

PURDUE UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ANALYSIS AND MINIMIZATION OF
MESSAGE ERROR IN PCM TELEMETRY

By

P. A. WINTZ

A. J. KURTENBACH

December, 1967
Lafayette, Indiana

Supported By:
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D. C.

Under

Grant NsG-553

ACKNOWLEDGEMENTS

We are most grateful to D. R. Anderson, J. C. Hancock, J. F. Hayes, and M. F. Neuts for their helpful discussion and comments during the course of this research.

We thank Phil Miller and Jan Wall for writing the computer programs used to obtain the computed and simulated results presented in this report. And special gratitude goes to Mrs. Sue Koehlinger for her pleasant attitude and patience while preparing the manuscript.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vii
LIST OF SYMBOLS	x
ABSTRACT	xv
CHAPTER I: INTRODUCTION	1
1.1 The Problem	1
1.2 Technical Review	3
1.2.1 History	3
1.2.2 The Telemetry Problem	4
1.2.3 System Model	5
1.2.4 Error Analysis	7
1.3 Approach and Contributions	10
CHAPTER II: PROCESS REPRESENTATION	12
2.1 Karhunen-Loeve Representation	13
2.2 Time Sampling	15
2.3 Hotelling's Principal Components	16
2.3.1 A Digital Filter for Data Compression	18
2.3.2 The Inverse Filter	19
2.4 An Example	21
CHAPTER III: DATA QUANTIZATION	26
3.1 The Optimum Quantizer	30
3.2 The Optimum Uniform Quantizer	32
3.3 Computational Procedures	34
3.4 An Example	36
3.4.1 Determination of the Quantizer Structures	39
3.4.2 Quantizer Performances - Natural Code	42
3.4.3 Quantizer Performances - Gray Code	50
3.4.4 Entropy, Channel Capacity, and Data Rate	53
3.4.5 Quantizer Performances - Nonreal Time Operation	56
3.5 Summary	60

	Page
CHAPTER IV: OVERALL SYSTEM PERFORMANCE - THE TRADEOFF PROBLEM . . .	63
4.1 K.L. Sampling	65
4.1.1 Optimum Bit Assignment	67
4.1.2 Suboptimum Bit Assignments	69
4.1.2.1 Equal Bit Assignments	69
4.1.2.2 Unequal Bit Assignments	70
4.1.3 Comparison of Bit Assignment Procedures	80
4.2 Time Sampling	82
4.2.1 No Data Compression	87
4.2.2 Data Compression	93
4.3 System Comparisons	96
4.4 Information Transmission Rates	98
4.5 Summary	103
CHAPTER V: CONCLUSIONS	105
LIST OF REFERENCES	108
APPENDIX A: RELATIONSHIP BETWEEN APPROXIMATE K.L. SAMPLES AND NORMALIZED PRINCIPAL COMPONENTS	114
APPENDIX B: SIMULATION PROCEDURE FOR ESTIMATING THE APPROXIMATION ERROR OF TIME SAMPLING	121
APPENDIX C: DERIVATION OF THE QUANTIZER DENSITY FUNCTIONS	127
APPENDIX D: DERIVATION OF THE QUANTIZER STRUCTURES	130
D.1 The Optimum Structure	130
D.2 The Optimum Uniform Structure	131
APPENDIX E: GENERATION OF CHANNEL MATRICES FOR USE IN ERROR ANALYSIS	135
APPENDIX F: TABLES 3.1, 3.3, 3.4, and 3.5	144
APPENDIX G: COMPUTER PROGRAMS USED FOR COMPUTING QUANTIZER STRUCTURES	150
G.1 Determining Max's Uniform Quantizer	150
G.2 Determining the Optimum Quantizer	154
G.2.1 Newton's Method	154
G.2.2 Alternate Method	156
G.3 Determining the Uniform Quantizer	156
APPENDIX H: FORMULATION OF THE TRADEOFF PROBLEM AS AN INTEGER NONLINEAR PROGRAMMING PROBLEM	160

LIST OF TABLES

Table		Page
3.1	Max's Nonuniform Quantizer Structure for $m = 2, \dots, 9$	144
3.2	Approximate Average Amount of IBM 7094 Computer Time Required to Determine an Optimum Quantizer Structure	40
3.3	The Optimum Quantizer Structure for PSK Reception and the Natural Code with $m = 2, \dots, 9$	146
3.4	The Quantizing Interval r of Max's Uniform Quantizer for $m = 1, \dots, 9$	149
3.5	The Optimum Quantizing Interval r of the Optimum Uniform Quantizer for PSK Reception and the Natural Code with $m = 5, 6, 7, 8$	149
3.6	Locations of the Error Asymptotes as β Approaches Zero	45
4.1	Optimum Bit Assignments for K.L. Sampler, Optimum Uniform Quantizer, Natural Code, and PSK Reception with $\Gamma = 100$	67
4.2	Mean-Square Error ϵ_*^2 for K.L. Samples, Optimum Uniform Quantizer, Unequal Bit Assignment, Natural Code, and PSK Reception with $\Gamma = 100$	74
4.3	Mean-Integral-Square Error for K.L. Sampler, Optimum Uniform Quantizer, Unequal Bit Assignment, Natural Code, and PSK Reception with $\Gamma = 100$	75
4.4	Mean-Integral-Square Error for K.L. Sampler, Max's Uniform Quantizer, Unequal Bit Assignment, Natural Code, and PSK Reception with $\Gamma = 100$	77
4.5	A Comparison of Bit Assignment Procedures for K.L. Samples Optimum Uniform Quantizer, Natural Code, and PSK Reception with $\Gamma = 100$	82

Appendix
Tables

Page

- A.1 Eigenvalues σ_i^2 of the Integral Equation and $\lambda_i T/N$
of the Matrix Equation for $R(\tau) = e^{-|\tau|}$
- A.2 Comparison of $\varphi_i(t_j)$ and β_{ji} for $i = 1, \dots, n$ and
 $j = 1, \dots, N$

117

118

LIST OF FIGURES

Figure		Page
1.1	Block Diagram of a PCM Telemetry System for Analog Data . . .	2
2.1	Block Diagram of the Sampling Subsystem which Utilizes Data Compression	20
2.2	Performance Curves of Sampling and Reconstruction Procedures	22
2.3	Approximation Waveform Error for Time Sampling	24
3.1	A Symmetric Nonuniform Quantizer ($M = 8$)	30
3.2	A Symmetric Uniform Quantizer for Data Distributed Gaussian, Zero Mean, Unit Variance ($M = 8$)	33
3.3	Probability of Bit Error vs. β with m as a Parameter	38
3.4	Optimum Quantizing Interval r vs. β with m as a Parameter . .	43
3.5	Performance Characteristics of the 2^m th Level Quantizer and the m th Extension of the BSC Using the Natural Code and PSK Reception	44
3.6	Performance Characteristics of the 2^m th Level Quantizer and the m th Extension of the BSC Using the Natural Code and FSK Reception	47
3.7	A Comparison of the Performances of Various Quantizer Structures Used Optimally (Natural Code)	48
3.8	A Comparison of Quantizer Performances for $m = 8$	51
3.9	The m.s.e. ϵ_*^2 vs. r with m as a Parameter (Natural Code, $\beta = 13$ db) .1.	51
3.10	Performance Characteristics of the 2^m th Level Quantizer and the m th Extension of the BSC Using the Gray Code and PSK Reception	52
3.11	A Comparison of Quantizer Performances for Gray Code vs. Natural Code	52

Figure		Page
3.12	Quantizer Entropy, Channel Capacity, and Information Rate for the Optimum Uniform Quantizer	55
3.13	A Comparison of Quantizer Performances under Nonreal Time Conditions	58
3.14	Quantizer Spread vs. m with Δ as the Parameter for Nonreal Time Conditions	59
3.15	The Error Terms: ϵ_q^2 , ϵ_c^2 , and ϵ_{m_1}	62
4.1	A Comparison of System Performances for Equal Bit Assignments	71
4.2	System Performance for K.L. Sampling and Unequal Bit Assignments with $\Gamma = 100, 200$, and 400	79
4.3	A Comparison of Bit Assignments	81
4.4	Block Diagram of Simulated PCM System	85
4.5	Approximation and Quantization Waveform Error for Time Sampling and Optimum Uniform Quantization	88
4.6	Approximation, Quantization and Channel Waveform Error for Time Sampling and Optimum Uniform Quantization	89
4.7	System Performance for Time Sampling and Equal Bit Assignments with $\Gamma = 100, 200$, and 400	92
4.8	Waveform Errors for Time Sampling, Digital Filtering, Equal and Unequal Bit Assignments, and Perfect and Noisy Channels	94
4.9	A Comparison of Different System Performances vs. n with $\Gamma = ST/N_0$ as a Parameter	97
4.10	A Comparison of Data Rate vs. ϵ^2 for both Practical and Theoretical Systems	101
Appendix		
Figures		
B.1	Block Diagram of Simulated Time Sampler and Data Compression System	122
G.1	The Data Deck Set-Up	153
G.2	Flow Diagram for Computing the Optimum Quantizer using Newton's Method	155

Appendix
Figures

Page

G.3	Flow Diagram for Computing the Optimum Quantizer using the Alternate Method	157
G.4	Flow Diagram for Computing Max's r	158
G.5	Flow Diagram for Computing the Optimum r	159

LIST OF SYMBOLS

$\beta = ST/N_o n$	= average channel energy/sample/noise spectral density
β_i	= eigenvectors of covariance matrix
$\beta = [\beta_{ji}]$	= N x N matrix of the N eigenvectors β_i
$\hat{\beta} = [\beta_{ji}]$	= N x n matrix of the first n eigenvector β_i
$r = ST/N_o$	= average channel energy/noise spectral density
$\Delta = ST/N_o m_b$	= average channel energy/channel bit/noise spectral density
$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$	= Kroneker's delta
$\epsilon^2 = \epsilon_a^2 + \epsilon_*^2$	= overall mean-integral-square error
ϵ_a^2	= approximation error
$\epsilon_*^2 = \epsilon_q^2 + \epsilon_c^2 + \epsilon_m$	= remaining system error
ϵ_q^2	= quantizer error only
ϵ_c^2	= channel error only
ϵ_m	= mutual quantizer and channel error
ϵ_s^2	= estimated error for time sampled system
ζ	= Lagrange multiplier
η	= additive noise
η_t^2	= error due to bandlimiting

η_s^2	= integral-square error for simulated sampling system
Λ	= diagonalized covariance matrix
λ_i	= eigenvalues of covariance matrix
u	= mean of random variable
ρ	= correlation between adjacent time samples
σ_i^2	= eigenvalues of integral equation
τ	= variable in correlation function of stationary process
$\Phi = [\frac{T}{\sigma_i N} \varphi_i(t_j)]$	= matrix of time sampled eigenfunctions
φ	= bandwidth parameter
φ	= normal, zero, one density function
$\varphi_i(t)$	= eigenfunctions of integral equation
$\varphi_i(t_j)$	= time samples of $\varphi_i(t)$
$\Psi = [\frac{1}{\sqrt{\lambda_i}} \beta_{ji}]$	= matrix of scaled eigenvectors
$\omega = 2\pi f$	= radian frequency
$A_r^{(m)} = [a_{ij}]$	= m^{th} extension of the channel matrix for an r -ary signal alphabet, $r \geq 3$
a_k	= probability that the quantizer output is v_k .
b_i	= roots of transcendental equations
C_i	= normalizing constant in the $\varphi_i(t)$
d_k	= computed terms in quantizer evaluation
$\varepsilon = ST/m_b$	= average channel energy/channel bit

erf	= error, function
e_k	= computed terms in quantizer evaluation
FSK	= frequency shift keying
f	= frequency, Hz
f_k	= computed terms in quantizer evaluation
g_k	= computed terms in quantizer evaluation
g.l.b.	= greatest lower bound
$k_k(r)$	= computed terms in quantizer evaluation
IID	= independent identically distributed
$K = 1/\sqrt{N/T}$	= normalizing interpolation
l.u.b.	= least upper bound
M	= number of quantization values
m	= number of channel bits/sample for equal bit assignments
m_i	= number of channel bits/ i^{th} sample for unequal bit assignments
m_b	= number of channel bits/analog sample function
N, n_1	= number of time samples/analog sample function
N	= number of points on simulated analog sample function (In Appendix B only)
$N(0,1)$	= Gaussian, zero mean, unit variance, data
N_o	= single sided noise spectral density
n	= number of samples transmitted/analog sample function
$P = [P_{k\ell}]$	= channel transition matrix
PSK	= phase shift keying

p	= probability of channel bit error
$p_{x,y,z}(X,Y,Z)$	= joint density of system variables
$q = 1 - p$	= probability of no channel bit error
R	= data rate (bits/second)
$R(\epsilon^2)$	= rate distortion function
$R(\tau)$	= autocorrelation function
r	= quantizing interval of uniform quantizer
S	= average transmitter power
SGM	= stationary Gaussian Markov
S.D.	= standard deviation
$S(f)$	= spectral density
s^2	= sample variance
T	= time interval of message waveform
T	= transpose of a matrix
t	= time
$u = [u_1, \dots, u_{M+1}]^T$	= quantizer transition levels
$v = [v_1, \dots, v_M]^T$	= quantization values
$W = [w_{k\ell}]$	= weights of channel error words
W_F, W_f	= frequency cutoffs of bandlimited filters
$w = [w_1, \dots, w_N]^T$	= principal components of time samples
$\hat{w} = [w_1, \dots, w_n]^T$	= compressed vector of principal components
$w' = [w'_1, \dots, w'_N]^T$	= normalized value of w
$x = [x_1, \dots, x_n]^T$	= vector of samples at quantizer input

$x_t = [x(t_1), \dots, x(t_N)]^T$	= vector of time samples
$x' = [x'_1, \dots, x'_N]^T$	= samples from discretized K.L. functions
$\bar{x}$	= sample mean
$x(t)$	= input waveform
$x_a(t)$	= approximation of input waveform
$\{x(t)\}$	= input process
$y = [y_1, \dots, y_n]^T$	= vector of quantized samples
$y^* = [y_1^*, \dots, y_n^*]^T$	= vector (in binary) of coded samples
ZOH	= zero order hold
$z = [z_1, \dots, z_n]^T$	= vector of decoded samples
$z^* = [z_1^*, \dots, z_n^*]^T$	= vector (in binary) of received samples
$z_t = [z(t_1), \dots, z(t_N)]^T$	= vector of samples from which output waveform is constructed
$z(t)$	= output waveform
$†$	= pseudo inverse of a matrix
$†, ††$	= footnotes

ABSTRACT

Transmission of telemetry data is a significant application of pulse code modulation (PCM) communication systems. For analog data it is necessary for the system to sample, quantize and code the data before transmission over a digital channel. At the destination a facsimile of the source waveform is reconstructed by the receiver. It differs from the source waveform because of error introduced by the sampler, quantizer, and the noisy channel. To evaluate the effect of these errors on the reconstructed waveform we use the mean-integral-square error criterion.

Our purpose is to minimize this error between sample functions of stationary, second order processes and their facsimiles presented to the user. Our approach is to tie together the existing specialized results adding necessary details and extensions where necessary in order to complete an analysis and optimization of the entire system.

We show analytically that the optimum quantizer is a function of the channel transition matrix and give the necessary conditions for a quantizer structure to be optimum. Suboptimum quantizer structures are also presented. Detailed numerical results are presented for the example of the stationary Gaussian Markov source

with $R(\tau) = e^{-|\tau|}$ and the m^{th} order extensions of binary symmetric memoryless channels.

A data compression scheme using Hotelling's method of principal components on time samples is evaluated by computer simulation. It is found to perform comparably to Karhunen-Loeve sampling which is known to be best for this class of sources. For both of these sampling procedures the samples are found to have unequal variances and hence have varying information content. It is demonstrated that the samples with greater variance should be represented by more channel bits so as to minimize the overall system error. An ad hoc method of obtaining the unequal bit assignments is evaluated and shown to compare favorably to optimum bit assignments obtained by an exhaustive search.

A conventional PCM telemetry system which utilizes time samples, no data compression, and includes a noisy channel is simulated. In this manner an estimate of its overall performance is obtained. This is compared to the performance of the system using a K.L. sampler. Both the system using a K.L. sampler and that using a time sampler followed by a digital filter are found to be superior to the conventional system. For the process considered the mean-integral-square error was decreased by $1/2$ while obtaining a bandwidth reduction of $1/3$.

When a digital filter is inserted after the time sampler it is also necessary to insert an inverse filter at the receiver. It is shown that the optimum filter is determined from the covariance matrix of the time samples.

CHAPTER I

INTRODUCTION

1.1 The Problem

Transmission of telemetry data is a significant application of pulse code modulation (PCM) communication systems. For analog data it is necessary for the system to perform three successive operations on the data before it is transmitted over the digital channel. These are commonly referred to as (i) sampling, (ii) quantizing, and (iii) coding. The received information is decoded and from this data an analog waveform is constructed. A block diagram of a PCM system for telemetering analog signals is shown in Figure 1.1.

The function of the system is to take the analog waveform presented to it by the source and supply a facsimile of it to the user. In order to evaluate the performance of the system, it is necessary to first define a criterion of goodness or index of performance. A meaningful and workable criterion is that the facsimile be close to the source output in the mean-integral-square sense, i.e., the error

$$\epsilon^2 = E\left\{\int_0^T [x(t) - z(t)]^2 dt\right\} \quad (1.1)$$

should be small.

The purpose of this work is not to analyze random processes provided by nature nor to analyze communication channels, but rather to provide a

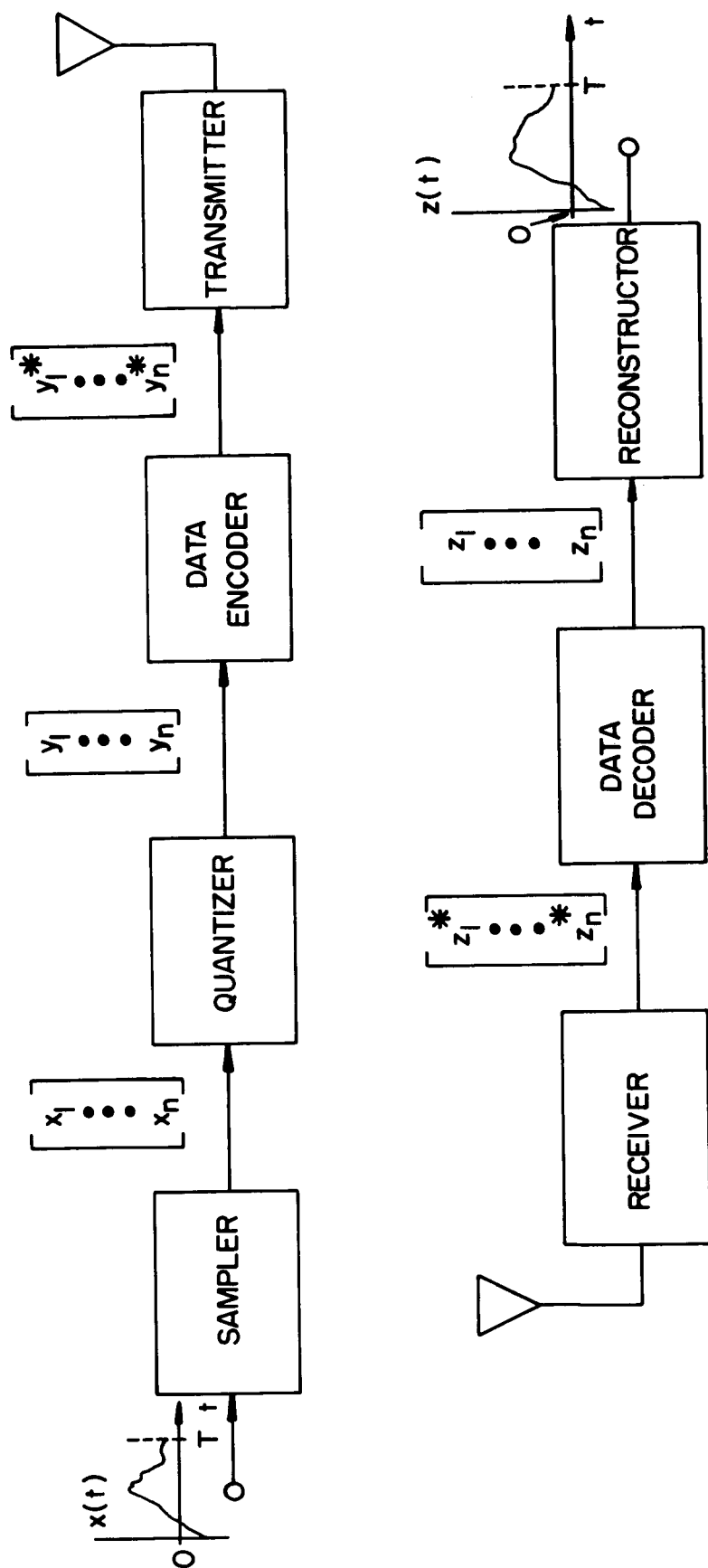


Figure 1.1 Block Diagram of a PCM Telemetry System for Analog Data

connecting link between the process and the channel. Therefore, given descriptions of a process and a channel, the problem is to determine sampling, quantizing and coding procedures which minimize the error in (1.1). Before proceeding with the analysis, let us review previous contributions toward the solution of the problem.

1.2 Technical Review

1.2.1 History. Twenty-nine years have passed since the concept of pulse code modulation (PCM) was patented by Alec H. Reeves [1]. In his patent disclosure Reeves clearly expressed that three successive steps were necessary: sampling, quantizing and coding. However, as explained 25 years later in a philosophical article [2] by the inventor, the invention was ahead of its time, and important, far reaching decisions are now being made which will greatly affect the future of PCM. Interesting reading on the historical background, progress, and some projections regarding PCM as a communications media are found in Reference [2] and an article, "The Philosophy of PCM", by Oliver, Pierce, and Shannon [3]. These articles also include extensive bibliographies.

Descriptions of many operating communications and telemetry systems are available in the literature, both at a theoretical level and a practical level. Two disciplines of electrical engineering which deal extensively with problems involved in implementing PCM systems are statistical communication theory and coding theory. A considerable amount of information regarding the state of the art with respect to operating systems has been presented at various conferences: e.g. Annual National Telemetering Conference [4], Adaptive Telemetry

Conference [5], etc.

1.2.2 The Telemetry Problem. The telemetry problem addresses the task of accurately monitoring a physical process, and accurately and efficiently transferring the information obtained from the point of measurement to some other point in space. Realizing that no information is obtained from a deterministic process, it is implied that sample functions representing the random phenomenon being measured are obtained at the output of the measuring apparatus. These sample functions are then transferred from the source to the user at an established rate, and subject to a certain amount of error.

To discuss the error of the telemetry system, it is necessary to first decide upon a criterion by which to measure the error. Various criteria have been proposed for systems [6-8]. The mean-square error criterion has received the most attention because it is intuitively satisfying and can be handled mathematically. Also, it is known [9] that minimizing the error with respect to the mean-square error criterion provides the minimum for other criteria, especially in the presence of Gaussian statistics. For these reasons, plus others presented later, the mean-integral-square error criterion will be used as the criterion of goodness in this study. However, it is recognized that if the user is to interpret the information directly (e.g., by listening to it or by observing it on a visual display), the best criterion of goodness for the system of interest (PCM telemetry in this case) should be determined through subjective testing. A research group at Massachusetts Institute of Technology, headed by Thomas S. Huang, is presently evaluating error criteria for application to

picture transmission systems [10,11]. Reference [11] contains an exhaustive bibliography of previous contributions. Bell Telephone Laboratories' researchers are active in the analysis of PCM transmission of both speech and pictures [12,13].

The modulation scheme used in the telemetry system is merely a technique for transferring information over a transmission channel. This channel consists of an interval in the frequency spectrum plus the associated transmitting and receiving equipment which represents a capital investment. Hence the user of this channel can minimize his expense by minimizing his demand on the channel while maintaining a specified accuracy. This can be accomplished by efficiently representing the data and properly preparing it for transmission.

If the mean-integral-square error criterion is used as a measure of the system error, it is possible to break the total error into parts associated with the operations of the system on the data sample function.

1.2.3 System Model. The total mean-integral-square error of the system is

$$\epsilon^2 = E\left\{\int_0^T [x(t) - z(t)]^2 dt\right\} \quad (1.1)$$

where $x(t)$, $t \in [0, T]$ is the analog signal presented by the source to the telemetry system and $z(t)$, $t \in [0, T]$ is the analog signal presented by the telemetry system to the user. For the case of real time transmission of an analog signal using an average and peak power limited transmitter, Totty [14], showed that (1.1) can be separated into three additive terms: the representation or approximation error, the quantization or round-off error, and the channel error; i.e.

$$\epsilon^2 = \epsilon_a^2 + \epsilon_q^2 + \epsilon_c^2 \quad (1.2)$$

These errors result from functional operations on the data by the system, as illustrated in Figure 1.1. This additive property is valid when the data in analog form is represented by a generalized Fourier expansion

$$x(t) = \sum_{i=1}^{\infty} x_i \varphi_i(t)$$

where the $\{\varphi_i(t)\}$ ($i = 1, 2, \dots$) is a complete orthonormal basis, and the x_i are the projections of $x(t)$ on the $\varphi_i(t)$, and when the quantization is performed in a manner which minimizes the mean-square quantization error [15,16]. Note that the errors are not independent, as can be seen by considering ϵ_q^2 and ϵ_c^2 . For example, if the number of quantization levels, M , is increased, ϵ_q^2 decreases. However, now the number of binary digits, m , required to represent each level also increases. Therefore, because of the time and power constraints mentioned previously, the energy per transmitted pulse will necessarily decrease, and ϵ_c^2 will increase.

The first unit of the telemetry system, the sampler, provides us with an approximate discrete representation of the sample function, $x(t)$, $t \in [0, T]$. This is accomplished by expressing the sample function as a truncated generalized Fourier expansion

$$x_a(t) = \sum_{i=1}^n x_i \varphi_i(t) \quad (1.3)$$

where

$$x_i = \int_0^T x(t) \varphi_i(t) dt, \quad i = 1, \dots, n \quad (1.4)$$

We see that the x_i ($i = 1, \dots, n$) are continuous random variables which have density functions $p_{x_i}(X_i)$ ($i = 1, \dots, n$) determined by the process

$\{x(t)\}$ and the sequence of linear operations performed on the process.

The sample function is now represented by a vector of dimension n whose components are quantized in the next stage of the system. The effect of quantization is to reduce the state space from an uncountable infinite set of real numbers to a finite set of rational numbers. Hence the components of the quantized data vector, y , are discrete random variables, and the continuous sample function on a continuous state space has now been reduced to a finite set of random variables each constrained to a finite state space.

The next step is to properly prepare this information for its precarious trip through an uncertain channel. This is done by properly coding each component of the data vector and by choosing signals best suited to the channel. These problems are known as the coding and signal design problems. Channel errors cause variations in the components of the decoded vector, z , from those in the quantized vector, y . We note that these errors are themselves discrete random variables while the errors incurred in quantization were continuous random variables.

1.2.4 Error Analysis. If the representation, quantization, and transmission problems are considered separately with no regard for their compatibility in the overall system, then a considerable amount of information is available regarding the nature and magnitude of the individual errors. In a recent article [17] Papoulis reviewed the various errors involved when representing a sample function using time samples. Of special interest are those discussions of the mean-square error of non-bandlimited functions. Articles to this effect appeared

in 1955-1956 [18-20] with Stewart obtaining the optimum non-realizable reconstruction filter. Then in 1960, Spilker [21] determined the optimum reconstruction filter and optimum pre-filter for a system which also filters before sampling. These and the truncated Karhunen-Loeve representation were compared in a recent article [22] for the case of a stationary, normal Markov process. Liff [23] obtained expressions for the mean-square error of various practical reconstructors and signal correlation functions.

The problem of quantizing optimally using the mean-square error criterion has also attracted the attention of several authors [24,15,16]. However, even for the single sample case, the optimum quantizer is difficult to implement, and is not always significantly better than selected sub-optimum quantizers [15]. More practical suboptimum approximations to optimal quantizers have subsequently been discussed [25-27]. Another interesting article [28] considers the quantization of a block or vector of correlated Gaussian random variables. These are first linearly transformed into uncorrelated, hence independent, random variables, and then quantized optimally by the method of Max into 2^{m_i} ($i = 1, \dots, n$) levels for each component of the n dimensional vector. The $\sum_{i=1}^n m_i$ bits are then transmitted over a noiseless channel, and the inverse operations performed at the receiver.

We now come to the channel which has been and still is studied by many researchers, but seldom using a mean-square error criterion. The first discussion of message error introduced by a binary channel and measured in the mean-square sense appeared in the Proceedings of IEE in 1955 [29]. Then in 1962 Bado [30], in a quite different

application -- that of storing analog information in binary form on magnetic tape -- derived some mean-square error expressions for the case of uniform quantization. The topic has received further attention since then [31-36], especially as associated with the transmission of speech and pictures. One important result obtained for input data uniformly distributed over the integers 0 to 2^m-1 is that the natural code minimizes the (total) average noise power [37]. These results were obtained using block codes; however, the case of redundancy coding has also received a small amount of attention [38-39]. One interesting result obtained by Clark and Totty [40] is that redundant coding may actually degrade the performance measured in the mean-square sense while improving it in the minimum probability of word error sense relative to natural block coding.

We have now reviewed individually the effects of representation, quantization, and channel errors. However, in an actual system all of these errors are present and quite possibly interrelated. The first discussion of the joint effects of the error sources was published in 1962 [41] and considered the case of time sampling and uniform quantizing. The optimum linear output filter for matching the system output to the desired output in a mean-square sense was obtained. Adding a further constraint, that the information be sent over a channel of fixed rate, Steiglitz [42] obtained the mean-square error for various reconstruction devices and power spectra of practical interest. Then in a series of letters, Goodman [43-45] discussed the minimization of total mean-square error. He made a significant step in the last article, in that he deviated from time samples and considered ordinary

Fourier coefficients as obtained by bandpass filters.

In an article published in 1964 [46], Viterbi considered the joint effects of quantizing and coding (non-redundant) on band limited signals with uniform amplitude distribution and sampled at the Nyquist rate. Then in his book [47] he presented a family of curves from which, given the channel signal-to-noise ratio and the message bandwidth, one can obtain the number of quantizing levels which optimizes the message signal-to-noise ratio. Finally, Totty [14] considered the simultaneous effects of all the sources of error. However, he does not propose that his expression for mean-integral-square error is minimum, only that the effects are additive in the mean-integral-square sense.

1.3 Approach and Contributions

Our purpose is to minimize the mean-integral-square error between sample functions of stationary, second order processes and their facsimiles presented to the user. Our approach is to tie together the existing specialized results adding necessary details and extensions where necessary in order to complete an analysis and optimization of the entire system. Then we seek suboptimum, but readily implementable, procedures which approach the optimum system in performance.

In Chapter II we discuss the Karhunen-Loeve representation of analog signals as well as the principal components of time samples taken on these signals. We present the digital filter and inverse filter which, used with time samples, minimize the error in (1.1).

In Chapter III we determine the optimum quantizer. It is shown that this quantizer is, in fact, a function of the channel. Hence,

the two must be analyzed together and not separately as has been common in the past. The suboptimum quantizer which has uniformly spaced intervals is determined.

The trade off problem between the number of data samples and number of quantizing intervals is considered in Chapter IV. The problem of determining the optimum trade off is formulated. An algorithm which provides a suboptimum solution is presented.

An example, that of the stationary Markov Gaussian process, is used to illustrate the results. The channels considered in the system are extensions of the binary symmetric channel where the bit errors result from the disturbing effects of additive white Gaussian noise.

