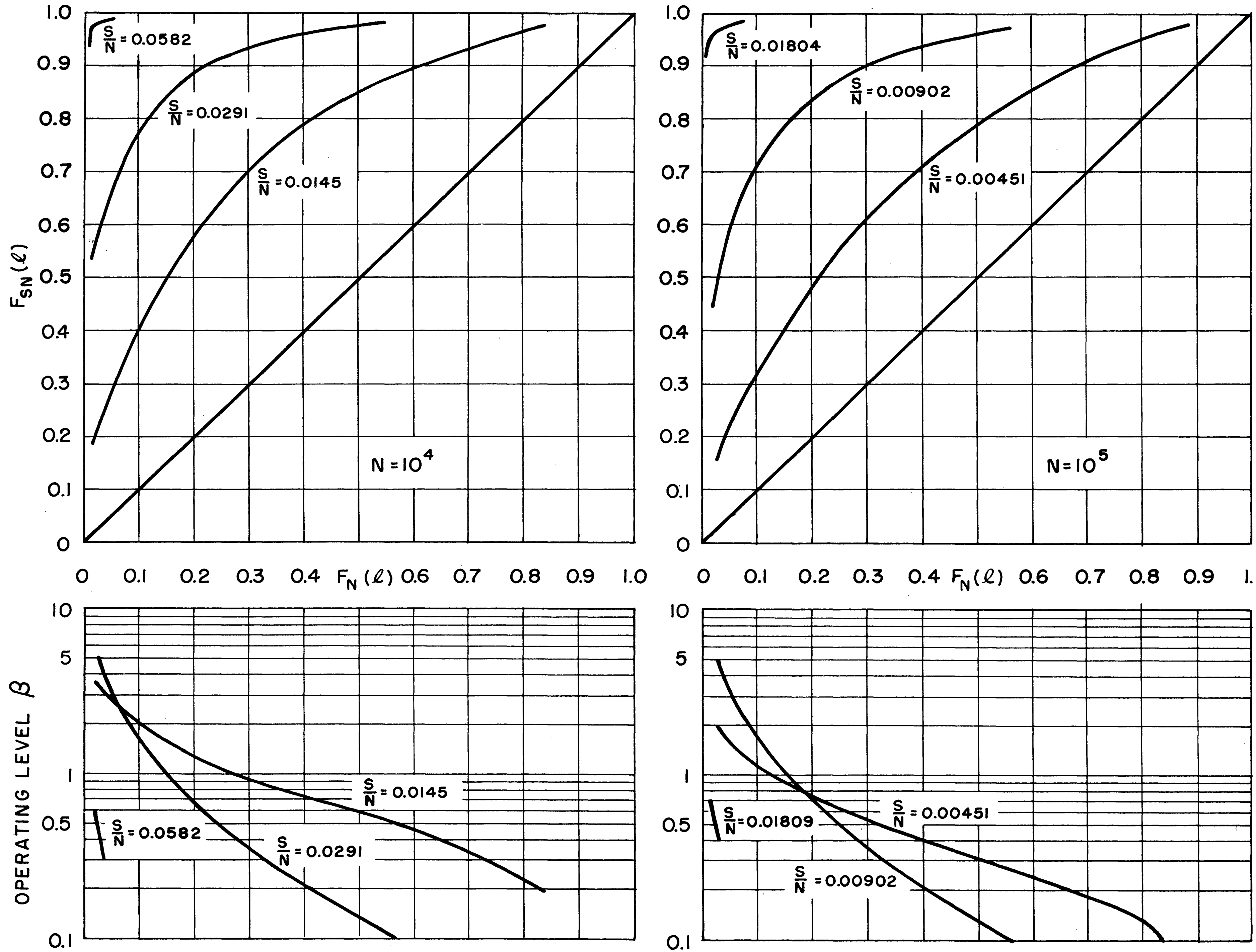

FIG. 4.7. RECEIVER OPERATING CHARACTERISTIC.
SIGNAL A SAMPLE OF WHITE GAUSSIAN NOISE.

M-970 A-G3-77 JEM 7-20-53

$$K_n(\alpha^2) \approx \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2\,\alpha^2} - \sqrt{2n-1}}^{\infty} \exp\left[-\frac{1}{2}y^2\right] dy \tag{4.35}$$

and

$$K_n\left(\frac{N}{N+S}\,\alpha^2\right) \approx \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\frac{2N\alpha^2}{N+S}} - \sqrt{2n-1}}^{\infty} \exp\left[-\frac{1}{2}y^2\right] dy$$

If the signal energy is small compared to that of the noise, $\sqrt{\frac{N}{N+S}}$ is nearly unity and both distributions have nearly the same variance. Then Fig. 4.1 applies to this case too, with the value of d given by

$$d = (2n-1)\left(1 - \sqrt{\frac{N}{N+S}}\right)^2 \quad . \tag{4.37}$$

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

$$1 - \sqrt{\frac{N}{N+S}} \approx \frac{1}{2}\frac{S}{N} \qquad \text{for } \frac{S}{N} << 1 \quad ,$$

$$d \approx \frac{nS^2}{2N^2} \tag{4.38}$$

Two signal to noise ratios, $\left(\frac{S}{N}\right)_1$ and $\left(\frac{S}{N}\right)_2$, will have approximately the same operating characteristic if the corresponding numbers of sample points, n_1 and n_2, satisfy

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

This can be verified for the three curves of Fig. 4.7 for $n = 10^5$, compared with Fig. 4.1 for $d = 1, 4, 16$.

The receiver specified is any device that produces the likelihood ratio of its input,

$$\ell(x) = \left(\frac{N}{N+S}\right)^{\frac{n}{2}} \exp\left[\frac{S}{N+S}\,\frac{1}{N}\,\sum x_i^{\,2}\right] \quad . \tag{4.31}$$

An energy detector has as its output

$$e_o(t) = \int_0^T \left[x(t)\right]^2 dt = \frac{1}{2W}\,\sum x_i^{\,2} \tag{4.40}$$

and this receiver can be calibrated so that its output at the end of the observation time, $e_o(t)$, will be read as

$$\ell(x) = \left(\frac{N}{N+S}\right)^{\frac{n}{2}} \exp\left[\frac{S}{N+S}\,\frac{e_o(T)}{N_o}\right] \tag{4.41}$$

4.5 Video Design of a Broad Band Receiver

The problem considered in this section is represented schematically in Fig. 4.8. The signals and noise are assumed to have passed through a band

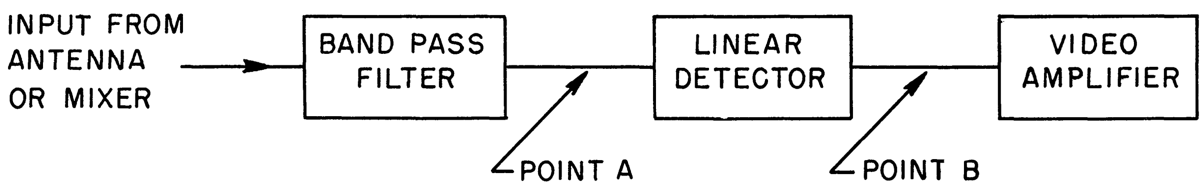

FIG. 4.8

BLOCK DIAGRAM OF A BROAD BAND RECEIVER.

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

$\frac{\omega}{2\pi} > \frac{W}{2}$. 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 $\frac{\omega}{2\pi} > \frac{W}{2}$. Then the alternate form of the sampling theorem can be used.[1] Any such function f(t) can be expanded as follows:

$$f(t) = x(t) \cos \omega t + y(t) \sin \omega t \tag{4.42}$$

where x(t) and y(t) are band limited to frequencies no higher than $\frac{W}{2}$, and hence can themselves be expanded by the sampling theorem:

$$f(t) = \sum_i \left[x\left(\frac{i}{W}\right) \psi_i(t) \cos \omega t + y\left(\frac{i}{W}\right) \psi_i(t) \sin \omega t \right] . \tag{4.43}$$

The function can be thought of as a point in a space of n = 2WT dimensions with coordinates $x\left(\frac{i}{W}\right) = x_i$ and $y\left(\frac{i}{W}\right) = y_i$. This is a rectangular coordinate

[1] See Appendix D.

system, since the family of functions $\psi_i(t) \cos \omega t$ and $\psi_i(t) \sin \omega t$ form an orthogonal system.

The amplitude of the function f(t) is

$$r(t) = \sqrt{\left[x(t)\right]^2 + \left[y(t)\right]^2} \tag{4.44}$$

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

$$r\left(\frac{i}{W}\right) = r_i = \sqrt{x_i^2 + y_i^2} \tag{4.45}$$

The angle

$$\theta_i = \arctan \frac{y_i}{x_i} = \arccos \frac{x_i}{r_i} \tag{4.46}$$

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. The r_i and θ_i are sample values of amplitude and phase, and form a sort of polar coordinate system in the space associated with the set of functions.

Let us denote by x_i, y_i, or r_i, θ_i, the coordinates or sample values for a receiver input after the filter (i.e., at point A in Fig. 4.8). Let a_i, b_i, or f_i, ϕ_i denote the coordinates for the signal as it would appear at point A if there were no noise. The envelope of the signal, hence the coordinates f_i, are assumed known. Let us denote by $F_S(\phi_1, \phi_2, \ldots, \phi_{\frac{n}{2}})$ the distribution function of the phase coordinates ϕ_i. The probability density function for the coordinates x_i, y_i when there is white Gaussian noise and no signal is

$$f_N(x, y) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2N} \sum_{i=1}^{n/2} x_i^2 + \sum_{i=1}^{n/2} y_i^2)\right] \tag{4.47}$$

and for signal plus noise

$$f_{SN}(x, y) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \int_R \exp\left[-\frac{1}{2N}\left(\sum_{i=1}^{n/2} (x_i-a_i)^2 + \sum_{i=1}^{n/2} (y_i-b_i)^2\right)\right] dP_S(a_i b_i) \quad . \tag{4.48}$$

Changing to the polar coordinates,

$$f_N(r, \theta) = \left(\frac{1}{2\pi N}\right)^{\frac{n}{2}} \prod_{i=1}^{n/2} r_i \exp\left[-\frac{1}{2N}\sum_{i=1}^{n/2} r_i^2\right] , \tag{4.49}$$

and

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

$$dF_S\left(\phi_1, \ldots, \phi_{\frac{n}{2}}\right) \quad . \tag{4.50}$$

The factors $\prod_{i=1}^{2WT} r_i$ are introduced because they are the Jacobian of the transformation from rectangular to polar coordinates.[1, 2]

The probability density function for r alone, i.e., the density function for the output of the detector, is obtained by simply integrating the density functions for r and θ with respect to θ.[3]

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

$$= \left(\frac{1}{N}\right)^{\frac{n}{2}} \prod_{i=1}^{n/2} r_i \exp\left[-\frac{1}{2N}\sum_{i=1}^{n/2} r_i^2\right] \tag{4.51}$$

---

[1]Cramér, Ref. 14, page 292.

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

[3]Cramér, Ref. 14, page 291.

and

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

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

$$= \left(\frac{1}{N}\right)^{\frac{n}{2}} \prod_{i=1}^{n/2} r_i\, I_o\left(\frac{r_i f_i}{N}\right) \exp\left[-\frac{1}{2N} \sum_{i=1}^{n/2} (r_i^{\,2} + f_i^{\,2}) \right] \qquad (4.52)$$

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 is

$$\ell(r) = \frac{f_{SN}(r)}{f_N(r)} = \exp\left[-\frac{1}{2N} \sum_{i=1}^{n/2} f_i^{\,2} \right] \prod_{i=1}^{n/2} I_o\left(\frac{r_i f_i}{N}\right). \qquad (4.53)$$

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

$$\frac{1}{2N} \sum_{i=1}^{n/2} f_i^{\,2} = \frac{W}{2N} \int \left[f(t)\right]^2 dt = \frac{E}{N_o}, \text{ and} \qquad (4.54)$$

$$\ln \ell(r) = -\frac{E}{N_o} + \sum_{i=1}^{n/2} \ln I_o\left(\frac{r_i f_i}{N}\right), \qquad (4.55)$$

which is approximately

$$\ln \ell\left(r(t)\right) = -\frac{E}{N_o} + W \int_0^T \ln I_o\left(\frac{r(t)\ f(t)}{N}\right) dt\ . \qquad (4.56)$$

The function $\ell n\ I_o(x)$ is plotted as a function of x in Fig. 4.9. This function is very nearly the parabola $\frac{x^2}{4}$ for small values of x and is approximately linear for large values of x. Thus, the expression for likelihood ratio might be approximated by

$$\ell n\, \ell\big(r(t)\big) = -\frac{E}{N_o} + \frac{W}{4N^2} \int_0^T \left[r(t)\right]^2 \left[f(t)\right]^2 dt \qquad (4.57)$$

for small signals, and by

$$\ell n\, \ell\big(r(t)\big) = C_1 + C_2 \int_0^T r(t)\ f(t)\ dt \qquad (4.58)$$

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 (4.57) and (4.58) can be interpreted as cross correlation. 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 $\big(f(t)\big)^2$, 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 threshold 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:

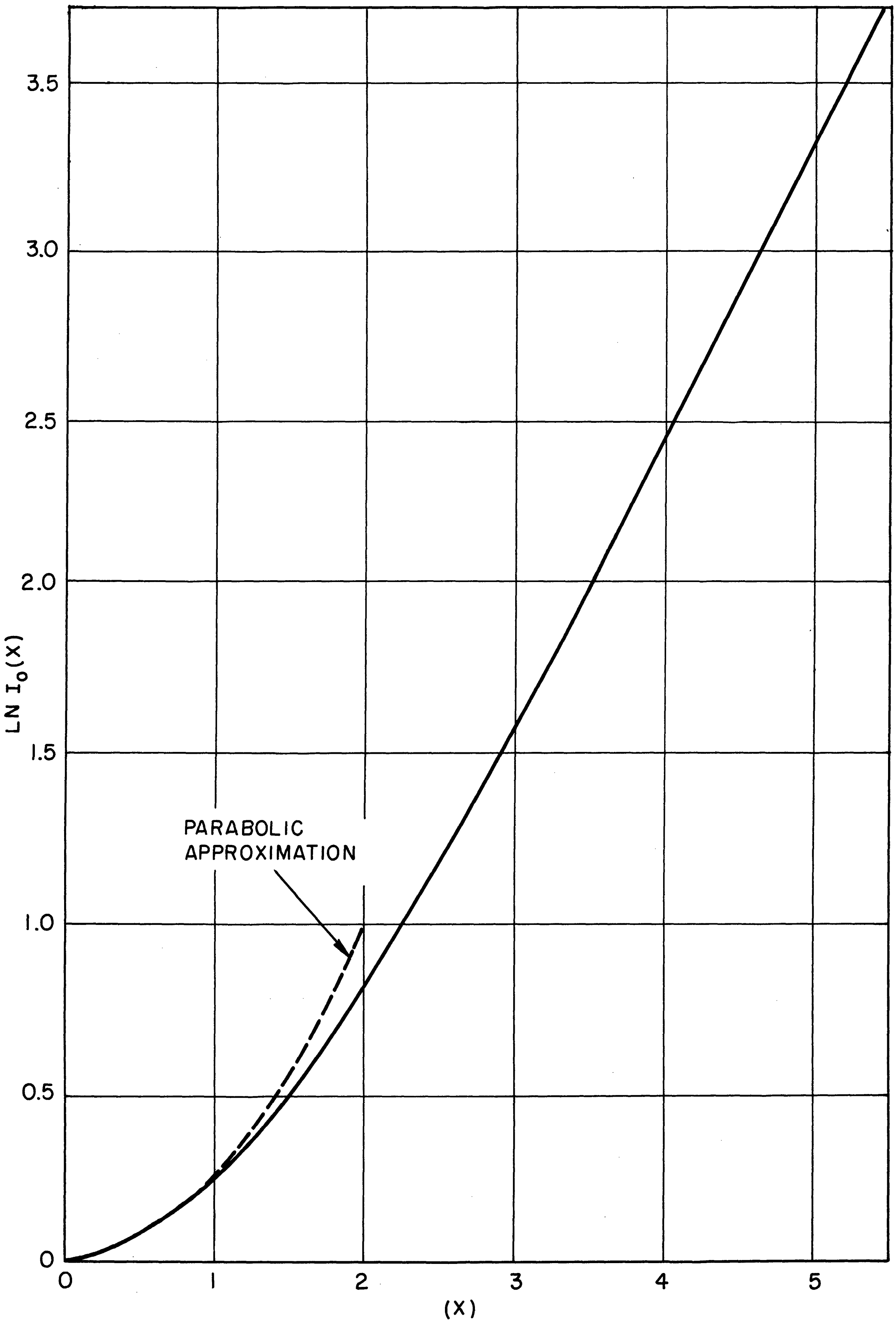

FIG. 4.9

GRAPH OF LN $I_0(X)$.

$$\ln \ell(r) \approx -\frac{1}{2N} \sum f_i^2 + \frac{1}{4N^2} \sum_{i=1}^{n/2} r_i^2 f_i^2 \tag{4.59}$$

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

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

$$g_N(r_i) = \frac{r_i}{N} \exp\left[-\frac{r_i^2}{2N}\right] . \tag{4.60}$$

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 $\frac{r_i^2 f_i^2}{N^2}$ in the sum in Eq (4.59) is

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

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

$$\mu_N\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{f_i^2}{4N} \int_0^{\infty} \frac{f_i^2}{N} g_N(r_i)\, dr_i = \frac{f_i^2}{4N} \int_0^{\infty} \frac{r_i^3}{N^2} \exp\left[-\frac{r_i^2}{2N}\right] dr_i \tag{4.61}$$

The second moment of each term $\frac{r_i^2 f_i^2}{4N^2}$ is

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

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

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

$$= \frac{f_i^4}{16N^2}\int_0^\infty \frac{r_i^5}{N^3}\;\exp\left[-\frac{r_i^2}{2N}\right] dr_i \qquad (4.62)$$

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

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

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

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 the book, Threshold Signals by Lawson and Uhlenbeck.[1] The results are

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

$$\mu_{SN}\left(\frac{r_i^4 f_i^4}{16N^4}\right) = \frac{1}{2}\;\frac{f_i^4}{N^2}\left(1 + \frac{f_i^2}{N} + \frac{f_i^4}{8N^2}\right) \qquad (4.64)$$

Since

$$\sigma^2(Z) = \mu(Z^2) - \left[\mu(Z)\right]^2 \quad , \qquad (4.65)$$

the variance of $\frac{r_i^2 f_i^2}{N^2}$ is

[1]Ref. 1., p. 174

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

$$\sigma^2_{N}\left(\frac{r_i^2 f_i^2}{4N^2}\right) = \frac{f_i^4}{4N^2} \tag{4.66}$$

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.

mean of $\ell n\ \ell(x)$ is

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

$$\mu_{N}\left(\ell n\ \ell(r)\right) = -\sum_{i=1}^{n/2}\frac{f_i^2}{2N} + \frac{1}{2}\sum_{i=1}^{n/2}\frac{f_i^2}{N} \tag{4.67}$$

$$= 0$$

and the variance of $\ell n\ \ell(r)$ is

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

$$\sigma^2_{N}\left(\ell n\ \ell(r)\right) = \sum_{i=1}^{n/2}\frac{f_i^4}{4N^2} \tag{4.68}$$

If the distribution functions $\ell n\ \ell(x)$ can be assumed to be normal, the distribution functions can be obtained immediately from the mean and standard deviation of the distribution. In some cases the normal distribution is a good approximation to the actual distribution.

Let us consider the case in which the incoming signal is a rectangular pulse which is $\frac{M}{W}$ seconds long.[1] The energy of the pulse is half its duration times the amplitude of its envelope, and therefore the amplitude has the value

$$f_i = \sqrt{\frac{2EW}{M}} \quad , \tag{4.69}$$

where E is the pulse energy. It has this value on M sample points and is zero at all others. For this case

$$\begin{aligned}
\mu_{SN}\left(\ln \ell(r)\right) &= \frac{1}{M}\,\frac{E^2}{{N_o}^2} \\
\mu_N\left(\ln \ell(r)\right) &= 0 \\
{\sigma_{SN}}^2\left(\ln \ell(r)\right) &= \frac{E^2}{M{N_o}^2}\left(1 + \frac{2}{M}\frac{E}{N_o}\right) \\
{\sigma_N}^2\left(\ln \ell(r)\right) &= \frac{E^2}{M{N_o}^2}
\end{aligned} \tag{4.70}$$

Also, for this case, the distribution of $\ln \ell(x)$ is approximately normal, if M is much larger than one. Since it is the sum of M independent random variables, all having the same distribution, it must, by the central limit theorem,[2] 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[3] for the $g_N(r_i)$ with itself any number of times can be

---

[1] The problem of finding the distribution for the sum of M independent random variables, each with a probability density function $f(x) = x \exp\left[-\frac{1}{2}(x^2+\alpha^2)\right] I_o(\alpha x)$ 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.

[2] Cramér, Ref. 14, p. 213 and 316.

[3] Cramér, Ref. 14, p. 188-9.

expressed in closed form. The density function for this distribution is plotted in Fig. 4.10 for several relatively small values of M. The distribution of $\ln \ell(x)$ for signal plus noise is more nearly normal than the distribution for 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. 4.11 using the normal distribution as approximation to the true distribution. In many cases it will be found that

$$\frac{1}{M} \cdot \frac{2E}{N_o} << 1 \quad . \tag{4.71}$$

In such a case the distributions have approximately the same variance. Assuming normal distribution then leads to the curves of Fig. 4.1, with

$$d = \frac{1}{4M} \left(\frac{2E}{N_o}\right)^2 \quad . \tag{4.72}$$

4.6 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:

$$s(t) = \sum_{m=0}^{M-1} f(t+m\tau) \cos (\omega t+\theta_i) \tag{4.73}$$

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

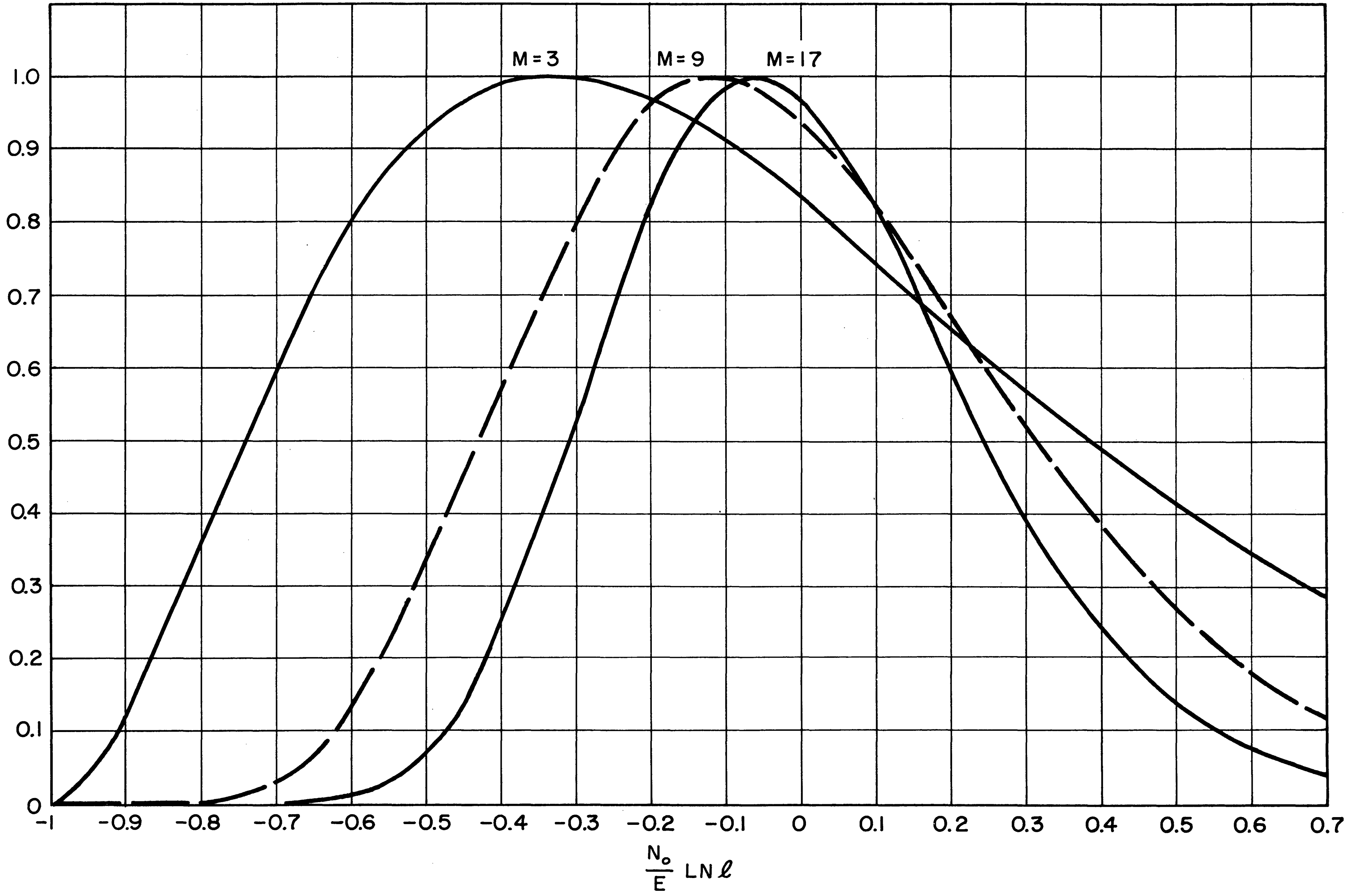

FIG. 4.10

$\frac{\text{PROBABLE DENSITY}}{\text{MAX. PROBABLE DENSITY}}$ OF $LN\ell$ VS. $\frac{N_o}{E} LN\ell$.

M-970 A-G3-79 JEM 7-21-53

M-970 A-G3-80 JEM 7-21-53

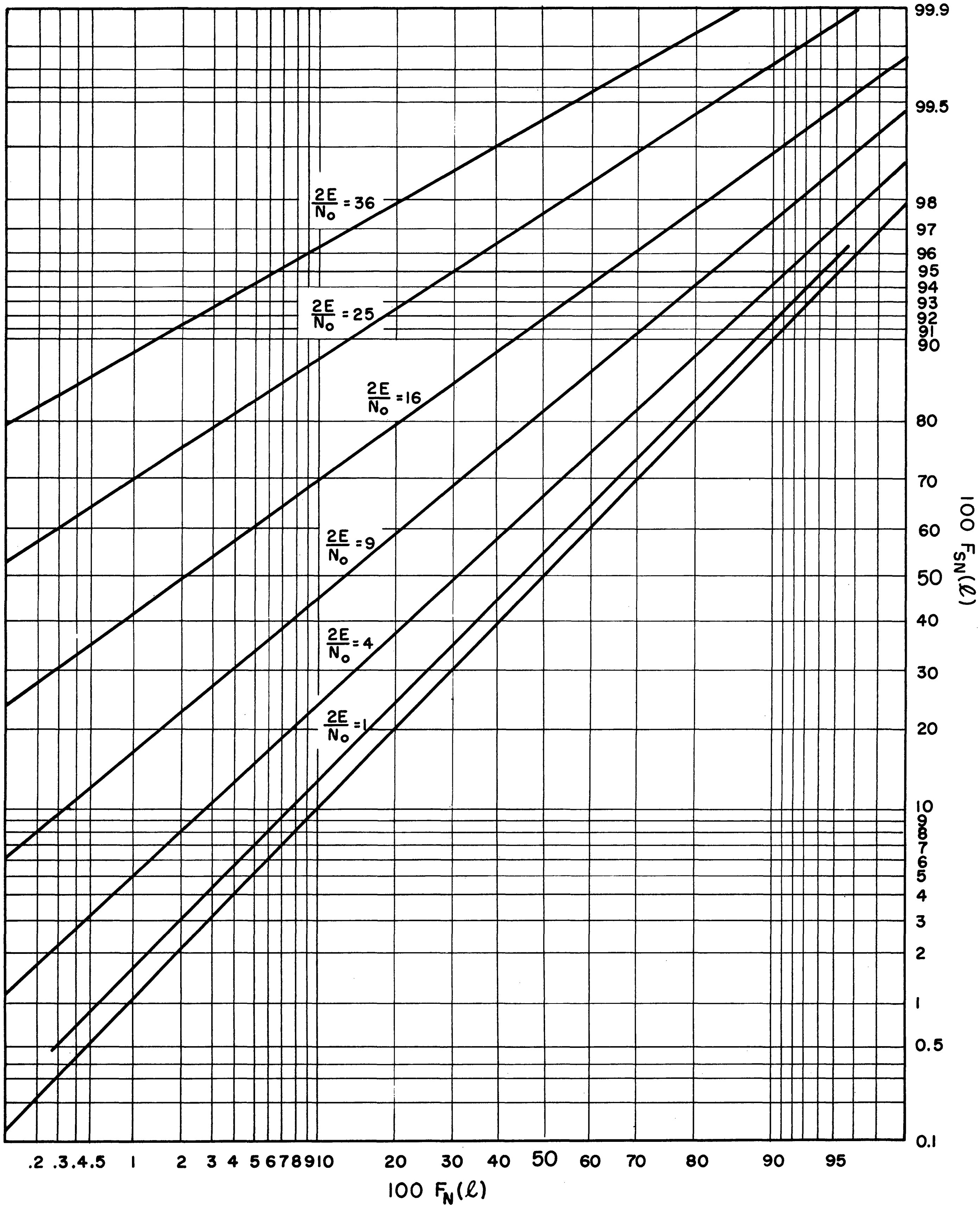

FIG. 4.11

RECEIVER OPERATING CHARACTERISTIC.

