

**NASA CONTRACTOR
REPORT**



NASA CR-412

0099615



TECH LIBRARY KAFB, NM

NASA CR-412

LOAN COPY - RETURN TO
AFWL (WLIL-2)
KIRTLAND AFB, N MEX

**ON SIGNAL AND NOISE LEVEL
ESTIMATION IN A COHERENT
PCM CHANNEL**

by R. B. Kerr

Prepared under Contract No. NAS 5-3815 by
DUKE UNIVERSITY
Durham, N. C.
for Goddard Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MARCH 1966



ON SIGNAL AND NOISE LEVEL ESTIMATION
IN A COHERENT PCM CHANNEL

By R. B. Kerr

Distribution of this report is provided in the interest of information exchange. Responsibility for the contents resides in the author or organization that prepared it.

Prepared under Contract No. NAS 5-3815 by
DUKE UNIVERSITY
Durham, N.C.

for Goddard Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION



Table of Contents

	Page
List of Figures	iv
Summary	v
Introduction	1
Estimation of A and N_0 Based Upon $y(t)$	3
Estimation of A and N_0 Based Upon ϕ_i	5
Confidence Intervals for \hat{S}	9
Conclusions	16
References	17

List of Figures:

	Page
Figure 1: Confidence Interval Curves, 4 Bits, 1 Data Word Observed.	11
Figure 2: Confidence Interval Curves, 4 Bits, 4 Data Words Observed.	12
Figure 3: Confidence Interval Curves, 4 Bits, 16 Data Words Observed.	13
Figure 4: Confidence Interval Curves, 6.65 Bits, 1 Data Word Observed.	14
Figure 5: Word-Error Probabilities, Phase Coherent Orthogonal Coding.	15

On Signal and Noise Level Estimation in a Coherent
PCM Channel

Summary

Joint maximum likelihood estimators are presented for the signal amplitude and noise power density in a coherent PCM channel with white Gaussian noise and a correlation receiver. The estimates are based upon the correlation coefficient outputs of the receiver. From these estimators, an estimator for the quantity

$$\frac{\text{(received signal energy) / bit}}{\text{(noise power) / (unit bandwidth)}},$$

upon which the error probabilities depend, is derived. This estimator is shown to be useful as:

- a) A point estimator for the signal-to-noise ratio for the higher values of this ratio (about 4 db. or greater).
- b) An easily calculated statistic upon which to base data acceptance or rejection criteria. The acceptance or rejection levels are obtained by the use of confidence interval curves in conjunction with word error probability data.

I Introduction

The advantages of phase-coherent PCM (and, as a special case, PFM) communications have been widely recognized^(1,2,3,4) and utilized in the recovery of information from scientific satellites. The optimum (Bayes) receiver for such a channel is one which computes the conditional probability for each of the possible code words, given the received (noisy) data word. If the transmission of each of the possible code words is a-priori equally likely, then maximizing the conditional probability is equivalent to maximizing the likelihood function. If the received words (i.e., the signal parts of the received waveforms) are all of equal energy, the likelihood function is maximized by choosing the word which yields the largest correlation coefficient when correlated with the received "data" word.^(5,6)

A matter of fundamental importance, particularly in the reception of scientific information from satellites, is the reliability of the information obtained. Presumably it is better in some instances to discard data rather than to use data for which the probability of error is large. For a coherent PCM channel with correlation reception, the error probabilities depend upon a single parameter,^(1,2)

$$\xi \triangleq \frac{S T}{N_0} \quad (1)$$

where

T = time per bit

S = received signal power

ST = received signal energy per bit

N_0 = noise power per unit bandwidth.

Since, in practice, it is not reasonable to assume that either the signal amplitude or the noise power density will be known, the question arises: Given the observed data, what is the best that one can do toward estimating these parameters, or, in particular, the single quantity ξ ? Since the error probability depends upon this parameter, a sequentially updated estimate of ξ (if the estimate is a "good" one) will provide the best information upon which a decision to use or reject the data can be based.

In the following discussion, the received data word will be denoted by

$$r_j(t) = A x_j(t) + n(t) \quad (2)$$

where $x_j(t)$ represents the j 'th. word waveform, normalized so that

$$\int_0^{T_w} x_j^2(t) dt = T_w \quad (3)$$

where

$$T_w \triangleq \text{time duration of word}$$

$$= T X \text{ (bits per word).}$$

A is an unknown amplitude factor, and $n(t)$ will be assumed white Gaussian noise of unknown power density. The correlation receiver generates a set of numbers

$$\phi_i = \frac{1}{T_w} \int_0^{T_w} r_j(t) x_i(t) dt \quad (4)$$

We also define:

- $N \triangleq$ number of words in the code set
 $=$ number of correlation filters for an orthogonal code.
 $= \log_2^{-1}$ (number of bits).
- $M \triangleq$ number of received data words up to the time at which the estimate is to be made. (It should be noted that the statistics in the following discussion will be considered stationary — this may impose an additional limitation on M in practice.)