CHAPTER II

PROCESS REPRESENTATION

Recall that the first operation performed by the telemetry system illustrated in Figure 1.1 is to convert the input waveform $x(t)$ $t \in [0, T]$ into a vector of finite dimension n . A method of obtaining this representation is to expand the function in a generalized Fourier expansion and then truncate this expansion, i.e.,

$$\begin{aligned}
 x_a &= \sum_{i=1}^n x_i \varphi_i(t) \\
 x_i &= \int_0^T x(t) \varphi_i(t) dt, \quad i = 1, \dots, n \\
 \int_0^T \varphi_i(t) \varphi_j(t) dt &= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}, \quad i, j = 1, \dots, n
 \end{aligned} \tag{2.1}$$

The waveform reconstructor in Figure 1.1 must construct a waveform $z(t)$, $t \in [0, T]$, from the samples z_i ($i = 1, \dots, n$) provided at the output of the decoder. Let us define this operation as

$$z(t) = \sum_{i=1}^n z_i \varphi_i(t) \tag{2.2}$$

where the φ_i ($i = 1, \dots, n$) are those of (2.1). Now the approximation error, that error due to the sampler, is defined as

$$\epsilon_a^2 = E \left\{ \int_0^T [x(t) - x_a(t)]^2 dt \right\} \tag{2.3}$$

Totter [14] showed that, under definitions (2.1), (2.2), and (2.3) the

total error (1.1) separates into two additive terms

$$\epsilon^2 = \epsilon_a^2 + \epsilon_*^2 \quad (2.4)$$

where

$$\epsilon_*^2 = E\left\{ \sum_{i=1}^n (x_i - z_i)^2 \right\} \quad (2.5)$$

The error ϵ_*^2 , which is due to the quantizer and channel, will be discussed in Chapter 3.

2.1 Karhunen-Loeve Representation

An orthogonal expansion with extremely useful properties exists for random processes which have finite and continuous covariance functions. These processes, known as second order processes, are discussed with rigor in Chapter 10 of Loeve's book [48]. Let us consider stationary second order processes with sample functions centered about their expectations, $E x(t)$, and let $E x(t) = 0$.

The orthogonal expansion, commonly referred to as the Karhunen-Loeve (K.L.) expansion, of a sample function $x(t)^\dagger$, $t \in [0, T]$, from such a process is

$$x(t) = \sum_{i=1}^{\infty} \sigma_i x_i \varphi_i(t) \quad (2.6)$$

with

$$\begin{aligned} E x_i x_j &= \delta_{ij}, & i, j &= 1, 2, \dots \\ \int_0^T \varphi_i(t) \varphi_j(t) dt &= \delta_{ij} & i, j &= 1, 2, \dots \end{aligned} \quad (2.7)$$

where the σ_i^2 are the eigenvalues and the $\varphi_i(t)$ are the eigenfunctions

[†]We will only consider real processes.

of the integral equation

$$\sigma_i^2 \varphi_i(t) = \int_0^T R(t,s) \varphi_i(s) ds \quad (2.8)$$

The x_i are the normalized projections

$$x_i = \frac{1}{\sigma_i} \int_0^T x(t) \varphi_i(t) dt \quad (2.9)$$

An approximation to $x(t)$ can now be obtained by truncating the expansion in (2.6)

$$x_a(t) = \sum_{i=1}^n x_i \sigma_i \varphi_i(t) \quad (2.10)$$

The approximate waveform (2.10) can be made as close to $x(t)$ as desired. This is due to the mean square convergence properties of the K.L. expansion and can be expressed as

$$[x(t) - \sum_{i=1}^n x_i \sigma_i \varphi_i(t)] \rightarrow 0 \text{ as } n \rightarrow \infty \quad (2.11)$$

uniformly on $[0, T]$. The approximation error ϵ_a^2 for the K.L. representation of a second order stationary process, obtained from (2.6), (2.10) and (2.3) by appealing to Mercer's theorem[†] is

$$\epsilon_a^2 = R(0) - \sum_{i=1}^n \sigma_i^2 \quad (2.12)$$

This representation has numerous advantages some of which are listed below.

1. The random variables x_i are uncorrelated all with zero mean and unit variance. Moreover, if the process is Gaussian, then they are independent.

[†]See Appendix 2 of Reference [6].

2. The basis functions are orthogonal and, therefore, it may be possible to obtain the samples by suitable filters.

3. By ordering the eigenvalues of (2.8) in a nonincreasing manner the approximation error ϵ_a^2 will always be minimum for any given n .

4. Of all sequences of functions $\varphi_i(t)$ ($i = 1, \dots, n$) defined, continuous, and orthonormal on $[0, T]$, those obtained from (2.8) minimize the approximation error ϵ_a^2 [49].

The K. L. expansion of the stationary Gaussian Markov (SGM) process with covariance function

$$R(\tau) = e^{-|\tau|}, \quad -\infty < \tau < \infty$$

has been obtained [6, 50]. Its performance is compared to those for other representation schemes in Section 2.4.

2.2 Time Sampling

In the previous section it was mentioned that filters might be realized to obtain K.L. samples. However, there are many practical difficulties involved in realizing these filters. For this reason it is common place to obtain the finite vector of samples at the output of the sampler by using time sampling. The principles of time sampling are well known and will not be discussed here. Recall, though, that no one has been able to analyze the system of Figure 1.1 under the mean-integral-square error criterion for time sampling of non-bandlimited functions.

The error

$$\epsilon_s^2 = E \int_0^T [x(t) - z(t)]^2 dt \quad (2.13)$$

incurred when time sampling sample functions from the SGM process and reconstructing $z(t)$ by linear interpolation (LI), by step functions (zero order hold ZOH), and by $\sin(\pi x)/\pi x$ functions was estimated by computer simulation. It is presented in Section 2.4 for comparative purposes.

2.3 Hotelling's Principal Components

Most processes of nature which are of interest to us are processes with a certain amount of memory. This is manifested in their statistical properties by nonzero cross correlations between the time samples taken on a sample function of the process. Ways of utilizing this relationship between time samples to somehow reduce the total number of samples required to represent a sample function are eagerly sought [51]. Schemes which accomplish this end are sometimes referred to as data compression or redundancy removal techniques. Typically, these methods operate in real time and utilize prediction techniques.

For applications where a fixed delay before receiving the samples is allowable a technique due to Hotelling [52] might be employed. This method, sometimes called the method of principal components [53], utilizes linear transformations to obtain uncorrelated random variables in a manner analogous to the Karhunen-Loeve expansion of sample functions.

Consider the case where the sample function $x(t)$, $t \in [0, T]$ of a zero mean random process is converted to an N dimensional vector by taking time samples at uniform intervals of T/N seconds. And

where the $N \times N$ covariance matrix Σ of the random vector is known.

For this case Hotelling's principal components are

$$w_j = \sum_{i=1}^N x(t_i) \beta_{ij} \quad (2.14)$$

where the β_{ij} are the entries of $\beta = (\beta_{ij})$, the orthogonal matrix which diagonalizes the covariance matrix, i.e.

$$\Lambda = \beta^T \Sigma \beta \quad (2.15)$$

The eigenvalues of the covariance matrix are the entries of the diagonal matrix Λ and the eigenvectors of the covariance matrix are the columns of β . It is assumed that the eigenvalues are ordered in a nonincreasing manner. The covariance matrix of the new samples is

$$\begin{aligned} E(w w^T) &= E(\beta^T x_t x_t^T \beta) \\ &= \beta^T \Sigma \beta = \Lambda \end{aligned} \quad (2.16)$$

where x_t is the N dimensional vector of time samples $x(t_i)$ ($i = 1, \dots, N$). From (2.16) we see that the w_i ($i = 1, \dots, N$) are uncorrelated and ordered in terms of decreasing variance.

The properties of the principal components obtained by (2.14) are shown compactly in (2.15) and (2.16). These properties expressed in an expanded form are

$$\begin{aligned} \beta_j^T \beta_k &= \delta_{jk}, & j, k &= 1, \dots, N \\ E\left(\frac{1}{\sqrt{\lambda_j} \sqrt{\lambda_k}} w_j w_k\right) &= \delta_{jk}, & j, k &= 1, \dots, N \end{aligned} \quad (2.17)$$

where β_j is the j^{th} eigenvector of the covariance matrix Σ and hence the j^{th} column of β and λ_j is the j^{th} diagonal entry of the matrix Λ .

Note the similarity between (2.7) and (2.17). Another point of similarity is between a time sampled version of (2.9), vis.

$$x_i' = \frac{1}{\sigma_i} \sum_{j=1}^N x(t_j) \varphi_i(t_j) (t_j - t_{j-1}) \quad (2.18)$$

and a normalized version of (2.14), vis.

$$w_i' = \frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^N x(t_j) \beta_{ji} \quad (2.19)$$

The relationship between the weights, $\varphi_i(t_j)/\sigma_i$ and $\beta_{ji}/\sqrt{\lambda_i}$, used to obtain the samples x_i' and w_i' is investigated in Appendix A. They are shown to be equivalent for N large.

2.3.1 A Digital Filter for Data Compression.[†] Since the information associated with a random variable is proportional to its variance, we see that the w_i provide less information as i increases. We then reason that, since it costs us in terms of channel requirements to send data samples, we will discard those w_i for $i > n$ where n is an arbitrary positive integer less than N . Call this n dimensional vector $\hat{w}$. It is obtained by projecting x_t on the n dimensional space generated by the first n columns of β . Call this projection matrix $\hat{\beta}$, then

$$\hat{w} = \hat{\beta}^T x_t \quad (2.20)$$

The matrix $\hat{\beta}$ defines the weights of the digital filter which compresses the data of the N component random vector x_t into the $n < N$ component random vector $\hat{w}$.

[†]For another discussion of digital filtering see Reference [22]

The covariance matrix of the random vector $\hat{w}$ is

$$\begin{aligned} E \hat{w} \hat{w}^T &= E \hat{\beta}^T x_t x_t^T \hat{\beta} \\ &= \beta^T \Sigma \beta = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix} \end{aligned} \quad (2.21)$$

Since the eigenvalues are ordered as

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq \dots \geq \lambda_N \quad (2.22)$$

we see that, for any n , the w_i ($i = 1, \dots, n$) have maximum variance of all selections of n of the random variables w_i $i \in \{1, \dots, N\}$. It is shown in [53] that the sum of the variances of the principal components is equal to the sum of the variances of the original random variables. Therefore, considering the transformation as a spectral decomposition of the input random variables, we see that the energy discarded is

$$\sum_{i=n+1}^N \lambda_i = NR(0) - \sum_{i=1}^n \lambda_i \quad (2.23)$$

If we divide through (2.23) by N we have the m.s.e. of the transformation [22].

2.3.2. The Inverse Filter. To avoid confusion let us review the problem at this point. We have at the input a sample function $x(t)$, this is time sampled to obtain the random vector x_t which is in turn transformed into the random vector $\hat{w}$ of dimension less than x_t . We must now supply an inverse filter to provide the vector z_t of dimension N from which the function $z(t)$ can be obtained. These operations are illustrated in Figure 2.1. When in actual operation, a quantizer and channel will be inserted between the digital filter

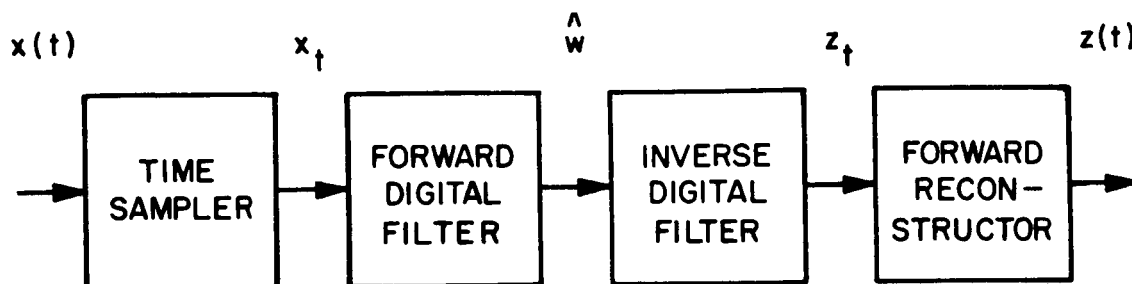


Figure 2.1 Block Diagram of the Sampling Subsystem which Utilizes Data Compression

and the inverse filter. If no data compression is realized, then obviously the best inverse filter is β . This would give

$$z_t = \beta w = \beta \beta^T x_t = x_t \quad (2.24)$$

However for data compression the inverse of the matrix $\hat{\beta}$ does not exist. We might ask for that inverse filter Γ which minimizes the mean square error between x_t and z_t , i.e., we want that Γ_o such that[†]

$$\min_{\Gamma} \|x_t - \Gamma \hat{\beta}^T x_t\|^2 = \|x_t - \Gamma_o \hat{\beta}^T x_t\|^2 \quad (2.25)$$

where

$$\|v\|^2 = E(v, v) = E \sum_{i=1}^N v_i^2$$

A slightly different but equivalent approach to the problem is:

Given the n tuple w and the $n \times n$ matrix $\hat{\beta}^T$ we ask for that x_o such that

[†]At this point it is advantageous to view the random variables as points in a Hilbert space and use algebraic notation. (See Appendix C of [54].)

$$\min_{x_t} \|\hat{\beta}^T x_t - w\|^2 = \|\hat{\beta}^T x_o - w\|^2 \quad (2.26)$$

or equivalently

$$\min_{x_t} \|\hat{\beta} \hat{\beta}^T x_t - \hat{\beta} w\| = \|\hat{\beta} \hat{\beta}^T x_o - \hat{\beta} w\| \quad (2.27)$$

It is known that^{††}

$$x_o = (\hat{\beta} \hat{\beta}^T)^{\dagger} \hat{\beta} w \quad (2.28)$$

where $A^{\dagger}$ is the psuedo inverse of the matrix A . Using a theorem of psuedo inverses [54,55], we have

$$x_o = \hat{\beta} \hat{\beta}^T \hat{\beta} w = \hat{\beta} w \quad (2.29)$$

Hence $\Gamma_o = \hat{\beta}$ is the desired inverse filter.

In Section 2.4 the performance of this technique is computed for the SGM process and compared to the performances of the other techniques discussed in the preceeding sections.

2.4 An Example

As an example to illustrate the sampling techniques discussed in the previous sections let us consider the stationary Gaussian Markov process with covariance function

$$R(\tau) = e^{-|\tau|}, \quad -\infty < \tau < \infty \quad (2.30)$$

This process, commonly referred to as the Ornstein-Uhlenbeck process, has been used to illustrate sampling and reconstruction techniques in previous works [20,22,23]. The approximation error obtained by various sampling and reconstruction techniques is shown in Figure 2.2

^{††}See Section 17, Appendix C of [54].

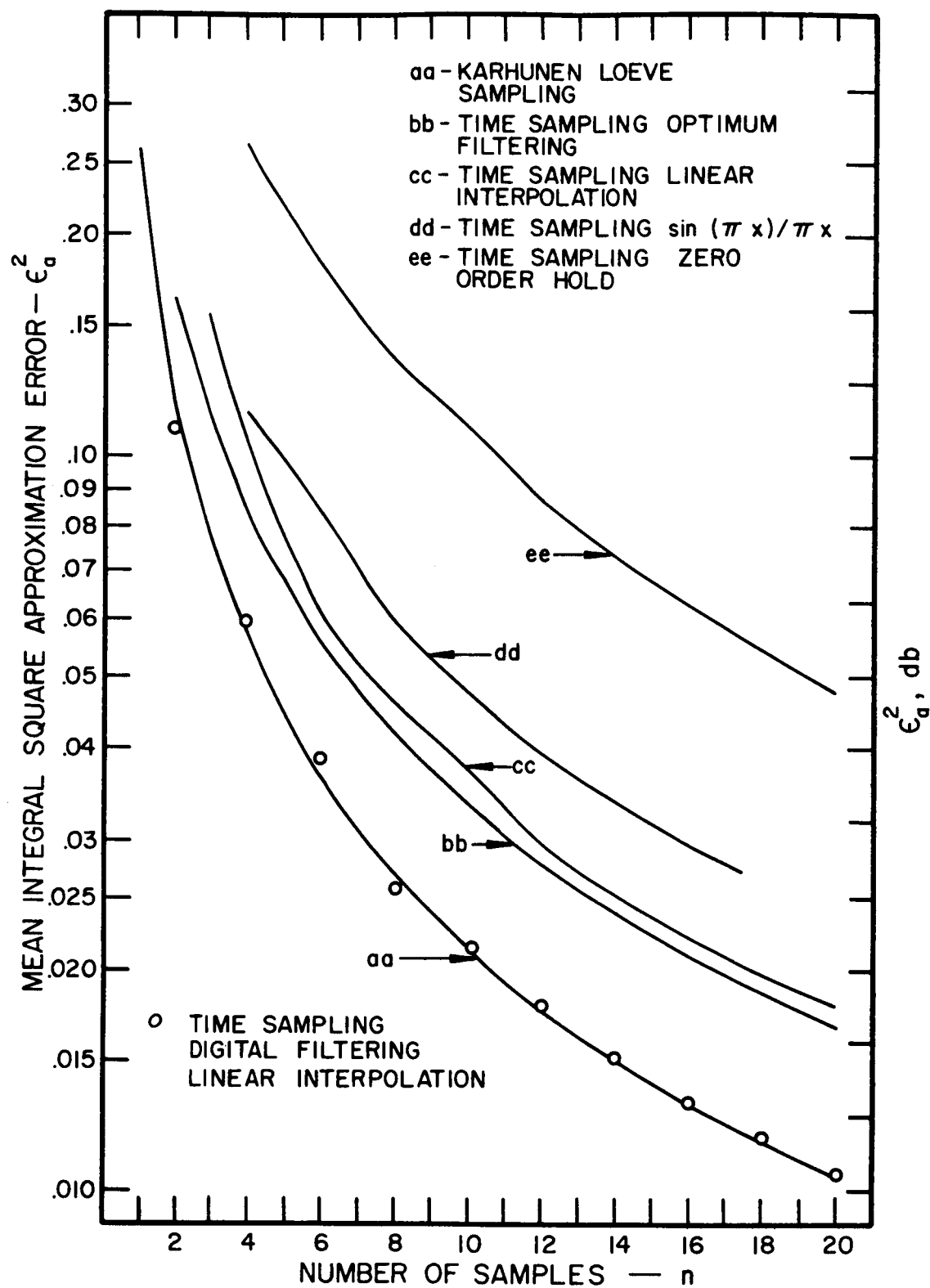


Figure 2.2 Performance Curves of Sampling and Reconstruction Procedures

Curve aa is a plot of the error for K.L. sampling as given by (2.12).

The eigenvalues and eigenfunctions for the process are given in Appendix A. Theoretical results are also used to plot Curve bb.

The equation for this curve, derived for time sampling and waveform reconstruction by the optimum nonrealizable filter by Stewart [20], is

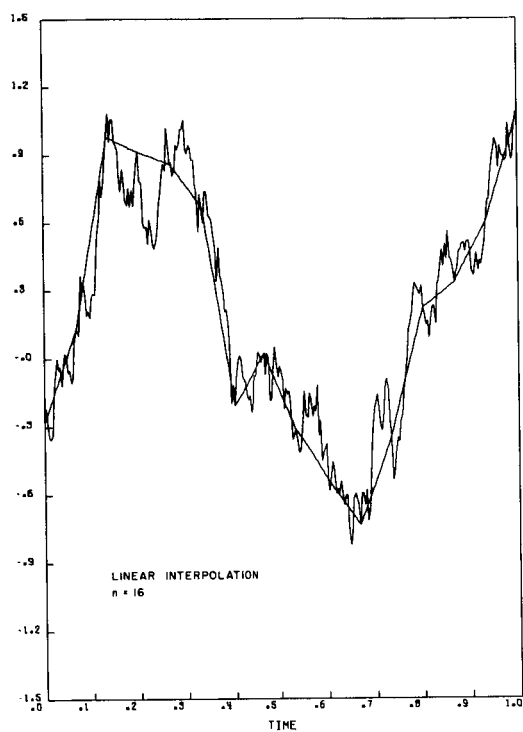
$$\epsilon_a^2 = \coth(T/n) - n/T \quad (2.31)$$

with $T = 1$.

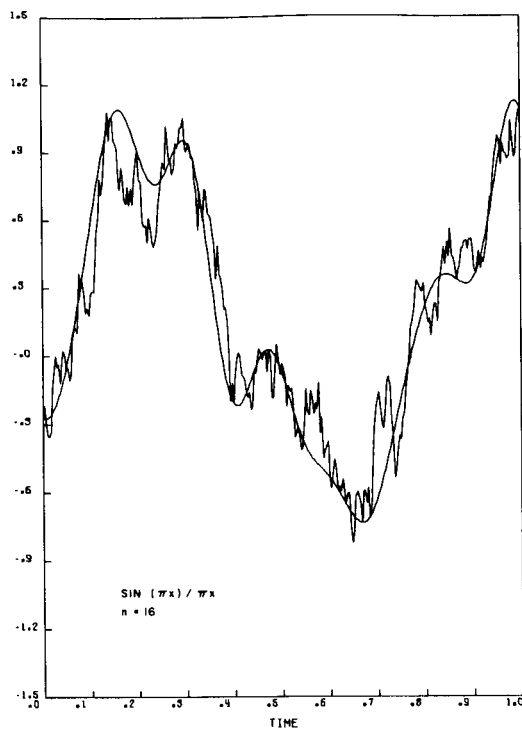
The remaining four curves in Figure 2.2 were obtained by computer simulation. A description of the simulation procedure is presented in Appendix B. Note that for no data compression, waveform reconstruction by linear interpolation is superior to that by $\sin(\pi x)/\pi x$ functions. This is so primarily because the linear interpolation technique takes advantage of the correlation between adjacent samples. Both linear interpolation and $\sin(\pi x)/\pi x$ are superior to the step function or zero order hold technique.

Finally note the excellent performance of the data compression technique illustrated in Figure 2.1 and discussed in Section 2.3. From the small circles in Figure 2.2 we see that this technique performs nearly as well as Karhunen-Loeve sampling. However, the hardware necessary to implement it is realizable today with inexpensive products of solid state technology, while the analog filters necessary to implement the K.L. technique are rather difficult to obtain.

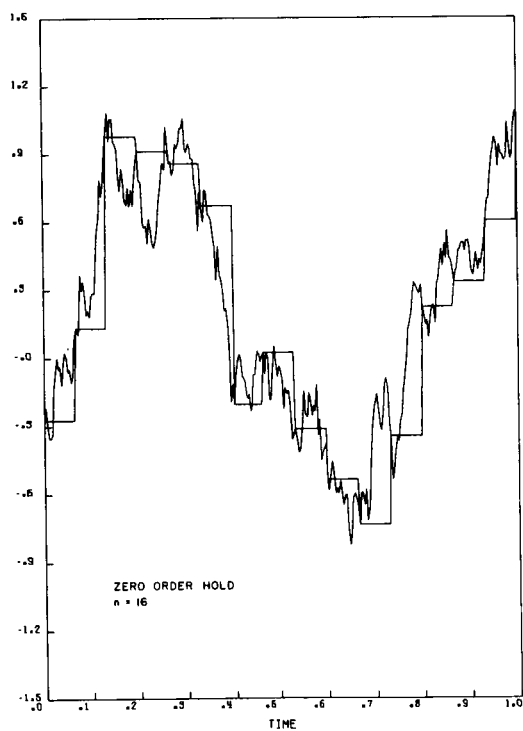
In order to obtain a better feel for the reconstruction techniques, the output waveforms obtained for a realization of the input random function are presented in Figures 2.3a, b, and c for LI, ZOH, and



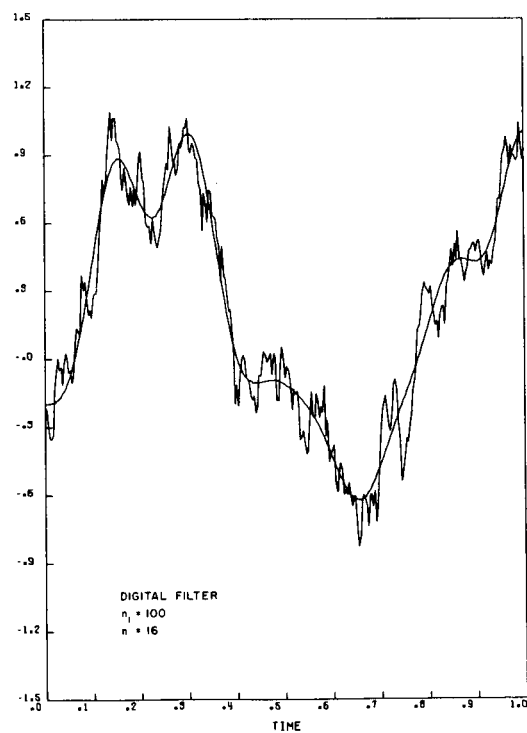
(a)



(c)



(b)



(c)

Figure 2.3 Approximation Waveform Error for Time Sampling

$\sin(\pi x)/\pi x$ respectively. The input waveform is obtained from a dense grid of 300 correlated points of which 16 are chosen as the time samples. We note by inspection of this figure that waveform reconstruction by linear interpolation is uniformly better than by zero order hold. This is not true for reconstruction by $\sin(\pi x)/\pi x$ functions. In fact, the error, for a single function, given by

$$\epsilon_1^2 = \int_0^T [x(t) - z(t)]^2 dt \quad (2.32)$$

was occasionally smaller for $\sin(\pi x)/\pi x$ functions than for linear interpolation.

Figure 2.3d illustrates the output waveform obtained by the system in Figure 2.1 for the same input random function. The input waveform is again a dense grid of 300 points of which 100 are selected as time samples. These are in turn reduced to 16 samples by the digital filter. For the input waveform of Figure 2.3 the integral-square approximation error (See Equation (2.32)) for the systems considered is

Digital filter with linear interpolation	.01291
Linear interpolation	.02232
$\sin(\pi x)/\pi x$	.02659
Step function or zero order hold	.06673

The fact that the reconstructed waveform in Figure 2.3d is so close to the input waveform is consistent with the performance curve presented in Figure 2.2.

CHAPTER III

DATA QUANTIZATION

In the previous chapter the error incurred when reducing a sample function of a random process to a finite vector of continuous random variables was discussed. It was pointed out in (2.4) that by taking the samples as projections on a set of orthonormal basis functions, the overall mean-integral-square error separated into two additive terms, i.e.

$$\epsilon^2 = \epsilon_a^2 + \epsilon_*^2 \quad (3.1)$$

The error ϵ_a^2 is due to the sampler alone and the error ϵ_*^2 is due to the remainder of the system. For the error ϵ_*^2 the mean-integral-squared error criterion reduces to the mean square error (m.s.e.) criterion, and we have

$$\epsilon_*^2 = E\left\{ \sum_{i=1}^n (x_i - z_i)^2 \right\} \quad (3.2)$$

Note that the error ϵ_*^2 is dependent upon the approximation scheme, but the approximation error ϵ_a^2 is independent of the remainder of the system. Therefore, for a given sampling procedure, the total error ϵ^2 is minimized by minimizing ϵ_*^2 .

Equation (3.2) can be expanded into the form

$$\epsilon_*^2 = \epsilon_q^2 + \epsilon_c^2 + \epsilon_m \quad (3.3)$$

where the quantization error ϵ_q^2 , the channel error ϵ_c^2 , and the mutual

error ϵ_m , expressed in terms of the variables shown in Figure 1.1, are

$$\epsilon_q^2 = E\left\{ \sum_{i=1}^n (x_i - y_i)^2 \right\} \quad (3.4)$$

$$\epsilon_c^2 = E\left\{ \sum_{i=1}^n (y_i - z_i)^2 \right\} \quad (3.5)$$

and

$$\epsilon_m = 2E\left\{ \sum_{i=1}^n (x_i - y_i)(y_i - z_i) \right\} \quad (3.6)$$

Totty and Clark [14] noted that if the quantizer structure is that proposed by Max [15], then the mutual error term ϵ_m is zero.

Max [15] gave necessary conditions for the quantizer structure which minimizes the quantization error ϵ_q^2 for one sample ($n = 1$) and a specified probability density $p_{x_1}(X_1)$. Fleischer [56] further showed that this quantizer is unique for data with Gaussian, Laplacian, or Rayleigh probability densities. Huang and Schultheiss [23] present a nonsingular transformation matrix which transforms a data vector $(x_1, \dots, x_n)$ of correlated zero mean, Gaussian random variables into a vector $(w_1, \dots, w_n)$ of uncorrelated Gaussian random variables. These variables are then quantized.

The m.s.e. due to bit errors at the receiver has been obtained for equally spaced quantizing levels and uniformly distributed input data ([29], [31], [35], [36]). Viterbi [47] determined both the quantization error and channel error for the case of a uniformly distributed input random variable and uniform (equally spaced) quantizing levels. For the case of uniformly distributed input data, the mutual error term in (3.3) is zero. However, in general, this is not true and it is necessary to consider all three of the terms in (3.3).

Since for a finite sum, the expectation and summation operators

commute, we write (3.2) as

$$\epsilon_*^2 = \sum_{i=1}^n E[(x_i - z_i)^2] \quad (3.7)$$

We can now proceed with the minimization of (3.7) by minimizing each of the terms and summing to obtain the error for the n samples which represent the waveform $x(t)$, $t \in [0, T]$. In some cases of interest the x_i ($i = 1, \dots, n$) are independent random variables. Three such cases are as follows:

- 1) When the expansion of Karhunen and Loeve[†] is used to represent the sample function and the resulting samples are normalized.
- 2) When the transformation of Huang and Schultheiss [28] is applied to correlated Gaussian random variables.
- 3) When (approximately) a Gaussian process is sampled at a rather slow rate so that the samples are only slightly correlated.

For independent, equal variance samples all n terms of (3.7) will be the same, and we have

$$\epsilon_*^2 = n E[(x_1 - z_1)^2] \quad (3.8)$$

Therefore, in this chapter we determine the quantizer structure which minimizes the m.s.e. for a single sample, i.e., we minimize

$$\epsilon_{*i}^2 = E[(x_i - z_i)^2] \quad (3.9)$$

We also determine the optimum uniform quantizer and make detailed comparisons between the error obtained using these structures and that obtained using Max's nonuniform and uniform quantizers. These comparisons are made for systems which have additive, white, Gaussian noise as the channel disturbance and utilize

[†]See Davenport and Root [6] Chapter 6, Section 4.

- 1) Coherent reception with anticorrelated signals (PSK).
- 2) Noncoherent reception with orthogonal signals (FSK).

Average and peak power constraints are placed on the transmitter, and the system is assumed to operate in real time.

The data encoder (See Figure 1.1) merely represents the quantization values v_ℓ ($\ell = 1, \dots, M$) in terms of binary variables. The use of redundancy encoding to reduce the average number of bit errors at the receiver can either increase or decrease the m.s.e. relative to uncoded data. Clark and Tott [40] give an example in which a Hamming code decreases the m.s.e. for sufficiently high signal-to-noise ratios, but increases the m.s.e. for low signal-to-noise ratios.

The two source encoding schemes considered are the natural code and the reflected binary or Gray code. For the natural code

v_1 is coded as 00 ... 00
 v_2 is coded as 00 ... 01
 v_3 is coded as 00 ... 10
 etc.

While for the Gray code

v_1 is coded as 00 ... 000
 v_2 is coded as 00 ... 001
 v_3 is coded as 00 ... 011
 v_4 is coded as 00 ... 010
 v_5 is coded as 00 ... 110
 v_6 is coded as 00 ... 111
 etc.

Hence, if for a particular realization of x_i , $y_i = v_3$, then using the natural code $y_i^* = 00 \dots 10$. If, for this realization, no bit errors

are made at the receiver, then $z_i^* = 00 \dots 10$ and $z_i = v_3$.