BROAD BAND RECEIVER WITH
OPTIMUM VIDEO DESIGN, M=16.

$$\int_0^T f(t+i\tau)\, f(t+j\tau)\, dt = \frac{2E}{M}\,\delta_{ij} \quad , \tag{4.74}$$

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 (4.74) states that the pulses are spaced far enough so that they are orthogonal, and that the total signal energy is E.[1] The function f(t) is also assumed to have no frequency components as high as $\frac{\omega}{2\pi}$.

The likelihood ratio can be obtained by applying Eq (3.7).

$$\ell(x) = \int_R \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{2}{N_o}\int_0^T s(t)\, x(t)\, dT\right] dP_S(s) \tag{4.75}$$

$$= \exp\left[-\frac{E}{N_o}\right] \int_0^{2\pi} \cdots \int_0^{2\pi} \exp\left[\frac{2}{N_o}\int_0^T \sum_{m=o}^{M-1} f(t+m\tau)x(t)\cos(\omega t+\theta_m)dt\; d\theta_o \ldots d\theta_{M-1}\right] \tag{4.76}$$

The integral can be evaluated, as in Section 4.3, and

$$\ell(x) = \exp\left[-\frac{E}{N_o}\right] \prod_{m=0}^{M-1} I_o\left(\frac{r_m}{N}\right) \quad , \tag{4.77}$$

where

$$\left(\frac{r_m}{N}\right)^2 = \left[\frac{2}{N_o}\int_0^T f(t+m\tau)x(t)\cos\omega t\,dt\right]^2 + \left[\frac{2}{N_o}\int_0^T f(t+m\tau)\, x(t)\sin\omega t\,dt\right]^2 \tag{4.78}$$

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 4.3. In fact, each r_m could be obtained in a receiver in the manner

---

[1]The factor 2 appears in (4.74) 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.

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

$$s_o(t) = f(t) \cos(\omega t+\theta) \tag{4.79}$$

for any value of the phase angle θ, and putting the output through a linear detector. The output will be $\frac{N_o}{2}\frac{r_o}{N}$ 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 $\frac{N_o}{2}\frac{r_m}{N}$.

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

$$\ln \ell(x) = -\frac{E}{N_o} + \sum_{m=o}^{M-1} \ln I_o\left(\frac{r_m}{N}\right) \tag{4.80}$$

Thus the $\ln I_o\left(\frac{r_m}{N}\right)$ must be found for each r_m, and these M quantities must be added. As in the previous section, $\frac{r_m}{N}$ will usually be small enough so that $\ln I_o(x)$ can be approximated by $\frac{x^2}{4}$.[1] The quantities $\frac{1}{4}\left(\frac{r_m}{N}\right)^2$ 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 i.f. amplifier with its passband matched to a single pulse,[2] a 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

$$\ln I_o\left(\frac{r_m}{N}\right) \approx \frac{r_m^2}{4N^2} \tag{4.81}$$

---

[1] See Fig. 4.9.

[2] It is usually most convenient to make the ideal filter (or an approximation to it) a part of the i.f. amplifier.

which is valid for small values of $\frac{r_m}{N}$,[1]

$$\ell n \ell \approx -\frac{E}{N_o} + \sum_{n=o}^{M-1} \frac{1}{4} \left(\frac{r_m}{N} \right)^2 . \tag{4.82}$$

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 (4.22)[2]

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

or

$$P_N\left(\frac{r_m}{N} \geqq a\right) = \exp\left[-\frac{a^2 N_o M}{2E}\right] , \tag{4.83}$$

and by (4.26),

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

or

$$P_{SN}\left(\frac{r_m}{N} \geqq a\right) = \frac{N_o M}{2E} \exp\left[-\frac{E}{N_o M}\right]\int_{a}^{\infty} a \exp\left(-\frac{a^2 N_o M}{4E}\right) I_o(a)\, da \tag{4.84}$$

The density functions can be obtained by differentiating (4.83) and (4.84):

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

$$g_{SN}\left(\frac{r_m}{N}\right) = \frac{MN_o}{2E}\left(\frac{r_m}{N}\right) \exp\left[-\frac{E}{MN_o}\right] \exp\left[-\left(\frac{r_m}{N}\right)^2\left(\frac{N_o M}{4E}\right)\right] I_o\left(\frac{r_m}{N}\right). \tag{4.85}$$

---

[1] See footnote 1, p.37.

[2] The M appears in the following equations because the energy of a single pulse is $\frac{E}{M}$ rather than E.

This is the same situation, mathematically, as appeared in the previous section on page 34. The standard deviation and the mean for the logarithm of the likelihood ratio can be found in the same manner, and they are

$$\mu_{SN}(\ln \mathcal{L}) = \frac{E^2}{MN_o{}^2}$$

$$\mu_N(\ln \mathcal{L}) = 0$$

$$\sigma_{SN}{}^2(\ln \mathcal{L}) = \frac{E^2}{MN_o{}^2}\left(1 + \frac{2E}{MN_o}\right)$$

$$\sigma_N{}^2(\ln \mathcal{L}) = \frac{E^2}{MN_o{}^2} \tag{4.86}$$

If the distributions can be assumed normal, they are completely determined by their means and variances. These formulas are identical with the formulas (4.70) on page 37 of the previous section. The problem is the same, mathematically, and the discussion and receiver operating characteristic curves at the end of Section 4.6 apply to both cases.

4.7 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 Fig. 4.1, 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 Fig. 4.1.

It was pointed out in Section 2.5.1 of Part I of this report that $F_{SN}(\mathcal{L})$ can be calculated if $F_N(\mathcal{L})$ is known. It was further pointed out that the n^{th} moment of the distribution $F_N(\mathcal{L})$ is the $(n-1)^{th}$ moment of the distribution $F_{SN}(\mathcal{L})$. 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 is $1 + \sigma_N^2$. Thus the difference between the means, and the variance with noise alone are the same number σ_N^2. This number probably characterizes the receiver reliability better than any other single number.

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

$$F_N(\ell) = \frac{1}{\sqrt{2\pi d}} \int_{\ln \ell}^{\infty} \exp\left[- \frac{(x-m)^2}{2d}\right] dx, \tag{4.87}$$

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:

$$\mu_N(\ell^n) = \int_0^{\infty} \ell^n \, dF_n(\ell) = \frac{1}{\sqrt{2\pi d}} \int_{-\infty}^{\infty} \exp\left[nx\right]\exp\left[- \frac{(x-m)^2}{2d}\right] dx \quad , \tag{4.88}$$

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

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

$$\mu_N(\ell^n) = \exp\left[\frac{n^2 d}{2} + nm \right] \quad . \tag{4.89}$$

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

$$\mu_N(\ell) = 1 = \exp\left[\frac{d}{2} + m \right] \quad , \tag{4.90}$$

and therefore

$$m = - \frac{d}{2} \quad , \tag{4.91}$$

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

$$\mu_N(\ell^2) = \left[\mu_N(\ell)\right]^2 + \sigma_N^2(\ell) = 1 + \sigma_N^2(\ell) \quad , \qquad (4.92)$$

and this must agree with (4.89).

$$\mu_N(\ell^2) = 1 + \sigma_N^2 = \exp\left[2d + 2m\right] = \exp\left[d\right] \qquad (4.93)$$

and therefore

$$d = \ln(1 + \sigma_N^2) \qquad (4.94)$$

The distribution of likelihood ratio with signal plus noise can be found by applying Theorem 8.[1]

$$dF_{SN}(\ell) = \ell\, dF_N(\ell) \quad ,$$

$$F_{SN}(\ell) = - \int_{\ell}^{\infty} dF_N(\ell) \quad . \qquad (4.95)$$

Substituting for $F_N(\ell)$ from (4.87), and letting $\ell = \exp x$ yields

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

$$= \frac{1}{\sqrt{2\pi d}} \int_{\ln \ell}^{\infty} \exp\left[- \frac{(x - \frac{d}{2})^2}{2d}\right] dx \quad . \qquad (4.96)$$

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

In summary, the variance σ_N^2 of the likelihood ratio probably measures the receiver reliability better than any other single number. If the logarithm of the likelihood ratio has a normal distribution, then this distribution, and

---

[1] See Part I, Section 2.4.

hence the signal plus noise distribution, are completely determined if σ_N^2 is given. Both distributions of $\ell n\ \ell(x)$ are normal with the same variance d, and the difference of the means is d. The receiver operating characteristic curves are those plotted in Fig. 4.1, with the parameter d related to σ_N^2 by the equation

$$d = \ell n\ (1 + \sigma_N^2) \quad . \tag{4.94}$$

In the case of a signal known exactly, this is the distribution which occurs. In the cases of Section 4.4, Section 4.5, and Section 4.6 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 receiver reliability 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 reliability is given approximately by the receiver operating characteristic curves of Fig. 4.1 with $d = \ell n\ (1 + \sigma_N^2)$.

4.8 Signal Which is One of M Orthogonal Signals

The following case has several applications, which will be discussed in Section 5.3. The importance of this case, and the one which follows it, lies in the fact that the uncertainty of the signal distribution can be varied by changing the parameter M.

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

$$\int_0^T s_k(t)\, s_q(t)\, dt = E\,\delta_{kq} \tag{4.97}$$

Then the likelihood ratio can be found from Eq (3.7) to be

$$\begin{aligned} \mathscr{l}(x) &= \sum_{k=1}^{M} \frac{1}{M} \exp\left[-\frac{E}{N_o}\right] \exp\left[\frac{1}{N} \sum_{i=1}^{n} x_i s_{ki}\right] \\ &= \frac{1}{M} \sum_{k=1}^{M} \exp\left[\frac{1}{N} \sum_{i=1}^{n} x_i s_{ki} - \frac{E}{N_o}\right] \end{aligned} \tag{4.98}$$

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

It should be clear that with noise alone, the terms $\frac{1}{N} \sum_{i=1}^{n} x_i s_{ki}$ have a Gaussian distribution with mean zero and variance $\sum_{i=1}^{n} \frac{s_{ki}^2}{N} = \frac{2E}{N_o}$.[1] Furthermore, the M different quantities $\frac{1}{N} \sum_{i=1}^{n} x_i s_{ki}$ are independent, since the functions $s_k(t)$ are orthogonal. It follows that the terms $\exp\left[\frac{1}{N} \sum_{i=1}^{M} x_i s_{ki} - \frac{E}{N_o}\right]$ are independent.

Since the logarithm of each term $Z = \exp\left[\frac{1}{N} \sum_{i=1}^{n} x_i s_{ki} - \frac{E}{N_o}\right]$ has a normal distribution with mean $-\frac{E}{N_o}$ and variance $\frac{2E}{N_o}$, the moments of the distribution can be found from Eq (4.89). The n^{th} moment is

$$\mu_N(Z^n) = \exp\left[n(n-1)\frac{E}{N_o}\right] . \tag{4.99}$$

[1]The reasoning is the same as that on page 9.

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

$$\sigma_N^2(Z) = \mu(Z^2) - \left[\mu(Z)^2\right] = \exp\left[\frac{2E}{N_o}\right] - 1 \quad . \tag{4.100}$$

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

$$\sigma_N^2(M\ell) = M\left[\exp\left(\frac{2E}{N_o}\right) - 1\right] \quad , \tag{4.101}$$

and it follows that the variance of the likelihood ratio is

$$\sigma_N^2(\ell) = \frac{1}{M}\left[\exp\left(\frac{2E}{N_o}\right) - 1\right] \quad . \tag{4.102}$$

It was pointed out in Section 4.7, page 47 that the receiver operating characteristic curves are approximately those of Figure 4.1, with

$$d = \ell n\,(1 + \sigma_N^2) = \ell n\left(1 - \frac{1}{M} + \frac{1}{M}\exp\left(\frac{2E}{N_o}\right)\right) \tag{4.103}$$

This equation can be solved for $\frac{2E}{N_o}$:

$$\frac{2E}{N_o} = \ell n\left[1 + M\,(e^d - 1)\right] \quad . \tag{4.104}$$

Curves of $\frac{2E}{N_o}$ for constant d are plotted in Fig. 4.12. They show how much the signal energy must be increased when the number of possible signals increases.