The possibility exists of basing the estimates either on $y(t)$ or the set of numbers ϕ_i - i.e., either on the entire received waveform or on the output coefficients of the correlating filters. These alternatives will now be compared.

II Estimation of A and N_0 Based Upon $y(t)$

It will now be assumed that $y(t) = A \mathcal{N}_j(t) + n(t)$ has been observed for M received words. The unknowns are j , A , and N_0 , and it is assumed that the noise is limited to bandwidth B at the receiver ($2B$ if the negative frequencies are included). Each $y(t)$ can then be represented (with the usual limitations) by $2BT_w$ samples, and each sample has a variance $\sigma^2 = 2BN_0$. It will be assumed that no a-priori information is available concerning the three unknowns, which will therefore be treated as parameters and the method of maximum likelihood employed.⁽⁷⁾ The likelihood function for the M observations is:

$$L(j, A, N_0) = \left(\frac{1}{2\pi\sigma^2} \right)^{BT_w M} \text{EXP} \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^M \left[\sum_{k=1}^{2BT_w} (y_k - A \mathcal{N}_{jk})^2 \right] \right\} \quad (5)$$

where $y_K - A \phi_{jK} = m_K$, the K^{th} sample of $n(t)$. Since, from the sampling theorem,

$$\sum_{K=1}^{2BT_w} m_K^2 = 2B \int_0^{T_w} n^2(t) dt \quad (6)$$

$$L = (4\pi B N_0)^{-BT_w M} \text{EXP} \left\{ -\frac{1}{2N_0} \sum_{i=1}^M \left[\int_0^{T_w} (y_i - A \phi_{j_i})^2 dt \right] \right\} \quad (7)$$

or

$$\text{Log } L = -BT_w M \text{Log}(4\pi B N_0) + \frac{1}{2N_0} \sum_{i=1}^M \left[2AT_w \phi_{j_i} - A^2 T_w - E_{y_i} \right] \quad (8)$$

where

$$E_{y_i} \triangleq \int_0^{T_w} y_i^2(t) dt \quad (9)$$

Since, after the data has been received, E_{y_i} is a constant, it is clear that regardless of the values of A and N_0 , $\text{Log } L$ (and therefore L) will be maximized by choosing the values j which correspond to the maximum ϕ_j 's. This is of course the principle of the correlation detector — however it is interesting to note that if one is attempting to jointly estimate not only which waveform was transmitted but also the signal amplitude and noise power density, it is not necessary to use some different technique to choose the signal waveform than would be used if this alone were required. Since our distribution is of the form admitting sufficient statistics,⁽⁷⁾ $\text{Log } L$ will have a single unique maximum depending on A and N_0 , which may be found by differentiation. Setting the

derivatives with respect to A and N_0 equal to zero yields the estimates

$$\hat{A} = \frac{1}{M} \sum_{K=1}^M \phi_K \quad (10)$$

$$\hat{N}_0 = \frac{1}{2BT_w M} \left(\sum_{K=1}^M (E_{y_k})_K - M T_w \hat{A}^2 \right) \quad (11)$$

where ϕ_K denotes the set of M maximum ϕ_j 's (one for each observed data word).

III Estimation of A and N_0 Based Upon ϕ_i

Since forming E_y implies instrumentation not normally incorporated in the correlation detector, we will consider now estimating A and N_0 from the outputs ϕ_i of the correlating filters. M of these outputs will consist of signal level A plus noise (and will be denoted by ϕ_j), while the remainder (ϕ_e) will consist of noise alone. For orthogonal codes the noise outputs will be independent (for "white" noise), and of mean zero and variance N_0/T_w . The likelihood function for these outputs is thus

$$L = (2\pi)^{-\frac{MN}{2}} \left(\frac{T_w}{N_0} \right)^{\frac{MN}{2}} \text{EXP} \left\{ -\frac{T_w}{2N_0} \left[\sum_{j=1}^M (\phi_j - A)^2 + \sum_{e=1}^{M(N-1)} \phi_e^2 \right] \right\} \quad (12)$$

and

$$\text{Log } L = (\text{const.}) - \frac{MN}{2} \text{Log } N_0 + \frac{T_w}{2N_0} \left[2A \sum_{j=1}^M \phi_j - MA^2 - \sum_{j=1}^M \phi_j^2 - \sum_{e=1}^{M(N-1)} \phi_e^2 \right]$$

$$= (\text{const.}) - \frac{MN}{2} \text{Log } N_0 + \frac{T_w}{2N_0} \left[2A \sum_{j=1}^M \phi_j - MA^2 - \sum_{i=1}^{MN} \phi_i^2 \right] \quad (13)$$

(i now indexes all of the filter outputs.)