3.1 The Optimum Quantizer

The quantizer in the telemetry system shown in Figure 1.1 operates in the following manner. A sample x_i , assumed to be a continuous random variable from a continuous probability density, is mapped into one of a finite set $\{v_1, \dots, v_M\}$ of rational numbers. As illustrated in Figure 3.1, this mapping follows the procedure

$$u_j < x_i \leq u_{j+1} \rightarrow y_i = v_j \quad (j = 1, \dots, M)$$

The u_j ($j = 2, \dots, M$) are the transition levels, with u_1 and u_{M+1} the greatest lower bound and least upper bound, respectively, to the input signal, and the v_j ($j = 1, \dots, M$) are the quantization values. The channel permutes the transmitted work into another work of the same set with probability

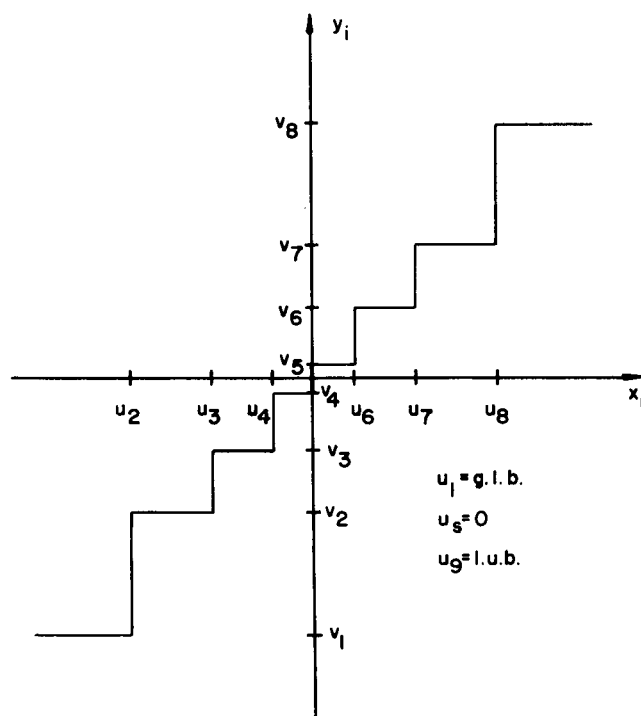


Figure 3.1 A Symmetric Nonuniform Quantizer ($M = 8$)

given by the channel transition matrix $P = [P_{k\ell}]$, where $P_{k\ell} = \text{Prob}[z_i = v_\ell | y_i = v_k]$, ($k, \ell = 1, \dots, M$). Note that this matrix is a stochastic matrix, i.e.

$$P_{k\ell} \geq 0, \quad k, \ell = 1, \dots, M$$

$$\sum_{\ell=1}^M P_{k\ell} = 1, \quad k = 1, \dots, M$$

To obtain a numerical value for the error ϵ_*^2 , it is necessary to carry out the expectation operation indicated in (3.9)[†]. Using the probability density functions obtained in Appendix C we have^{††}

$$\begin{aligned} E(x_i - z_i)^2 &= E(x_i^2) - 2E(x_i z_i) + E(z_i^2) \\ &= \int_{-\infty}^{\infty} x_i^2 p_{x_i}(x_i) dx_i \\ &\quad - 2 \sum_{\ell=1}^M v_\ell \sum_{k=1}^M P_{k\ell} \int_{u_k}^{u_{k+1}} x_i p_{x_i}(x_i) dx_i \\ &\quad + \sum_{\ell=1}^M v_\ell^2 \sum_{k=1}^M P_{k\ell} \int_{u_k}^{u_{k+1}} p_{x_i}(x_i) dx_i \end{aligned} \tag{3.10}$$

We note that the m.s.e. is a function of (i) the probability density, $p_{x_i}(x_i)$, (ii) the channel matrix, $P = [P_{k\ell}]$, (iii) the transition levels, u_k ($k = 2, \dots, M$), (iv) the quantization values, v_ℓ ($\ell = 1, \dots, M$), and (v) the number of these values, M . For a given probability density, channel matrix, and number of quantization values, we desire those transition levels and quantization values which minimize the m.s.e., $E[(x_i - z_i)^2]$.

[†]The error ϵ_*^2 is also evaluated in Totty and Clark [14]. However, they expand ϵ_*^2 as $\epsilon_*^2 = \epsilon_{a_i}^2 + \epsilon_{q_i}^2 + \epsilon_{m_i}^2$ and evaluate the three terms on the right-hand side.

^{††}The lower case letters x_i, y_i, z_i denote random variables while the upper case letters X_i, Y_i, Z_i denote real variables.

This quantizer structure, obtained in Appendix D, is

$$v_j = \frac{\sum_{k=1}^M P_{kj} \int_{u_k}^{u_{k+1}} x_i p_{x_i}(X_i) dX_i}{\sum_{k=1}^M P_{kj} \int_{u_k}^{u_{k+1}} p_{x_i}(X_i) dX_i}, \quad j = 1, \dots, M \quad (3.11)$$

$$u_j = \frac{\frac{1}{2} \sum_{\ell=1}^M v_\ell^2 (P_{j\ell} - P_{j-1\ell})}{\sum_{\ell=1}^M v_\ell (P_{j\ell} - P_{j-1\ell})}, \quad j = 2, \dots, M \quad (3.12)$$

For a noiseless channel, the channel matrix becomes the identity matrix or Kronker's delta, $\delta_{k\ell}$, and (3.11) and (3.12) reduce to

$$v_j = \frac{\int_{u_j}^{u_{j+1}} x_i p_{x_i}(X_i) dX_i}{\int_{u_j}^{u_{j+1}} p_{x_i}(X_i) dX_i}, \quad j = 1, \dots, M \quad (3.13)$$

$$u_j = \frac{1}{2} (v_j + v_{j-1}), \quad j = 2, \dots, M \quad (3.14)$$

We note that these results are precisely those obtained by Max [15]

3.2 The Optimum Uniform Quantizer

We next consider a suboptimum quantizer which has the advantage that the quantization values and transition levels are uniformly spaced which makes the quantizer easier to realize. We call this uniform distance between the transition levels, the quantizing interval r , and require that the quantization values be at the midpoint in the interval. This is illustrated for the Gaussian distribution with zero mean and unit variance in Figure 3.2. In this case, instead of having a $2M-1$ dimensional problem, we have a 2 dimensional problem. We need only determine the quantizing interval r and any particular quantization

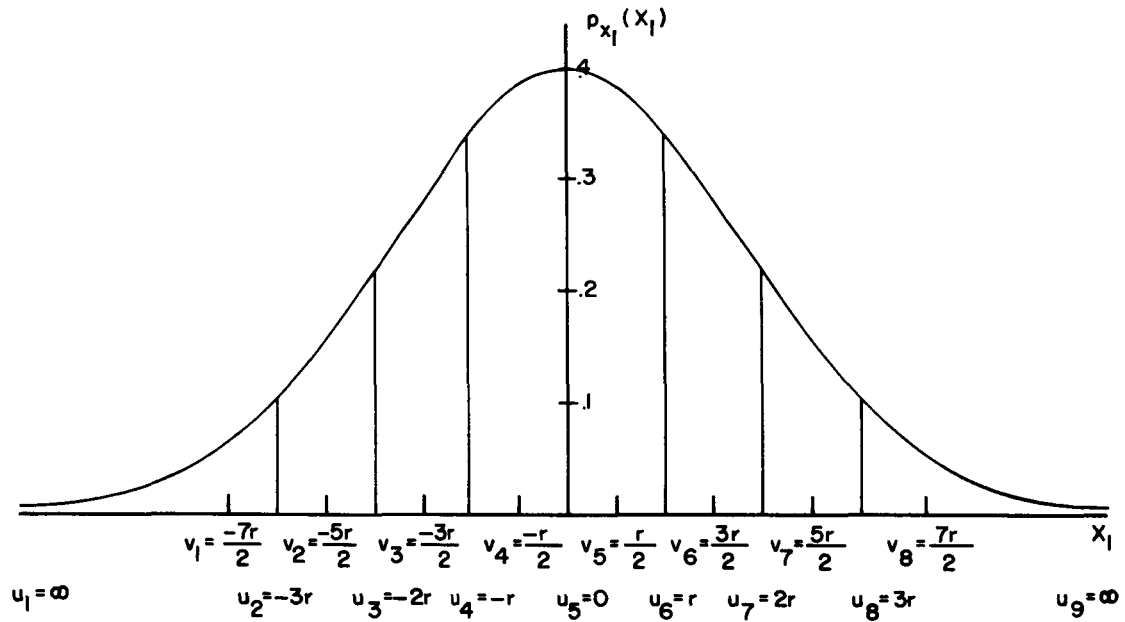


Figure 3.2 A Symmetric Uniform Quantizer for Data Distributed Gaussian, Zero Mean, Unit Variance ($M = 8$)

value y_j . If the input probability density $p_{X_i}(X_i)$ is symmetric and a symmetric answer is desired, then the problem is one dimensional. For a symmetric probability density and M , the number of quantization values, even, we have

$$\begin{aligned}
 v_\ell &= \left(\ell - \frac{M+1}{2}\right)r, \quad \ell = 1, \dots, M \\
 u_k &= \left(k - 1 - \frac{M}{2}\right)r, \quad k = 2, \dots, M \\
 u_1 &= \text{greatest lower bound on } X_i \\
 u_{M+1} &= \text{least upper bound on } X_i
 \end{aligned} \tag{3.15}$$

In Appendix C equation (C.6), obtained by substituting (3.15) into (3.10), is presented. This equation again reduces to that of Max for the Perfect channel. An implicit expression from which r can be

obtained by search techniques is also derived. In the next section we present methods for obtaining the u and v vectors of (3.11) and (3.12) and the optimum interval r of (D.6).

3.3 Computational Procedures

To solve (3.13) and (3.14) Max [15] suggested an iterative scheme in which one assumes a v_1 then obtains the remaining u_j ($j = 2, \dots, M$) and v_j ($j = 2, \dots, M$) by alternating between (3.13) and (3.14). After v_M has been determined from (3.14), the assumption on v_1 is checked by using (3.13). If the magnitude of the difference

$$\left| v_M - \frac{\int_{u_M}^{u_{M+1}} X_i p_{X_i}(X_i) dX_i}{\int_{u_M}^{u_{M+1}} p_{X_i}(X_i) dX_i} \right| \quad (3.16)$$

is too great, a new v_1 is taken and the process is repeated.

The solution of (3.11) and (3.12) is slightly more complicated. We note, however, that a certain separability exists in that the v_j ($j = 1, \dots, M$) are only dependent upon the u_j ($j = 2, \dots, M$) and not the remaining v_k ($k = 1, \dots, j-1, j+1, \dots, M$). The same is true for the u_j ($j = 2, \dots, M$). Two methods for solving (3.11) and (3.12) suggest themselves. One is to set up the equations as a system of $2M-1$ simultaneous nonlinear equations in $2M-1$ unknowns and solve these using numerical techniques. The second is to assume a v vector (i.e., a set v_j ($j = 1, \dots, M$)) and solve for the u vector using (3.12). Use this u vector in (3.11) to obtain a new v vector and continue this procedure until, hopefully, they converge to the solution. Both of these methods appear rather distasteful; however, the channel matrix may have nice properties which can be exploited. Consider, for example, the m^{th} extension of the binary symmetric channel (BSC) and

let $M = 2^m$. If the quantization value v_k ($k = 1, \dots, 2^m$), is assigned the binary representation of $k-1$ ($k = 1, \dots, 2^m$) as a code word, then every entry of the channel matrix is of the form $p^j(1-p)^{m-j}$ where p is probability of a bit error and j is the number of bit errors per code word. For $m = 2$, the channel matrix is

$$P = \begin{bmatrix} (1-p)^2 & p(1-p) & p(1-p) & p^2 \\ p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{bmatrix} \quad (3.17)$$

For this case we proceed by first letting $p = 0$ and, as indicated above, (3.11) and (3.12) reduce to (3.13) and (3.14), which we solve using Max's algorithm. We then increase p slightly and, using the u and v vector obtained by Max's algorithm along with either of the two methods described above, obtain a solution of (3.13) and (3.14). Then p is again increased and, using the most recently obtained u and v vectors as a starting point, the next solution of (3.13) and (3.14) is obtained, etc. This procedure is used in the example presented in Section 3.4.

For the optimum uniform quantizer we proceed in a similar manner. First let $p = 0$ and solve (D.19) for r . This can be done by searching for the minimum in (D.7) or by differentiating (D.7) and obtaining that r for which the resulting equation is zero. After r is obtained for this case of the perfect channel, p is increased slightly and that value of r which minimizes (D.6) is obtained. This procedure is repeated, always using the most recently obtained value of r as the starting point in the search for the next value. An example illustrating the results obtained above is now presented.

3.4 An Example

As an example we consider that part of the telemetry system in Figure 1.1 from the input to the quantizer to the output of the data decoder. Because of the arguments preceding (3.8), we consider only one, say x_1 , of the random data samples in the collection $x_i (i = 1, \dots, n)$. We further assume a Gaussian, zero mean, unit variance probability distribution[†] on this sample. Various quantizer structures are determined and evaluated for each value of $M = 2^m (m = 1, \dots, 9)$. The number of quantization values M is restricted to powers of two because we only consider m^{th} order extensions of the binary symmetric channel (BSC). In one case the data encoder assigns the binary representation of $j-1$ as the channel code word of the quantization value v_j . For example, for the quantizer structure in Figures 3.1 or 3.2 (with $m = 3$), if $u_6 < x_1 \leq u_7$, then $y_1 = v_6$ and $y_1^* = 101$. The merits of this assignment procedure for equally likely, uniformly spaced quantization values are discussed in [35] and [36]. The data decoder performs the inverse operation assigning v_j to the binary representation of $j-1$. The Gray code is also considered as a data encoding scheme and its performance compared to that obtained for the natural code.

The channel disturbance is assumed to be additive, white Gaussian noise with one-sided power spectral density N_0 . As receiver structures we consider

[†]The probability density for this distribution is

$$p_{x_1}(x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2}$$

- 1) coherent reception of anticorrelated channel signals (PSK).
- 2) noncoherent reception of orthogonal channel signals (FSK).

For 1) the bit error is [57]

$$p = \frac{1}{2}(1 - \operatorname{erf} \sqrt{\varepsilon/N_0})^\dagger \quad (3.18)$$

and for 2)

$$p = \frac{1}{2} \exp(-\varepsilon/2N_0) \quad (3.19)$$

where ε is the transmitted energy per channel bit. In terms of the variables (i) average transmitted power S , (ii) time interval over which the analog waveform is taken $[0, T]$, (iii) number of digital data samples per waveform n , and (iv) number of channel bits per digital data sample m , the energy per channel bit is

$$\varepsilon = \frac{ST}{nm} \quad (3.20)$$

We define a parameter

$$\beta \triangleq \frac{ST}{N_0 n} \quad (3.21)$$

which is the ratio of the average energy per sample to the noise spectral density N_0 . By using m ($m = 1, \dots, 9$) as a parameter, and varying β , we obtain from (3.18) and (3.19) families of curves for the channel bit error p . These are presented in Figure 3.3. We note that the channel matrix (3.17) is completely determined by p .

The channel matrix shown in (3.17) is for $M = 4$ or $m = 2$. For $m = 9$ the channel matrix is a 512×512 matrix and would require 262, 144

[†]We use the National Bureau of Standards definition for the error function, i.e.,

$$\operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt$$

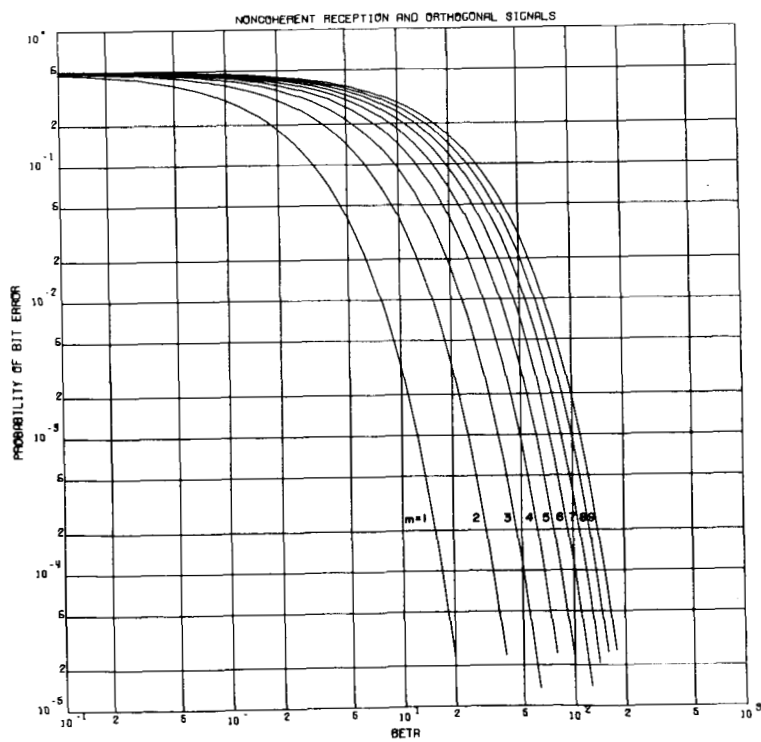
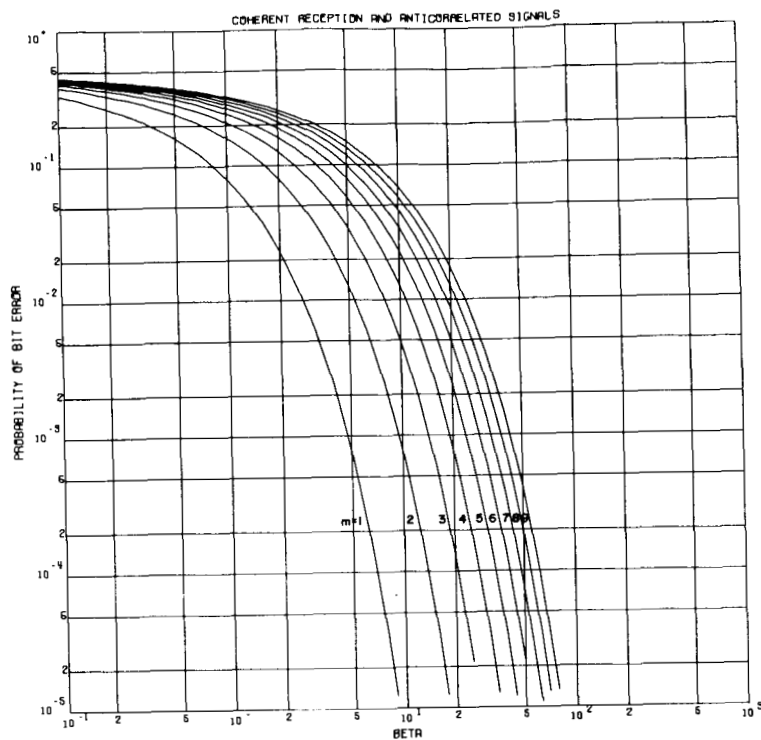


Figure 3.3 Probability of Bit Error vs. β as a Parameter

computer storage locations, a formidable amount. To get around this problem some interesting properties of the m^{th} extension of the BSC were exploited. These are discussed in Appendix E.

3.4.1 Determination of the Quantizer Structures. Under the added constraints of the example, the results above, (3.10), (3.11), and (3.12) become

$$\epsilon_{*1}^2 = E[(x_1 - z_1)^2] = 1 + \sum_{k=1}^{2^m} [-2b_k f_k + c_k g_k] \quad (3.22)$$

$$v_j = \frac{\sum_{k=1}^{2^m} P_{kj} f_k}{\sum_{k=1}^{2^m} P_{kj} g_k}, \quad j = 1, \dots, 2^m \quad (3.23)$$

$$u_j = \frac{1}{2} \frac{c_j - c_{j-1}}{b_j - b_{j-1}}, \quad j = 2, \dots, 2^m - 1 \quad (3.24)$$

where

$$b_k = \sum_{\ell=1}^{2^m} v_{\ell} P_{k\ell}, \quad k = 1, \dots, 2^m \quad (3.25)$$

$$c_k = \sum_{\ell=1}^{2^m} v_{\ell}^2 P_{k\ell}, \quad k = 1, \dots, 2^m \quad (3.26)$$

$$f_1 = -\frac{1}{\sqrt{2\pi}} \exp(-u_1^2/2) \quad (3.27)$$

$$f_k = \frac{1}{\sqrt{2\pi}} [\exp(-u_k^2/2) - \exp(u_{k+1}^2/2)], \quad k = 2, \dots, 2^m - 1 \quad (3.28)$$

$$f_{2^m} = \frac{1}{\sqrt{2\pi}} \exp(-u_{2^m}^2/2) \quad (3.29)$$

$$g_1 = \frac{1}{2} [1 + \operatorname{erf}(u_1/\sqrt{2})] \quad (3.30)$$

$$g_k = \frac{1}{2} [\operatorname{erf}(u_{k+1}/\sqrt{2}) - \operatorname{erf}(u_k/\sqrt{2})], \quad k = 2, \dots, 2^m - 1 \quad (3.31)$$

$$g_{2^m} = \frac{1}{2} [1 - \operatorname{erf}(u_{2^m}/\sqrt{2})] \quad (3.32)$$

For the uniform quantizer, (3.22) and (3.25) through (3.32) are unchanged; however, the $v_\ell (\ell = 1, \dots, 2^m)$ and $u_k (k = 2, \dots, 2^m)$ take on the values in (3.15) with $M = 2^m$.

Using the natural code four general quantizer structures are determined. These are

- A - Optimum Nonuniform
- B - Max's Nonuniform
- C - Optimum Uniform
- D - Max's Uniform

All four of the structures are dependent upon m while structures A and C also depend upon β . All structures are for a Gaussian, zero mean, unit variance probability distribution on x_1 , and therefore all of the structures are symmetric. The optimum structure, that which minimizes ϵ_{*1}^2 , is structure A. As pointed out above, this structure is obtained by assuming a perfect channel and then allowing the channel to become progressively more noisy. The nonuniform quantizer structure for the perfect channel is Max's nonuniform structure. Max [15] determined this structure for $m = 1, \dots, 5$. We extend his results to $m = 6, \dots, 9$. All of these results are tabulated in Table 3.1 in Appendix F. Since the structures are symmetric, only the non-negative transition levels $u_k (k = 2^{m-1} + 1, \dots, 2^m)$ and quantization values $v_\ell (\ell = 2^{m-1} + 1, \dots, 2^m)$ are presented.

By starting with these structures and decreasing the variable β , (which we see from Figure 3.3 corresponds to increasing the average bit error rate), we obtain a series of optimum quantizer structures. For

$m = 1, 2$, these were obtained for β 's ranging down to 0.01(-30 db). For $m = 3, 4, 5, 6$, lack of precision in the program limited the number of β 's for which quantizer structures were obtained. The time required to determine a structure increases exponentially with m . The approximate average amount of IBM 7094 computer time required to determine an optimum quantizer structure as a function of m is shown in Table 3.2. The number of quantizer structures determined for $m = 7, 8, 9$ is limited because of the computer time required per structure. In Table 3.3 (See Appendix F), one of the optimum quantizer structures obtained for

Table 3.2

Approximate Average Amount of IBM 7094 Computer Time
Required to Determine an Optimum Quantizer Structure

m	IBM 7094 Computer Time
5	2 sec.
6	10 sec.
7	2 min.
8	5 min.
9	12 min.

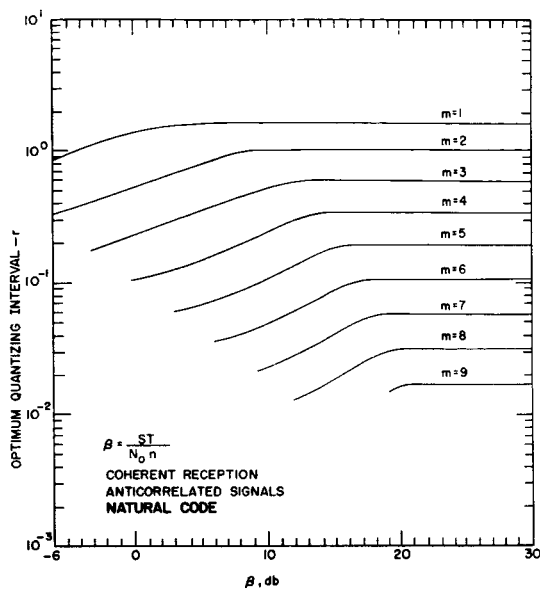
each value of m is presented. A comparison of the structure for the noiseless channel to that for the noisy channel shows that the structure for the noisy channel is more tightly grouped about the origin.

For the uniform quantizer, the solution procedure and results are analogous to those for the nonuniform quantizer. Max [15] also presented the optimum uniform quantizer for the noiseless channel for $m = 1, \dots, 5$. We extend these to $m = 6, \dots, 9$. These results are presented in Table 3.4 in Appendix F. We also determine the optimum uniform quantizers for various β 's corresponding to each m ($m = 1, \dots, 9$)

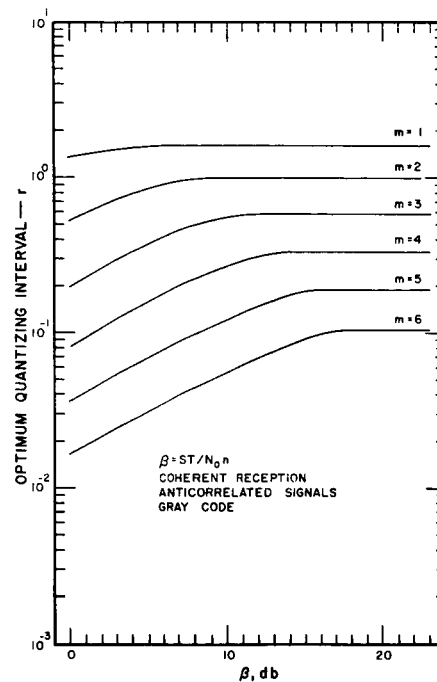
These are presented in Table 3.5 in Appendix F. The optimum quantizing interval r as a function of β with m as a parameter is plotted in Figure 3.4a for receiver structure 1) and in Figure 3.4b for receiver structure 2).

The optimum uniform quantizer structure was also obtained using the Gray code to represent the quantization values and coherent reception of anticorrelated signals (receiver structure 1). The procedure is the same as that described in the previous paragraph. However, for this case, m was limited to values of six or less. The optimum quantizing interval r as a function of β is plotted in Figure 3.4c. For $m = 1$ the Gray code is the same as the natural code and hence this curve in Figure 3.4c is the same as for $m = 1$ in Figure 3.4a. The computational algorithms utilized to perform the numerical computations are presented in Appendix G.

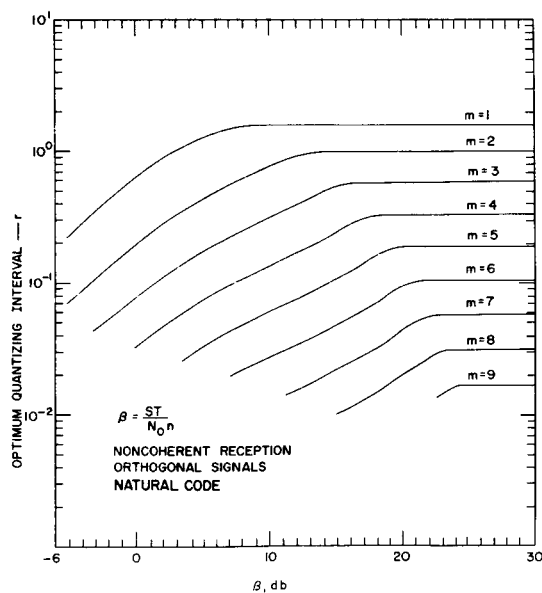
3.4.2 Quantizer Performances - Natural Code. The performance of each of the four quantizer structures determined for the natural code is evaluated using (3.22) and (3.25) to (3.32) for receiver structures 1) and 2). The m.s.e. for a single sample is plotted as a function of β with m as the parameter. These plots are shown in Figure 3.5 for coherent reception of anticorrelated signals (receiver structure 1). The four graphs a,b,c, and d of Figure 3.5 correspond respectively, to quantizer structures A, B, C, and D. The curves approach horizontal asymptotes for β both large and small. The error at the right-hand asymptotes is due entirely to the quantizer, since for large values of β the channel is essentially noiseless. Given m , the location of the asymptote is determined by



(a)

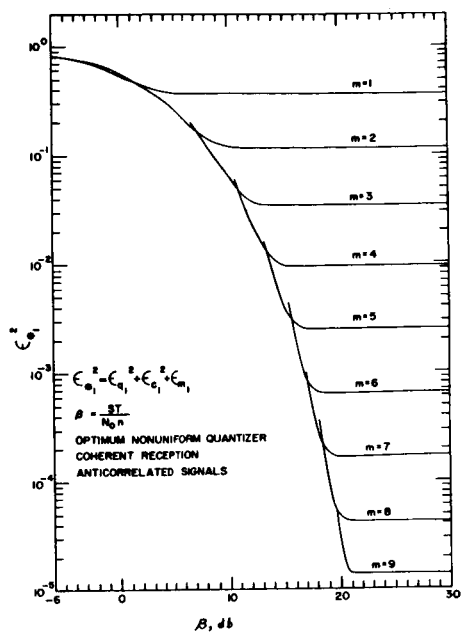


(c)



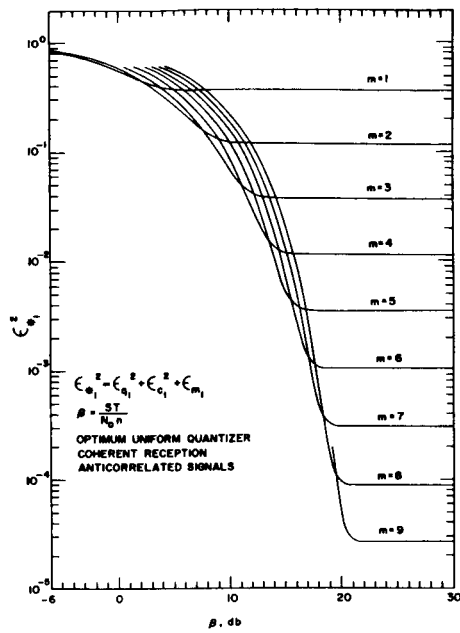
(b)

Figure 3.4 Optimum Quantizing Interval r vs. β with m as a Parameter



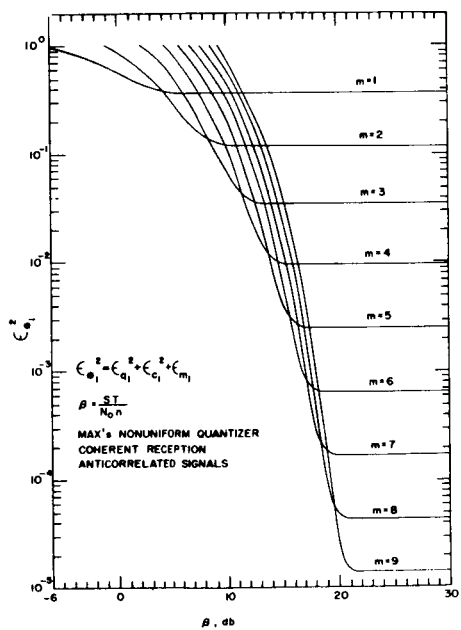
(a)

(a)

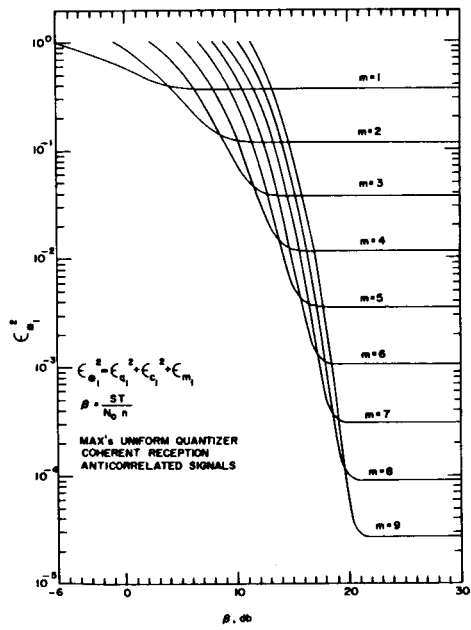


(c)

(c)



(b)



(d)

Figure 3.5 Performance Characteristics of the 2^m th Level Quantizer and the m^{th} Extension of the BSC Using the Natural Code and Receiver Structure No. 1

$$\epsilon_{*1R}^2 = E[(x_1 - y_1)^2] = \sum_{k=1}^{2^m} \int_{u_k}^{u_{k+1}} (x_1 - v_k)^2 p_{x_1}(x_1) dx_1 \quad (3.33)$$

where the u and v vectors are those for a noiseless channel. On the other hand, as $\beta \rightarrow 0$ we see from (3.18), (3.19), (3.20), and (3.21) that $p \rightarrow 1/2$. Therefore every entry in the channel matrix (3.17) is $1/2^m$. And, for a symmetric distribution on x_1 , (3.10) reduces to

$$\epsilon_{*1L}^2 = E(x_1^2) + \frac{1}{2^m} \left[\sum_{\ell=1}^{2^m} v_\ell^2 \right] \quad (3.34)$$

For the uniform quantizer (3.34) becomes

$$\epsilon_{*1L}^2 = E(x_1^2) + \frac{r^2}{6} 2^m (2^m - 1) \quad (3.35)$$

For the process under consideration, we have $E(x_1^2) = 1$. Also, for the optimum and optimum uniform quantizers, as $\beta \rightarrow 0$ the structures tend to group around the origin. Therefore the location of the left-hand asymptote will be at unity. This has not been shown analytically; however, no contradictory results have been obtained. The locations of left-hand asymptotes in Figure 3.5b and d are tabulated in Table 3.6.