4.9 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

$$s_k(t) = f_k(t)\cos(\omega t + \theta) \quad . \tag{4.105}$$

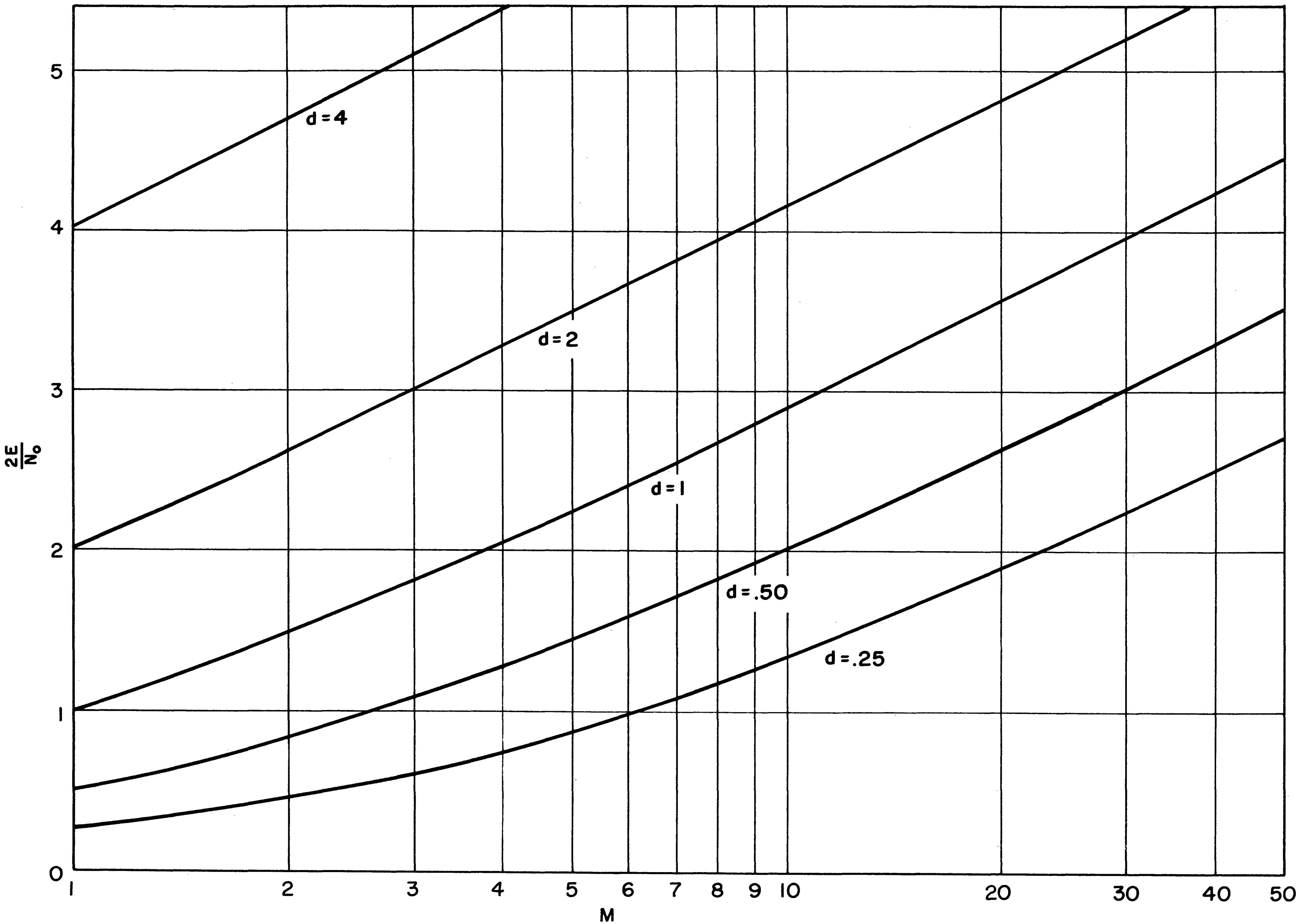

FIG. 4.12. SIGNAL ENERGY AS A FUNCTION OF M AND d.
SIGNAL ONE OF M ORTHOGONAL SIGNALS.

M-970 A-G3-81 JEM 7-21-53

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

$$\int_0^T f_k(t)\, f_q(t)\, dt = 2E\,\delta_{kq} \quad , \tag{4.106}$$

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 ω. 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 $\frac{1}{M}$.

With these assumptions, the likelihood ratio can be obtained from Eq (3.7), and it is

$$\ell(x) = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{2\pi}\int_0^{2\pi} \exp\left[\frac{1}{N}\sum_{i=1}^{n} x_i s_{ki} - \frac{E}{N_o}\right] d\theta \tag{4.107}$$

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 can be obtained from Eq (4.19)

$$\ell(x) = \frac{1}{M}\sum_{k=1}^{M} \exp\left[-\frac{E}{N_o}\right] I_o\left(\frac{r_k}{N}\right) \quad , \tag{4.108}$$

where

$$r_k = \sqrt{\left(\sum_i x_i\, f_k(t_i) \cos\omega t_i\right)^2 + \left(\sum_i x_i\, f_k(t_i) \sin\omega t_i\right)^2} \quad . \tag{4.109}$$

Now the problem is to find $\sigma_N^2(\ell)$. The variance of each term in the sum in Eq (4.108) can be found, since the distribution function with noise alone can be found as in Section 4.3. Since the $f_k(t)$ are orthogonal, the distributions of the r_k are independent, and the terms in the sum in Eq (4.107) 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. [1]

The distribution function for each term $\exp\left[-\frac{E}{N_o}\right] I_o\left(\frac{r_k}{N}\right)$ is given in Section 4.3 by Eq (4.21) and (4.22). If α is defined by the equation

$$\beta = \exp\left[-\frac{E}{N_o}\right] I_o\left(\alpha\sqrt{\frac{2E}{N_o}}\right) , \tag{4.10}$$

then the distribution function in the presence of noise for each term in Eq (4.108) is

$$F_N^{(k)}(\beta) = \exp\left[-\frac{\alpha^2}{2}\right] . \tag{4.111}$$

The mean value of each term is

$$\mu^{(k)}(\beta) = \int_0^\infty \beta dF_N^{(k)}(\beta) = \int_0^\infty \exp\left[-\frac{E}{N_o}\right] I_o\left(\sqrt{\frac{2E}{N_o}}\,\alpha\right)\alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha . \tag{4.112}$$

This can be evaluated,[2] and the result is that $\mu^{(k)}(\beta) = 1$.

The second moment of each term is

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

$$= \int_0^\infty \exp\left[-\frac{2E}{N_o}\right]\left[I_o\left(\alpha\sqrt{\frac{2E}{N_o}}\right)\right]^2 \alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha \tag{4.113}$$

---

[1] Cramér, Ref. 14, p. 188.

[2] Lawson and Uhlenbeck, Ref. 1, p. 174.

The integral is evaluated in Appendix E, and the result is

$$\mu_N^{(k)}(\beta^2) = I_o\left(\frac{2E}{N_o}\right) \quad . \tag{4.114}$$

The variance is

$$\left[\sigma_N^{(k)}(\beta)\right]^2 = \mu^{(k)}(\beta^2) - \left[\mu^{(k)}(\beta)\right]^2 = I_o\left(\frac{2E}{N_o}\right) - 1 \; . \tag{4.115}$$

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

$$\sigma_N^2\,(M\ell) = M\left[I_o\left(\frac{2E}{N_o}\right) - 1\right] \quad , \text{ and} \tag{4.116}$$

$$\sigma_N^2(\ell) = \frac{1}{M}\left[I_o\left(\frac{2E}{N_o}\right) - 1\right] \quad , \tag{4.117}$$

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

If the approximation described in Section 4.7 is used, the receiver operating characteristic curves are approximately those of Fig. 4.1, with

$$d = \ell n\,(1 + \sigma_N^2) = \ell n\left(1 - \frac{1}{M} + \frac{1}{M}\, I_o\left(\frac{2E}{N_o}\right)\right). \tag{4.118}$$

Curves of $\frac{2E}{N_o}$ vs M for constant d are plotted in Fig. 4.13.

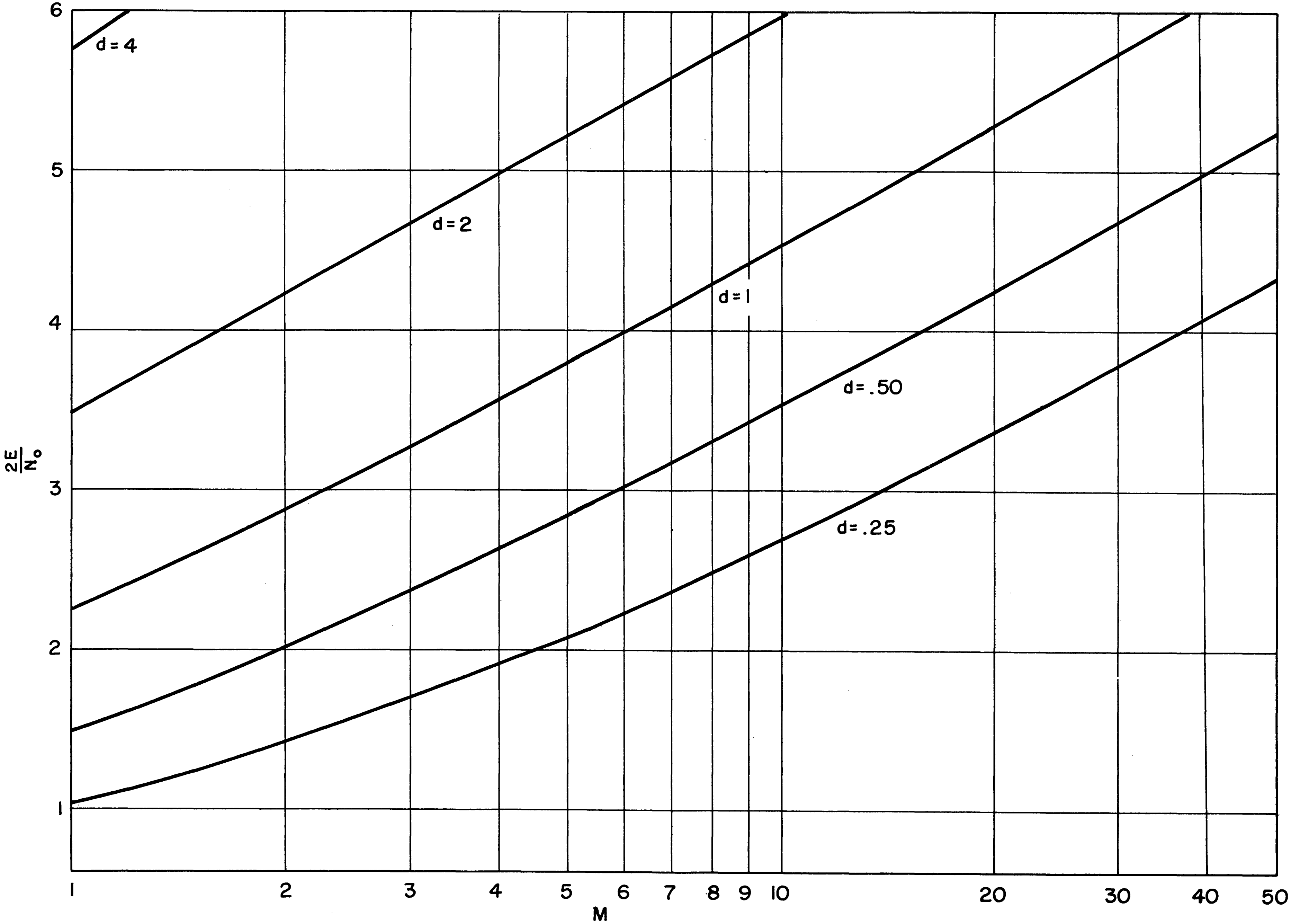

FIG. 4.13. SIGNAL ENERGY AS A FUNCTION OF M AND d.
SIGNAL ONE OF M ORTHOGONAL SIGNALS
KNOWN EXCEPT FOR PHASE

M-970 A-G3-82 JEM 7-21-53

5. DISCUSSION OF THE SPECIAL CASES

5.1 Receiver Evaluation

5.1.1 Introduction. In Section 2.5 it was shown that the receiver reliability can be determined from the distribution functions for likelihood ratio. In particular an optimum criterion receiver operating at the level β of likelihood ratio has false alarm probability $P_N(A) = F_N(\beta)$, and probability of detection $P_{SN}(A) = F_{SN}(\beta)$. The functions $F_N(\beta)$ and $F_{SN}(\beta)$ are calculated in Section 4 for a number of special cases.

For the purpose of discussing receiver reliability it is sufficient to have the receiver operating characteristic in which $F_{SN}(\beta)$ is plotted as a function of $F_N(\beta)$. In this discussion β plays only a secondary role.

The receiver operating characteristic shown in Figure 5.1 applies to several cases. Among them is the case of the signal known exactly, with the parameter d equal to $\frac{2E}{N_o}$, twice the ratio of signal energy to noise power per unit bandwidth.[1] Thus, for example, if the signal is a voltage which is a known function of time, and if the signal energy is twice the noise power per unit bandwidth, theoretically a receiver could be built with false alarm probability of 0.25 and a probability of detection 0.90. If the false alarm probability is required to be no greater than 0.10, the probability of detection can be made no greater than 0.76. If the false alarm probability is required to be no greater than 0.023 and the probability of detection is to be at least 0.98, the signal energy must be at least eight times the noise power per unit bandwidth.

5.1.2 Comparison of the Simple Cases. Several curves for the case of a signal known except for phase are shown in Fig. 5.2 for some of the same values

[1]See Section 4.2.