Again, differentiating with respect to A and N_0 yields the estimates:

$$\hat{A} = \frac{1}{M} \sum_{k=1}^M \phi_k \quad (14)$$

$$\hat{N}_0 = \frac{T_w}{MN} \left[\sum_{i=1}^{MN} \phi_i^2 - M \hat{A}^2 \right] \quad (15)$$

where, as before, the ϕ_k represent the maximum correlator outputs (one for each observed data word).

It is interesting to note the correspondence between these estimates and those of equations 10 and 11. The difference is that in deriving the estimates the signals were represented in two different orthogonal signal spaces. If the dimensionality of these two spaces is the same — i.e., if $N = 2BT_w$, then we would expect our estimates to be identical. In this case, since $y(t)$ is completely characterized by the ϕ_i components,

$$E_y = \int_0^{T_w} y^2(t) dt = T_w \sum_{m=1}^N \phi_m^2 \quad (16)$$

and \hat{N}_o from equation (11) becomes:

$$\begin{aligned} \hat{N}_o &= \frac{1}{2BT_w M} \left[\sum_{K=1}^M (E_y)_K - MT_w \hat{A}^2 \right] = \frac{1}{MN} \left[\sum_{K=1}^M T_w \sum_{m=1}^N \phi_m^2 - MT_w \hat{A}^2 \right] \\ &= \frac{T_w}{MN} \left[\sum_{i=1}^{MN} \phi_i^2 - M \hat{A}^2 \right] \quad (17) \end{aligned}$$

which corresponds to equation (15).

As an example of this equivalence, consider a PFM channel with N possible orthogonal words. The minimum frequency difference (for orthogonality) between adjacent frequencies is $1/2T_w$ C.P.S.,⁽²⁾ requiring therefore a bandwidth of approximately $B_{\min.} \approx N/2T_w$, or

$N \approx 2B_{\min} \cdot T_w$. Hence if the noise is pre-filtered such that the resultant bandwidth is just sufficient to pass the signals, the ϕ_i coefficients in this case contain all of the information pertinent to estimating A and N_o that was contained in the entire continuous (filtered) $y(t)$. Furthermore, since A and N_o are sufficient statistics, none of this information is lost in forming the estimates from the ϕ_i .

Maximum likelihood estimates, if they are sufficient, guarantee the property of minimum variance, but are not necessarily unbiased. A (eq. 14) is unbiased, but N_o (eq. 15) has the expected value

$$\begin{aligned} \mathcal{E}\{\hat{N}_o\} &= \frac{T_w}{MN} \left[\sum_{i=1}^{MN} \mathcal{E}\{\phi_i^2\} - \frac{1}{M} \mathcal{E}\left\{\sum_{K=1}^M \phi_K\right\}^2 \right] \\ &= \frac{T_w}{MN} \left[M\left(A^2 + \frac{N_o}{T_w}\right) + M(N-1)\frac{N_o}{T_w} - MA^2 - \frac{N_o}{T_w} \right] = \frac{MN-1}{MN} N_o \quad (18) \end{aligned}$$

Hence the minimum variance unbiased estimator for N_o is given by

$$\hat{N}_o = \frac{T_w}{MN-1} \left[\sum_{i=1}^{MN} \phi_i^2 - M\hat{A}^2 \right] \quad (19)$$

Since the error probability depends only on \mathcal{J} , where

$$\mathcal{J} \triangleq \frac{ST}{N_o} = \frac{TA^2}{N_o} \quad (20)$$

and since A and N_0 are jointly optimum in the sense previously discussed, are easy to calculate, and are heuristically "natural" estimators for the respective parameters, it is natural to investigate

$$\hat{\xi} \triangleq \frac{T\hat{A}^2}{\hat{N}_0} = \frac{MN-1}{\text{Log}_2 N} \frac{\hat{A}^2}{\sum_{i=1}^{MN} \phi_i^2 - M\hat{A}^2} \quad (21)$$

$$\hat{A} = \frac{1}{M} \sum_{K=1}^M \phi_K$$

as an estimator for ξ . This was done experimentally on a high speed digital computer, and results are presented in the form of confidence interval curves.

IV Confidence Intervals for $\hat{\xi}$

The primary object in attempting to estimate ξ is to present information about the validity of the data obtained. This may be done by the use of confidence curves such as those presented (by way of example) in this section.

Figures 1-3 show 0.90 factor confidence interval curves for a 16 filter (4 bit) system, for 1 data word observed, 4 data words observed, and 16 data words observed respectively ($M = 1, 4, 16$). If a horizontal line is drawn at the value of the ordinate given by the estimate ($\hat{\xi}$), the portion of that line lying between the two confidence curves defines an interval which will contain, with probability 0.90, the true value of ξ .⁽⁸⁾ Also, the true value will lie, with probability 0.95, to the left of the point of intersection with the right-most confidence curve, and with probability 0.95 to the right of the left-most confidence curve. The curves for a 100 filter (6.65 bit) system are shown in Figure 4 (for 1 observed data word).