From Figure 3.5 with β large we see that the advantage of the nonuniform quantizer over the uniform quantizer starts with $m = 2$ and increases as m increases. And from Table 3.6 and Figures 3.5b and d, we see that for β small Max's nonuniform quantizer significantly outperforms Max's uniform quantizer. Both of these effects are due to the fact that the nonuniform quantizer structure has more quantization values near the origin. This is evident from Tables 3.1 and 3.4 in Appendix F. For example, for $m = 9$ the equal interval r of Max is $r = 0.0165$ so that for the tenth transition level to the right of the origin we have $u_{267} = 0.165$. While from Table 3.1, for the nonuniform quantizer $u_{267} = 0.076$.

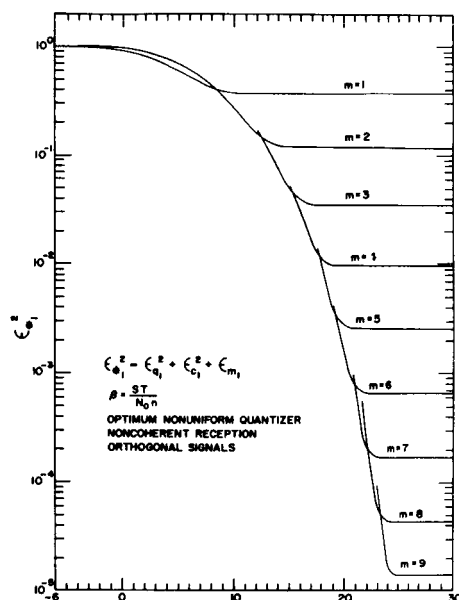
Table 3.6

Locations of the Error Asymptotes as β Approaches Zero

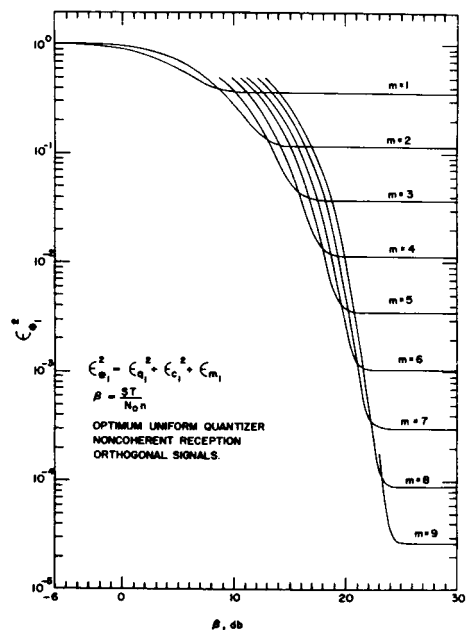
Number of Quantization Values $M = 2^m$		Mean Square Error ϵ_{*1}^2	
m		Max's Nonuniform Quantizer	Max's Uniform Quantizer
1		1.6366	1.8488
2		2.2426	2.9808
3		2.7673	4.2052
4		3.1791	5.4944
5		3.4784	6.8522
6		3.6777	8.2767
7		3.8050	9.7652
8		3.8759	11.2944
9		3.8281	12.8716

In Figure 3.6 graphs of the m.s.e. ϵ_{*1}^2 vs. β with m as a parameter are presented for noncoherent reception of orthogonal signals (receiver structure 2). These are similar in form to the graphs in Figure 3.5 with each curve shifted approximately 3 db to the right for β large. For β small the shift is more pronounced.

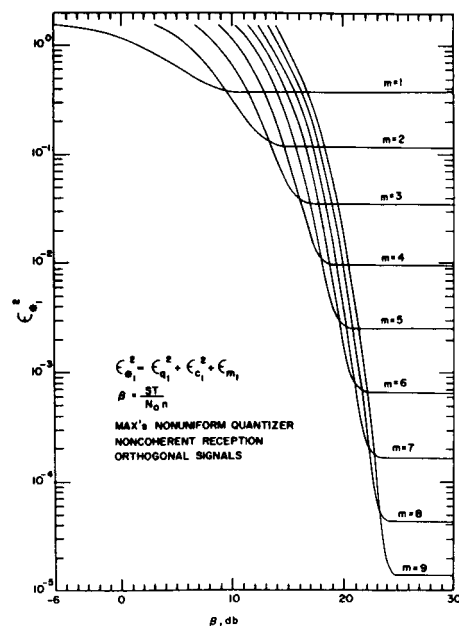
In order to get a better comparison of the performances of the quantizers, we have plotted the lower envelopes of the curves in Figure 3.5a, c, and d on a single graph. These are presented in Figure 3.7a. The lower envelopes of the curves in Figure 3.6a, b, and d are presented in Figure 3.7b. Due to lack of precision in the computer program, portions of the curve for the optimum quantizer in Figure 3.7 were not obtained. Note that the optimum uniform quantizer is identical to the optimum quantizer for $m = 1$ and to Max's uniform quantizer for β large. As a second comparison, all eight curves of



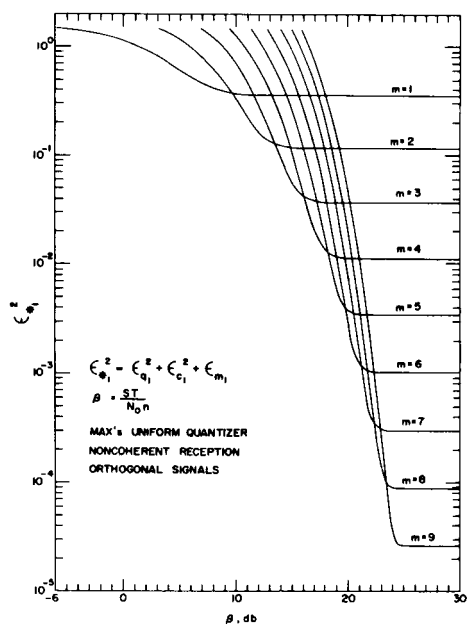
(a)



(c)



(b)



(d)

Figure 3.6 Performance Characteristics of the 2^m th Level Quantizer and the m^{th} Extension of the BSC Using the Natural Code and Receiver Structure No. 2

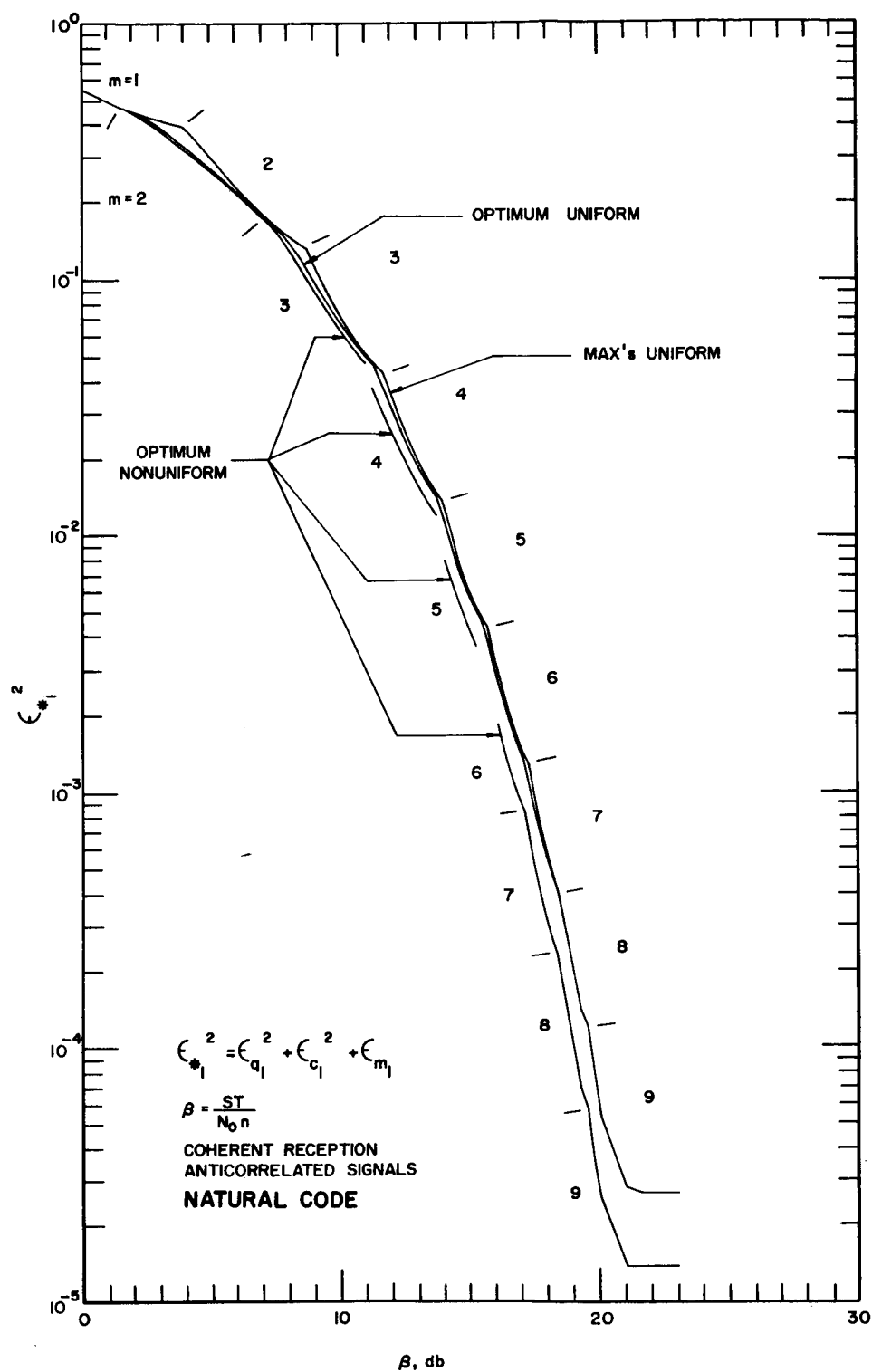


Figure 3.7(a) A Comparison of the Performance of the Quantizer Structure Used Optimally for PSK Reception

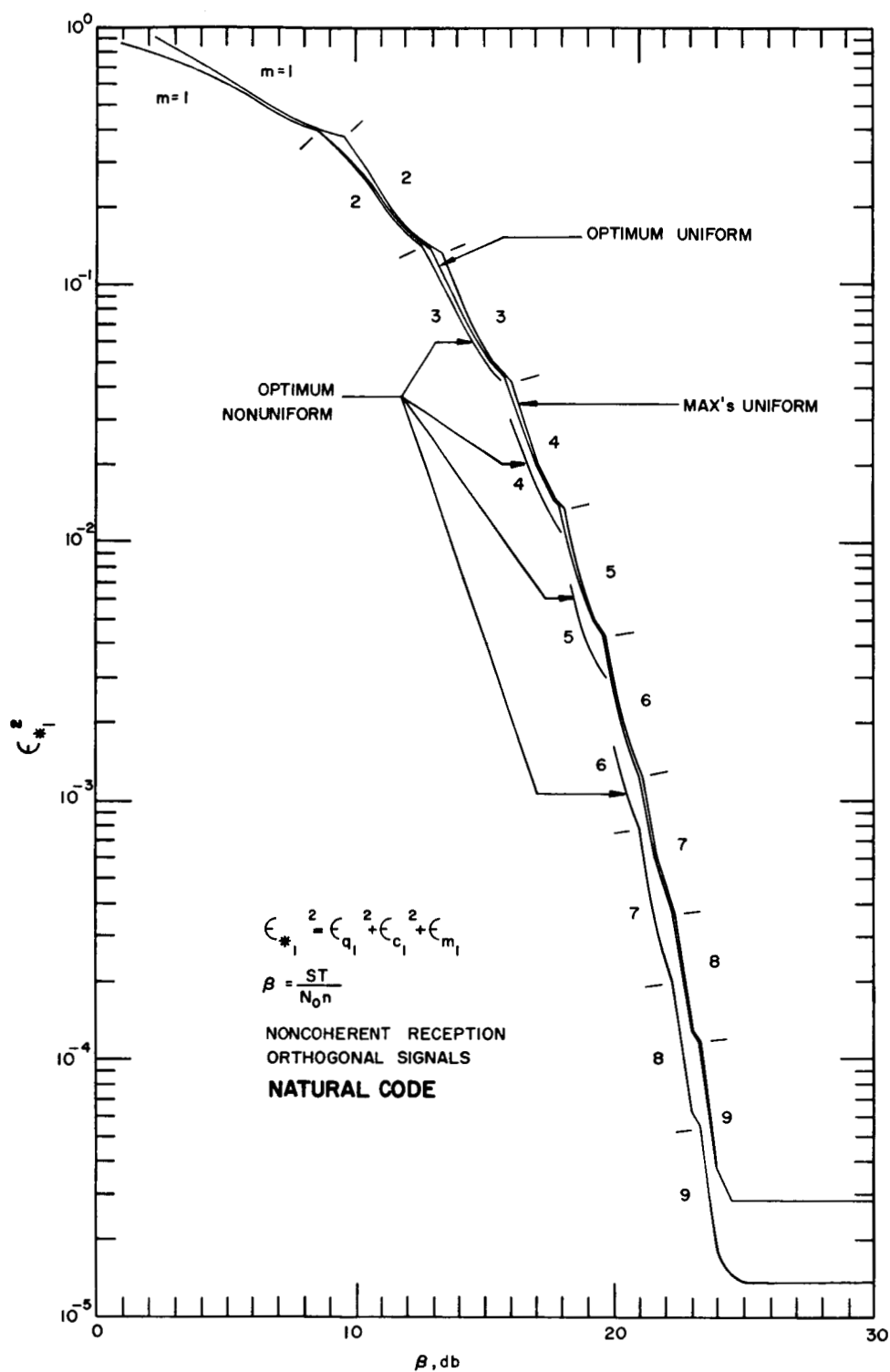


Figure 3.7(b) A Comparison of the Performances of the Quantizer Structures Used Optimally for FSK Reception

ϵ_{*1}^2 vs. β for $m = 8$ are shown in Figure 3.8.

The sensitivity of the m.s.e. ϵ_{*1}^2 to the quantizing interval r has also been investigated. This is plotted in Figure 3.9 for receiver structure 1) with $m = 2, \dots, 8$ as a parameter and $\beta = 20(13 \text{ db})$.

Figures 3.4a and b are used in conjunction with Figures 3.5c and 3.6c, respectively, to determine the optimum quantizer for the receiver structure and β given.

This is done in two steps. First, for example, determine the optimum value of m from Figure 3.5c and then the optimum quantizing interval corresponding to that m from Figure 3.4a.

3.4.3 Quantizer Performance - Gray Code. The performance of the optimum uniform quantizer structure determined for the Gray code and receiver structure 1) is presented graphically in Figure 3.10. These results appear quite similar to those presented for the natural code in Figure 3.5c. In order to make a closer comparison both curves for $m = 2, 4$ and 6 are plotted in Figure 3.11. We note, from Figure 3.11, that the Gray code outperforms the natural code for β large and that the natural code outperforms the Gray code for β small. This demonstration is significant since, as pointed out in Section 1.2.4, appreciable effort has been expended in order to prove that the natural code is uniformly best when the data presented to the quantizer is distributed uniformly over an interval of the real line.

The crossover effect, illustrated in Figure 3.11, is due to three things (i) the nature of the codes, (ii) the shape of the distribution, and (iii) the quantizer structure. The Gray code emphasizes the channel error which either changes v_1 to v_M or v_M to v_1 , however, the

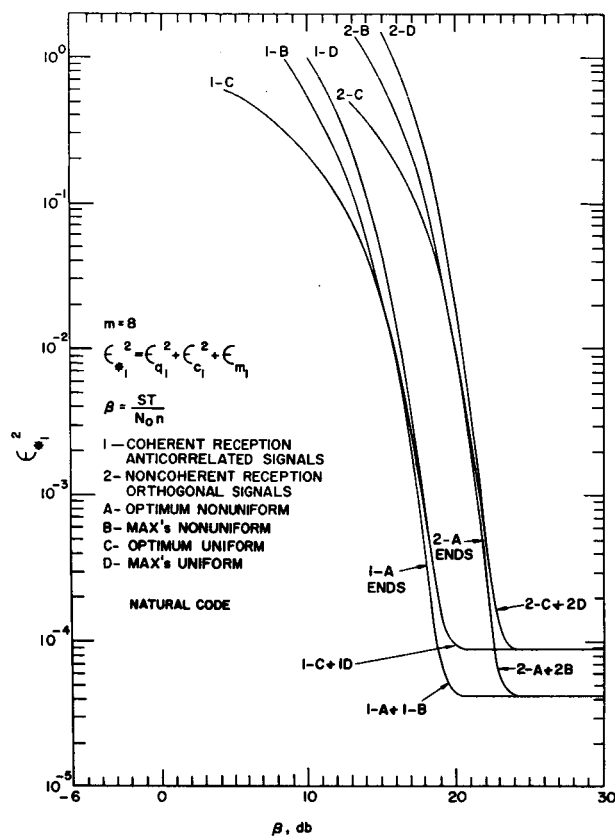


Figure 3.8 A Comparison of Quantizer Performances for $m = 8$

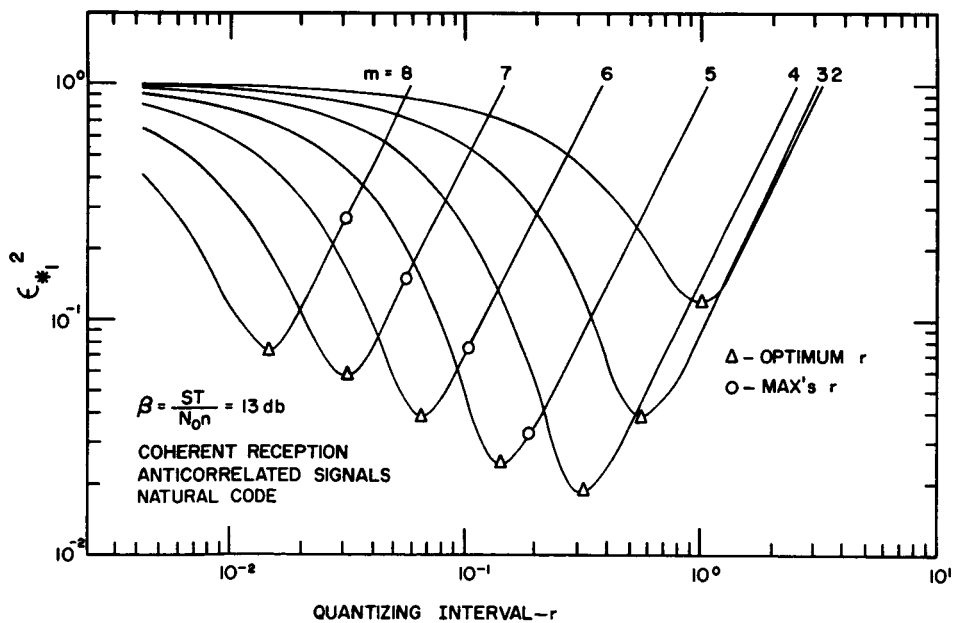


Figure 3.9 The m.s.e. ϵ_{*1}^2 vs. r with m as a Parameter ($\beta = 13 \text{ db}$)

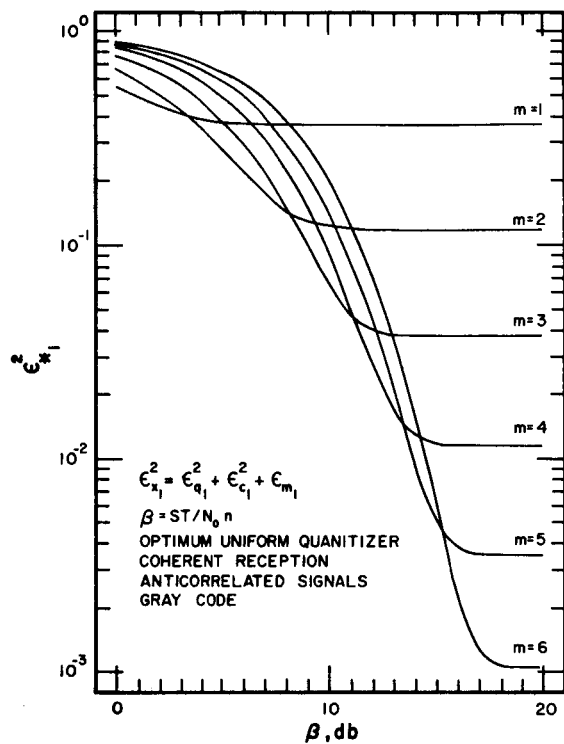


Figure 3.10 Performance Characteristics of the 2^m th Level Quantizer and the m th Extension of the BSC Using the Gray Code and PSK Reception

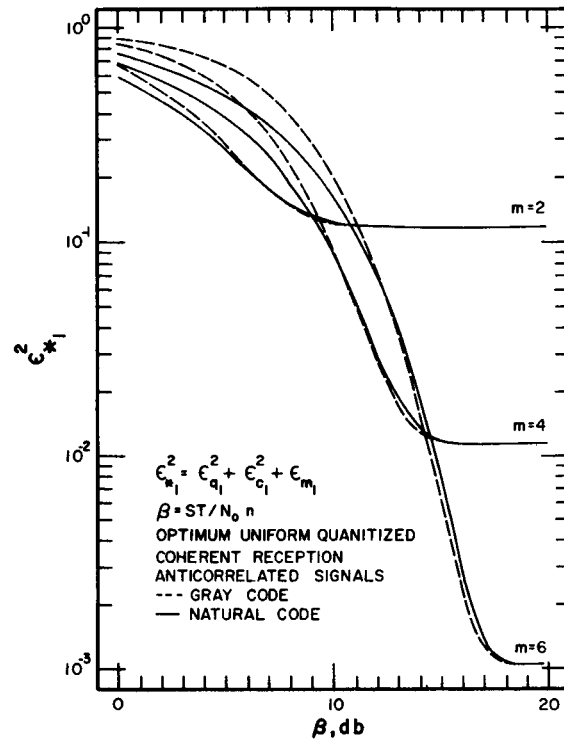


Figure 3.11 A Comparison of Quantizer Performances for Gray Code vs. Natural Code

quantizer structure determines the probability of its output taking on these values, v_1 and v_M . For Gaussian data and the signal-to-noise ratio ρ large these probabilities are small (see Figure 3.2), but, for ρ small the quantizer structure tends to cluster about the origin, and the probabilities become large. Hence neither code performs uniformly better than the other.

3.4.4 Entropy, Channel Capacity, and Data Rate. As a by-product of determining the quantizer structures, most of the necessary calculations for computing three information theoretic quantities were completed. These are (i) the entropy at the output of the quantizer, (ii) the channel capacity of the m^{th} extension of the BSC, and (iii) the rate distortion function or source information rate.

It is possible to treat the quantizer output as a discrete data source for which the entropy is easily computed as

$$H_Q(V) = - \sum_{k=1}^M P(v_k) \log_2 P(v_k) \quad \text{bits/sample} \quad (3.36)$$

where

$$\begin{aligned} P(v_k) &= \text{Prob } [y_1 = v_k] \\ &= \text{Prob } [u_k \leq X_1 < u_{k+1}] \\ &= \int_{u_k}^{u_{k+1}} p_{x_1}(x_1) dx_1 \end{aligned} \quad (3.27)$$

And, for the Gaussian distributed data of the example with the m^{th} extension of the BSC, we have by (3.30) through (3.32) that

$$P(v_k) = g_k, \quad k = 1, \dots, 2^m \quad (3.38)$$

The channel matrix for the m^{th} extension of the BSC is such that the

terms in every row and every column are permutations of the terms in the first row. A channel with this property is sometimes referred to as a uniform channel [58]. The capacity of the uniform channel with 2^m input and output symbols is independent of the row k , i.e.

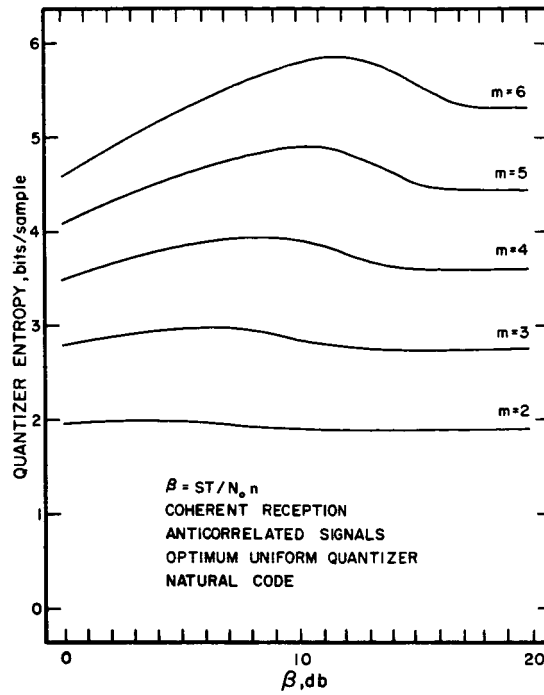
$$C = m + \sum_{\ell=1}^{2^m} P_{k\ell} \log_2(P_{k\ell}) \quad \text{bits/sample} \quad (3.39)$$

The information rate for a source producing independent Gaussian samples for a m.s.e. fidelity criterion between the source data and its facsimile presented to the user was given by Shannon [59]. In terms of the m.s.e. ϵ_{*1}^2 , the information rate is

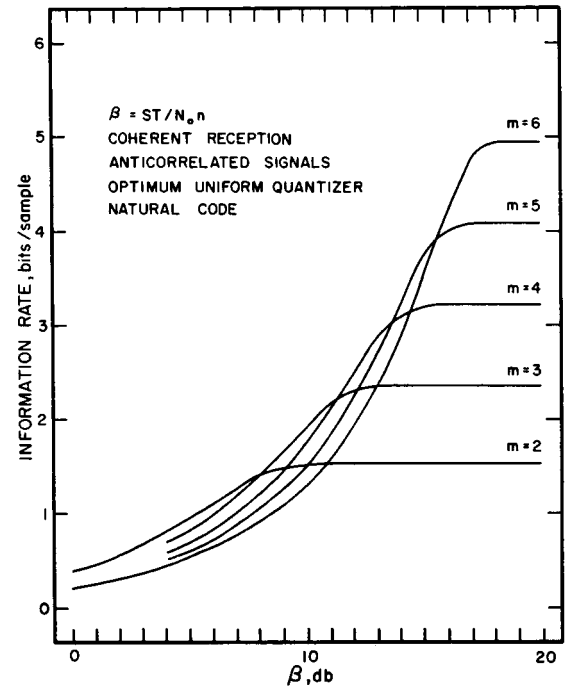
$$R(\epsilon_{*1}^2) = \frac{1}{2} \log_2 (\sigma_1^2 / \epsilon_{*1}^2) \quad \text{bits/sample} \quad (3.40)$$

where $\sigma_1^2 = 1$ for our example. The significance of the information rate is that it is impossible to have $R(\epsilon_{*1}^2) > C$ with m.s.e. $\leq \epsilon_{*1}^2$.

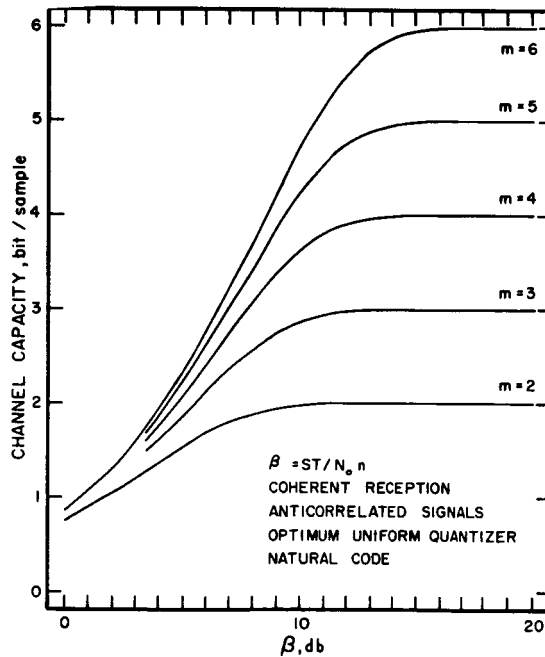
The quantizer entropy, channel capacity, and information rate are plotted in Figures 3-12a, b and c respectively. The conditions for the plotted results are those of the example with coherent reception, anticorrelated signals, optimum uniform quantizer, the natural code, and $m = 1, \dots, 6$. These three quantities are compared in Figure 3-12d for both the natural and Gray codes with $m = 4$ and 6. The channel capacity remains invariant between codes because the rows of the channel matrix for the Gray code are permutations of those for the natural code. Further insight regarding the effect of optimizing the quantizer interval when using the Gray code as compared to the natural code can be obtained by studying Figure 3-12d. The Gray code tends to protect itself by keeping r , the quantizer interval, large and hence $P(v_1) = P(v_{2^m})$, the probabilities of the end quantizer values, small. As pointed out



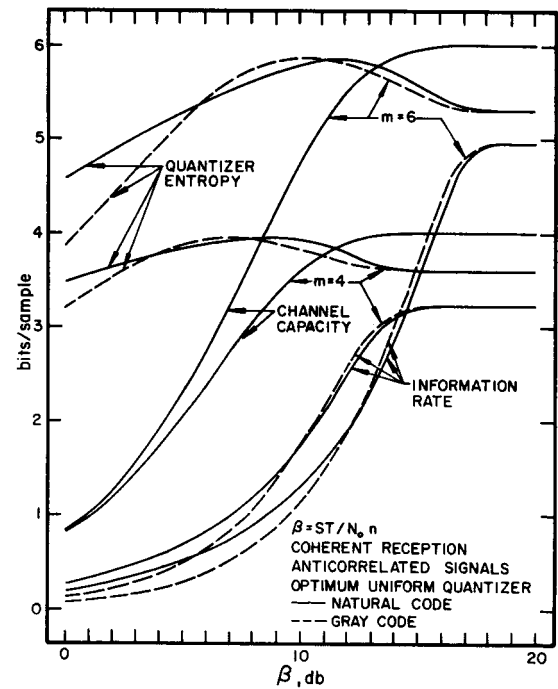
(a)



(c)



(b)



(d)

Figure 3.12 Quantizer Entropy, Channel Capacity, and Information Rate for the Optimum Uniform Quantizer

previously, word errors resulting from single bit errors which cause the transitions from v_1 to v_{2^m} and v_{2^m} to v_1 are severely penalized by the m.s.e. criterion. The curves for the information rate behave similar to those for m.s.e., ϵ_{*1}^2 , as would be expected from (3.40). The result indicates that it might have been used as a measure instead of m.s.e.