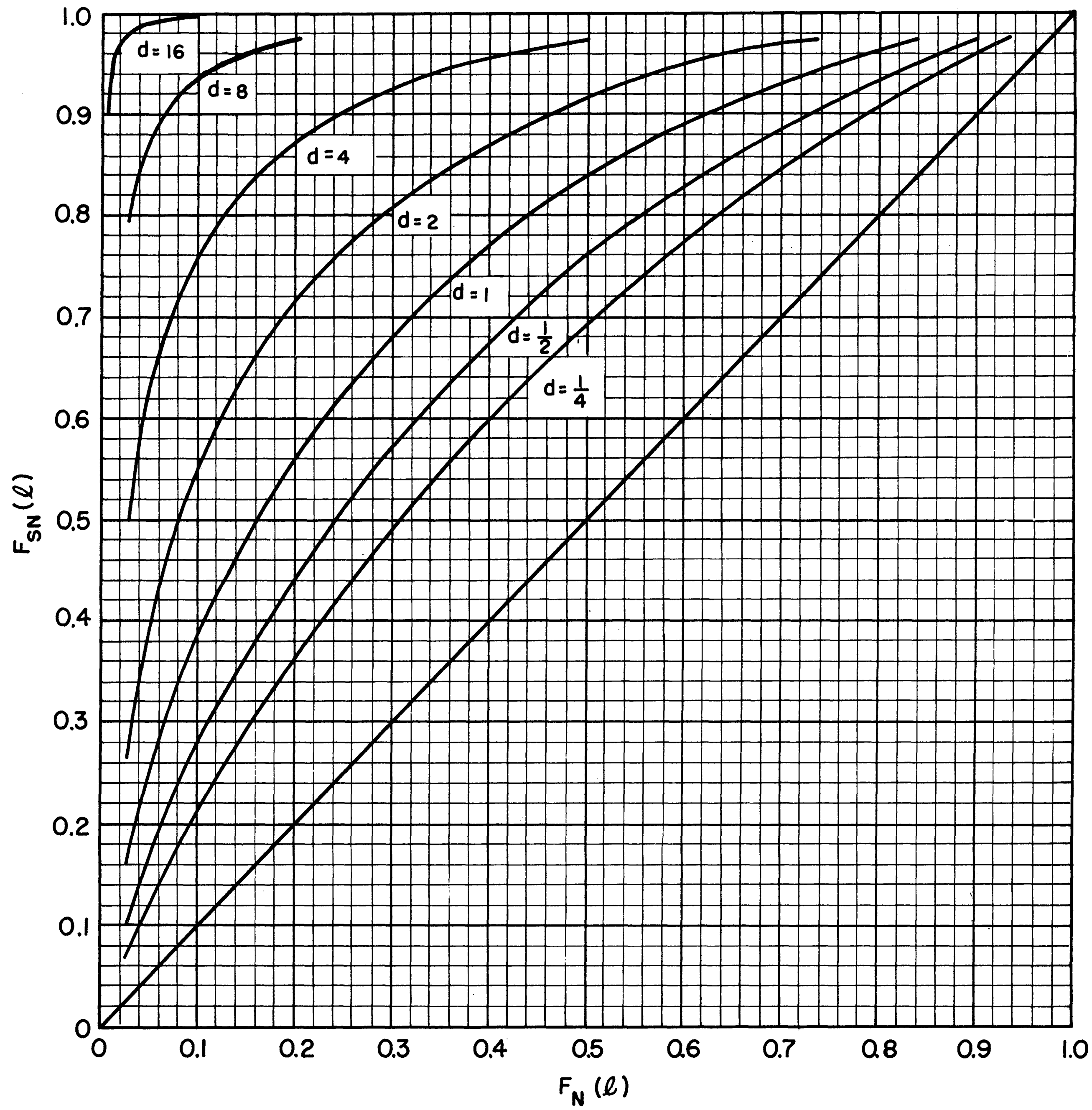

FIG. 5.1

RECEIVER OPERATING CHARACTERISTIC.

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

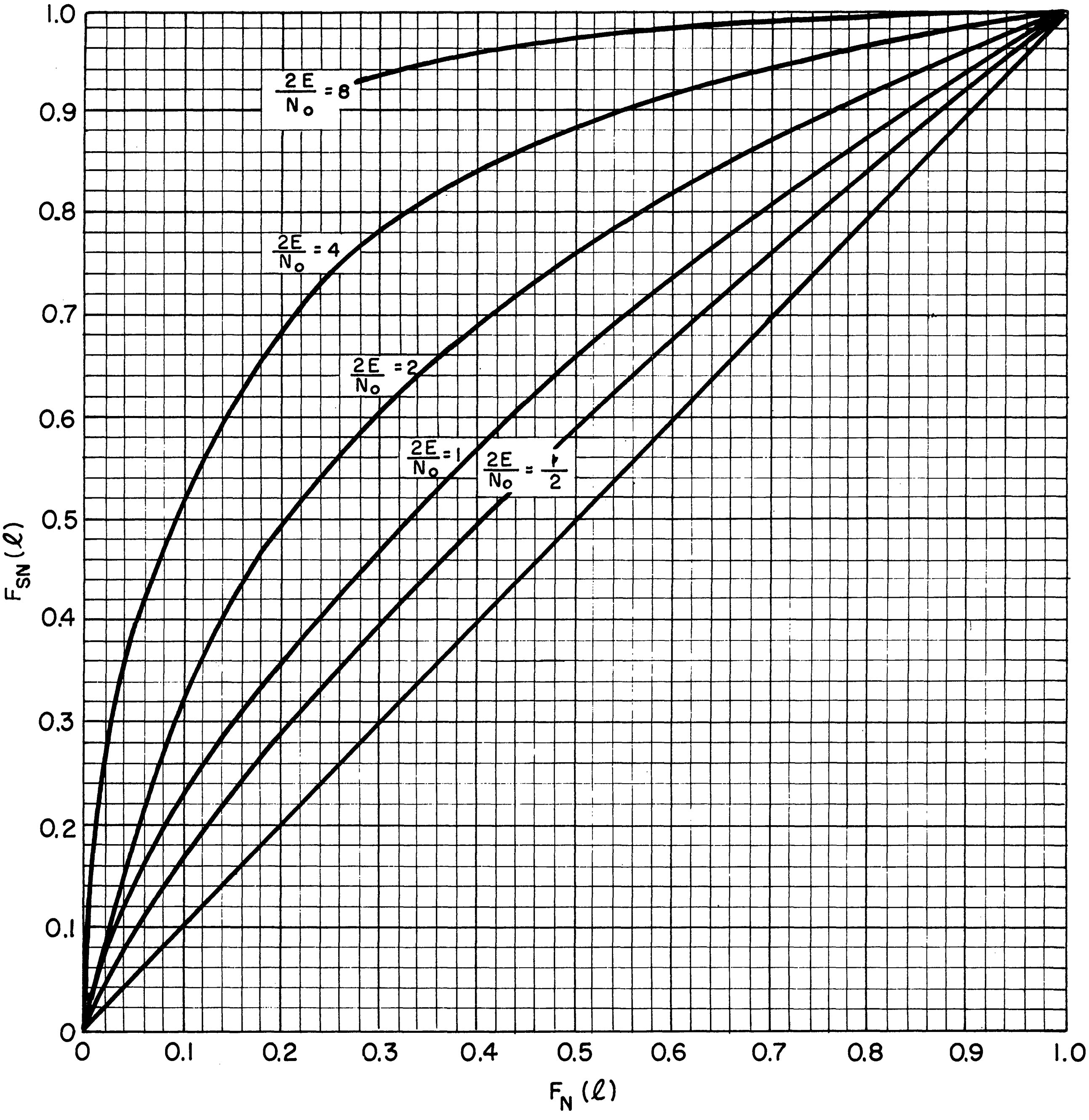

FIG. 5.2

RECEIVER OPERATING CHARACTERISTIC.

SIGNAL KNOWN EXCEPT FOR PHASE.

of the ratio $\frac{2E}{N_o}$ as appear in Fig. 5.1. The curves for a given energy lie below those for the case of the signal known exactly; with a given false alarm probability and a given value of $\frac{2E}{N_o}$, one cannot achieve as high a probability of detection if the carrier phase of the signal is unknown.

It was found that in several cases the distribution of $\ell n\, \ell(x)$ approached a normal distribution as a limiting case, and that in the limit the variance with signal plus noise and the variance with noise alone are equal. In any such case the curves of Fig. 5.1 apply, and a comparison of these cases is simplified. For example, in the case of a signal which is a sample of white Gaussian noise, it was found that if the number of sample points is large and the signal to noise ratio is small, then this approximation applies, with

$$d = (2n-1)\left(1 - \sqrt{\frac{N}{N+S}}\right)^2 \approx \frac{n}{2}\left(\frac{S}{N}\right)^2 \quad . \tag{4.37}$$

Other curves for this case, some with small sample number and moderate signal to noise ratio, are given in Figs. 4.6 and 4.7. The exact equations for the distribution are Eqs (4.33) and (4.34).

The following two cases lead to the same receiver operating characteristic in the approximation considered in Sections 4.5 and 4.6: (1) the broad band receiver with optimum video design, with a pulse signal, and (2) the optimum receiver for a train of pulses with incoherent phase. In the first case the parameter M was taken as the product of the total bandwidth of the receiver and the pulse width of the signal. In the case of the train of pulses, M is the number of pulses. In each case E is the total energy of the signals. Approximate receiver operating characteristics are plotted in Fig. 4.10. Small signal to noise ratio and large M lead to the distributions for which Fig. 5.1 is plotted, this time with

$$d = \frac{1}{4M}\left(\frac{2E}{N_o}\right)^2 \quad . \tag{4.71}$$

5.1.3 An Approximate Evaluation of Optimum Receivers. Some simpler evaluation of receivers was needed because of the difficulty in solving directly for the distribution function of likelihood ratio in any cases more complicated than the ones already mentioned. It seemed reasonable to approximate the actual receiver operating characteristic by the curves given in Fig. 5.1, finding in some manner the value of the detection index d which leads to the best fit of the approximate curve to the real curve. This is suggested by the occurrence of the curves of Fig. 5.1 in four of the five cases already discussed. Also, any receiver operating characteristic must have in common with the curves of Fig. 5.1 that its slope is positive and its second derivative is negative, and that it must start at the lower left hand corner and end at the upper right hand corner of the graph.

It is shown in Section 2.5.2 that the variance σ_N^2 of the likelihood ratio when there is noise alone is the same as the difference of the means of likelihood ratio with noise alone and with signal plus noise. This parameter σ_N^2 seems to characterize signal detectability better than any other single number. In Section 4.7, it is shown that if σ_N^2 is given and the logarithm of the likelihood ratio is assumed to have a normal distribution with noise alone, then it follows that the logarithm of the likelihood ratio with signal plus noise also has a normal distribution with the same variance, and thus the receiver operating characteristic is that of Fig. 5.1. The index d is given by

$$d = \ln\left(1 + \sigma_N^2\right) \quad . \tag{4.94}$$

It seems reasonable that the curves be fitted on this basis, i.e., that σ_N^2 be determined for the actual situation and the approximate receiver operating characteristic graph be taken as the curve of Fig. 5.1 with index d given by the above Eq (4.94).

5.1.4 The Signal One of M Orthogonal Signals. The methods of the previous section have been applied to the case where the operator knows that the signal, if it occurs, will be one of M orthogonal functions of equal energy. Orthogonal, of course, means that the functions have zero cross correlation, i.e., f(t) and g(t) are orthogonal if

$$\int_0^T f(t)\, g(t)\, dt = 0 \tag{5.1}$$

where the integration is over the observation interval. The value obtained for σ_N^2 is[1]

$$\sigma_N^2 = \frac{1}{M}\left[\exp\left(\frac{2E}{N_o}\right) - 1\right] \tag{4.102}$$

and so the approximate receiver operating characteristic is that of Fig. 5.1 with

$$d = \ln\left[1 - \frac{1}{M} + \frac{1}{M}\exp\left(\frac{2E}{N_o}\right)\right] \tag{4.103}$$

The value of σ_N^2 was also found for the case where each of the M orthogonal signals is known except for phase, and the phase angle has a uniform distribution.[2] For this case

$$\sigma_N^2 = \frac{1}{M}\left[I_o\left(\frac{2E}{N_o}\right) - 1\right] \text{, and hence} \tag{4.117}$$

$$d = \ln\left[1 - \frac{1}{M} + \frac{1}{M} I_o\left(\frac{2E}{N_o}\right)\right] . \tag{4.118}$$

[1] See Section 4.6. [2] See Section 4.7.

These two cases are the basis for the best approximation available to the problem of a signal of unknown time origin or a signal of unknown frequency or both. For example, 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 starting time is random and distributed uniformly over a time interval. However, if the problem is changed slightly, so that the starting time is restricted to times spaced approximately a pulse width apart, then pulses starting at different times would be approximately orthogonal, and the case of the signal one of M orthogonal signals known except for phase could be applied. Eq (4.118) should be used with M equal to the ratio of observation time to pulse width. A similar argument applies to the case in which a signal is a pulse known except for phase and center frequency. Eq (4.118) should be used with M taken as the ratio of total bandwidth to signal bandwidth. It should be pointed out that it is not the same to assume that the signal can appear in only a finite number of different positions, even though the positions are close to each other, as to say that the signal can appear anywhere in an interval. There is more uncertainty in the latter case, and the signal cannot be detected as easily.

5.1.5 The Broad Band Receiver and the Ideal Receiver. One common method of detecting pulse signals in a frequency band is to build a receiver whose bandwidth is the entire frequency band. The receiver operating characteristic for such a receiver with a pulse signal of known starting time is calculated in Section 4.4. This is not a truly ideal receiver, and it would be interesting to compare it with an ideal receiver. This can be done using the approximation of the preceding paragraph for the ideal receiver. Since the bandwidth of a pulse is approximately the reciprocal of the pulse width, the parameter M of Section 4.4 and the parameter M in Eq (4.118) are both equal to the ratio of total bandwidth to pulse bandwidth. Curves showing $\frac{2E}{N_o}$ as a function of d are given in Fig. 5.3 for the approximate ideal receiver and the broad band receiver for several values of M. The expression used for d is Eq (4.71) which holds for large values of M.

5.1.6 Uncertainty and Signal Detectability. In the two cases discussed in Section 5.1.4, where the signal considered is one of M orthogonal signals, the uncertainty of the signal is a function of M. This gives us an opportunity to study the effect in these two cases of uncertainty on signal detectability. In the approximate evaluation of the receiver built to detect the presence of a signal when the signal is one of M orthogonal functions, the curves of Fig. 5.1 are used with the detection index d given by

$$d = \ell n \left[1 - \frac{1}{M} + \frac{1}{M} \exp\left(\frac{2E}{N_o}\right) \right] \tag{4.103}$$

This equation can be solved for the signal energy.

$$\frac{2E}{N_o} = \ell n \left[1 - M + Me^d \right]$$

$$\approx \ell n \ M + \ell n \ (e^d - 1) \quad , \tag{5.2}$$

the approximation holding for large $\frac{2E}{N_o}$.[1] 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.[2]

[1] If $\frac{2E}{N_o} > 3$, the error is less than 10%.