Curves such as these may be used in the following way, in conjunction with already published error-probability curves. (Figure 5 is reproduced, with consent of the authors, from Reference 3.) Suppose that the four bit system is being used, and that the noise is sufficiently stationary that it can be assumed to be stationary over intervals of four word-transmission times. We would then make use of the curves of Figure 2 (using four word groups to form our estimates). If our estimate turned out to be 0 db. or less, we know from the curves that with probability 0.95 the true signal-to-noise ratio is less than 1 db. Going then to the error-probability curve ($n = 4$ bits), we know that with probability 0.95 our error frequency will be greater than one in ten. We might desire, for example, to base our "rejection" level on this error probability, as indicated on the curves. If, on the other hand, our estimate was 6.7 db. or greater, we know from the curves that with probability 0.95 the true ξ will be greater than 4 db., corresponding to an error frequency of less than one in one-hundred. Placing, for example, our "valid data" level at this value, we have then a criterion for labeling the data as valid, questionable, or to be rejected, depending on the value obtained for $\hat{\xi}$. (These levels of course are purely arbitrary, and are indicated on the curves simply as an illustration of how they can be used.)

$\hat{\xi}$, as a point estimator for ξ , is only good for higher signal-to-noise ratios, as can be seen from the confidence curves. For $\xi \geq 4$ db. approximately, $\hat{\xi}$ is for practical purposes unbiased and consistent, while for the lower signal-to-noise ratios $\hat{\xi}$ is badly biased, this bias in fact causing the unfavorable "bend" in the confidence interval curves. However, the fact that the estimator is poor as a point estimator for low values of ξ does not interfere with the use of the confidence curves to establish a "rejection" threshold as was done in the preceding paragraph.