3.4.5 Quantizer Performance - Nonreal Time Operation. So far we have restricted the system's operation to real time operation, In this section we will remove this restriction and consider the problem of transmitting numerical data over a noisy channel. Let the data be distributed $N(0,1)$ as above and the channel be a BSC (memoryless) with probability of bit error p corresponding to a channel signal-to-noise ratio

$$\Delta = \frac{\mathcal{E}}{N_0} \quad (3.36)$$

where

$$\mathcal{E} = \frac{ST}{m_b} \quad (3.37)$$

with S the average transmitted power and m_b the number of channel bits transmitted in the time interval $[0,T]$. Hence the data rate is

$$R = \frac{m_b}{T} \quad (3.38)$$

For the case of coherent reception with anticorrelated signals and the natural code we have p given by (3.18) and for two channel bits per data sample ($m = 2$) the channel matrix given by (3.17).

Since the real time restriction has been removed, p is no longer a function of m . We determine the optimum uniform quantizer and evaluate

its performance under the m.s.e. criterion. The performance curves as a function of m with Δ the running parameter are presented as the solid lines in Figure 3.13. Note that for a fixed value of Δ the performance curves approach a horizontal asymptote as m increases. This agrees with intuition since it makes sense that using more channel bits should not cause a deterioration in performance.

Also, plotted in Figure 3.13 using dashed lines is the performance of Max's uniform quantizer in conjunction with the noisy channel. Note that its performance does deteriorate with increasing m for Δ fixed. This result is rather contrary to intuition. The cause of it is the manner by which the m.s.e. criterion emphasizes bit error combinations which cause large realizations of the random variable

$$(z - y)^2$$

To better appreciate this, consider the uniform quantizer structure presented in Figure 3.2. We define the difference between v_{2^m} and v_1 to be the spread of the quantizer (QS) which by (3.15) with $M = 2^m$ is

$$\begin{aligned} \text{QS} &= v_{2^m} - v_1 \\ &= (2^m - 1)r \end{aligned} \tag{3.39}$$

For the optimum uniform quantizer the quantizer spread is a function of the channel where Max's quantizer is the optimum quantizer for the noiseless channel. In Figure 3.14 we plot the quantizer spread as a function of m with Δ as the running parameter. Max's uniform quantizer corresponds to $\Delta = \infty$. For the five curves with Δ finite the spread of quantizer tends to approach a maximum value. While for the case of the perfect channel no such maximum is evident. The conclusion is that,

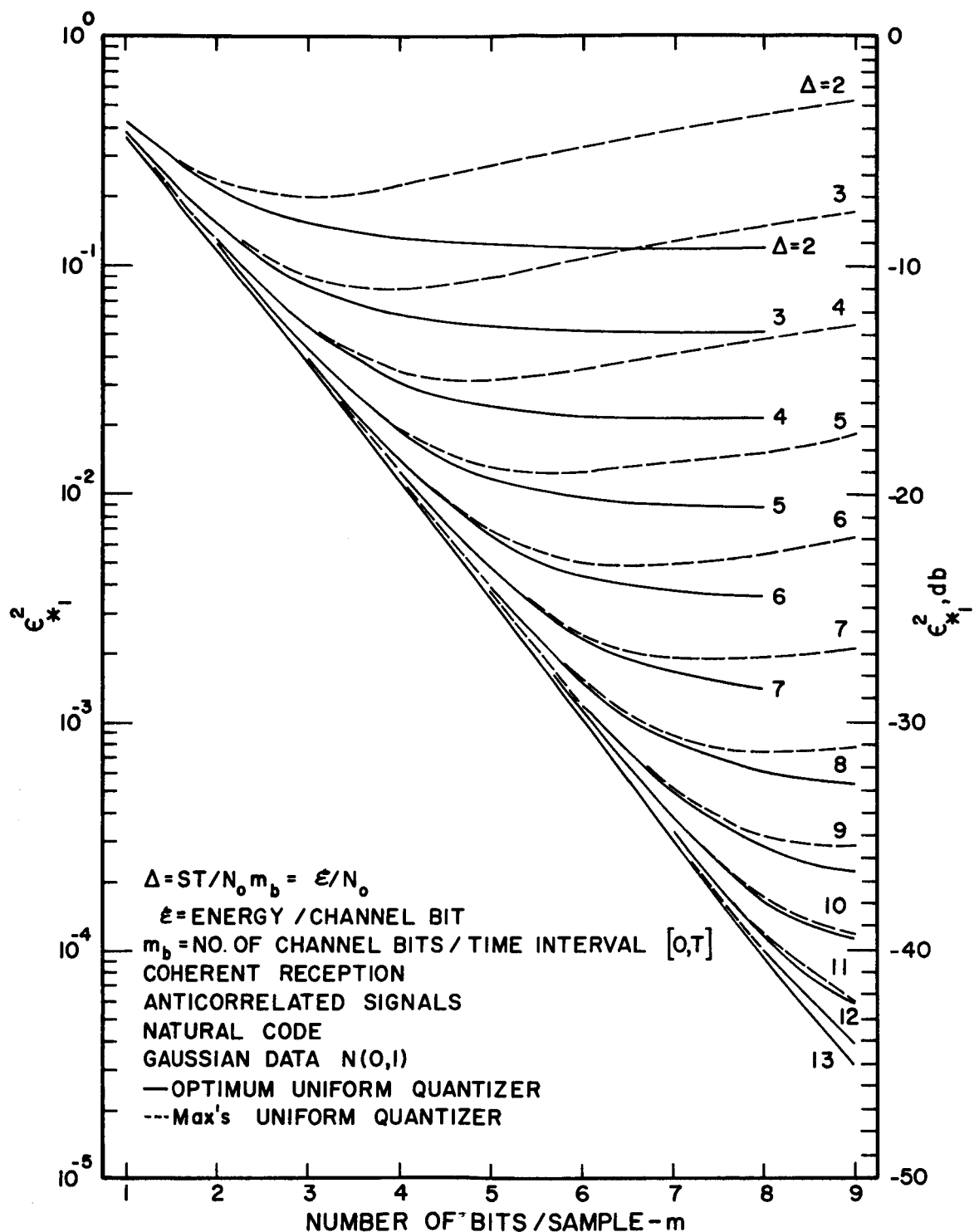


Figure 3.13 A Comparison of Quantizer Performance under Normal Time Conditions

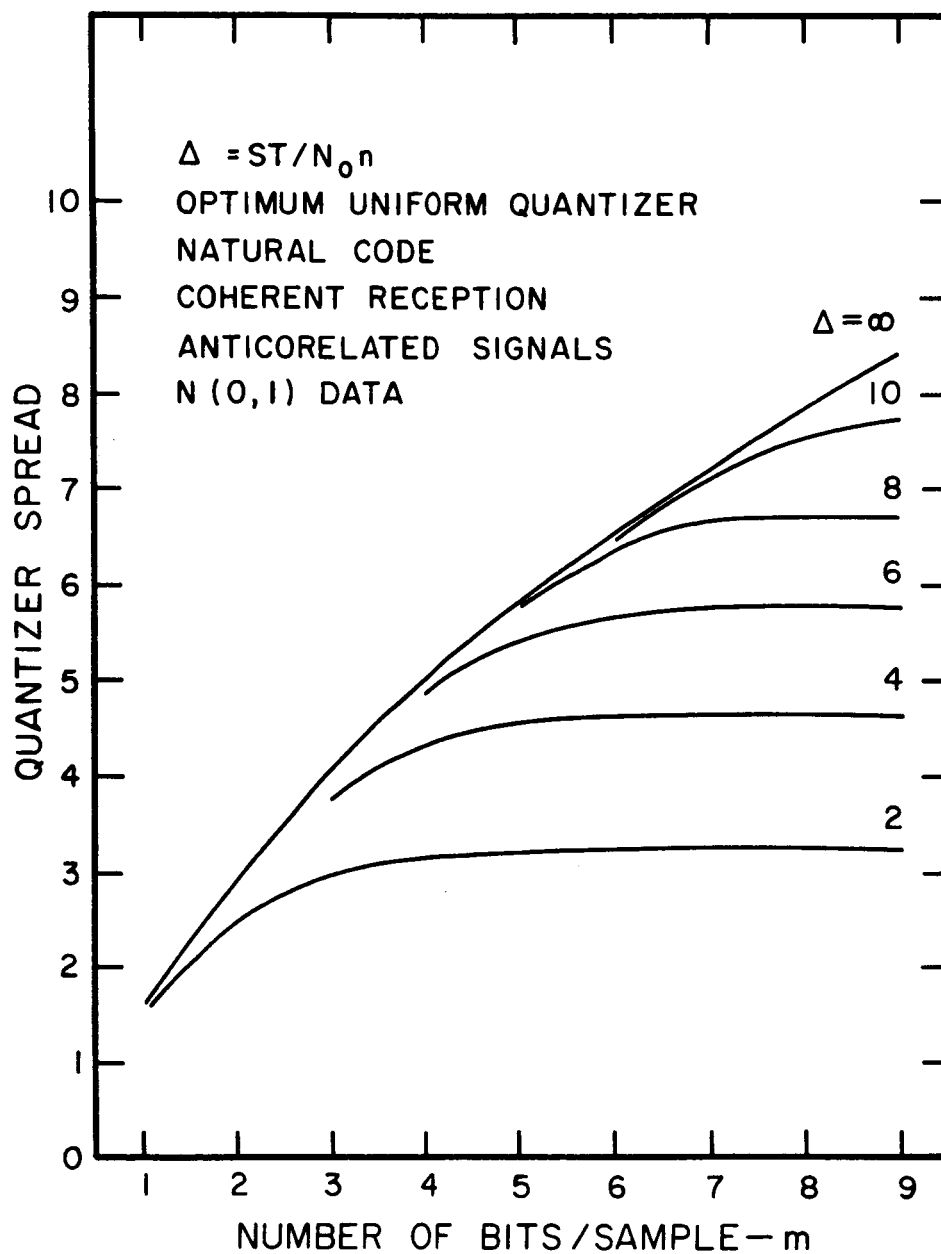


Figure 3.14 Quantizer Spread vs. m with Δ as the Parameter for Normal Time Conditions

even for a system not operating in real time, it is important to take the channel into consideration when designing the quantizer.

3.5 Summary

In this chapter we have determined the optimum and optimum uniform quantizer structures for digital data with a known continuous probability density. To quantize optimally it is necessary to take into account the adverse effects of the channel over which the data is to be transmitted. This is clear from (3.11) and (3.12).

Both the optimum and optimum uniform quantizer structures were obtained for the case of numerical data distributed $N(0,1)$ with the channel disturbance that of additive, white, Gaussian noise. Both PSK and FSK receivers are evaluated for the quantized data source encoded using the natural code. Source encoding using the Gray code with PSK reception is also evaluated.

We note from Figure 3.7 that if the optimum number of channel bits is selected then the uniform quantizer of Max performs within 1 db of the optimum quantizer. However, it is evident from Figure 3.8 that, for the number of channel bits fixed, the choice of the quantizing interval is most significant. In applications where adaptive capabilities may be programmed into a system, either of these cases may be exploited to improve system performance.

We point out that it costs no more to build the optimum quantizer than a suboptimum one. The difficulty lies in determining the distribution of the data and the channel matrix. We have provided a design procedure for the optimum uniform quantizer when the data distribution and channel matrix above are used. We have demonstrated

that neither the natural code nor the Gray code uniformly out perform the other for Gaussian data. We have also pointed out that one must design the quantizer for the channel even for nonreal time operation.

Early in the chapter we showed the total error expanded into three terms, i.e.

$$\epsilon_*^2 = \epsilon_q^2 + \epsilon_c^2 + \epsilon_m^2 \quad (3.3)$$

These terms are plotted vs. β in Figure 3.15 for the optimum uniform quantizer with $m = 4$. Note that for β large (an essentially perfect channel) all the m.s.e. is due to the quantizer, then as β decreases, the channel error dominates, and for β small the quantizer error plus the mutual error term are most significant. The mechanism which causes the mean-square channel error to peak out and then decrease for an increasingly noisy channel is the form of the optimum quantizer. Observe from Figure 3.4 that the quantizing interval decreases with β . Hence from (3.15) we see that the penalty for making a channel error

$$(v_\ell - v_k)^2 = (\ell - k)^2 r^2 \leq M^2 r^2$$

decreases with r^2 for M fixed. We earlier conjectured that as $\beta \rightarrow 0$ $r \rightarrow 0$ in which case $\epsilon_{c_1}^2 \rightarrow 0$. From (3.6) we also see that $\epsilon_{m_1} \rightarrow 0$ and the total error ϵ_{*1}^2 is again due entirely to the quantizer.

We finally note that the bandwidth of the channel was required to change as m varied. Constraining the bandwidth would add to the complexity of the problem.

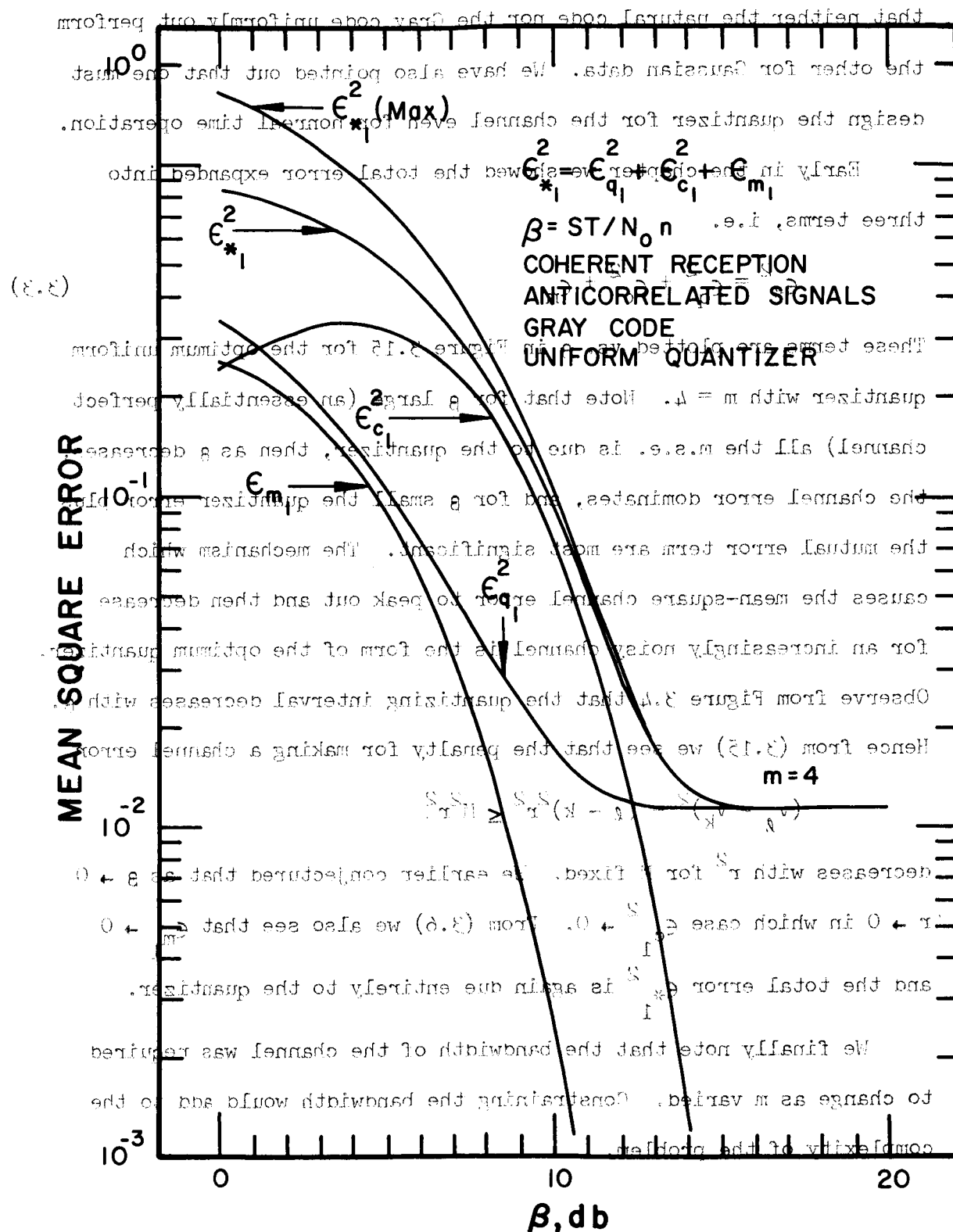


Figure 3.15 The Error Terms: ϵ_{q1}^2 , ϵ_{c1}^2 , and ϵ_{m1}^2

CHAPTER IV

OVERALL SYSTEM PERFORMANCE - THE TRADEOFF PROBLEM

Our objective, that of minimizing the mean-integral-square error

$$\epsilon^2 = E\left\{\int_0^T [x(t) - z(t)]^2 dt\right\} \quad (4.1)$$

has not yet been attained. We have used Totty's result [14]

$$\epsilon^2 = \epsilon_a^2 + \epsilon_*^2 \quad (4.2)$$

which is valid for orthonormal expansions of the source process, then cited Brown's result [49] that, for second order processes, ϵ_a^2 is minimized by using the expansion of Karhunen and Loeve. In Chapter III we determined the quantizer structure which minimizes

$$\epsilon_{x_i}^2 = E(x_i - z_i)^2 \quad (4.3)$$

where x_i and z_i are components of the x and z vectors of Figure 1.1 and indicated that for the x_i independent, identically distributed (IID)

$$\epsilon_*^2 = n \epsilon_{x_1}^2 \quad (4.4)$$

More generally

$$\epsilon_*^2 = \sum_{i=1}^n E(x_i - z_i)^2 \quad (4.5)$$

We found that the approximation error ϵ_a^2 is a function of only the sampler. Hence, given an expansion, ϵ_a^2 is a function of n (the number of

samples per sample function). The remaining system error ϵ_*^2 is a function of the quantizer and channel. Consider the case of real time transmission using both an average and peak power limited transmitter. Then as m_b (the number of channel bits per sample function) is increased, the energy per channel bit ϵ decreases. Therefore, more bit errors are made which causes an increase in ϵ_*^2 .

For the x_i IID, we have

$$m_b = nm$$

where we assign m bits to each of the n samples. This case requires but one quantizer which is the optimum quantizer for all of the n samples. From Chapter III we know the best quantizer given m . Hence, we desire that couple (n, m) which minimizes ϵ_*^2 .

From Chapter II we know that the K.L. samples and the principal components of Hotelling are not identically distributed. Therefore, we might expect that not all n of the samples should be assigned the same number of channel bits. In which case, we desire that couple (n, m_b) and that rule which assigns the m_b bits to the n samples such that the error ϵ_*^2 is minimized.

In Section 4.1 we determine and evaluate various bit assignment procedures for K.L. samples. This is done for the SGM process and the BSC with an average power constraint and for a limited number of values for m_b and n . In Subsection 4.1.1 we determine optimum bit assignments using a search procedure. Then in Subsection 4.1.2 we submit an algorithm which provides nearly optimum bit assignments. In Section 4.2 we evaluate time sampled systems with and without digital filtering and for equal and nearly optimum bit assignments. The overall

performance of these systems are compared in Section 4.3. Finally, in Section 4.4 we compare the data rates required by the various procedures in order to perform at specified error levels.

4.1 K.L. Sampling

When representing $x(t)$ by a truncated K.L. expansion under the criterion of (4.1), (4.2) becomes

$$\epsilon^2 = \epsilon_a^2 + \sum_{i=1}^n \sigma_i^2 E(x_i - z_i)^2 \quad (4.6)$$

where ϵ_a^2 is given by (2.12), the σ_i^2 are the eigenvalues of the integral equation (2.8), n is the number of samples, and the $E(x_i - z_i)^2$ terms are evaluated by (3.10).

The eigenvalues σ_i^2 ($i = 1, \dots, n$) which appear in (4.6) have an important effect on the optimization procedure and deserve a bit of discussion. Their significance lies in the fact that the value of a random variable with large variance is less certain and hence provides more information to the user. For the telemetry system to transfer as much information as possible it is necessary for the designer to take these variances into consideration by designing the quantizer to assign more bits to the samples with the larger variances.

A quantizer which operates on a vector of K.L. samples in a manner such that m_i bits are assigned to the i^{th} sample ($i = 1, \dots, n$) with not all m_i equal can be realized in two ways. One consists of a bank of quantizers operating in parallel with each quantizer designed for a different sample of the vector. The second is to first normalize each sample by dividing by the square root of the variance and using a sequential operation to quantize all samples with the same quantizer.

The latter realization appears more economical, but both methods have the same m.s.e. performance.

Returning to (4.6) we recall that the approximation error ϵ_a^2 is a function only of the number of samples n . One approach to the minimization of (4.6) is to select an n and minimize

$$\epsilon_*^2 = \sum_{i=1}^n \sigma_i^2 E(x_i - z_i)^2 \quad (4.7)$$

for this value of n . Then vary n and again minimize (4.7), etc. For n small ϵ_a^2 is large; however, m_b will be small and, therefore, more energy can be applied to each bit and ϵ_*^2 will be small. While for n large ϵ_a^2 is small but then m_b is large so that less energy can be applied to each bit. It follows that ϵ_*^2 is then large. Therefore, it is reasonable to expect that ϵ^2 will pass through a minimum for some moderate value of n . Another approach is to consider that we have at our disposal a BSC (memoryless) which operates at a bit rate of m_b bits per T seconds with a bit error probability p . In this case we have m_b bits to assign to n samples which represent the waveform $x(t)$ on the time interval $[0, T]$. We now determine that value of n and that assignment of the m_b bits to these n samples which minimizes ϵ^2 given by (4.6). Then assume another m_b and determine the n and corresponding assignment which minimizes ϵ^2 , etc. We use the first of these methods to obtain the optimum bit assignments presented in the following sections.

4.1.1 Optimum Bit Assignment. In this section we consider the problem of determining the assignment of m_b bits to n samples so as to minimize ϵ_*^2 given by (4.7). This problem is formulated as an integer nonlinear programming problem with linear constraints and a separable objective function in Appendix H[†]. In order to formulate the problem it was necessary to impose further specifications on the system. Those imposed were $N(0,1)$ data, optimum uniform quantizer, the natural code, a particular signal-to-noise ratio, and PSK reception. Unfortunately, general techniques for solving this problem are not available. Two alternatives suggest themselves. The first is to make an exhaustive search through all possible bit assignments for the best one and the second is to devise an algorithm which consistently provides nearly optimum bit assignments.

An exhaustive search was made for reasonable values of n and m_b . These results are presented in Table 4.1. To obtain the numerical values for ϵ_*^2 , the channel signal-to-noise ratio $\Gamma = ST/N_0$ was set at 100. The procedure followed was to determine that m_b and that bit assignment which minimizes ϵ_* of (4.7) for each n considered. Note that as the number of samples n increases, the optimum number of bits m_b increases; ϵ_*^2 also increases while ϵ_a^2 and ϵ^2 decrease. It is interesting that m_b does not increase nearly as fast as n . In fact, the ratio m_b/n , decreases from 5 for $n = 3$ to 2.5 for $n = 8$. Also note that the number of bits m_1 assigned to the sample having the largest variance decreases as the bits are spread over more samples.

[†]The terminology used here is that of Hadley [60].

Table 4.1
 Optimum Bit Assignments for K.L. Sampler, Optimum Uniform Quantizer
 Natural Code, and PSK Reception with $\Gamma = 100$

Number of Samples n	Optimum Number of Channel Bits m_b	Approximation Error ϵ_a^2	Quantizer and Channel Error ϵ_*^2	Overall System Error ϵ^2	Optimum Bit Assignments $m_1 m_2 m_3 m_4 m_5 m_6 m_7 m_8$
3	15	.07810	.00332	.08141	6 5 4 0 0 0 0 0
4	16	.05677	.00699	.06373	6 4 4 3 0 0 0 0
5	18	.04449	.01083	.05532	6 4 3 3 2 0 0 0
6	19	.03654	.01471	.05125	5 4 3 3 2 2 0 0
7	20	.03099	.01841	.04940	5 4 3 3 2 2 1 0
8	20	.02690	.02160	.04850	5 4 3 2 2 2 1 1

This is the effect of the real time and average power constraints on the optimum value of m_b and the bit assignment of these m_b bits to the n samples.

4.1.2 Suboptimum Bit Assignments. In the last section we determined the optimum number and assignment of channel bits for $n = 3, \dots, 8$. This required less than 10 minutes of IBM 7094 computer time; however approximately 30 minutes is required for $n = 9$ and the amount continues to increase exponentially with n . Because of this difficulty n could not be increased often enough to obtain the minimum value of ϵ^2 .

Therefore, suboptimum assignment procedures are sought. We obtain two. The first simply assigns an equal number of bits to each sample even though the samples vary in significance. The second uses an unequal assignment obtained by making simplifying assumptions and applying methods of calculus. As we shall see the second method performs excellently.

4.1.2.1 Equal bit assignment. For this procedure we first normalize the variance of the K.L. samples by dividing by their standard deviations and then quantize all n samples using the same optimum uniform quantizer[†]. A quantizer satisfying these constraints is obtained from Figure 3.5c by taking that m corresponding to the given $\beta = \Gamma/n$. For $\Gamma = 100$ and $n = 5$ we have $\beta = 13$ db and $m = 4$ or a 16 level quantizer. Both PSK and FSK reception for both the

[†]This method requires that $m_b = nm$, i.e. m_b is restricted to integral multiples of n .

natural and Gray codes were evaluated. These results are presented in Figure 4.1.

A number of other results are also evident in this figure. First note that as the signal-to-noise ratio Γ increases the optimum value of n increases. Note also that the Gray code performs uniformly better than the natural code and that its minimum occurs at larger values of n . Another result is observed when comparing the curves for PSK to those for FSK. For Γ fixed we see that the value of ϵ^2 for the optimum value of n and PSK reception is very close to one-half the value of ϵ^2 for the optimum n and FSK reception. For example, let $\Gamma = 800$ then for the natural code and PSK reception $\epsilon_{\min}^2 = 0.013$ at $n = 20$ while for FSK reception $\epsilon_{\min}^2 = 0.026$ at $n = 11$. This corresponds to 3 db of error energy. Observe though that an increase of more than 3 db in average transmitted energy is required for FSK to perform as well as PSK. In the remainder of the system performance measurements we consider only PSK reception and the natural code.

4.1.2.2 Unequal bit assignments. From Figure 4.2 for the natural code with PSK reception and $\Gamma = 100$ we see that $\epsilon_{\min}^2 = 0.064$ at $n = 5$. This is considerably larger than the minimum value $\epsilon^2 = 0.04850$ at $n = 8$ given in Table 4.1. Hence, we seek another suboptimum rule which performs closer to the optimum rule.

Observe that the curve for $\Delta = 13$ in Figure 3.13 is nearly the straight line described by the equation[†]

[†]The procedure followed here is that of Huang and Schultheiss [28].

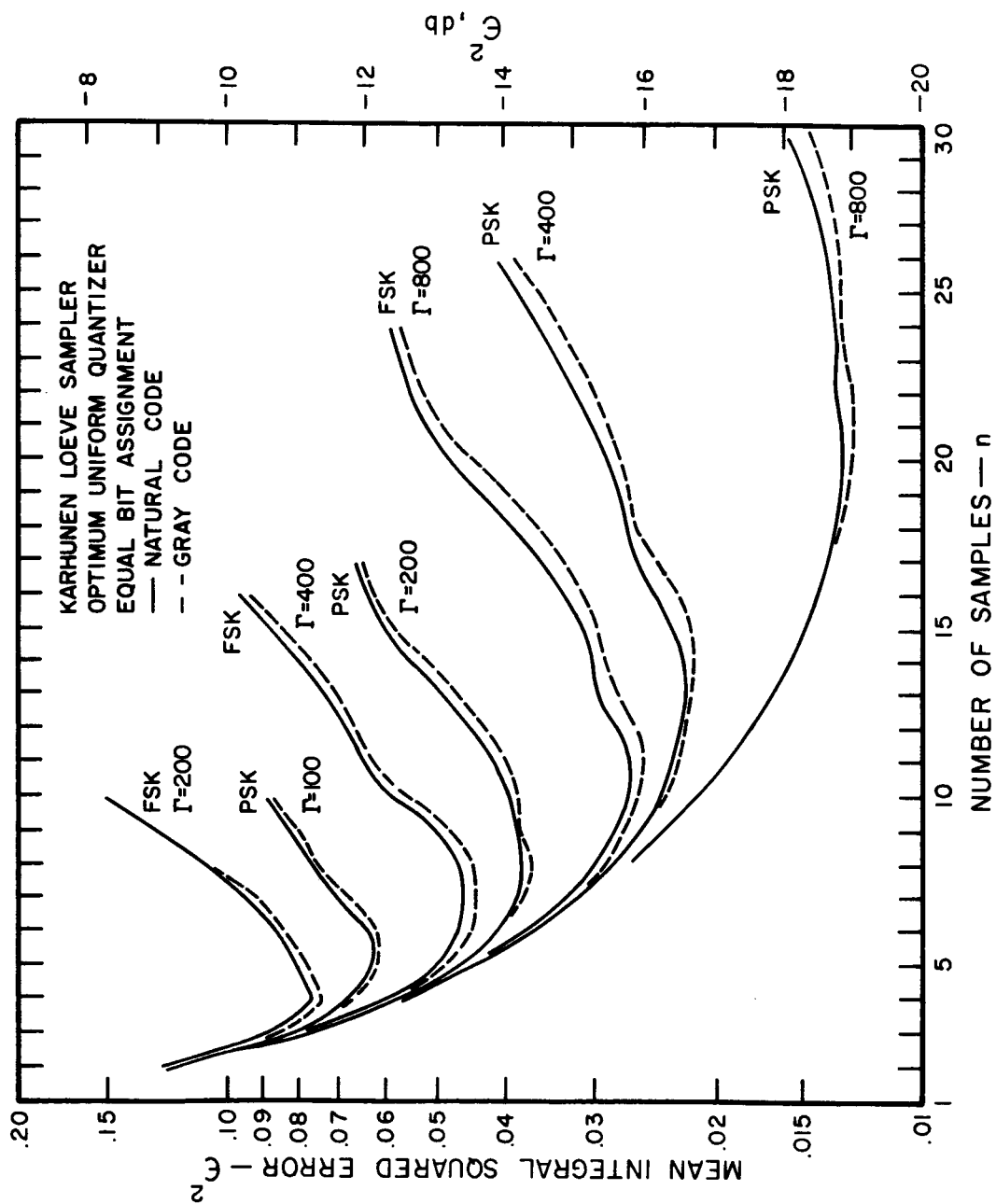


Figure 4.1 A Comparison of System Performances for Equal Bit Assignments

$$\log_{10}[\epsilon_{*1}^2(m)] = -m/2 \quad (4.8)$$

Replacing m by m_i in (4.8) and solving for $\epsilon_{*1}^2(m_i)$ we have

$$\epsilon_{*i}^2(m_i) = e^{-\frac{1}{2} m_i \ln 10} \quad (4.9)$$

Using this approximation in (4.7) we obtain

$$\epsilon_* = \sum_{i=1}^n \sigma_i^2 e^{-\frac{1}{2} m_i \ln 10} \quad (4.10)$$

Imposing the constraint

$$m_b = \sum_{i=1}^n m_i \quad (4.11)$$

then treating the m_i as continuous variables and introducing the Lagrange multiplier ξ we have

$$\begin{aligned} \frac{\partial}{\partial m_i} \left[\sum_{i=1}^n \sigma_i^2 e^{-\frac{1}{2} m_i \ln 10} + \xi \sum_{i=1}^n m_i \right] \\ = -\frac{1}{2} \sigma_i^2 \ln 10 e^{-\frac{1}{2} m_i \ln 10} + \xi = 0, \\ i = 1, \dots, n \end{aligned}$$

or

$$\sigma_i^2 e^{-\frac{1}{2} m_i \ln 10} = 2 \xi / \ln 10 = \text{const} \quad (4.12)$$

Solving for m_i we have

$$m_i = \frac{2}{\ln 10} [\ln(\text{const}) - \ln \sigma_i^2]$$

Then applying the constraint gives

$$m_b = \sum_{i=1}^n m_i = \frac{2}{\ln 10} [n \ln(\text{const}) - \sum_{i=1}^n \ln \sigma_i^2]$$

or

$$\ln(\text{const}) = \frac{1}{n} \left[\sum_{i=1}^n \ln \sigma_i^2 - \frac{\ln 10}{2} m_b \right]$$

Therefore

$$m_i = \frac{m_b}{n} + \frac{2}{\ln 10} \left(\ln \sigma_i^2 - \frac{1}{n} \sum_{i=1}^n \ln \sigma_i^2 \right) \quad (4.13)$$

The values for the m_i ($i = 1, \dots, n$) determined by (4.13) are next rounded to the nearest integer. From (H.7) we have

$$m_1 > m_2 \geq \dots \geq m_n$$

So, should the constraint (4.11) not be satisfied, we arbitrarily adjust some of the m_i according to the following rules:

- (i) If $m_b < \sum_{i=1}^n m_i$ take that m_i corresponding to the largest i such that $m_i > 1$ and replace it by $m_i - 1$.
- (ii) If $m_b > \sum_{i=1}^n m_i$ take that m_j corresponding to the smallest j such that $m_j = m_n$ and replace it by $m_j + 1$.