[2] It might be suspected that $\frac{2E}{N_o}$ 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. This is not the case, except when all the p_i are equal. The expression which occurs in this more general case is:

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

M-970 A-G3-83 JEM 7-21-53

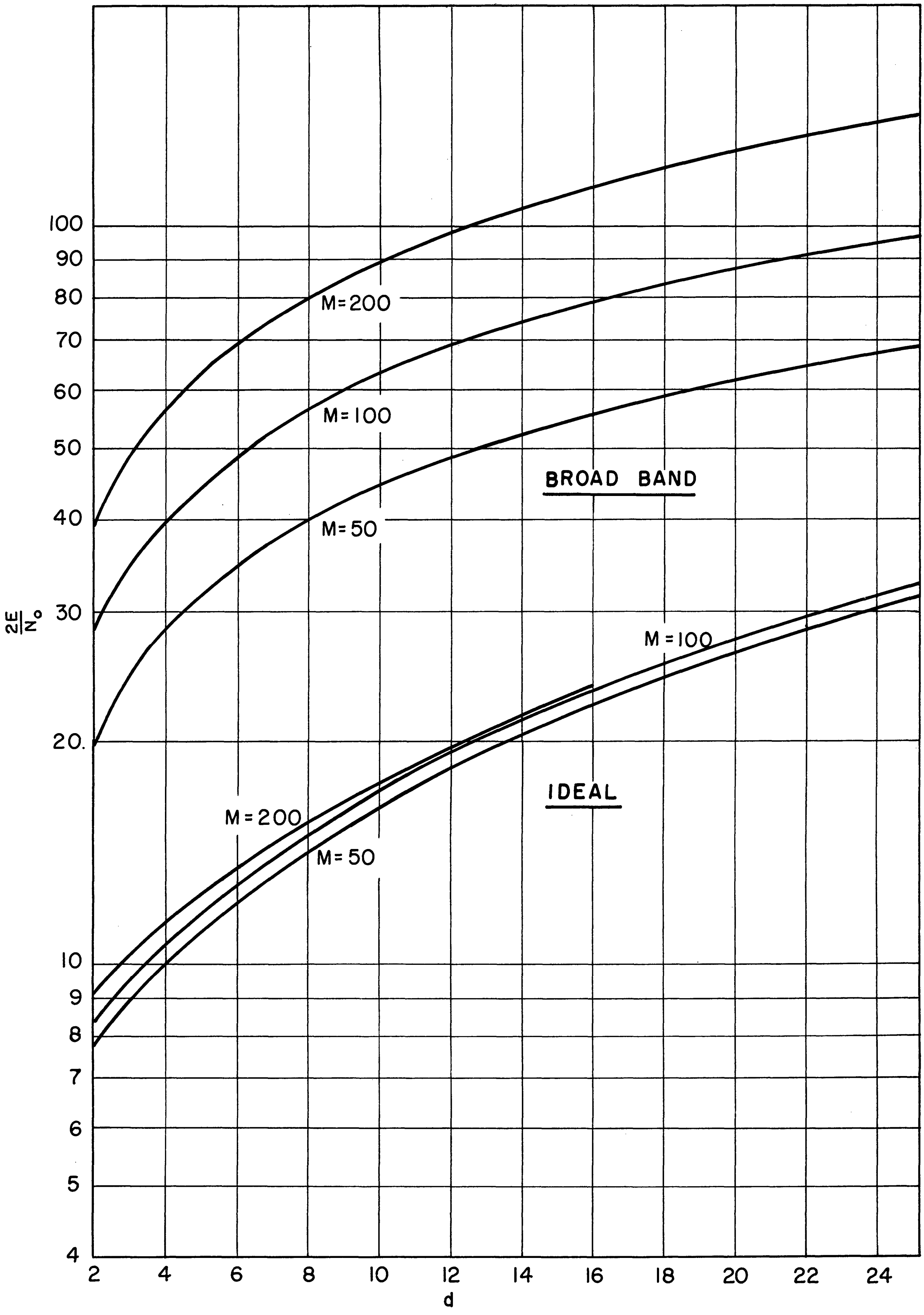

FIG. 5.3. COMPARISON OF IDEAL AND BROAD BAND RECEIVERS.

5.2 Receiver Design. There are a few cases when the receiver design is simple to specify if the noise is Gaussian. If, for example, not only the noise, but also the signal are Gaussian, and both have a uniform spectrum over their bandwidth, then the optimum receiver simply measures the energy which comes in during the observation period. The simple relation between energy and likelihood ratio is given by Eq (4.41) of Section 4.4.

The simplest remaining case is that in which the signal is known exactly. Then the theory specifies that the receiver find the cross correlation between the expected signal and the receiver input, i.e.,

$$\int_0^T s(t)\, x(t)\, dt, \tag{5.3}$$

where s(t) is the expected signal and x(t) is the receiver input, and the observation interval is from t = 0 to t = T. The ratio of this cross correlation to the noise power per unit bandwidth is one-half the natural logarithm of the likelihood ratio.[1] Several elaborate correlating devices have been built recently.[2]

There is, in this case, a simple means of obtaining the correlation, if the signal is simple in form, for example, a pulse. If a filter can be designed with impulse response

$$\begin{aligned} h(t) &= s(T-t) \qquad \text{if } 0 \leqq t \leqq T, \\ &= 0 \qquad\qquad \text{otherwise,} \end{aligned} \tag{4.10}$$

and the receiver input applied to the filter, then the output at time T will be

[1]See page 9, Eq (4.1b).

[2]Harrington and Rogers, Ref. 16; Harting and Meade, Ref. 17; Lee Cheatham, and Wiesner, Ref. 18; Levin and Reintzes, Ref. 19.

$$\int_{-\infty}^{T} x(\tau)\ h(T-\tau)\ d\tau \ = \ \int_{0}^{T} x(\tau)\ s(\tau)\ d\tau \quad , \tag{4.11}$$

which is the required correlation. It turns out that this is the same filter specified by Middleton, Van Vleck, Wiener, North, and Hansen as the filter which maximizes signal-to-noise ratio.[1]

If the signal being sought is an amplitude modulated signal known except for carrier phase, then the ideal receiver has a filter like the one specified in the previous paragraph designed for any particular phase. The receiver input is applied to this filter, and the output is an rf (or more likely, if) voltage. It turns out that the envelope of this voltage is the required quantity. Its relation to likelihood ratio is derived in Section 4.3 and presented in Eqs (4.19) and (4.29).

A look at the general equation for likelihood ratio

$$\ell(x) \ = \ \int_R \exp\left[-\frac{E(s)}{N_o}\right] \exp\left[\frac{2}{N_o} \int_0^T x(t)\ s(t)\ dt\right] dP_S(s) \quad . \tag{3.7b}$$

suggests the following method for designing the optimum receiver for signal detection. First find the correlation as described above, between the receiver input and each possible expected signal. Next, divide each by N_o, the noise per unit bandwidth, and find the exponential function of each. Finally, find the weighted average of all these quantities. The hard part is to find the cross correlation between each expected signal and the receiver input. This means that the ideal filter and associated amplifiers are needed for each expected signal, or essentially a separate receiver for each expected signal. In most

---

[1] Lawson and Uhlenbeck, Ref. 1, p. 206; North, Ref. 11.

cases this is out of the question. In the cases studied in Sections 4.2, 4.3, 4.4, and 4.6, some peculiarity of the set of expected signals made a simpler ideal receiver possible.

There is another noteworthy case. If the signal is known except for starting time, then it is sufficient to look at the same ideal filter at different times rather than to have a different filter for each starting time.

For even a simple square pulse, it is impossible to synthesize the ideal filter exactly. Just how critical, then, is the design of the ideal filter? This can be answered by finding how well signal detection can be accomplished with an approximation to the ideal filter.

For simplicity, consider the case of the signal known exactly. The results for this will follow with little modification for the other cases where the ideal filter is used. The theory specifies that the response of a certain filter to the receiver input be observed at a certain instant. Once it is known that the ideal receiver has this form, it is clear that this filter must be the one which maximizes the instantaneous signal output voltage (or power), the noise rms voltage (or average power) being kept fixed. This is the reason the filter which other authors have found maximizes signal-to-noise ratio is the one which is the absolute optimum <u>for this case</u>.[1]

If a filter can be built for which the output ratio of peak signal to rms noise is nearly the same as that obtained with an ideal filter, then this filter will give results nearly as good as the ideal filter. The noise power at the output of a filter with transfer function $H(\omega)$ is equal to

$$N = \frac{N_0}{2} \int_{-\infty}^{\infty} H(\omega) \overline{H(\omega)} \, d\omega \qquad (5.4)$$

[1] See Footnote, p. 65.

where N_o is the noise power per unit bandwidth of the input noise. By Parseval's theorem,[1] and the fact that h(t), the impulse response, is the Fourier transform of $H(\omega)$.

$$N = \frac{N_o}{2} \int_{-\infty}^{\infty} H(\omega)\, \overline{H(\omega)}\, d\omega$$

$$= \frac{N_o}{2} \int_{-\infty}^{\infty} h(t)\, \overline{h(t)}\, dt. \qquad (5.5)$$

In the case of the ideal filter, Eq. 4.10 can be applied, and the result is

$$N = \frac{N_o}{2} \int_0^T s(T-\tau)^2\, d\tau = \frac{N_o E}{2} \qquad (5.6)$$

where E is the signal energy. The peak voltage output if there is signal but no noise is

$$\int_0^T s(t)^2\, dt = E \quad , \qquad (5.7)$$

and hence the peak signal power at the output is E^2. The ratio of peak signal power to average noise power is thus $\frac{2E}{N_o}$ for the ideal case.

For the particular case of the signal consisting of a single rectangular pulse, if an RC filter is used with time constant 80% of the pulse duration, the receiver operating characteristic will be the same as if the ideal filter were used and the signal reduced 0.90 db. This is derived in Appendix F. Several other pulse cases have been treated and the results for the best filter of each type are summarized in the following table:

---

1 Titchmarsh, An Introduction to the Theory of Fourier Integrals, Oxford University Press, 1937, p. 50.

TABLE II

Pulse	Filter	Equivalent Loss in Signal Strength
Gaussian	Rectangular Passband	0.98 db[1]
Rectangular	Gaussian Passband	
Rectangular	Rectangular Passband	0.83 db[1]
Rectangular	Simple RC Filter (or Single Tuned Circuit)	0.90 db
Rectangular	 impulse response	0.51 db
	 impulse response	1.62 db
 (Exponential Decay)	Simple RC Filter (or Single Tuned Circuit)	2.67 db

The minimum equivalent loss was obtained by adjusting the bandwidth of the filter. Thus in detecting pulses the form of the filter passband is relatively unimportant. However, it is important to have the correct filter bandwidth. This is essentially the present-day attitude in building receivers for receiving pulses of known frequency.

5.3 Conclusions

Part II of The Theory of Signal Detectability consists of the application of the theory presented in Part I to some special cases of signal detection problems in order to obtain information on (1) the design of optimum receivers for the detection of signals, and (2) the performance of these receivers.

---

[1]These cases are derived in Lawson and Uhlenbeck, Ref. 1, p. 206.

The special 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.

TABLE I

Section	Description of Signal Ensemble	Application
4.2	Signal Known Exactly	Coherent radar with a target of known range and character
4.3	Signal Known Except for Phase	Ordinary pulse radar with no integration and with a target of known range and character.
4.4	Signal a Sample of White Gaussian Noise	Detection of noise-like signals; detection of speech sounds in Gaussian noise.
4.5	Video Design 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.
4.6	A Radar Case (A train of pulses with incoherent phase)	Ordinary pulse radar with integration and with a target of known range and character.
4.8	Signal One of M Orthogonal Signals	Coherent radar where the target is at one of a finite number of non-overlapping positions.
4.9	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.

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.

It should be borne in mind that this report discusses the detection of signals in noise; the problem of obtaining information from signals or about signals, except as to whether or not they are present, is not discussed. Furthermore, in treating the special cases, the noise was assumed to be Gaussian.[1]

In addition to general remarks on receiver design,[2] most sections on special cases include specific information describing the simplest design for the optimum receiver for the case considered in those sections.

For the simple cases, the design indicated corresponds closely to the design indicated by the type of analysis in which signal to noise ratio is maximized. For the more complicated cases, the design suggested is usually impractical. For some problems it may never be practical to attempt to build an optimum system. For others, however, engineers equipped with a good understanding of statistical methods and their application to the problem of signal detectability, and to communication theory in general, will undoubtedly invent systems which approach the optimum system.

For each special case treated in this report, at least an approximation is given for the receiver performance. Receiver performance received primary emphasis because it has generally been slighted in previous work. It is

---

[1]See the footnote on page 4 with reference to the spectrum of the assumed noise.

[2]See Section 5.2.

important to know the performance which could be obtained from an optimum receiver even if an optimum receiver cannot be built, since this gives an upper bound on the performance which can be obtained with any receiver in a given situation, and since this also gives an upper bound on what can possibly be accomplished by improvements in receiver design.

APPENDIX D

The Sampling Theorem

Suppose f(t) is a measureable function which is defined for $0 \leqq t \leqq T$. Then f(t) can be expanded in a Fourier series in this interval. The frequency of any term in the series is an integral multiple of 1/T. Suppose there are no terms in the series with frequency above W.[1] This makes the function band limited. Denote by $\psi_m(t)$ the function

$$\psi_m(t) = \frac{\sin\left[\pi(2WT)\left(\frac{t}{T} - \frac{m}{2WT}\right)\right]}{(2WT)\sin\left[\pi\left(\frac{t}{T} - \frac{m}{2WT}\right)\right]} \qquad (D.1)$$

Then

$$f(t) = \sum_{m=1}^{2WT} f\left(\frac{m}{2WT}\right)\psi_m(t) \qquad (D.2)$$

Furthermore, the functions ψ_m are orthogonal on the interval $0 \leq t \leq T$,

$$\int_0^T \psi_m(t)\,\psi_k(t)\,dt = \frac{\delta_{km}}{2W} \qquad (D.3)$$

and

$$\int_0^T \psi_m(t)\,dt = \frac{1}{2W}, \qquad (D.4)$$

where δ_{km} is the Kronecker delta function, which is zero if $k \neq m$ and unity if $k = m$.

[1] We shall assume 2WT is an odd integer. This equivalent to choosing the limit of the band half way between the frequency of the last non zero term in the Fourier series and the frequency of the next term (which, of course, has a zero coefficient).

It follows from Eq (D.2) and Eq (D.3) that

$$\int_0^T \left[f(t)\right]^2 dt = \frac{1}{2W} \sum_{M=1}^{2WT} \left[f\left(\frac{m}{2W}\right)\right]^2 \tag{D.5}$$

and from Eqs (D.2) and Eq (D.4)

$$\int_0^T f(t)\, dt = \frac{1}{2W} \sum_{m=1}^{2WT} f\left(\frac{m}{2W}\right) \tag{D.6}$$

Thus the 2WT functions ψ_m have the same properties for the finite interval which Shannon's interpolation functions have on the infinite interval.[1] It is interesting to note that when 2WT is large, these functions, except the ones near the ends of the interval, are approximately the same as Shannon's.