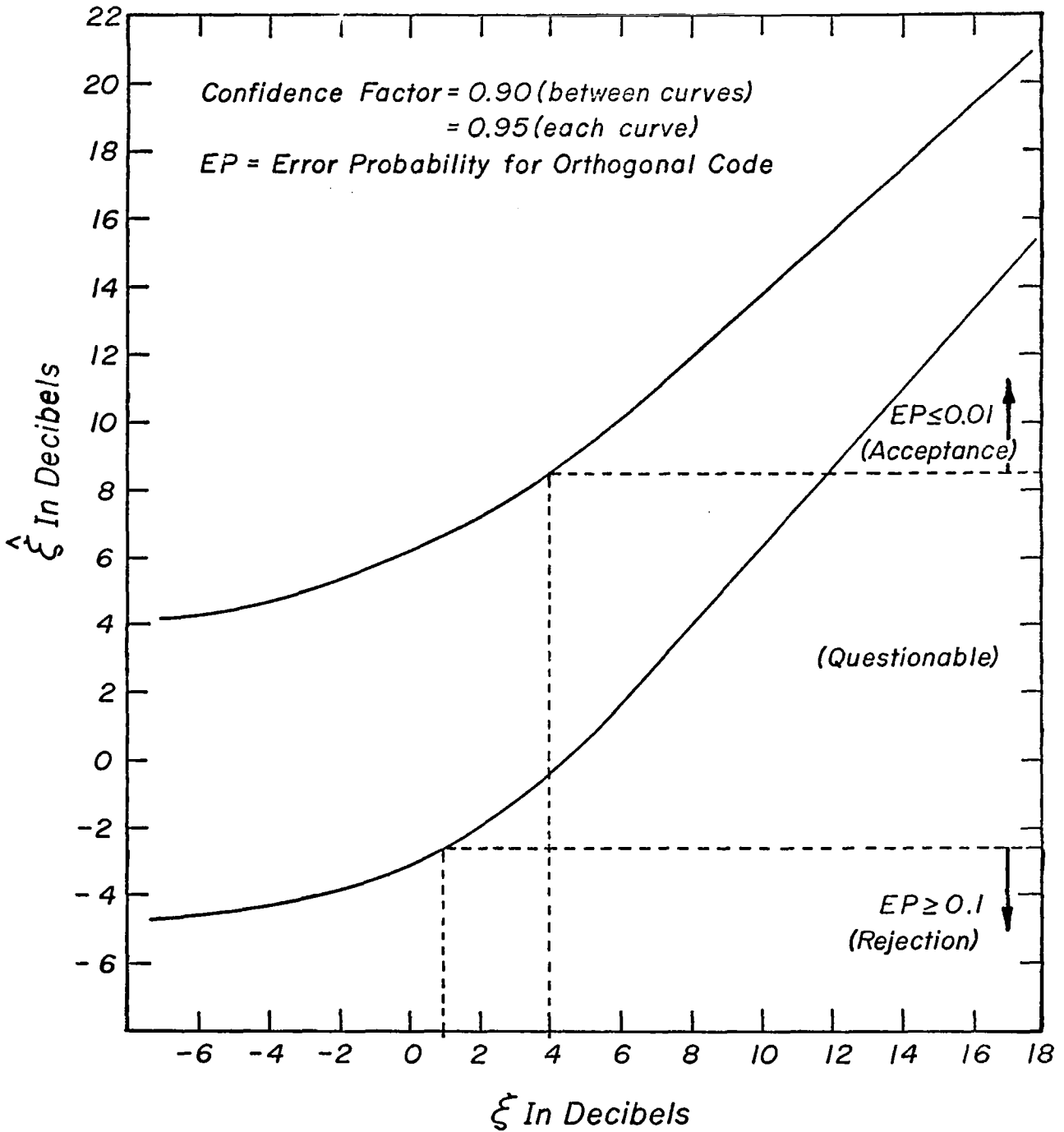


Figure 1 Confidence Interval Curves, 4 Bits, 1 Data Word Observed

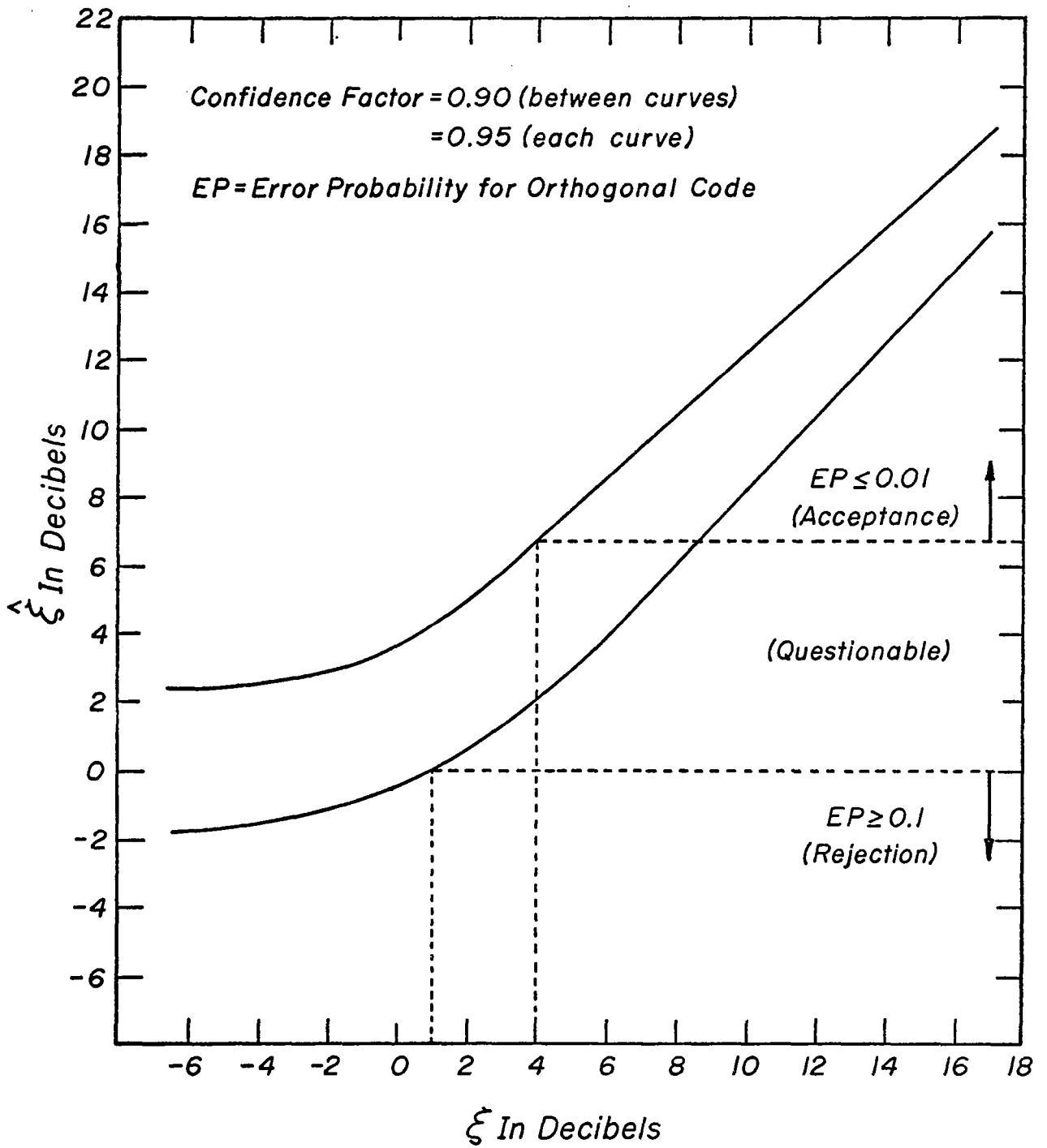


Figure 2 Confidence Interval Curves, 4 Bits, 4 Data Words Observed

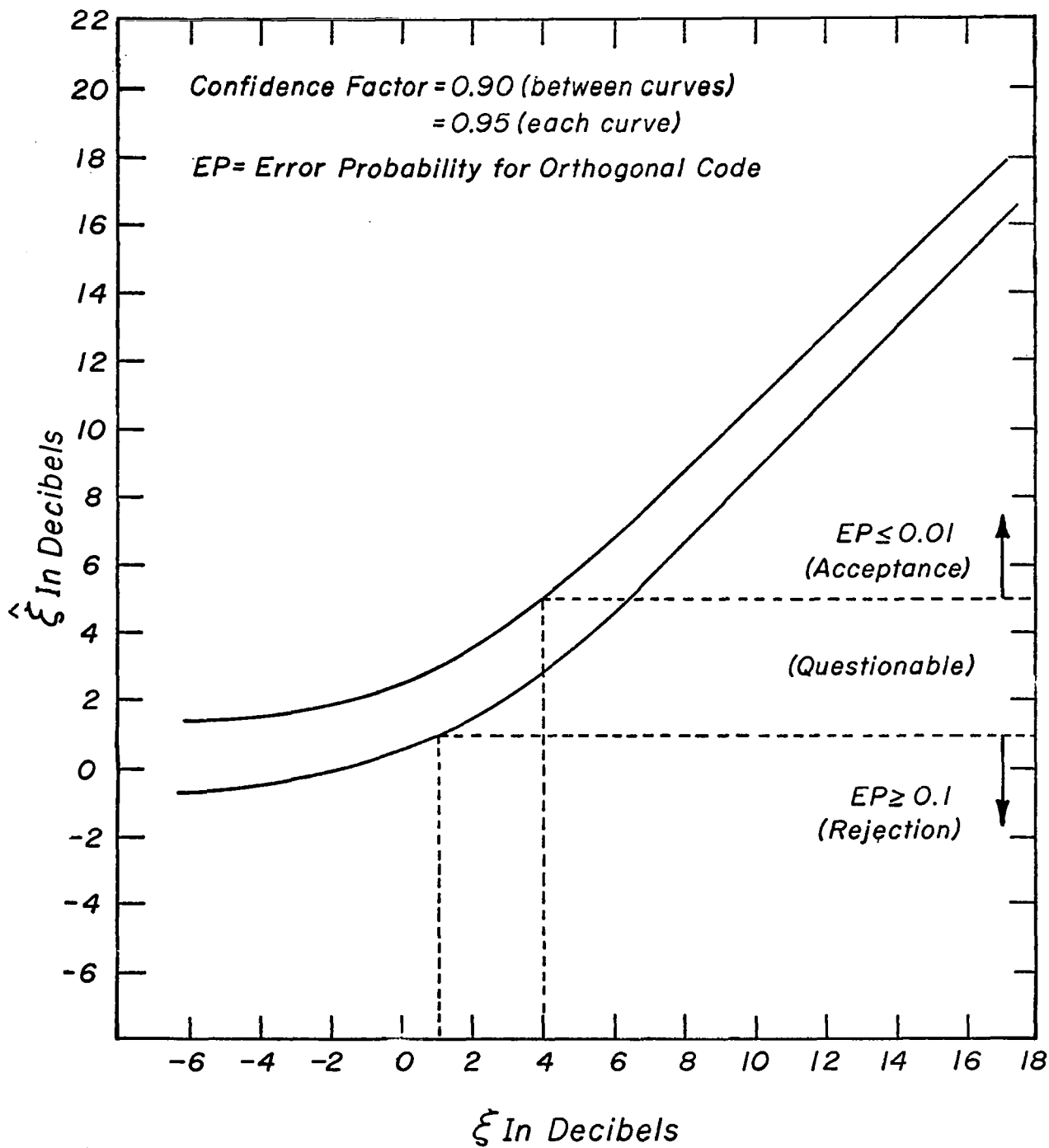


Figure 3 Confidence Interval Curves, 4 Bits, 16 Data Words Observed

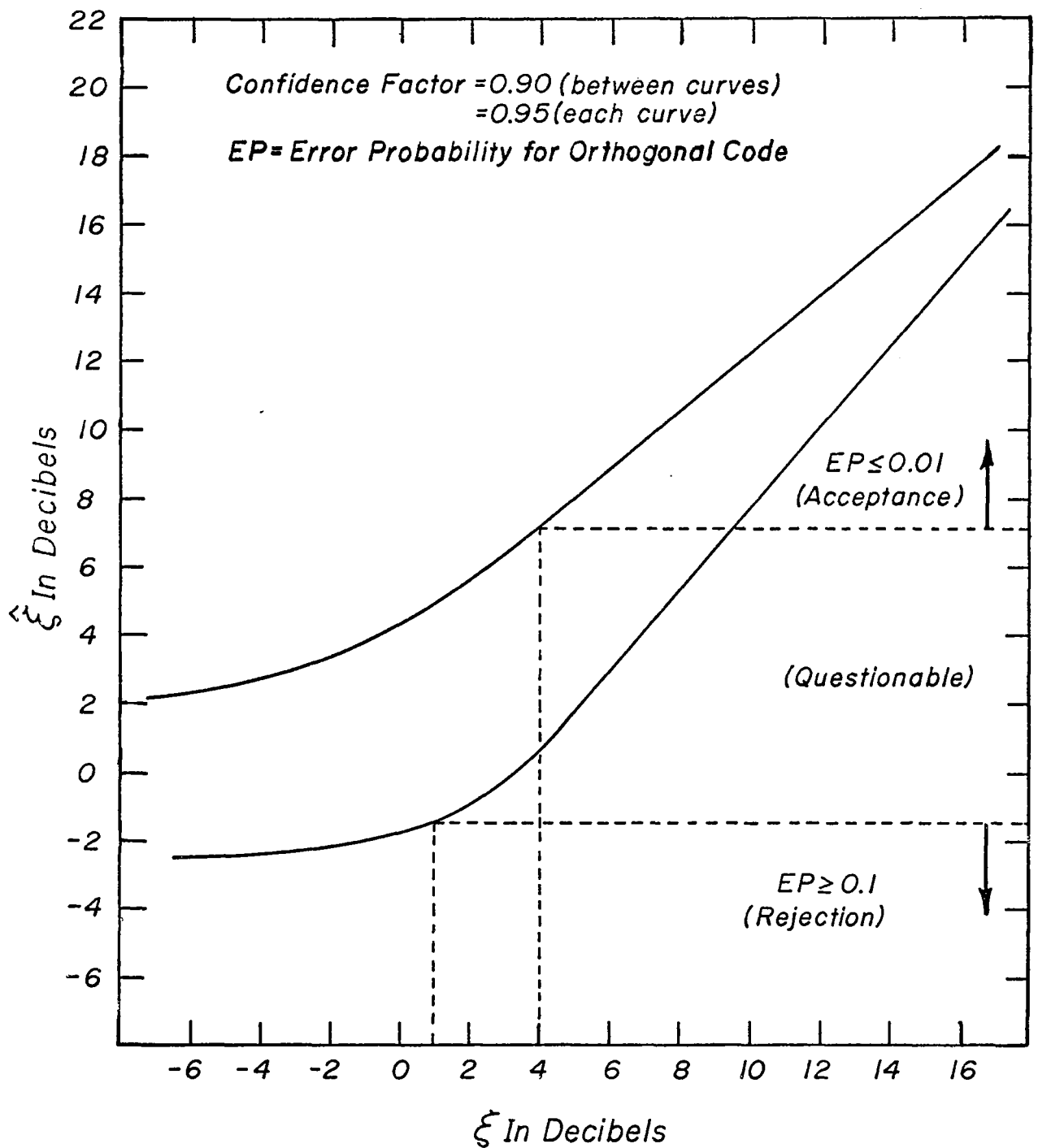


Figure 4 Confidence Interval Curves, 6.65 Bits, 1 Data Word Observed

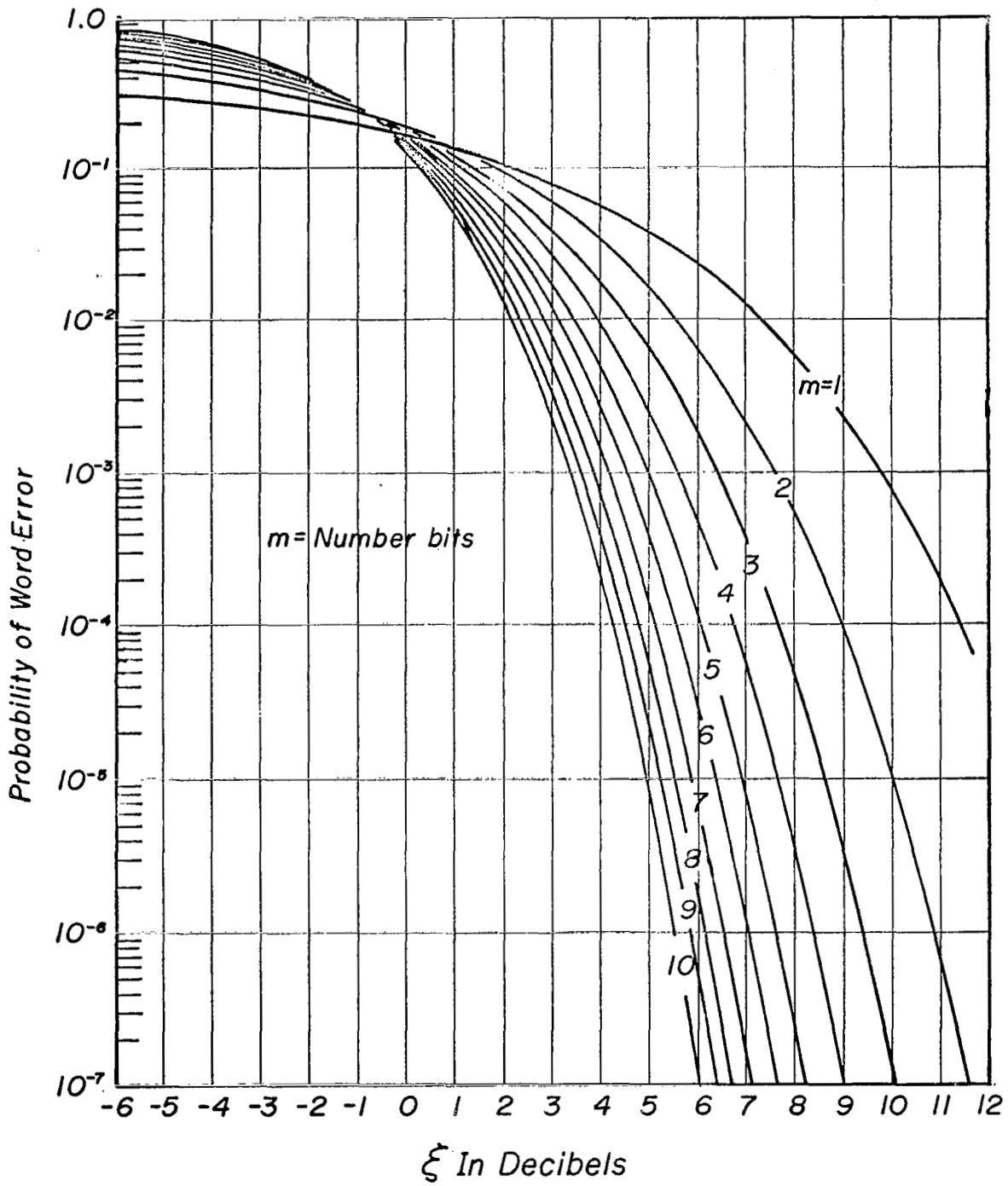


Figure 5 Word Error Probability-Phase-Coherent Orthogonal Coding

V Conclusions

The optimum "detector" for a PCM channel with white Gaussian noise and equal a-priori signal probabilities is a correlation detector. If it is desired not only to choose the most probable signal, but also to calculate the most probable values for