Equation (4.13) together with rules (i) and (ii) constitute a (suboptimum) algorithm which provides an unequal assignment of the m_b bits to the n samples.

This suboptimum bit assignment algorithm was programmed for the IBM 7094 computer and used to determine nearly optimum values of n and m_b . In this case we first selected a value for m_b , say m_b' ; then for each value[†] of $n \in \{[m_b'/5], \dots, [3m_b'/5]\}$, the unequal bit assignment

[†]The notation $[a/b]$ means the largest integer less than the real number a/b .

was obtained and $\epsilon_*^2(n, m_b^1)$ and $\tilde{\epsilon}^2(n, m_b^1)$ were evaluated. Then m_b was changed and the procedure repeated.

The values of ϵ_*^2 and $\tilde{\epsilon}^2$ actually represent a rectangular net above the n, m_b plane. In order to visualize this net tables were prepared with n as the ordinate, m_b as abscissa, and $\epsilon_*^2(n, m_b)$ or $\tilde{\epsilon}^2(n, m_b)$ on the field. These are presented in Tables 4.2 and 4.3 respectively. The system evaluated used a K.L. sampler, optimum uniform quantizer, natural encoder, and PSK reception with $\Gamma = 100$. Observe from Table 4.2 that for m_b fixed $\epsilon_*^2(n, m_b)$ increases monotonically with n . Also note that for n fixed there exists an m_b^* such that $\epsilon_*^2(n, m_b^*)$ is a minimum. These values are underscored in Table 4.2. With one exception ($n = 6$) we see that m_b^* is a nondecreasing function of n . The exception is probably due to the suboptimality of the assignment algorithm.

In Table 4.3 $\tilde{\epsilon}^2(n, m_b)$ is presented. This differs from Table 4.2 in that here the approximation error ϵ_a^2 is also accounted for. Again we see that for n fixed there exists an m_b^* such that $\tilde{\epsilon}^2(n, m_b^*)$ is a minimum. Note that the m_b^* are the same in Tables 4.2 and 4.3. This follows from the fact that the approximation error is not a function of the remainder of the system, i.e.

$$\tilde{\epsilon}^2(n, m_b) = \epsilon_a^2(n) + \epsilon_*^2(n, m_b) \quad (4.14)$$

Therefore, with n fixed $\tilde{\epsilon}^2$ varies only with m_b through ϵ_*^2 . In this case we see that for m_b fixed $\tilde{\epsilon}^2(n, m_b)$ does not increase monotonically with n . Rather there exists an n^* such that $\tilde{\epsilon}^2(n^*, m_b)$ is a minimum. We see that the overall minimum occurs for $(n, m_b) = (9, 21)$ with

Table 4.2

Mean-Square Error ϵ_*^2 for K.L. Samples, Optimum Uniform Quantizer, Unequal Bit Assignment,
Natural Code, and PSK Reception with $\Gamma = 100$

m_b n	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	.0024	.0014	.0010	<u>.0010</u>	.0013											
3	.0131	.0096	.0061	.0046	.0039	<u>.0033</u>	.0037	.0045	.0054	.0070						
4	.0268	.0217	.0159	.0141	.0092	.0080	<u>.0071</u>	.0077	.0072	.0082	.0097	.0117	.0131	.0152	.0175	
5	.0423	.0313	.0255	.0205	.0179	.0142	.0119	.0113	<u>.0108</u>	.0113	.0119	.0129	.0147	.0165	.0185	.0210
6	.0564	.0453	.0343	.0287	.0238	.0213	.0178	.0158	.0150	.0150	<u>.0150</u>	.0159	.0180	.0185	.0195	.0224
7		.0474	.0417	.0310	.0263	.0250	.0250	.0208	.0190	<u>.0184</u>	.0187	.0189	.0197	.0215	.0232	.0246
8				.0435	.0330	.0285	.0275	.0275	.0234	.0219	<u>.0216</u>	.0221	.0226	.0236	.0252	.0271
9					.0452	.0348	.0323	.0298	.0260	.0247	<u>.0247</u>	<u>.0247</u>	.0254	.0261	.0274	.0292
10							.0441	.0343	.0321	.0286	<u>.0275</u>	<u>.0275</u>	.0277	.0900	.0296	.0311
11									.0364	.0345	.0312	<u>.0304</u>	.0314	.0323	.0332	
12										.0387	.0370	<u>.0339</u>	.0349	.0358	.0357	
13													.0396	.0406	<u>.0380</u>	.0391

Table 4.3

Mean-Integral-Square Error for K.L. Sampler, Optimum Uniform Quantizer, Unequal Bit Assignment,
Natural Code, and PSK Reception with $\Gamma = 100$

m_b n	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	.1256	.1246	.1242	<u>.1242</u>	.1245											
3	.0912	.0877	.0842	.0827	.0820	<u>.0814</u>	.0818	.0826	.0835	.0851						
4	(.0836)	.0784	.0726	.0708	.0660	.0648	<u>.0632</u>	.0644	.0640	.0649	.0665	.0684	.0699	.0719	.0743	
5	.0868	(.0758)	(.0700)	(.0650)	.0624	.0587	.0564	.0558	<u>.0553</u>	.0558	.0564	.0574	.0592	.0610	.0630	.0655
6	.0930	.0818	.0709	.0652	(.0603)	.0578	(.0544)	.0523	.0515	.0515	<u>.0515</u>	.0525	.0545	.0550	.0570	.0590
7	.0784	.0727	.0620	(.0573)	.0560	(.0518)	(.0500)	<u>.0494</u>	.0497	.0499	.0507	.0525	.0542	.0556		
8		.0704	.0599	.0554	.0544	.0544	.0503	(.0488)	<u>.0485</u>	.0490	.0490	.0505	.0521	.0540		
9			.0689	.0586	.0560	.0560	.0535	.0498	(.0484)	<u>(.0484)</u>	.0492	(.0499)	.0511	.0529		
10						.0654	.0556	.0534	.0498	<u>.0488</u>	(.0490)	.0502	(.0509)	(.0523)		
11							.0557	.0537	.0505	<u>.0496</u>	.0506	.0515	.0524			
12								.0563	.0546	<u>.0515</u>	.0525	.0534	.0533			
13									.0558	.0568	<u>.0524</u>	.0553				

$\epsilon^2(9,21) = 0.0484$. Considering the entries $\epsilon^2(n, m_b)$ as a net above the n, m_b plane we see that a trough extends from the upper left portion of the table down through the lower right portion. Table 4.4 is analogous to Table 4.3; the only exception is that Max's uniform quantizer was employed. The entries in these tables are the same for m_b small, which corresponds to a high channel signal-to-noise ratio Δ . However, as m_b increases Δ decreases, and the ability of the optimum uniform quantizer to combat the noisy channel is evident. We point out that even when using Max's uniform quantizer we choose the m_i corresponding to the quantizer which performs best for that noise level. For the optimum uniform quantizer we go one step further and also select the optimum interval r .

The data presented in Tables 4.2, 4.3, and 4.4 is all for $\Gamma = ST/N_0 = 100$.

As the average channel signal energy is increased which corresponds to increasing Γ the system performance improves. This is illustrated in Figure 4.2. The unequal (suboptimum) bit assignment algorithm was employed. Results for the optimum uniform quantizer are shown by the dashed lines for $\Gamma = 100, 200$. For Max's uniform quantizer results were also obtained for $\Gamma = 400$. These are shown by the solid lines. Note that doubling Γ (increasing Γ by 3 db) reduces ϵ^2 by approximately 2.5 db (a reduction factor of .55). The effect of using the optimum quantizing interval r can be seen from the curve for $\Gamma = 200$. The minimum using Max's quantizer is

$$\epsilon_{\min}^2(n, m_b) = \epsilon^2(14, 34) = 0.0277$$

Table 4.4

Mean-Integral-Square Error for K.L. Sampler, Max's Uniform Quantizer, Unequal Bit Assignment
Natural Code, and PKS Reception with $\Gamma = 100$

m_b n	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	.1256	.1246	.1242	<u>.1242</u>	.1248											
3	.0912	.0877	.0842	.0827	.0820	<u>.0816</u>	.0823	.0835	.0850	.0872						
4	(.0836)	.0784	.0726	.0708	.0660	.0648	<u>.0642</u>	.0650	.0650	.0666	.0693	.0720	.0748	.0784	.0826	
5	.0868	(.0758)	.0700	(.0650)	.0624	.0587	.0565	.0564	<u>.0563</u>	.0573	.0589	.0610	.0636	.0668	.0705	.0755
6	.0930	.0818	.0709	.0652	(.0603)	.0579	(.0545)	.0525	<u>.0520</u>	.0530	.0538	.0578	.0595	.0617	.0658	.0680
7		.0784	.0727	.0620	(.0573)	.0561	(.0520)	(.0504)	<u>.0501</u>	.0520	.0533	.0554	.0591	.0628	.0667	
8			.0704	.0599	.0555	.0546	.0507	(.0495)	.0496	.0523	.0541	.0568	.0603	.0651		
9				.0689	.0587	.0562	.0539	.0505	(.0496)	(.0501)	.0538	(.0561)	.0593	.0633		
10					.0656	.0560	.0540	.0509	<u>.0505</u>	(.0515)	.0564	(.0590)	(.0627)			
11						.0564	.0548	.0521	<u>.0521</u>	.0568	.0596	.0628				
12							.0574	.0562	<u>.0540</u>	.0587	.0615	.0636				
13								.0582	.0629	.0622	.0656					

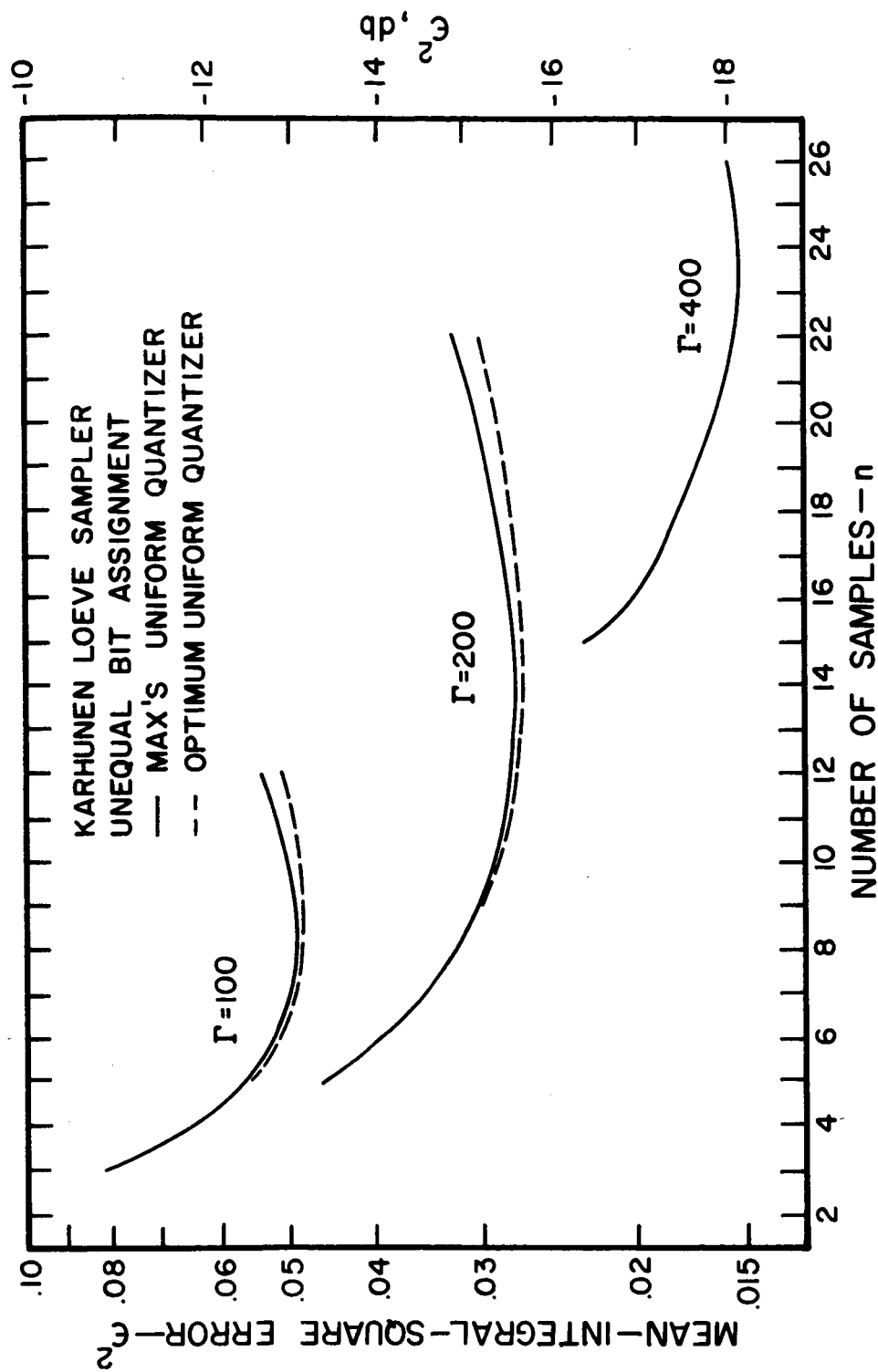


Figure 4.2 System Performance for K.L. Sampling and Unequal Bit Assignments with $\Gamma = 100, 200$, and 400

while for the optimum quantizer

$$\epsilon_{\min}^2(n, m_b) = \epsilon^2(15, 35) = 0.0271$$

and the improvement in performance is but 0.06 db. Corresponding to $\Gamma = 100, 200$, and 400 the optimum values of n are 8, 14, and 24 and of m_b are 19, 34, and 60 respectively. The reason for the increase is that the channel is less noisy and hence can handle a higher bit rate with no increase in the bit error rate.

So far in this chapter, we have determined the optimum bit assignments for K.L. sampling with n (the number of samples) small. We have also presented and evaluated two suboptimum bit assignment procedures. In the next section we will compare the performances of the suboptimum assignments to the optimum assignment.

4.1.3 Comparison of Bit Assignment Procedures. The performance of the optimum bit assignment procedure and these two suboptimum procedures are compared in Figure 4.3 for the example of Chapters II and III. Here the performance under the mean-integral-square error criterion is plotted as a function of the number of samples. Also plotted are ϵ_a^2 , which remains invariant with bit assignment, and the three curves for ϵ_*^2 corresponding to the three bit assignment procedures. Note that the equal bit assignment procedure yields significantly poorer performance especially for large values of n . For $n = 12$ the error ϵ^2 (equal bit assignment) exceeds ϵ^2 (unequal bit assignment) by 2.8 db. From Figure 4.3 we observe that the suboptimum unequal bit assignment procedure performs nearly identically to the optimum assignment procedure.

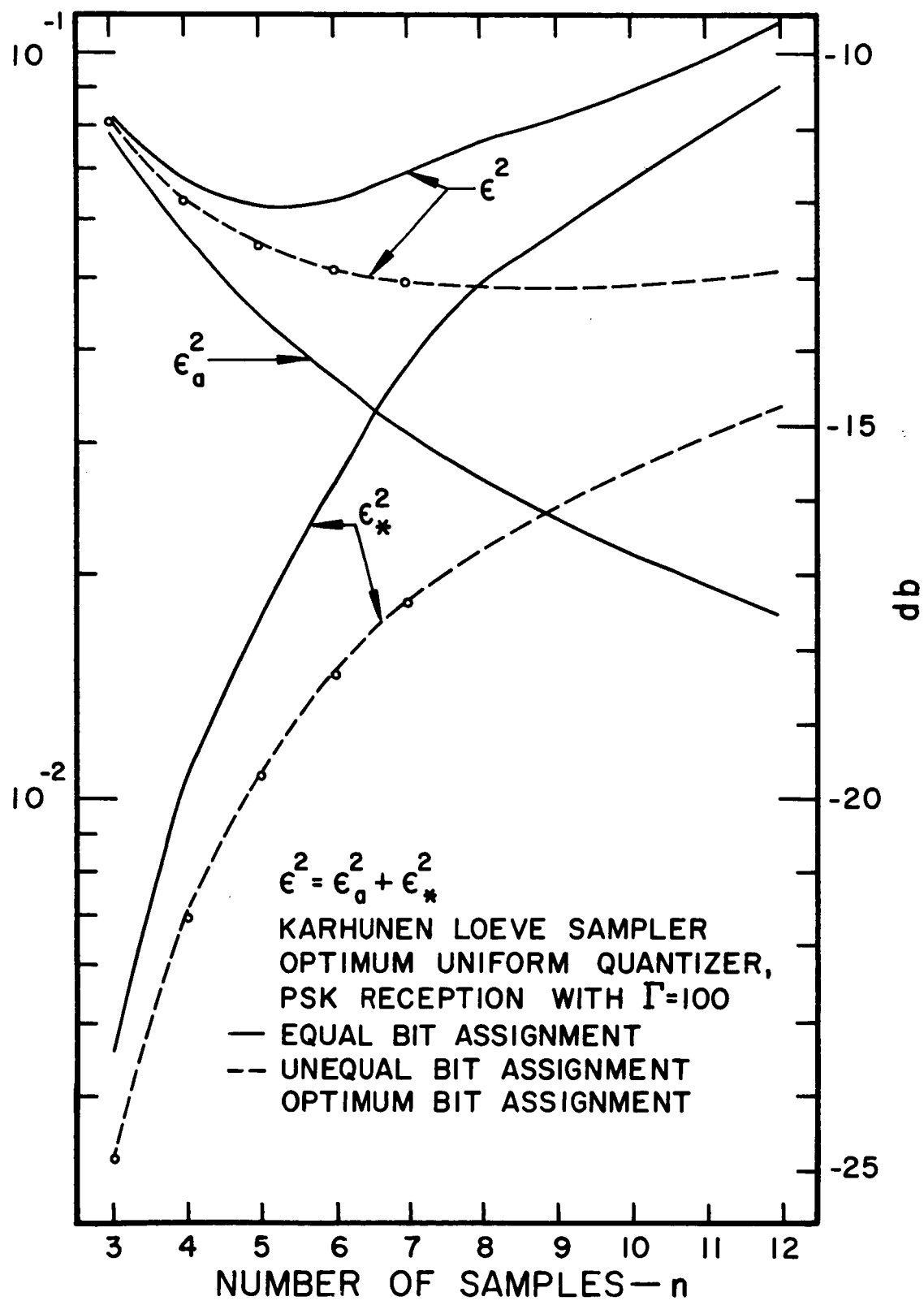


Figure 4.3 A Comparison of Bit Assignments

In order to further point out the fine performance of the suboptimum bit assignment algorithm presented in Subsection 4.1.2.2, we prepared Table 4.5. Note that for $n = 3, 5, 7$, and 8 the suboptimum unequal and optimum assignments have the same performance to 3 significant figures. For $n = 4$ and 6 the performances are also nearly the same.

The bit assignments given by the algorithm for $n = 3, 5$, and 8 were identical to the optimum assignments presented in Table 4.1. However, for $n = 4, 6$, and 7, m_b differs between the assignment procedures and consequently the assignments are different. The fact that different assignments perform so nearly the same indicates that the performance is not extremely sensitive to the bit assignment.

We have discussed the problem of choosing the best couple (n, m_b) so as to minimize the overall system error ϵ^2 for K.L. sampling. Numerical results for the example of the SGM source and the m^{th} extension of the BSC were obtained. We presented these using a number of graphs and tables. In this manner we also illustrated the sensitivity of the overall system performance to such things as receiver, quantizer, and encoder structures, bit assignment procedures, and average channel signal-to-noise ratio. In the next section we evaluate the overall performance of systems using time sampling.

4.2 Time Sampling

It is quite probable that all telemetry systems, telemetering analog signals, in operation today use time samplers. This is because a time sampler is so easy to realize while it is, in general, not feasible to build a K.L. sampler. Therefore, overall performance

Table 4.5

A Comparison of Bit Assignment Procedures for K.L. Samples
Optimum Uniform Quantizer, Natural Code, and PSK Reception with $\Gamma = 100$

Number of Samples per Sample Function n	Number of Bits per Sample Function = m_b					
	Equal Bit Assignment		Unequal Bit Assignment		Optimum Bit Assignment	
	m_b	ϵ^2	m_b	ϵ^2	m_b	ϵ^2
3	15	.0827	15	.0814	15	.0814
4	20	.0676	16	.0639	17	.0637
5	20	.0623	18	.0553	18	.0553
6	24	.0631	20	.0515	19	.0513
7	28	.0699	19	.0494	20	.0494
8	24	.0762	20	.0485	20	.0485
9	27	.0812	21	.0484		
10	30	.0885	21	.0488		
11	33	.0983	22	.0496		
12	36	.1106	22	.0515		

figures on conventional PCM telemetry, using time sampling, are valuable for comparative purposes.

Recall from Subsection 1.2.4, that numerous attempts have been made to analyze this system under the mean-integral-square error criterion. Viterbi's [47] analysis included all three of the error sources: sampler, quantizer, and channel. However, he assumed a bandlimited source process and uniformly distributed channel noise. Therefore, by Nyquist's criterion the sampler was errorless; also, Max's quantizer was optimum. For nonbandlimited sources the analysis is somewhat more complicated. Limited results have been obtained by inserting an ideal bandlimiting filter just ahead of the sampler [21].

We appreciate the difficulty of this problem and hence simulate the extended BSC and the SGM source on a digital computer. We estimate the performance of a system which employs time sampling and optimum uniform quantization. This system is illustrated in Figure 4.4. The waveforms of the source process are simulated in the same manner as for the example discussed in Section 2.4. This method is presented in Appendix B. The sampler, quantizer, data encoder, and user decoder were easily simulated requiring only a small interplay of real and integer arithmetic in the programming language, Fortran IV.

The natural code was used because of its compatibility with the computer. The additive channel disturbance was simulated by looking individually at the bits required for the binary representation of the integers $k-1$ ($k = 1, \dots, 2^m$) corresponding to the encoded version of the v_k which the random variable takes on. For each bit a biased coin was flipped and the outcome observed. On this coin a head

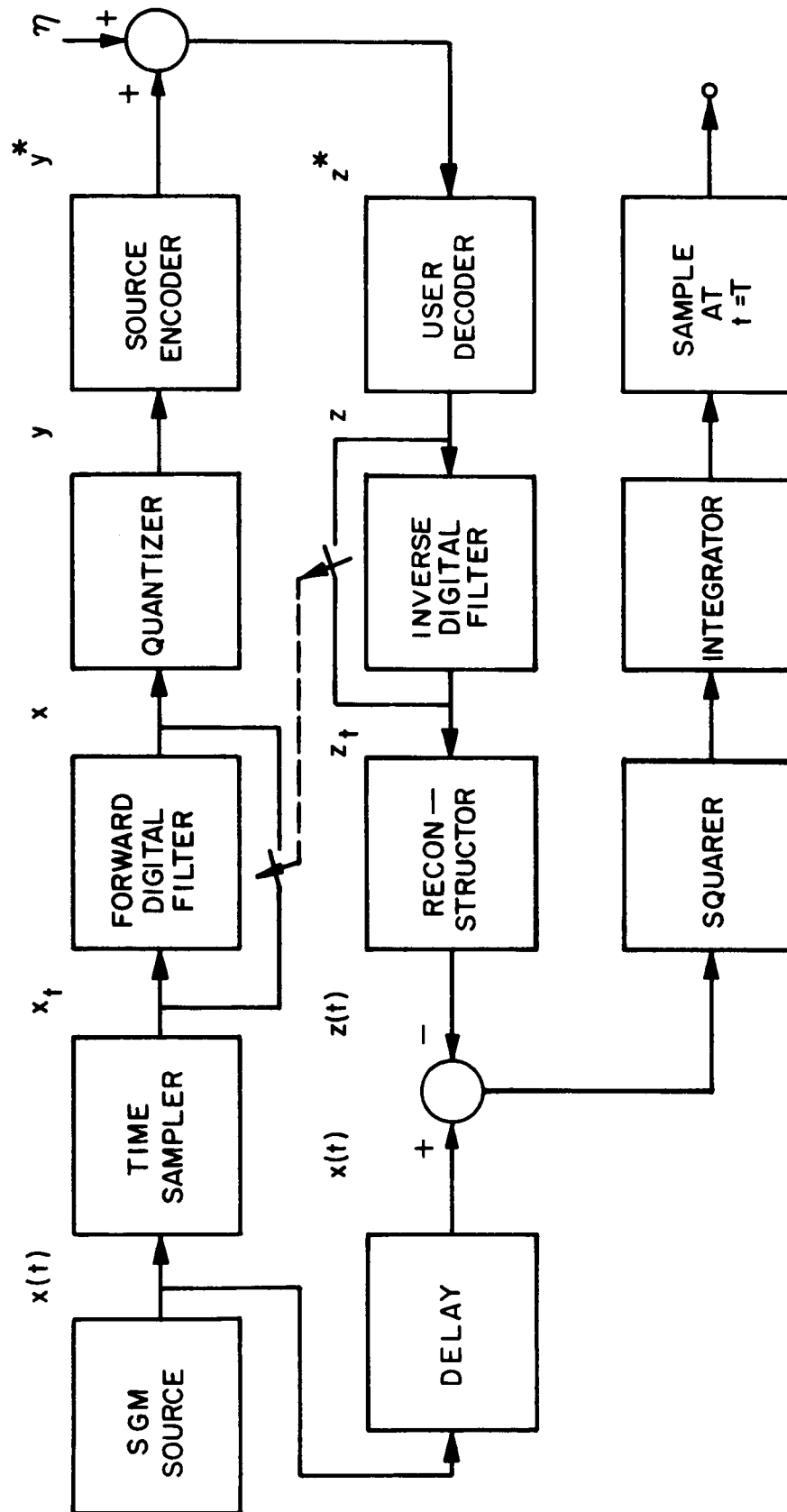


Figure 4.4 Block Diagram of Simulated PCM Telemetry System

appears with probability p and a tail with probability $(1-p)$. If a head was observed we considered that a channel error had occurred and the bit was changed to the other state. The coin flipping experiment was simulated by calling a psuedo-random number distributed uniformly over the interval $(0,1)$. If the called number was less than p a channel error was assumed to have occurred.

For the system without data compression (the switches in Figure 4.4 are closed) we reconstructed the waveform using linear interpolation (LI), $\sin(\pi x)/\pi x$, and zero order hold (ZOH) reconstruction. The performance of the conventional system (time sampling and no data compression) for these three waveform reconstruction techniques is presented in Subsection 4.2.1.

Recall that in Section 2.3, we presented a digital filtering technique for data compression of second order processes with known autocovariance functions. We saw that, for the approximation error ϵ_a^2 only, the time sampler, digital filter system performed nearly as well as the K.L. sampler. Hence, the overall performance of a system using this data compression technique is of interest. By opening the switches in Figure 4.4 we add the forward and inverse digital filters to the conventional system. Since these filters are realized as special purpose digital computers, it is a simple matter to simulate their operation on a general purpose digital computer. The overall performance of the system including the filters is presented in Subsection 4.2.2

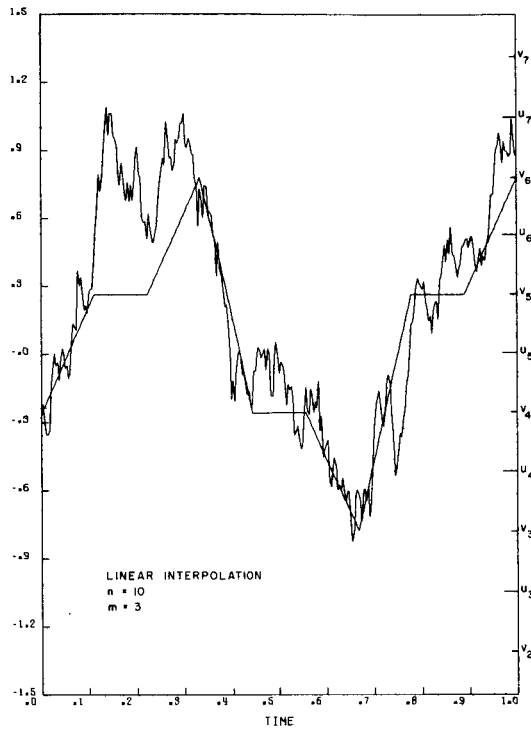
4.2.1 No Data Compression. Again, as in Section 2.4, to gain a better appreciation of the system operations we present sample functions of the input and output processes. In Figure 4.5 we present waveforms which illustrate the effect of the time sampler, quantizer, and reconstruction device (no channel errors). The input waveform is the same as the one presented in Figure 2.3. Ten samples taken on the input waveform were quantized using the optimum uniform eight level quantizer corresponding to a signal-to-noise ratio per channel bit of 3.33 ($\Delta = 3.33$).