The Fourier series for $\psi_m(t)$ has no terms with frequency above W. It is, in exponential form,

$$\psi_m(t) = \frac{1}{2WT} \sum_{n=-\left(WT-\frac{1}{2}\right)}^{WT-\frac{1}{2}} \exp\left[j\, \frac{-2\pi\, mn}{2WT}\right] \exp\left[j\, \frac{2\pi\, nt}{T}\right] \tag{D.7}$$

This can be shown by expressing the sine functions in Eq (D.1) as exponentials and using the algebraic identity

$$\frac{a^{n+1} - a^{-n-1}}{a - a^{-1}} = \sum_{k=-n}^{n} a^k \tag{D.8}$$

Formula (D.4) can be proved by integrating Eq (D.7) directly. Note that the only term which contributes to the integral is the term for which $n = 0$.

[1] See Shannon, Ref. 21.

Formula (D.3) can best be proved also by using the Fourier series.

$$\int_0^T \psi_m(t)\,\overline{\psi_k(t)}\,dt$$

$$= \frac{1}{(2WT)^2}\int_0^T \sum_{n=-\left(WT-\frac{1}{2}\right)}^{WT-\frac{1}{2}} \exp\left[j\,\frac{-2\pi mn}{2WT}\right]\exp\left[j\,\frac{2\pi nt}{T}\right]\sum_{p=\left(WT-\frac{1}{2}\right)}^{WT-\frac{1}{2}} \overline{\exp\left[j\,\frac{-2\pi kp}{2WT}\right]}\,\overline{\exp\left[j\,\frac{2\pi pt}{T}\right]}\,dt$$

$$= \frac{1}{(2WT)^2}\sum_{n=-\left(WT-\frac{1}{2}\right)}^{WT-\frac{1}{2}}\ \sum_{p=-\left(WT-\frac{1}{2}\right)}^{WT-\frac{1}{2}} \exp\left[j\,\frac{-2\pi(mn-kp)}{2WT+1}\right]\int_0^T \exp\left[j\,\frac{2\pi t(n-p)}{T}\right]dt$$

$$= \frac{1}{(2WT)^2}\sum_{n=-\left(WT-\frac{1}{2}\right)}^{WT-\frac{1}{2}}\ \sum_{p=-\left(WT-\frac{1}{2}\right)}^{WT-\frac{1}{2}} \exp\left[j\,\frac{-2\pi(mn-kp)}{2WT}\right]\,T\,\delta_{np}$$

$$= \frac{T}{(2WT)^2}\sum_{n=-\left(WT-\frac{1}{2}\right)}^{WT-\frac{1}{2}} \exp\left[j\,\frac{-2\pi n(m-k)}{2WT}\right] .$$

If $m = k$, each of the 2WT terms in the sum is unity. If $m \neq k$, the terms in the sum are equally spaced around the unit circle in the complex plane and must sum to zero. Thus

$$\int_0^T \psi_m(t)\,\psi_k(t)\,dt = \frac{\delta_{km}}{2W} ,$$

which was to be proved.

The validity of the expansion in equation (D.2) follows from the fact that the functions $\psi_m(t)$ are 2WT linearly independent linear combinations of the 2WT functions

$$\exp\left[j \ \frac{2\pi nt}{T}\right] \ \frac{1}{2} - WT \leq n \leq WT - \frac{1}{2}$$

Which are used in the Fourier series expansion. Thus any function which can be expanded in a Fourier series with only the first 2WT **terms** can also be expanded in a unique way in terms of the functions ψ_m.

There is an alternate form of the sampling theorem for band limited signals. With this form the signal function can be described by giving sample values of the envelope and phase of the signal, and hence this form is often convenient to use in describing rf signals.

Suppose the function f(t), when expanded in a Fourier series on the interval $0 \leqq t \leqq T$ has only a finite number of terms in its expansion, and suppose they are included in the terms ranging from frequency f_1 to frequency f_2.[1] The bandwidth then could be defined as

$$W = f_2 - f_1 + \frac{1}{T} , \tag{D.9}$$

and the center frequency is

$$\frac{\omega}{2\pi} = \frac{f_2 + f_1}{2} \tag{D.10}$$

Then the Fourier series can be written

$$f(t) = \sum_{-m}^{m} a_k \cos\left[\left(\omega + \frac{2\pi}{T} k\right) t\right] + b_k \sin\left[\left(\omega + \frac{2\pi}{T} k\right) t\right] \tag{D.11}$$

$$f(t) = R\left\{ \sum_{-m}^{m} \left(a_k - ib_k\right) \exp\left[i\left(\omega + \frac{2\pi k}{T}\right) t\right]\right\} \tag{D.12}$$

where R means "the real part of", and $m = \frac{1}{2}$ (WT - 1).

[1]We shall assume $Tf_2 - Tf_1$ is an even integer and that $Tf_1 \geqq 1$.

$$f(t) = R\left\{\exp\left[i\omega t\right]\sum_{-m}^{m}\left(a_k - ib_k\right)\exp\left[i\frac{2\pi kt}{T}\right]\right\}$$

$$\begin{aligned} f(t) &= R\left\{\exp\left[i\omega t\right]\left(\sum_{-m}^{m} a_k \exp\left[i\frac{2\pi kt}{T}\right] + i\sum_{-m}^{m} b_k \exp\left[i\frac{2\pi kt}{T}\right]\right)\right\} \\ &= R\left\{\exp\left[i\omega t\right]\left(x(t) - iy(t)\right)\right\} \\ &= x(t)\cos\omega t + y(t)\sin\omega t \end{aligned} \tag{D.13}$$

where

$$x(t) = \sum_{-m}^{m}\frac{a_k}{2}\exp\left[i\frac{2\pi kt}{T}\right], \text{ and}$$

$$y(t) = -\sum_{-m}^{m}\frac{b_k}{2}\exp\left[i\frac{2\pi kt}{T}\right] . \tag{D.14}$$

The functions x(t) and y(t) meet the conditions of the first form of the sampling theorem, for a signal with frequencies no higher than $\frac{W}{2}$. They can be expressed therefore in the form

$$x(t) = \sum_{k=1}^{WT} x\left(\frac{k}{W}\right)\psi_k(t)$$

$$y(t) = \sum_{k=1}^{WT} y\left(\frac{k}{W}\right)\psi_k(t) \tag{D.15}$$

Where the ψ functions are defined for a signal with no frequencies above $\frac{W}{2}$. Thus the original function can be written as

$$f(t) = \sum_{k=1}^{WT} x\left(\frac{k}{W}\right)\psi_k(t)\cos\omega t + \sum_{k=1}^{WT} y\left(\frac{k}{W}\right)\psi_k(t)\sin\omega t , \tag{D.16}$$

and the function f(x) can be represented by giving the sample values $x\left(\frac{k}{W}\right)$ and $y\left(\frac{k}{W}\right)$.

Since f(t) can be expressed in the form

$$f(t) = x(t) \cos \omega t + y(t) \sin \omega t \qquad (D.13)$$

and x(t) and y(t) are limited to frequencies less than $\frac{W}{2}$ which is less than $\frac{\omega}{2\pi}$, the envelope of f(t) is

$$r(t) = \sqrt{x(t)^2 + y(t)^2} \qquad (D.17)$$

The angle $\theta(t)$ defined by

$$\cos \theta(t) = \frac{x(t)}{r(t)}$$

$$\sin \theta(t) = - \frac{y(t)}{r(t)} \qquad (D.18)$$

can be considered as the phase of the signal, since

$$f(t) = r(t) \cos \theta(t) \cos wt - r(t) \sin \theta\,(t) \sin wt$$

$$= r(t) \cos \left[\omega t + \theta(t)\right] \qquad (D.19)$$

Note that the sample values x_i and y_i can be obtained from sample values of r and θ,

$$x_i = x\left(\frac{i}{W}\right) = r\left(\frac{i}{W}\right) \cos \left[\theta\left(\frac{i}{W}\right)\right] = r_i \cos \theta_i$$

$$y_i = r_i \sin \theta_i \qquad (D.20)$$

Thus the function f(t) may be represented by giving the sample values of its amplitude and phase at points spaced $\frac{1}{W}$ apart through the observation interval.

APPENDIX E

The integral

$$\exp(-b^2)\int_0^\infty \left[I_o(b\alpha)\right]^2 \alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha \tag{E.1}$$

is required.

The integral

$$\int_0^\infty \alpha^{2n+1} I_o(b\alpha) \exp\left[-\frac{\alpha^2}{2}\right] d\alpha$$

$$= n!2^n \exp\left[\frac{b^2}{2}\right] F(-n, 1; -\frac{b^2}{2})$$

$$= n!2^n \exp\left[\frac{b^2}{2}\right] \sum_{k=0}^{n} \frac{n!b^{2k}}{(n-k)!k!k!2^k}, \tag{E.2}$$

where $F(-n, 1; -\frac{b^2}{2})$ is the confluent hypergeometric function.[1] The function $I_o(b\alpha)$ can be expanded in a power series

$$I_o(b\alpha) = \sum_{n=0}^{\infty} \frac{b^{2n}\alpha^{2n}}{2^{2n}\, n!n!} \tag{E.3}$$

Then the integral (E.1) can be written

$$\exp(-b^2)\int_0^\infty \left[I_o(b\alpha)\right]^2 \alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha = \tag{E.4}$$

(Substituting (E.3) for $I_o(b\alpha)$)

$$= \exp(-b^2)\int_0^\infty \sum_{n=0}^{\infty} \frac{b^{2n}\alpha^{2n}}{2^{2n}n!n!} I_o(b\alpha)\, \alpha \exp\left[-\frac{\alpha^2}{2}\right] d\alpha \tag{E.5}$$

[1] Lawson and Uhlenbeck, Ref. 1, p. 174.

$$= \exp(-b^2) \sum_{n=0}^{\infty} \int_0^{\infty} \frac{b^{2n} \alpha^{2n+1}}{2^{2n} n! n!} I_o(b\alpha) \exp\left[-\frac{\alpha^2}{2}\right] d\alpha \qquad (E.6)$$

(Substituting from (E.2))

$$= \exp(-b^2) \sum_{n=0}^{\infty} \frac{b^{2n}}{2^{2n}\, n! n!}\; n! 2^n \exp\left[\frac{b^2}{2}\right] \sum_{k=0}^{n} \frac{n!\, b^{2k}}{(n-k)! k! k! 2^k} \qquad (E.7)$$

$$= \exp\left[-\frac{b^2}{2}\right] \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{b^{2n+2k}}{2^{n+k}(n-k)! k! k!} \qquad (E.8)$$

(Rearranging the terms in the double sum)

$$= \exp\left[-\frac{b^2}{2}\right] \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} \frac{b^{2n+2k}}{2^{n+k}(n-k)! k! k!} \qquad (E.9)$$

$$= \exp\left[-\frac{b^2}{2}\right] \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} \frac{b^{4k}\, b^{2n-2k}}{2^{2k}\, 2^{n-k}\, k! k!\, (n-k)!} \qquad (E.10)$$

(Letting $m = n-k$)

$$= \exp\left[-\frac{b^2}{2}\right] \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \frac{b^{4k}\, b^{2m}}{2^{2k}\, k! k! m! 2^m} \qquad (E.11)$$

$$= \exp\left[-\frac{b^2}{2}\right] \sum_{k=0}^{\infty} \frac{b^{4k}}{2^{2k}\, k! k!} \sum_{m=0}^{\infty} \frac{b^{2m}}{m! 2^m} \qquad (E.12)$$

$$= \exp\left[-\frac{b^2}{2}\right] I_o(b^2) \exp\left[\frac{b^2}{2}\right] = I_o(b^2) \qquad (E.13)$$

The steps in this derivation which must be justified are interchanging the order of integration and summation at step (E.6) and rearranging the double sum, at steps (E.9) and (E.12). It is easy to show that the integral (E.4) exists. The integrands in (E.6) are uniformly bounded by the integrand in (E.4). Thus the integrals in (E.6) converge uniformly, and the order of integration and summation can be interchanged. As for rearranging double sums, this is possible since all the terms are positive, and hence the convergence is absolute.

APPENDIX F

Let us consider a simple case of approximating the ideal filter by some other filter. Suppose s(t) is a rectangular pulse of energy E and width d. Then

$$s(t) = \sqrt{\frac{E}{d}} \text{ if } 0 \leqq t \leqq d \qquad \text{(F.1)}$$
$$= 0 \text{ otherwise}$$

Suppose the filter is made up of a single resistor and a single condenser, with an amplifier or attenuator, whichever is needed to make the noise power at the output $\frac{N_0 E}{2}$ as in the ideal case. Then the impulse response is of the form

$$h(t) = h_o e^{-\frac{t}{\tau}} \text{ if } t \geqq 0$$
$$= 0 \text{ otherwise} \qquad \text{(F.2)}$$

where τ is the time constant of the filter and h_0 is a constant depending on the gain of the amplifier or attenuator. The requirement that the noise power at the output be $\frac{N_0 E}{2}$ is, by (5.5), equivalent to requiring that

$$E = \int_{-\infty}^{\infty} \left[h(t)\right]^2 dt, \qquad \text{(F.3)}$$

or

$$E = \int_0^{\infty} h_o^2 e^{-\frac{2t}{\tau}} dt = \frac{h_o^2 \tau}{2} \qquad \text{(F.4)}$$

which yields

$$h_o^2 = \frac{2E}{\tau}, \text{ and} \qquad \text{(F.5)}$$

$$h(t) = \sqrt{\frac{2E}{\tau}}\, e^{-\frac{t}{\tau}} \text{ if } t \geqq 0$$
$$= 0 \text{ otherwise} \qquad \text{(F.6)}$$

The response V(t) of this filter to the pulse s(t) is, by (4.11),

$$V(t) = \int_{-\infty}^{T} s(\lambda)\, h(t-\lambda)\, d\lambda. \qquad \text{(F.7)}$$

Substitutions from (F.1) and (F.6) for s(t) and h(t) give

$$V(t) = \int_0^t \sqrt{\frac{E}{d}} \cdot \sqrt{\frac{2E}{\tau}}\, e^{-\frac{(t-\lambda)}{\tau}}\, d\lambda \quad \text{if } 0 \leqq t \leqq d$$

$$= \int_0^d \sqrt{\frac{E}{d}} \sqrt{\frac{2E}{\tau}}\, e^{-\frac{(t-\lambda)}{\tau}}\, d\lambda \quad \text{if } t > d. \tag{F.8}$$

These integrals can be evaluated easily, and

$$V(t) = E\sqrt{\frac{2\tau}{d}}\left(1 - e^{-\frac{t}{\tau}}\right) \quad \text{if } 0 \leqq t \leqq d$$

$$V(t) = E\sqrt{\frac{2\tau}{d}}\left(1 - e^{-\frac{d}{\tau}}\right) e^{-\frac{(t-d)}{\tau}} \quad \text{if } t > d \quad . \tag{F.9}$$

V(t) increases with time if $t < d$ and decreases with time if $t > d$, so it must have its maximum value at $t = d$. That maximum value is

$$V_{max} = E\sqrt{\frac{2\tau}{d}}\left(1 - e^{-\frac{d}{\tau}}\right). \tag{F.10}$$

In Fig. F.1, V_{max}/E is plotted as a function of $\frac{\tau}{d}$. It is seen that at $\frac{\tau}{d} = 0.8$ approximately, V_{max} has a maximum, and at this point V_{max} is approximately 0.9E.