signal amplitude and noise power density, this can be done with no additional instrumentation (if the estimates are based upon the outputs of the correlators), and with no modification of the correlation detection scheme itself. These estimates, which are minimum variance (and can be made unbiased), are heuristically "natural" ones and are easy to calculate. They may be used to form an estimator $\hat{\xi}$ for the signal-to-noise ratio ξ , although this estimator will no longer be optimum in a minimum variance unbiased sense. In fact, for signal-to-noise ratios $\xi \geq 4$ db., $\hat{\xi}$ is essentially unbiased and probably very close to optimum, while for lower values of ξ , the bias renders $\hat{\xi}$ a very poor point estimator for ξ .

The ultimate aim in attempting to estimate the signal-to-noise ratio is to establish acceptance or rejection criteria for the received data. This can be done through the use of confidence interval curves, based on $\hat{\xi}$, in conjunction with already published error probability curves. The validity of using $\hat{\xi}$ for this purpose is not affected by the poor performance of $\hat{\xi}$ as a point estimator at the low signal-to-noise levels. Thus, for this purpose, $\hat{\xi}$ is a good choice of a statistic, as it is easy to calculate and has a clear physical interpretation.

References

- (1) A. J. Viterbi "On Coded Phase-Coherent Communications"
I.R.E. Trans. on Space Electronics and Telemetry, Vol. 7,
March, 1961.
- (2) R. W. Rochelle "Pulse Frequency Modulation" NASA Technical
Report, TR R-189, January, 1964.
- (3) J. P. Strong and T. V. Saliga "Comparison of Phase Coherent
and Non-Phase Coherent Coded Communications" -to be presented,
1965 International Space Electronics Symposium, Miami Beach,
Fla., Nov. 2, 3, 4, 1965.
- (4) A. J. Viterbi "Optimum Detection and Signal Selection for
Partially Coherent Binary Communication" IEEE Trans. on
Information Theory, Vol. 11, #2, April, 1965.
- (5) P. M. Woodward "Probability and Information Theory, with
Applications to Radar" Pergamon Press, New York, 1964.
- (6) W. B. Davenport and W. L. Root "Random Signals and Noise"
McGraw-Hill, Inc., New York, 1958.
- (7) R. B. Kerr "Statistical, Structural, and Regression Models
in Linear Process Identification" Proceedings, 1965
Joint Automatic Control Conference, Troy, N. Y.
- (8) A. M. Mood and F. A. Graybill "Introduction to the Theory of
Statistics" McGraw-Hill, Inc., New York, 1963.