For LI the output waveform (See Figure 4.5a) was obtained by connecting the adjacent values of the z_t vector (without data compression $z_t = z$). For ZOH each value of the z_t vector is held until the next arrives. While for $\sin(\pi x)/\pi x$ the points on the output waveform are obtained from the equation

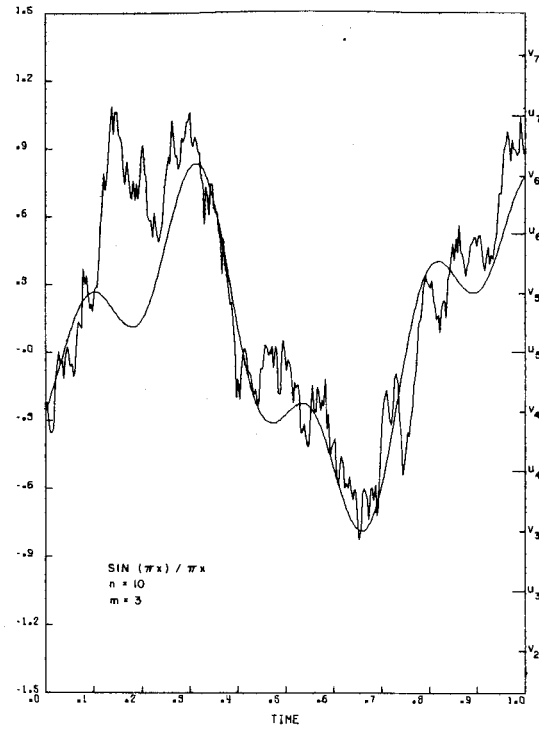
$$z(t_j) = \sum_{i=1}^{n_1} z_{t_i} \frac{\sin\left[\frac{\pi N}{T} \left(j \frac{T}{N} - i \frac{T}{n_1}\right)\right]}{\frac{\pi N}{T} \left(j \frac{T}{N} - i \frac{T}{n_1}\right)}, \quad j = 1, \dots, N \quad (4.15)$$

where in Figure 4.5 $N = 300$ and $n_1 = n = 10$.

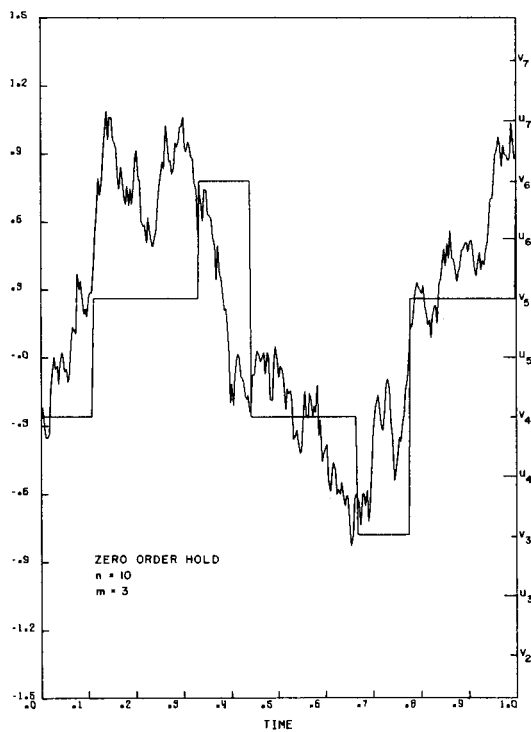
The absolute error at $t = 0.22$ sec. results from a slight dip of the input waveform below the transition level u_6 which occurred at that instant. The quantization interval of $r = 0.5195$ was used in this system. It was obtained from Figure 3.4a for $m = 3$ and $\beta = \Gamma/n = 10$. For the input waveform of Figure 4.5 the integral-square error corresponding to the different reconstructors is



(a)

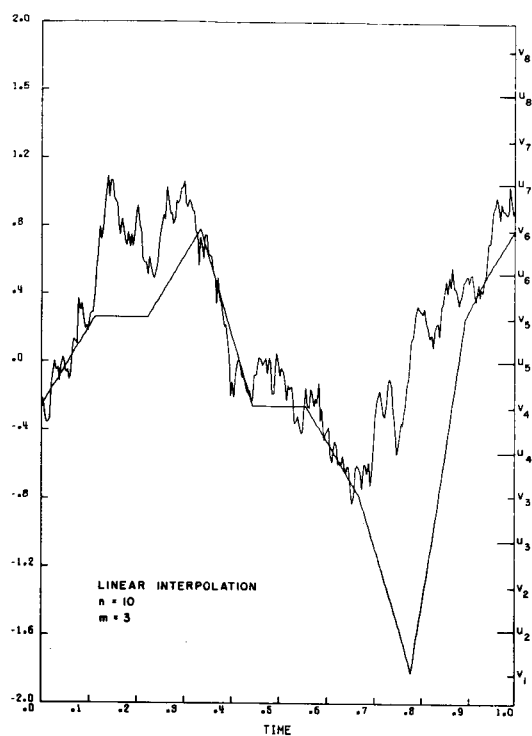


(c)

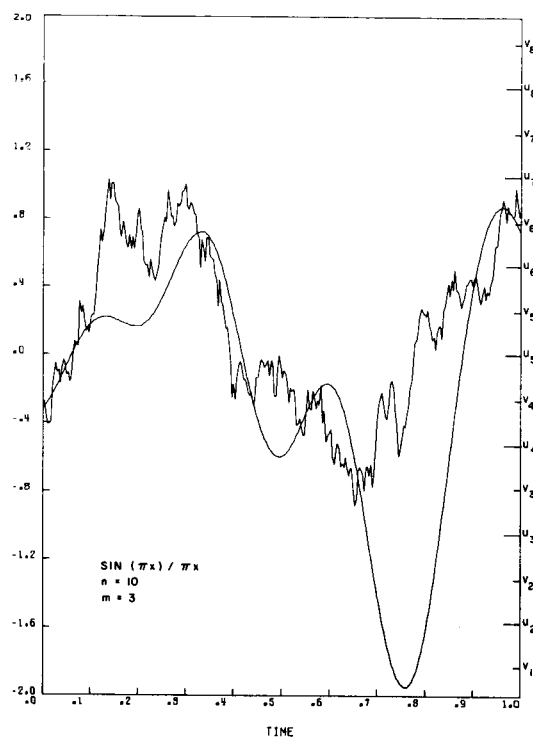


(b)

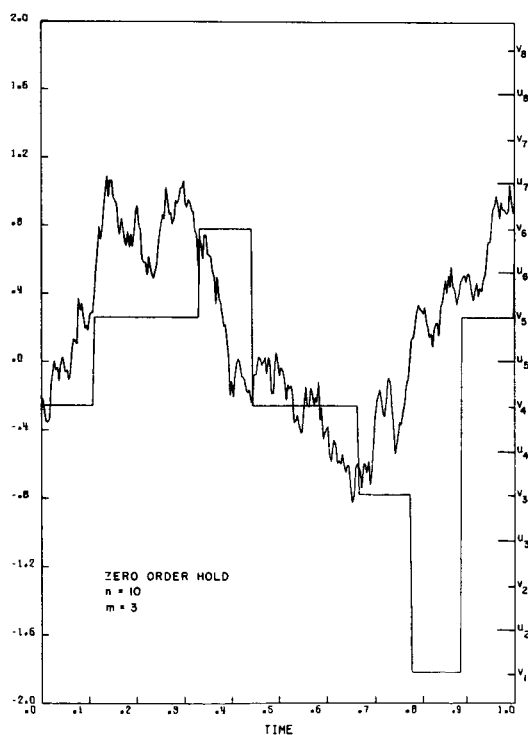
Figure 4.5 Time Sampling, Optimum Uniform Quantization, and Waveform Reconstruction



(a)



(c)



(b)

Figure 4.6 Time Sampling, Optimum Uniform Quantization, Noisy Channel, and Waveform Reconstruction

Linear interpolation	.06656
$\text{Sin}(\pi x)/\pi x$	.07588
Zero order hold	.19333

Recall that no bit errors were made during the transmission of the waveform presented in Figure 4.5.

In Figure 4.6 we present the output waveform obtained for a bit error on the 22nd bit. This is the first and most significant bit for the eighth sample (The time sample at $t = 0.78$ sec.). Note that the bit error affects the output waveform over one sample period for ZOH, two sample periods for LI, and over several sampling periods for $\text{sin}(\pi x)/\pi x$. For the same input waveform as in Figure 4.5 the integral-square errors of Figure 4.6 are

Linear interpolation	.34540
$\text{Sin}(\pi x)/\pi x$	.45865
Zero order hold	.69678

By comparing Figures 4.5 and 4.6 we note that the effect of a bit error is more pronounced than the effect of the quantization error for the situation illustrated. However, as is evident from Figure 3.3a, bit errors occur rather infrequently. Therefore, to assure reasonable estimates of the mean-integral-square error ϵ^2 using the Monte Carlo technique, it is necessary that the transmission of a large number of waveforms be simulated. We see from Figure 3.3a that for $\beta = 10$ and $m = 3$ the probability of a bit error $p = 0.005$. If 1000 different waveform transmissions are simulated then a total of 30,000 bits are required and we would expect 150 to be in error. This corresponds to five errors for each of the channel bits. The

time samples used to represent the source waveforms are all equally significant. Therefore we expect 50 bit errors for each of the 3 bit positions. We required that the expected number of errors per significant bit position be at least ten when estimating ϵ^2 . A different waveform was generated for each simulated transmission.

The estimates of ϵ^2 obtained by this simulation procedure are presented in Figure 4.7. The dashed line is for ZOH reconstruction with $\Gamma = 100$ and the three dotted lines are for LI reconstruction with $\Gamma = 100, 200$, and 400 . As is indicated by (4.15) considerably more computer time is required to reconstruct using the $\sin(\pi x)/\pi x$ method. For this reason estimates of ϵ^2 with $\sin(\pi x)/\pi x$ reconstruction were not obtained. Note that for $\Gamma = 100$ LI outperforms ZOH by a factor of $1/2(3 \text{ db})$. For LI we see that doubling the average signal-to-noise ratio Γ reduces ϵ^2 between 2 and 2.5 db. Also as Γ is increased the minimum value of ϵ^2 occurs at larger values of n . The reason being that more channel bits can be used with about the same bit error rate, hence more samples can be used and the approximation error is significantly decreased.

In this subsection the overall performance of a conventional PCM telemetry system has been estimated using the Monte Carlo method for a simulated system. Results were obtained for three different waveform reconstruction techniques.

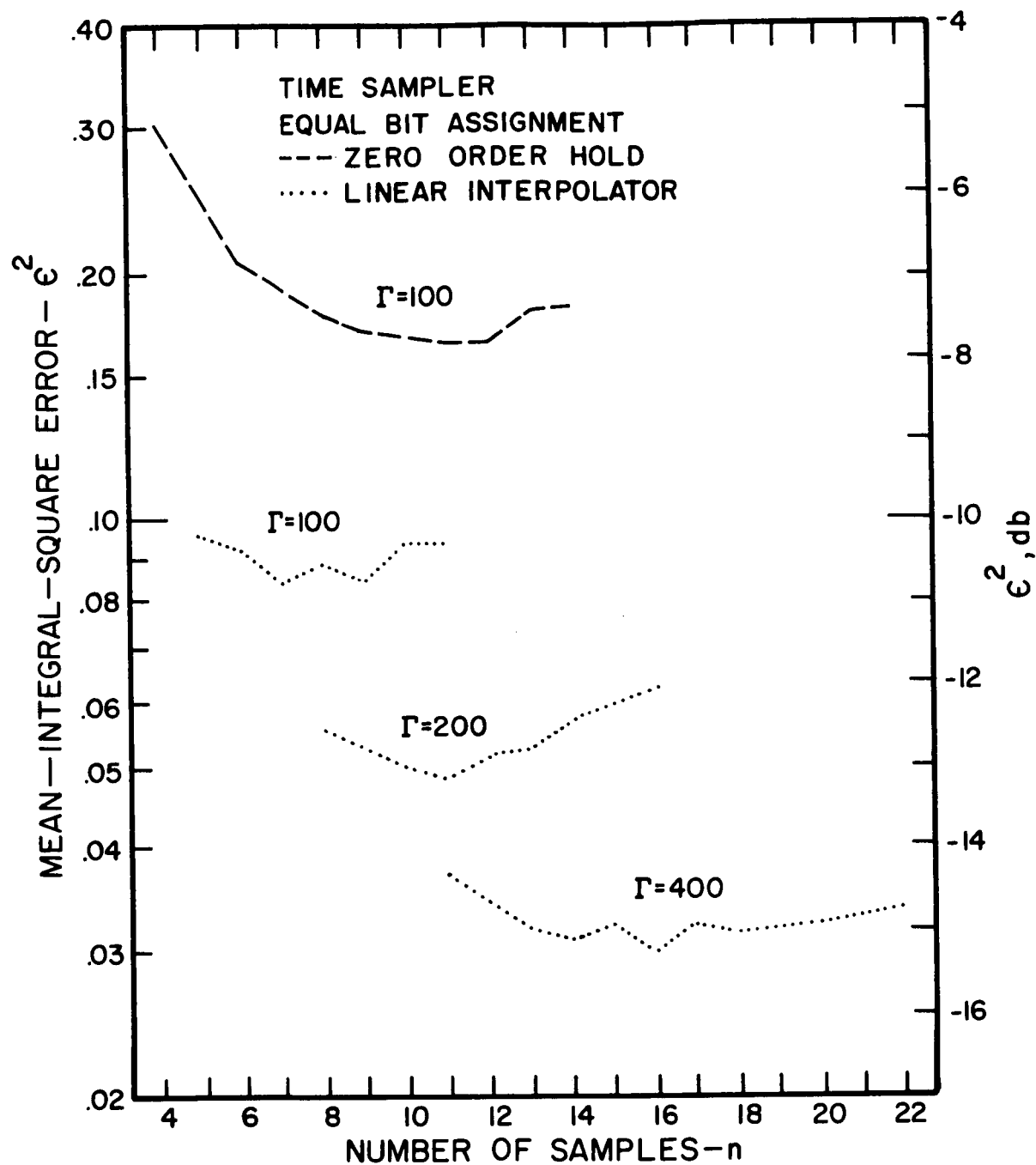
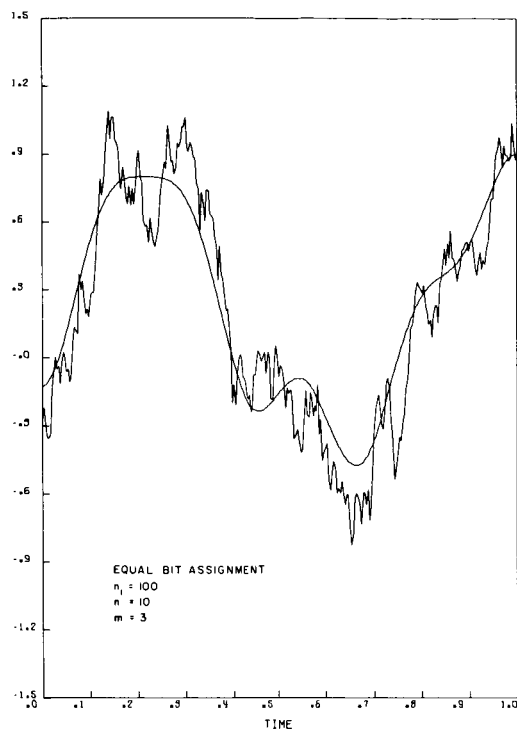


Figure 4.7 System Performance for Time Sampling and Equal Bit Assignments with $\Gamma = 100, 200$, and 400

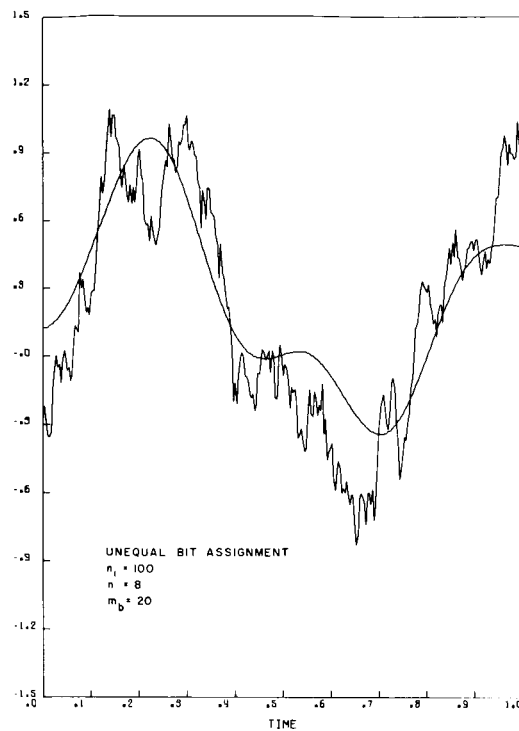
4.2.2 Data Compression. In this subsection we evaluate the performance of the system shown in Figure 4.4 incorporating the forward and inverse digital filters. The filters consist of matrix operations and would be realized as small digital computers so simulation on the digital computer was straight forward. Only linear interpolation is used for waveform reconstruction following the inverse filter.

Recall from Section 2.3 that the samples at the output of the forward filter are uncorrelated and can be ordered according to variance in a nonincreasing manner. This we do and, therefore, as in Section 4.2, the question of the correct bit assignment arises. We consider both equal and unequal assignments of the m_b channel bits per sample function to the n samples per sample function. The assignments used were those obtained for K.L. sampling in Section 4.1. These are presented in Table 4.1 for small values of n .

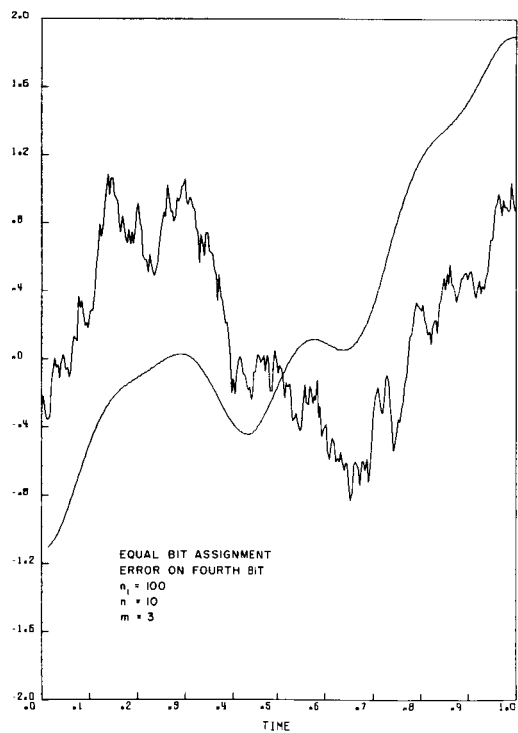
The effect of the quantizer on a single sample function for equal and unequal bit assignments is illustrated in Figures 4.8a and c. In each case the input waveform was obtained by connecting 300 points of which 100 were taken as time samples. In Figure 4.8a the 100 time samples are reduced to 10 by the filter and then all quantized using an eight level quantizer. While in Figure 4.8c the 100 time samples are reduced to 8 which are quantized using one 32, one 16, one 8, three 4, and two 2 level quantizers. This requires a total of 20 channel bits compared to 30 for the equal bit assignment above. These values of m_b and their assignments are optimum for the prevailing channel conditions. For the waveforms in both Figure 4.8a and 4.8c, the error is entirely due to sampling, filtering, and quantizing;



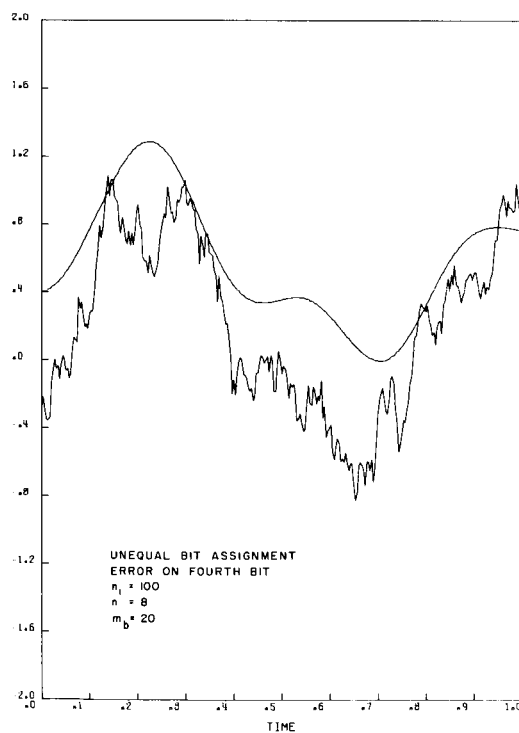
(a)



(c)



(b)



(d)

Figure 4.8 Time Sampling, Digital Filtering, Equal and Unequal Bit Assignments, and Perfect and Noisy Channel

no bit errors were made.

The integral-square errors in Figures 4.8a and c corresponding to the bit assignments are

Equal bit assignment ($n = 10$)	.03094
Optimum bit assignment ($n = 8$)	.06229

In general, even though 10 samples and 30 bits are used with the bits assigned equally, one would expect poorer performance than for 8 samples and 20 bits with the bits assigned optimally. The reason that the equal assignment procedure performed so well for this waveform is that the first and most significant sample was very close to a quantization value.

The effects of a bit error on the reconstructed waveform is illustrated in Figures 4.8b and d. The fourth bit was in error. This is the most significant bit of the second most significant sample for the equal bit assignment procedure. Had it been the very first bit its effect would be still more marked. For the assignment the bit in error was the fourth of five assigned to the first sample. The integral-square errors for the realizations in Figures 4.8b and d are

Equal bit assignment	.67729
Optimum bit assignment	.19801

The positions of the bit errors whose effects are illustrated in Figures 4.6 and 4.8 were selected arbitrarily. One could obtain waveforms illustrating the effect of a bit error in any position by overriding the decisions in the simulated channel. This, of course, was not done when estimating average performances.

4.3 System Comparisons

The average results for time sampling followed by digital filtering are presented in Figure 4.9. Also superimposed on these results are those presented in Figures 4.2 and 4.7. Note that the performance of the time sampling, digital filtering system using unequal bit assignments (The circles in Figure 4.9) is comparable to that for K.L. sampling and unequal bit assignments. The same system using equal bit assignments performs approximately 1 db poorer. These results are consistent with those for K.L. sampling presented in Figure 4.3. We also see that the error level, $\epsilon^2 = 0.05$, is obtained using only half the transmitter power required by the conventional system (time samples without the digital filters). This amounts to a tradeoff of average transmitter power for added system complexity consisting of two digital filters. Note also that the K.L. sampler with unequal bit assignments with $\Gamma = 200$ outperforms the conventional system with $\Gamma = 400$. For the former system

$$\epsilon_{\min}^2(n, m_b) = \epsilon^2(14, 34) = 0.0277$$

and for the latter system

$$\epsilon_{\min}^2(n, m_b) = \epsilon^2(16, 80) = 0.0291$$

Note that the number of channel bits m_b required by the conventional system is 2.35 times that for the system using K.L. sampling and unequal bit assignments. When this saving in bit rate is translated to bandwidth it amounts to a bandwidth reduction of 0.425. For $\Gamma = 100$ the optimum bit rates are 19 and 44 for K.L. sampling and conventional time sampling, respectively. From Figure 4.9, we see that the time

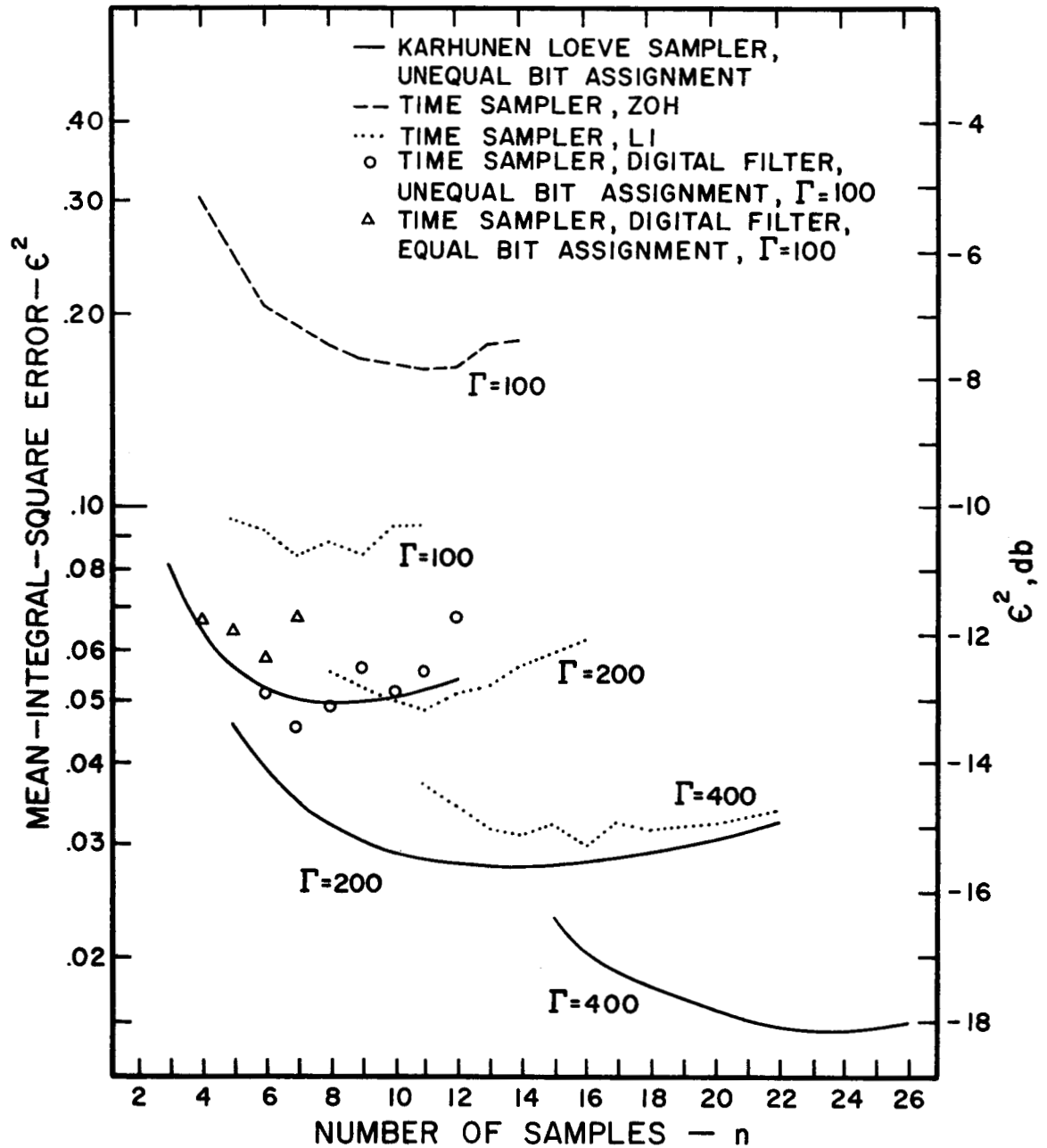


Figure 4.9 A Comparison of Different System Performances vs. n
with $\Gamma = ST/N_0$ as a Parameter

sampler followed by a digital filter performs comparable to the K.L. sampler. Hence, we conclude that adding the forward and inverse digital filters to the conventional system provides improvement in ϵ^2 while requiring only 1/2 as much transmitter power and channel bandwidth.

4.4 Information Transmission Rates

With the exception of Section 3.4.4 we have considered the mean-integral-square error criterion as our sole measure of performance. Another measure of performance, related to the mean-square error criterion, is Shannon's [59] rate distortion function $R(\epsilon^2)$. It gives the information in bits/second required to reconstruct discrete data with distortion less than or equal to ϵ^2 . Recall from (3.40) that for discrete $N(0,1)$ data

$$R(\epsilon_{*1}^2) = \frac{1}{2} \log_2(1/\epsilon_{*1}^2) \quad \text{bits/sample} \quad (4.16)$$

In this section we apply this measure to the systems discussed in the previous sections of this chapter. We compare their performances to the theoretical limit and to the performance of an idealized system.

The rate distortion concept has also been applied to the transmission of random processes ([61] - [63]). Gobllick [62] defines the quantity $R_T(\epsilon^2)$ as

$$R_T(\epsilon^2) = \min I(x_T; z_T) \quad (4.17)$$

$I(x_T; z_T)$ is the average mutual information between the input $x(t)$ and the output $z(t)$ $t \in (0, T)$ of a communication system. And the minimization in (4.17) is carried out over all possible communication systems. We

may interpret $R_T(\epsilon^2)$ as the minimum information capacity of the system necessary to have mean-integral-square error less than ϵ^2 for the interval $[0, T]$. The minimum required information capacity is then defined as

$$R(\epsilon^2) = \underset{T}{\text{g.l.b.}} R_T(\epsilon^2) \quad (4.18)$$

For analog waveforms $R(\epsilon^2)$ is the minimum rate of transmission of information necessary to achieve a mean-integral-square error, ϵ^2 .

When considering stationary Gaussian sources, the quantities ϵ^2 and $R(\epsilon^2)$ can be expressed in terms of a bandwidth parameter, φ .

Consider a source with a monotone decreasing power spectral density $S_X(f)$. Define the noise power spectral density $N(f)$ as

$$N(f) = \begin{cases} S_X(f), & |f| > \varphi \\ S_X(\varphi), & |f| \leq \varphi \end{cases} \quad (4.19)$$

Then the parametric equations for $R(\epsilon^2)$ and ϵ^2 are [61], [62]

$$R(\varphi) = \int_0^\varphi \log_2 \frac{S_X(f)}{S_X(\varphi)} df \quad (4.20)$$

$$\epsilon^2(\varphi) = 2\varphi S_X(\varphi) + 2 \int_\varphi^\infty S_X(f) df, \quad \varphi \geq 0 \quad (4.21)$$

For the SGM source with $R(\tau) = e^{-|\tau|}$, we have

$$S_X(f) = \frac{2}{1 + (2\pi f)^2} \quad (4.22)$$

so that (4.20) and (4.21) become

$$R(\varphi) = \frac{1}{\pi \ln 2} (2\pi\varphi - \tan^{-1} 2\pi\varphi) \quad (4.23)$$

$$\epsilon^2(\varphi) = 1 + \frac{2}{\pi} \left(\frac{2\pi\varphi}{1 + (2\pi\varphi)^2} - \tan^{-1} 2\pi\varphi \right) \quad (4.24)$$

The curve given by (4.23) and (4.24) is plotted in Figure 4.10. It gives the theoretical lower limit at which information must be transmitted to achieve the corresponding error performance.

Results are also available for a digitization system consisting of i) a bandlimiting filter, ii) a time sampler, iii) a uniform quantizer, iv) an entropy encoder and v) a noiseless channel (abbreviated as the BSQC encoder or digitizer) [63]. These results are presented in Figure 4.10 as the family of solid lines for the source process of our example. The parameter W_F is the bandwidth in Hz of the bandlimiting filter.