For this particular case, if the RC filter with time constant $\tau = 0.8d$ is used in place of the ideal filter, the reliability of signal detection will be the same as if the ideal filter were used and the signal amplitude were reduced to ninety per cent, or 0.90 decibel.

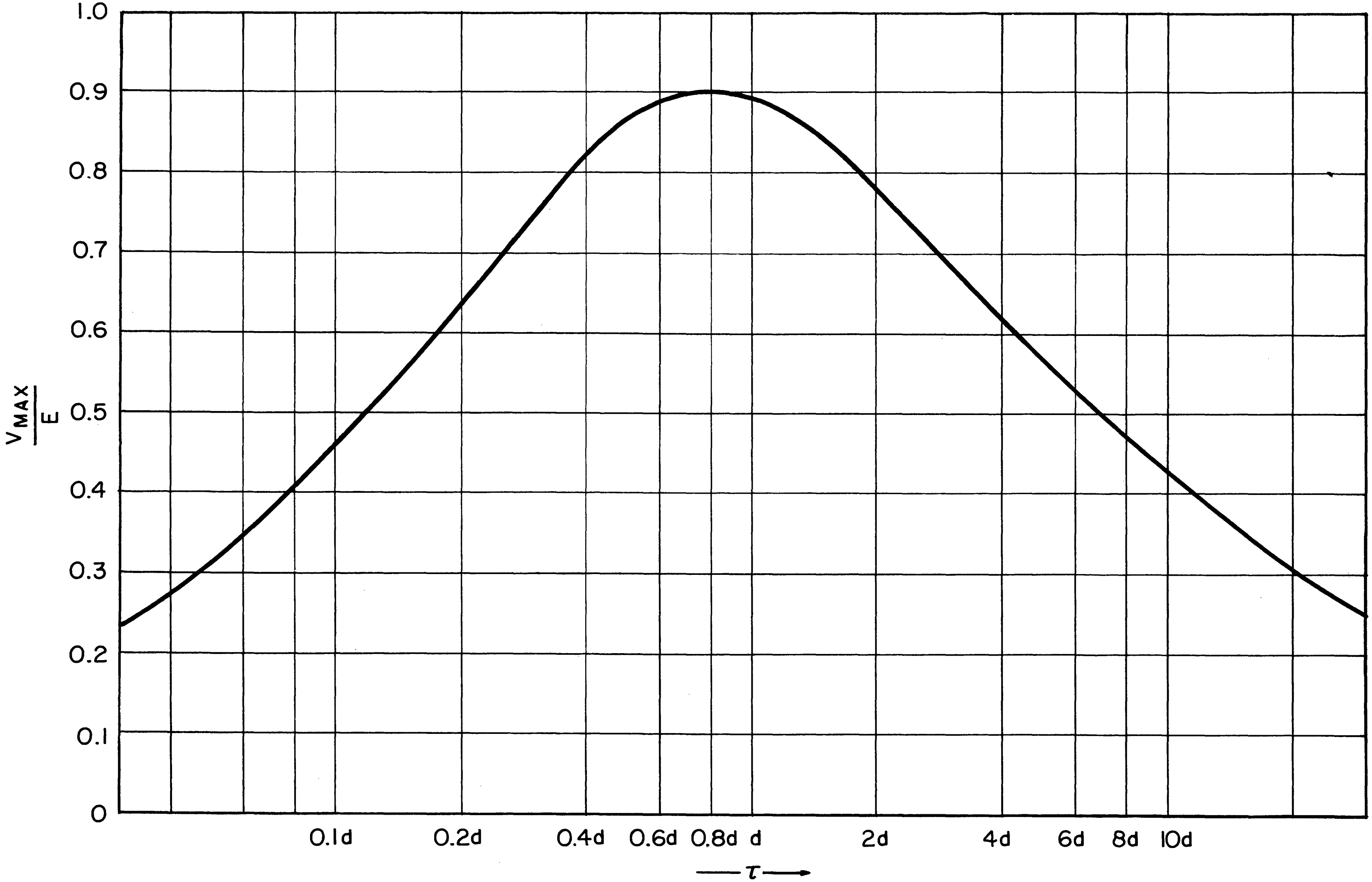

FIG. F.1. MAXIMUM RESPONSE OF R.C. FILTER TO A RECTANGULAR PULSE AS A FUNCTION OF FILTER TIME CONSTANT.

1970 A-G3-129 JEM 4-23-54

BIBLIOGRAPHY

On Statistical Approaches to the Signal Detectability Problem:

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

   This book is certainly the outstanding reference on threshold signals. It presents a great variety of both theoretical and experimental work. Chapter 7 presents a statistical approach of the criterion type for the signal detection problem, and the idea of a criterion which minimizes the probability of an error is introduced. (This is a special case of an optimum criterion of the first type.)

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

3. Woodward, P. M., and Davies, I. L., "Information Theory and Inverse Probability in Telecommunication," Proc. I.E.E. (London), Vol. 99, Part III, p. 37, March, 1952.

4. Woodward, P. M., and Davies, I. L., "A Theory of Radar Information," Phil. Mag., Vol. 41, p. 1001, 1950.

5. Woodward, P. M., "Information Theory and the Design of Radar Receivers," Proc. I.R.E., Vol. 39, p. 1521.

   Woodward and Davies have introduced the idea of a receiver having a posteriori probability as its output, and they point out that such a receiver gives a maximum amount of information. They have handled the case of an arbitrary signal function known exactly or known except for phase with no more difficulty than other authors have had with a sine wave signal. Their methods serve as a basis for the second part of this report.

6. Reich, E., and Swerling, P., "The Detection of a Sine Wave in Gaussian Noise," Jour. App. Phys., Vol. 24, p. 289, March, 1953.

   This paper considers the problem of finding an optimum criterion (of the second type presented in this report) for the case of a sine wave of limited duration, known amplitude and frequency, but unknown phase in the presence of Gaussian noise of arbitrary autocorrelation. The method probably could be extended to more general problems. On the other hand, the methods of this report can be applied if the signals are band limited even in the case of non-uniform noise by putting the signals and noise through an imaginary filter to make the noise uniform before applying the theory. See The Theory of Signal Detectability, Part II, Section 3.2.

7. Middleton, D., "Statistical Criteria for the Detection of Pulsed Carriers in Noise, Jour. Appl. Phys., Vol. 24, p. 371, April, 1953.

A thorough discussion is given of the problem of detecting pulses (of unknown phase) in Gaussian noise. Both types of optimum criteria are discussed, but not in their full generality. The sequential type of test is discussed also. Middleton's equation (6.1) does not hold for the sequential test, and as a result, his calculations for the minimum detectable signal with a sequential test are incorrect. The discussion of the tests is not clear. The comparison of the tests, which are designed to optimize different quantities, seems inappropriate; each test accomplishes its own task in the best possible way.

8. Slattery, T. G., "The Detection of a Sine Wave in Noise by the Use of a Non-Linear Filter," Proc. I.R.E., Vol. 40, p. 1232, October, 1952.

This article considers the problem of detecting a sine wave of known duration, amplitude, and frequency, but unknown phase in uniform Gaussian noise. The article contains several errors, and the results are not clearly presented.

9. Hanse, H., "The Optimization and Analysis of Systems for the Detection of Pulsed Signals in Random Noise," Doctoral Dissertation (MIT), January, 1951.

10. Schwartz, M., "A Statistical Approach to the Automatic Search Problem," Doctoral Dissertation (Harvard), June, 1951.

These dissertations both consider the problem of finding the optimum receiver of the criterion type for radar type signals.

11. North, D. O., "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems," RCA Laboratory Report PTR-6C, 1943.

The ideas of false alarm probability and probability of detection are introduced. North argues that these probabilities will be most favorable when peak signal to average noise ratio is largest. The ideal filter, which maximizes this ratio, is derived. (This commentary is based on second-hand knowledge of the report.)

12. Kaplan, S. M., and Fall, R. W., "The Statistical Properties of Noise Applied to Radar Range Performance," Proc. I.R.E., Vol. 39, p. 56, January, 1951.

The ideas of false alarm probability and probability of detection are introduced and an example of their application to a radar receiver is given.

Marcum, J. I., "A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix," Rand Corporation Report R-113, July 1, 1948.

This report contains a careful, thorough study of the mathematical problem which it considers.

On Statistics:

13. Neyman, J., and Pearson, E. S., "On the Problem of the Most Efficient Tests of Statistical Hypotheses," Phil. Trans. Roy. Soc., Vol. 231, Series A, p. 289, 1933.

14. Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, 1951.

On Related Topics:

15. Dwork, B. M., "Detection of a Pulse Superimposed on Fluctuation Noise," Proc. I.R.E., Vol. 38, p. 771, July, 1950.

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

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

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

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

20. Rice, S. O., "Mathematical Analysis of Random Noise," B.S.T.J., Vol. 23, p. 282-332 and Vol. 24, p. 46-156, 1945-6.

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

LIST OF SYMBOLS

A	The event "The operator says there is signal plus noise present," or a criterion, i.e., the set of receiver inputs for which the operator says there is a signal present.
$A_1(\beta)$	Any criterion A which maximizes $P_{SN}(A) - \beta P_N(A)$, i.e., an optimum criterion of the first type.
$A_2(k)$	Any criterion A for which $P_N(A) \leqq k$, and $P_{SN}(A)$ is maximum, i.e., an optimum criterion of the second type.
CA	The event "The operator says there is noise alone."
d	A parameter describing the ability of a receiver to detect signals. (See Section 5.1 and Fig. 5.1.)
E, $E(s)$	The signal energy.
E^n	The n-dimensional Euclidean space.
$f_N(x)$	The probability density for points x in R if there is noise alone.
$f_{SN}(x)$	The probability density for points x in R if there is signal plus noise.
$F_N(\beta)$, $F_N(\ell)$	The complementary distribution function for likelihood ratio if there is noise alone, i.e., $F_N(\beta)$ is the probability that the likelihood ratio will be greater than β if there is noise alone.
$F_{SN}(\beta)$, $F_{SN}(\ell)$	The complementary distribution function for likelihood ratio if there is signal plus noise.
k	A symbol used primarily for the upper bound placed on false alarm probability $P_N(A)$ in the definition of the second kind of optimum criterion.
$\ell(x)$	The likelihood ratio for the receiver input x. $\ell(x) = \dfrac{f_{SN}(x)}{f_N(x)}$.
n	The dimension of the space of receiver inputs. $n = 2WT$.
N	The event "There is noise alone," or the noise power.
N_o	The noise power per unit bandwidth. $N_o = N/W$.
$P_N(A)$	The probability that the operator will say there is signal plus noise if there is noise alone, i.e., the false alarm probability.

$P_{SN}(A)$	The probability that the operator will say there is signal plus noise if there is signal plus noise, i.e., the probability of detection.
$P_x(SN)$	The a posteriori probability that there is signal plus noise present. (See Sections 1.3 and 2.3.)
$P_S(s)$	The probability measure defined on R for the set of expected signals.
R	The space of all receiver inputs. (The set of all possible signals is the same space.)
s	A signal s(t), which may also be considered as a point s in R with coordinates $(s_1, s_2, \ldots, s_n)$.
SN	The event "There is signal plus noise."
t	Time.
T	The duration of the observation.
W	The bandwidth of the receiver inputs.
x	A receiver input x(t), which may also be considered as a point x in R with coordinates $(x_1, x_2, \ldots, x_n)$
β	A symbol usually used for the likelihood ratio level of an optimum criterion.
$\mu_{SN}(z)$	The mean of the random variable z if there is signal plus noise.
$\mu_N(z)$	The mean of the random variable z if there is noise alone.
$\sigma_N^2(z)$	The variance of the random variable z if there is noise alone.
σ_N^2	The variance of likelihood ratio if there is noise alone.

Note: The terms "normal distribution" and "Gaussian distribution" have been used interchangeably in this report.

DISTRIBUTION LIST

1 copy	Director, Electronic Research Laboratory Stanford University Stanford, California Attn: Dean Fred Terman
1 copy	Commanding Officer Signal Corps Electronic Warfare Center Fort Monmouth, New Jersey
1 copy	Chief, Engineering and Technical Division Office of the Chief Signal Officer Department of the Army Washington 25, D. C. Attn: SIGGE-C
1 copy	Chief, Plans and Operations Division Office of the Chief Signal Officer Washington 25, D. C. Attn: SIGOP-5
1 copy	Countermeasures Laboratory Gilfillan Brothers, Inc. 1815 Venice Blvd. Los Angeles 6, California
1 copy	Commanding Officer White Sands Signal Corps Agency White Sands Proving Ground Las Cruces, New Mexico Attn: SIGWS-CM
1 copy	Signal Corps Resident Engineer Electronic Defense Laboratory P. O. Box 205 Mountain View, California Attn: F. W. Morris, Jr.
75 copies	Transportation Officer, SCEL Evans Signal Laboratory Building No. 42, Belmar, New Jersey For - Signal Property Officer Inspect at Destination File No. 25052-PH-51-91(1443)

1 copy	W. G. Dow, Professor Dept. of Electrical Engineering University of Michigan Ann Arbor, Michigan
1 copy	H. W. Welch, Jr. Engineering Research Institute University of Michigan Ann Arbor, Michigan
1 copy	Document Room Willow Run Research Center University of Michigan Willow Run, Michigan
10 copies	Electronic Defense Group Project File University of Michigan Ann Arbor, Michigan
1 copy	Engineering Research Institute Project File University of Michigan Ann Arbor, Michigan