We have computed the overall mean-integral-square error (including the channel error) incurred in the transmission of sample functions from the same process over the noisy channel described above and recorded m_b (the optimum number of channel bits per sample function) corresponding to that error. The parameter m_b as used previously is the data rate (in bits/sec.) required, to transmit the source functions with mean-integral-square error ϵ^2 . These values are plotted as the three dotted lines in Figure 4.10 for the system employing time sampling, optimum uniform quantization, natural encoding, unequal bit assignments, and PSK reception with $\Gamma = 100, 200$, and 400. Note that for the low signal-to-noise ratio of $\Gamma = 100$ this system outperforms Goblick's BSQC digitizer. However, these results include the m.s.e. caused by bit errors on a noisy channel. Also, the entropy encoder of the BSQC digitizer or encoder assumes that information at the output

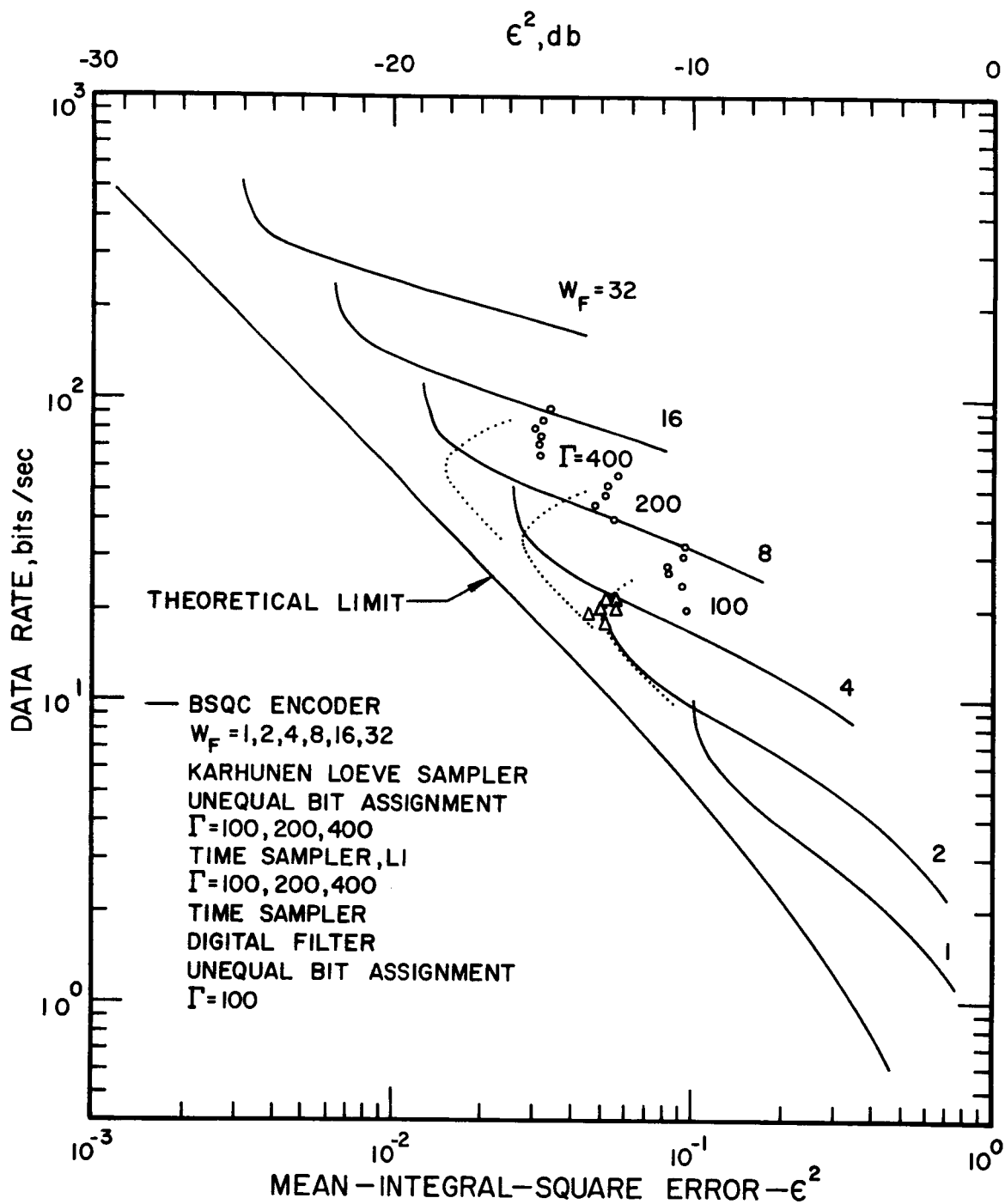


Figure 4.10 A Comparison of Data Rate vs. ϵ^2 for both Practical and Theoretical Systems

of the quantizer can be transmitted at a rate equal to its entropy. Such an encoder has never been physically realized. As Γ increases the data rate of the system whose performance we have evaluated appears to converge toward the theoretical limit while that for the BSQC encoder diverges. A reason for the apparent convergence is that the number of available channel bits for the same bit error rate is increased and the unequal bit assignment procedure uses these bits more efficiently than the BSQC encoder or digitizer. As an illustration, we see that for $\Gamma = 400$ and $\epsilon^2 = 2 \times 10^{-2}$ (-17 db) the theoretical limit is 28 bits/sec., the rate for the system we describe is 40 bits/sec., and the rate for the BSQC encoder is 52 bits/sec. This amounts to reducing the necessary channel bandwidth by a factor of 0.77.

The three sets of small circles in Figure 4.10 illustrate the data rates required to obtain the corresponding performance levels for the simulated system which utilizes time sampling and LI waveform reconstruction with no data compression. These are also for $\Gamma = 100, 200$, and 400. The improvement in both ϵ^2 and data rate of the system using K.L. sampling with unequal bit assignments over the conventional system for the same signal-to-noise ratio Γ is clear. The simulation results for time sampling, data compression, and unequal bit assignments are shown by triangles for $\Gamma = 100$. Again note how close these results are to those for K.L. sampling with unequal bit assignments.

Figure 4.10 together with Figure 4.9 summarize the results of system performance vs. the data rate and the number of samples.

We also must note from Figure 4.10 that, even though the system using K.L. samples and unequal bit assignments performs better than the others, it still requires a bit rate that is about 40% greater than the theoretical limit. For example with $\epsilon^2 = 0.02$ (-17 db),

$$R(\text{theoretical limit}) = 28$$

and

$$R(\text{K.L. sampler, unequal assign, } \Gamma = 400) = 40$$

which amounts to 43% more bits.

4.5 Summary

In this chapter we have considered the tradeoff between the number of samples per sample function and the number of bits per sample function. Our results are restricted to the SGM process with $R(\tau) = e^{-|\tau|}$. We formulated the optimum tradeoff problem as an integer nonlinear programming problem and then submitted a suboptimum assignment rule which assigns the m_b channel bits to the n samples. We evaluated the performance of this assignment procedure and compared it to the performance of the optimum assignment obtained by an exhaustive search procedure.

Estimated performance figures obtained by computer simulation for a conventional PCM system which utilizes time sampling and equal bit assignments are compared to those above obtained for K.L. sampling and unequal bit assignments. Simulation results were also used to demonstrate that time sampling followed by digital filtering can be utilized to approach the performance of K.L. sampling.

The data rates required to achieve the errors obtained were compared to theoretical limits determined using Shannon's rate

distortion theory. Finally, it was shown that the system we describe performs better using less transmitter power and less channel bandwidth than the conventional system.

CHAPTER V

CONCLUSIONS

The PCM telemetry system whose basic operations consist of sampling an analog waveform, quantizing the samples, and transmitting binary representations of the quantized samples over a noisy channel was analyzed under the mean-integral-square error criterion. Design criteria were established which minimize this error when the second order source statistics and the transition matrix of the channel are known.

Sampling procedures were discussed in Chapter II. The sampler was considered to provide a truncated generalized Fourier expansion of an input waveform to the system. Two samplers were given special attention: (i) the Karhunen Loeve (K.L.) sampler which is known to be best in the mean-integral-square error sense and (ii) the time sampler which is most practical. Then a "sampler" consisting of a time sampler followed by a digital filter was presented. These three samplers were evaluated for the stationary Gaussian Markov (SGM) process with $R(\tau) = e^{-|\tau|}$. It was shown that the "sampler"-time sampler, digital filter-we presented performs within 4% of the K.L. sampler. Moreover, it can be realized using existing hardware developed for digital computer fabrication.

In Chapter III we determined the optimum and optimum uniform quantizer structures for digital data with a known continuous probability density. To quantize optimally it is necessary to take into account the adverse effects of the channel over which the data is to be transmitted. This is clear from (3.11) and (3.12).

Both the optimum and optimum uniform quantizer structures were obtained for the case of numerical data distributed $N(0,1)$ with the channel disturbance that of additive, white, Gaussian noise. Both PSK and FSK receivers were evaluated for the quantized data source encoded using the natural code. Source encoding using the Gray code with PSK reception was also evaluated.

We note from Figure 3.7 that if the optimum number of channel bits is selected then the uniform quantizer of Max performs within 1 db of the optimum quantizer. However, it is evident from Figure 3.8 that, for the number of channel bits fixed, the choice of the quantizing interval is most significant. In applications where adaptive capabilities may be programmed into a system, either of these cases may be exploited to improve system performance.

We point out that it costs no more to build the optimum quantizer than a suboptimum one. The difficulty lies in determining the distribution of the data and the channel matrix. We have provided a design procedure for the optimum uniform quantizer when the data distribution and channel matrix above are used. We have demonstrated that neither the natural code nor the Gray code uniformly out performs the other for Gaussian data. We have also pointed out that one must design the quantizer for the channel even for nonreal time operation.

In Chapter IV we considered the tradeoff between the number of samples per sample function and the number of bits per sample function. Our results are restricted to the SGM process with $R(\tau) = e^{-|\tau|}$. We formulated the optimum tradeoff problem as an integer nonlinear programming problem and then submitted a suboptimum assignment rule which assigns the m_b channel bits to the n samples. We evaluated the performance of this assignment procedure and compared it to the performance of the assignment obtained by an exhaustive search procedure. The suboptimum assignments performed within 0.5% of the optimum assignments.

Estimated performance figures obtained by computer simulation for a conventional PCM system which utilizes time sampling and equal bit assignments were compared to those above, obtained for K.L. sampling and unequal bit assignments. The latter system yielded an improvement in overall system error of 2.5 to 3 db for the same average transmitted channel energy per source waveform. Moreover, this error reduction was accomplished using only $3/4$ of the channel bandwidth required by the conventional system.

Simulation results were also used to demonstrate that time sampling followed by digital filtering can be utilized to approach the performance of K.L. sampling. Therefore, we conclude that the system which results when adding forward and inverse digital filters to the conventional PCM telemetry systems and then taking into account the comparative significance of the resulting samples performs better using less transmitter power and less channel bandwidth than the conventional PCM system.

LIST OF REFERENCES

- [1] A. H. Reeves, U.S. Patent 2,272,070, assigned to International Standard Electric Corp.; February 3, 1942. Also, French Patent 852,183; October 23, 1939.
- [2] E. Maurice Deloraine and Alec H. Reeves, "The 25th Anniversary of Pulse Code Modulation," IEEE Spectrum, Vol. 2, No. 5, pp. 59-63, May, 1965.
- [3] B. M. Oliver, J. R. Pierce, and C. E. Shannon, "The Philosophy of PCM," Proceedings of IRE, Vol. 36, No. 11, pp. 1324-1331, November, 1948.
- [4] Proceedings of the 196n National Telemetering Conference, n = 0, 1, 2, 3, 4, 5, 6.
- [5] Proceedings of the Adaptive Telemetry Conference Goddard Space Flight Center, February 15-16, 1966.
- [6] W. B. Davenport, Jr. and W. L. Root, An Introduction to the Theory of Random Signals and Noise. New York: McGraw-Hill Book Company, Inc., 1958, p. 220, 99.
- [7] J. H. Laning, Jr. and R. H. Battin, Random Processes in Automatic Control. New York: McGraw-Hill Book Company Inc., 1956, pp. 179-180.
- [8] J. G. Truxal, Automatic Feedback Control System Synthesis. New York: McGraw-Hill Book Company Inc., 1955, p. 413.
- [9] S. Sherman, "Non-mean-square Error Criteria," IRE Transactions on Information Theory, Vol. IT-4, pp. 125-126, September, 1958.
- [10] Thomas S. Huang, "PCM Picture Transmission," IEEE Spectrum, Vol. 2, No. 12, pp. 57-63, December, 1965.
- [11] Thomas S. Huang, "Digital Picture Coding," Proceedings of the Annual National Electronics Conference, pp. 793-797, 1966.

- [12] Bernard Smith, "Instantaneous Companding of Quantized Signals," The Bell System Technical Journal, Vol. 36, pp. 653-709, May, 1957.
- [13] J. B. O'Neal, Jr., "Predictive Quantizing Systems (Differential Pulse Code Modulation) for the Transmission of Television Signals," The Bell System Technical Journal, pp. 689-721, May-June, 1966.
- [14] R. E. Totty and G. C. Clark, "Reconstruction Error in Waveform Transmission," IEEE Transactions on Information Theory, Vol. IT-13, April, 1967.
- [15] Joel Max, "Quantizing for Minimum Distortion," IRE Transactions on Information Theory, Vol. IT-6, No. 1, pp. 7-12, March, 1960.
- [16] S. P. Lloyd, "Least Square Quantization in PCM," Bell Telephone Laboratory Memo, Murray Hill, N.J. (unpublished).
- [17] A. Papoulis, "Error Analysis in Sampling Theory," Proceedings of the IEEE, Vol. 54, No. 7, pp. 947-955, July, 1966.
- [18] Gene Franklin, "Linear Filtering of Sampled Data," IRE Convention Record, Part 4, 1955.
- [19] S. P. Lloyd and B. McMillen, "Linear Least Squares Filtering and Prediction of Sampled Signals," presented at 1955 Symposium on Network Theory sponsored by the Polytechnical Institute of Brooklyn.
- [20] R. M. Stewart, "Statistical Design and Evaluation of Filters for the Restoration of Sampled Data," Proceedings of IRE, Vol. 44, pp. 253-257, February, 1956.
- [21] J. J. Spilker, Jr., "Theoretical Bounds on the Performance of Sampled Data Communications Systems," IRE Transactions on Circuit Theory, Vol. CT-7, pp. 335-341, September, 1960.
- [22] C. J. Palermo, R. V. Palermo, and H. Horwitz, "The Use of Data Omission for Redundancy Removal," Records of the International Space Electronics and Telemetry Symposium, pp. (11)D1 - (11)D16, 1965.
- [23] Allan I. Liff, "Mean-square Reconstruction Error," IEEE Transactions on Automatic Control, Vol. AC-10, No. 3, pp. 370-371, July, 1965.
- [24] V. A. Garmash, "Quantization of Signals with Non-uniform Steps," Electrosvyaz, Vol. 10, pp. 11-13, October, 1957.
- [25] A. I. Velichkin, "Optimum Characteristics of Quantizers," Radio Engineering, No. 2, pp. 1-7, February, 1963.

- [26] G. M. Roe, "Quantizing for Minimum Distortion," IEEE Transactions on Information Theory, (correspondence), Vol. IT-10, pp. 284-385, October, 1964.
- [27] Vidal R. Algazi, "Useful Approximations to Optimum Quantizations," IEEE Transactions on Communication Technology, Vol. COM-14, No. 3, pp. 297-301, June, 1966.
- [28] J. J. Y. Huang and P. M. Schultheiss, "Block Quantization of Correlated Gaussian Random Variables," IEEE Transactions on Communications Systems, Vol. CS-11, No. 3, pp. 289-296, September, 1963.
- [29] V. L. Yates-Fish and E. Fitch, "Signals Research and Development Establishment Signal/Noise Ratio in PCM," Proceedings of the Institution of Electrical Engineers, Vol. 102, Part B, pp. 204-210, March, 1955.
- [30] Joeseef Bado, "Coding for Least RMS Error in Binary PCM Channels," Wescon, TR62-3, May, 1962.
- [31] S. Karp, "Noise in Digital-to-analog Conversion Due to Bit Errors," IEEE Transactions on Space Electronics and Telemetry, Vol. Set-10, No. 3, p. 124, September, 1964.
- [32] Ian T. Young, "Least Mean-Square Error Analysis of PCM Transmission," Physical sciences research papers No. 62, Data sciences laboratory project 8600, Air Force Cambridge Research Laboratories, Office of Aerospace Research, USAF, L.G. Hanscom Field, Mass. AFCRL-64-864, October, 1964.
- [33] I. Dostis, "The Effects of Digital Errors on PCM Transmission of Companded Speech," The Bell Systems Technical Journal, pp. 2227-2243, December, 1965.
- [34] Y. Yamagushi and T. S. Huang, "Optimum Binary Fixed Length Block Codes," M.I.T. Research Laboratory of Electronics Quarterly Report, No. 78, pp. 231-233, July 15, 1965.
- [35] Y. Yamagushi and T. S. Huang, "Optimum Binary Code," M.I.T. Research Laboratory of Electronics Quarterly Research Report No. 79, pp. 214-217, October 15, 1965.
- [36] T. S. Huang, "Optimum Binary Code," M.I.T. Research Laboratory of Electronics Quarterly Research Report No. 82, pp. 223-224, July 15, 1966.
- [37] T. Crimmins, H. Horwitz, C. J. Palermo, and R. V. Palermo, "Minimization of Mean Square Error for Data Transmitted via Group Codes," International Symposium on Information Theory, 1967.

- [38] M. B. Rudin and J. Siegel, "Telemetry Systems Study Final Report," Aeronutronic Publication No. U-743, Vol. 1, Appendix 2, December, 1959,
- [39] Ye. V. Mitryayev, "Transmission of Telemetry Data by Means of Group Codes," Radio Engineering and Electronics Physics, Vol. 8, No. 6, pp. 935-940, June, 1963.
- [40] George C. Clark and R. E. Totty, "PCM Transmission with Minimum Mean-square Error," International Telemetry Conference Proceedings, Vol. 2, 1966.
- [41] Jacob Katzenelson, "On Errors Introduced by Combined Sampling and Quantization," IRE Transactions on Automatic Control, pp. 58-68, April, 1962.
- [42] K. Steiglitz, "Transmission of an Analog Signal over a Fixed Bit-rate Channel," IEEE Transactions on Information Theory, Vol. 12, No. 4, pp. 469-474, October, 1966.
- [43] L. M. Goodman, "Optimum Rate Allocation for Encoding Sets of Analog Messages," Proceedings of IEEE (correspondence), Vol. 53, pp. 1776-1777, November, 1965.
- [44] L. M. Goodman and P. R. Drouilhet, Jr., "Asymptotically Optimum Pre-emphasis and De-emphasis Networks for Sampling and Quantizing," Proceedings of IEEE (letters) Vol. 54, pp. 795-796, May, 1966.
- [45] L. M. Goodman, "Channel Encoders," Proceedings of IEEE (Letters) Vol. 55, pp. 127-128, January, 1967.
- [46] A. J. Viterbi, "Lower Bounds on Maximum Signal-to-noise Ratio for Digital Communication over the Gaussian Channel," IEEE Transactions on Communications Systems, Vol. CS-12, No. 1, pp. 10-17, March, 1964.
- [47] A. J. Viterbi, Principles of Coherent Communications, New York: McGraw-Hill Book Company, 1966, p. 259.
- [48] Michel Loeve, Probability Theory, 3rd Edition, Princeton, N. J., D. Van Nostrand Co., Inc., 1963.
- [49] J. L. Brown, Jr., "Mean-square Truncation Error in Series Expansions of Random Functions," J. Soc. Industrial Applied Mathematics, Vol. 8, No. 1, pp. 18-32, March, 1960.
- [50] D. Youla, "The Solution of a Homogeneous Weiner-Hopf Integral Equation Occurring in the Expansion of Second Order Stationary Random Functions," IRE Transactions on Information Theory, Vol. IT-3, pp. 187-193, September, 1957.

- [51] "Special Issue on Redundancy Reduction," Proceedings of IEEE, Vol. 55, No. 3, pp. 251-401, March, 1967.
- [52] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," J. of Educ. Psychology, Vol. 24, pp. 417-441, 498-520, 1933.
- [53] T. W. Anderson, An Introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons, Inc., 1962.
- [54] L. A. Zadah and C. A. Desoer, Linear System Theory--The State Space Approach. New York: McGraw-Hill Book Company, Inc., 1963.
- [55] T. N. E. Greville, "Some Applications of the Pseudo Inverse of a Matrix," Soc. Indust. Appl. Math., Vol. 8, No. 1, pp. 181-217, 1960.
- [56] P. E. Fleischer, "Sufficient Conditions for Achieving Minimum Distortion in a Quantizer," IEEE International Convention Record, Part 1, pp. 104-111, 1964.
- [57] J. C. Hancock and P. A. Wintz, Signal Detection Theory. New York: McGraw-Hill Book Company, Inc., 1966, pp. 51, 73.
- [58] N. Abramson, Information Theory and Coding. New York: McGraw-Hill Book Company, Inc., 1963, Chapter 5.
- [59] C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," IRE National Convention Record, Part 4, pp. 142-163, 1959.
- [60] G. Hadley, Nonlinear and Dynamic Programming, Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1964.
- [61] R. A. McDonald and P. M. Schultheiss, "Information Rates of Gaussian Signals under Criteria Constraining the Error Spectrum," Proceedings of IEEE, Vol. 52, pp. 415-416, April, 1964.
- [62] T. J. Gablick, Jr., "Theoretical Limitations on the Transmission of Data from Analog Sources," IEEE Transactions on Information Theory, Vol. IT-11, pp. 558-567, October, 1965.
- [63] T. J. Gablick, Jr. and J. L. Hoslinger, "Analog Source Digitization: A Comparison of Theory and Practice," IEEE Transactions on Information Theory, Vol. IT-13, No. 2, pp. 323-326, April, 1967.

- [64] T. J. Lynch, "Performance Measures for Compressed and Coded Space Telemetry Systems," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-3, No. 5, pp. 784-795, September, 1967.
- [65] A. J. Kurtenbach and P. A. Wintz, "Optimum Quantization," supplement to IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-3, No. 6, pp. 563-530, November, 1967.
- [66] T. P. Stanely and B. Liu, "Data Compression for Discrete Markov Signals," Proceedings of the National Electronics Conference, Vol. 23, pp. 534-539, 1967.
- [67] R. Bellman, Introduction to Matrix Analysis, New York: McGraw-Hill Book Company, Inc., 1960.

APPENDIX A

RELATIONSHIP BETWEEN APPROXIMATE K.L. SAMPLES
AND NORMALIZED PRINCIPAL COMPONENTS

The K.L. samples are the projections of the sample function to be represented on the eigenfunctions of the integral equation,

$$\sigma_i^2 \varphi_i(t) = \int_0^T R(t,s) \varphi_i(s) ds \quad (A.1)$$

These projections are given exactly by

$$x_i = \frac{1}{\sigma_i} \int_0^T x(t) \varphi_i(t) dt \quad (A.2)$$

and approximately by

$$x'_i = \frac{1}{\sigma_i} \frac{T}{N} \sum_{j=1}^N x(t_j) \varphi_i(t_j) \quad (A.3)$$

where the N samples are taken uniformly each T/N seconds. Equation (A.3) is merely a rectangular approximation to the integral in (A.2).

Let x_t be the vector of time samples and Φ be the $N \times n$ matrix with $n < N$ such that

$$\begin{aligned} \Phi_{ji} &= \frac{T}{\sigma_i N} \varphi_i(t_j), & j &= 1, \dots, N \\ & & i &= 1, \dots, n \end{aligned} \quad (A.4)$$

then

$$x' = \Phi^T x_t \quad (A.5)$$

is a vector of dimension n .

Consider also the first n samples obtained from (2.19) rewritten

here as

$$w_i' = \frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^N x(t_j) \beta_{ji} \quad (\text{A.6})$$

Again let x_t be the vector of time samples and define the $N \times n$ matrix

ψ as

$$\psi_{ji} = \frac{1}{\sqrt{\lambda_i}} \beta_{ji}, \quad \begin{array}{l} j = 1, \dots, N \\ i = 1, \dots, n \end{array} \quad (\text{A.7})$$

so that

$$w' = \psi^T x_t \quad (\text{A.8})$$

is again a vector of dimension n . The question at hand is the relationship between the components of x' and w' . By (A.3), (A.4), (A.6), and (A.7) we have

$$x_i' - w_i' = \sum_{j=1}^N x(t_j) (\phi_{ji} - \psi_{ji}) \quad (\text{A.9})$$

so that we need only to compare the weights of the two matrices ϕ and ψ as given by (A.4) and (A.7). Consider an approximate version of (A.1) evaluated at t_j , viz.

$$\sigma_i^2 \varphi_i(t_j) \doteq \frac{T}{N} \sum_{k=1}^N R(t_j, s_k) \varphi_i(s_k) \quad (\text{A.10})$$

which for stationary processes can be written as

$$\frac{N \sigma_i^2}{T} \varphi_i(t_j) = \sum_{k=1}^N R(t_j - t_k) \varphi_i(t_k) \quad (\text{A.11})$$

Compare (A.11) to the solution of the matrix equation[†]

$$\lambda_i \beta_i = \Sigma \beta_i \quad (\text{A.12})$$

for the components of the eigenvectors β_i , viz.

[†]For Σ the covariance matrix, we have $\Sigma = R(t_j - t_k)$
 $_{jk}$

$$\beta_{ji} \lambda_i = \sum_{k=1}^N R(t_j - t_k) \beta_{ki} \quad (\text{A.13})$$

For N large the approximation in (A.11) is good and hence by the uniqueness properties of the spectral representation of non-negative definite matrices we have

$$\frac{N \sigma_i^2}{T} \rightarrow \lambda_i \text{ as } N \rightarrow \infty \quad (\text{A.14})$$

and

$$K \varphi_i(t_j) \rightarrow \beta_{ji} \text{ as } N \rightarrow \infty \quad (\text{A.15})$$

where K is a normalizing constant. For the difference in (A.9) to approach zero

$$\bar{\Phi}_{ji} - \Psi_{ji} = \frac{T}{\sigma_i^2 N} \bar{\varphi}_i(t_j) - \frac{1}{\sqrt{\lambda_i}} \beta_{ji} \quad (\text{A.16})$$

must approach zero. Hence the normalizing constant is

$$K = 1/\sqrt{N/T}$$

Consider the example of the SMG process with covariance function

$$R(\tau) = e^{-|\tau|}$$

over the time interval $[0,1]$. This process has eigenvalues

$$\lambda_i = \frac{2}{1+b_i^2} \quad i = 1, 2, \dots \quad (\text{A.17})$$

and eigenfunctions

$$\varphi_i(t) = \frac{1}{C_i} \left(\frac{1}{b_i} \sin b_i t + \cos b_i t \right), \quad \begin{array}{l} t \in [0,1] \\ i = 1, 2, \dots \end{array} \quad (\text{A.18})$$

where the b_i are the solutions of

$$\tan b/2 = 1/b$$

and

$$\cot b/2 = -1/b \quad (\text{A.19})$$

with the normalizing constant squared

$$c_i^2 = \frac{1}{b_i^2} \sin^2 b_i + \frac{b_i^2 - 1}{4b_i^3} \sin 2b_i + \frac{1+b_i^2}{2b_i^2} \quad (\text{A.20})$$

Let the number of samples after the matrix operations (digital filter) be 5 ($n = 5$) and the number of time samples be 5 and 10 ($N = 5, 10$). The first five eigenvalues of the integral equation (A.1) and of the matrix equation (A.12) with Σ of dimension 5 x 5, 10 x 10, and 100 x 100 are presented in Table A.1.

Table A.1

Eigenvalues σ_i^2 of the Integral Equation and $\lambda_i T/N$ of the Matrix Equation for $R(\tau) = e^{-|\tau|}$

Source i	Integral Equation	Matrix Equation Σ 5x5	Matrix Equation Σ 10x10	Matrix Equation Σ 100x100
1	0.7388	0.7490	0.7414	0.7388
2	0.1380	0.1464	0.1401	0.1380
3	0.0451	0.0529	0.0469	0.0451
4	0.0213	0.0297	0.0232	0.0213
5	0.0123	0.0219	0.0141	0.0123
SUM	0.9555	0.9999	0.9657	0.9555

The vectors obtained by time sampling the eigenfunctions given by (A.18) compared to the eigenvectors obtained from (A.12) for $n = 5$ and $N = 5, 10$ are presented in Table A.2.

Table A.2

Comparison of $\varphi_i(t_j)$ and β_{ji} for $i=1, \dots, n$ and $j=1, \dots, N$

A.2.a $n = 5, N = 5$

$$\varphi_i(t_j)/\sqrt{N/T}$$

$\begin{matrix} i \\ j \end{matrix}$	1	2	3	4	5
1	0.3463	0.5200	0.5560	0.5659	0.5697
2	0.4130	0.4283	0.0423	-0.3810	-0.5710
3	0.4360	0.0000	-0.5624	0.0000	0.5715
4	0.4130	-0.4283	0.0423	0.3810	-0.5710
5	0.3460	-0.5200	0.5560	-0.5659	0.5697

$$B_{ji}$$

$\begin{matrix} i \\ j \end{matrix}$	1	2	3	4	5
1	0.4149	-0.5871	0.5334	0.3941	-0.2083
2	0.4621	-0.3941	-0.1599	-0.5871	0.5108
3	0.4782	0.0000	-0.6164	0.0000	-0.6256
4	0.4621	0.3941	-0.1599	0.5871	0.5108
5	0.4149	0.5871	0.5334	-0.3941	-0.2083

Table A.2 (cont'd.)

A.2.b $n = 5, N = 10$

$$\varphi_i(t_j)/\sqrt{N/T}$$

$\begin{smallmatrix} i \\ j \end{smallmatrix}$	1	2	3	4	5
1	0.2565	0.3852	0.4119	0.4192	0.4220
2	0.2822	0.3952	0.3483	0.2394	0.0988
3	0.3020	0.3403	0.1064	-0.1894	-0.3911
4	0.3154	0.2294	-0.1899	-0.4212	-0.2212
5	0.3222	0.0809	-0.3891	-0.2149	0.3219
6	0.3222	-0.0809	-0.3891	0.2149	0.3219
7	0.3154	-0.2294	-0.1899	0.4212	-0.2212
8	0.3020	-0.3403	0.1064	0.1894	-0.3911
9	0.2822	-0.3952	0.3483	-0.2394	0.0988
10	0.2565	-0.3852	0.4119	-0.4192	0.4220

$$\beta_{ji}$$

$\begin{smallmatrix} i \\ j \end{smallmatrix}$	1	2	3	4	5
1	0.2821	0.4175	-0.4298	0.4097	-0.3745
2	0.3041	0.3325	-0.2915	0.0982	0.1171
3	0.3210	0.3325	-0.0318	-0.2974	0.4438
4	0.3324	0.2190	0.2412	-0.4384	0.2458
5	0.3382	0.0764	0.4137	-0.2045	-0.3575
6	0.3382	-0.0764	0.4137	0.2045	-0.3575
7	0.3324	-0.2190	0.2412	0.4384	0.1438
8	0.3210	-0.3325	-0.0318	0.2974	0.4438
9	0.3041	-0.4017	-0.2915	-0.0982	0.1171
10	0.2821	-0.4175	-0.4208	-0.4097	-0.3745

First note the close agreement of the eigenvalues given in Table A.1. For the integral equation, only the first five of a countably infinite set are given and for the 10 x 10 and 100 x 100 covariance matrices the first five of 10 and 100; while for the 5 x 5 covariance matrix all five are given. This accounts for the behavior of the sums. The fact that the first few eigenvalues of the integral equation and the 100 x 100 matrix equation are the same to 4 significant places points out the very small penalty to be paid for the convenience of using matrix techniques. Compare the digital filter weights obtained by time sampled eigenfunctions as compared to the normalized eigenvectors of the covariance matrix as illustrated in Table A.2. Again we see that as N is increased the similarity increases. For the case of N = 100, the weights for i = 1 and j = 1 and 100 differ by less than 0.005. With j = 50 the difference

$$|\phi_i(t_j)/\sqrt{N/T} - \beta_{ji}| < .00001$$

These comparisons illustrate the speed of convergence for (A.14) and (A.15). They also point out that when designing a digital filter for data compression one need not solve an integral equation in order to derive the benefits of the discretized K.L. functions.

APPENDIX B

SIMULATION PROCEDURE FOR ESTIMATING THE
APPROXIMATION ERROR OF TIME SAMPLING

We desire a method of evaluating the "closeness" of a time waveform constructed from time samples taken from a random waveform to the actual random waveform itself. Many analytical techniques ([17] - [23]), both exact and approximate, have been suggested for certain special situations. We have simulated the system of Figure 2.1 on an IBM 7094 Digital Computer and evaluated the "closeness" of the two waveforms under the mean-integral-squared error criterion using Monte Carlo methods. The analog operations which were digitally simulated are illustrated in Figure B.1. The most difficult operation to simulate was that of generating a random waveform. Actually no waveform was generated. Rather a large array ($N \times 1$) of correlated numbers representing a very dense set of time samples taken on a sample function of the SMG process with autocorrelation was generated.

$$R(\tau) = e^{-|\tau|} \quad (B.1)$$

This, of course, does not include all possible realizations of the process, but only those which would be available at the output of an ideal filter bandlimited to $W_f = N/2T$ where N is the number of correlated variables in the large array and T the time interval for the sample function. The mean-integral-squared error resulting from this limitation is given by

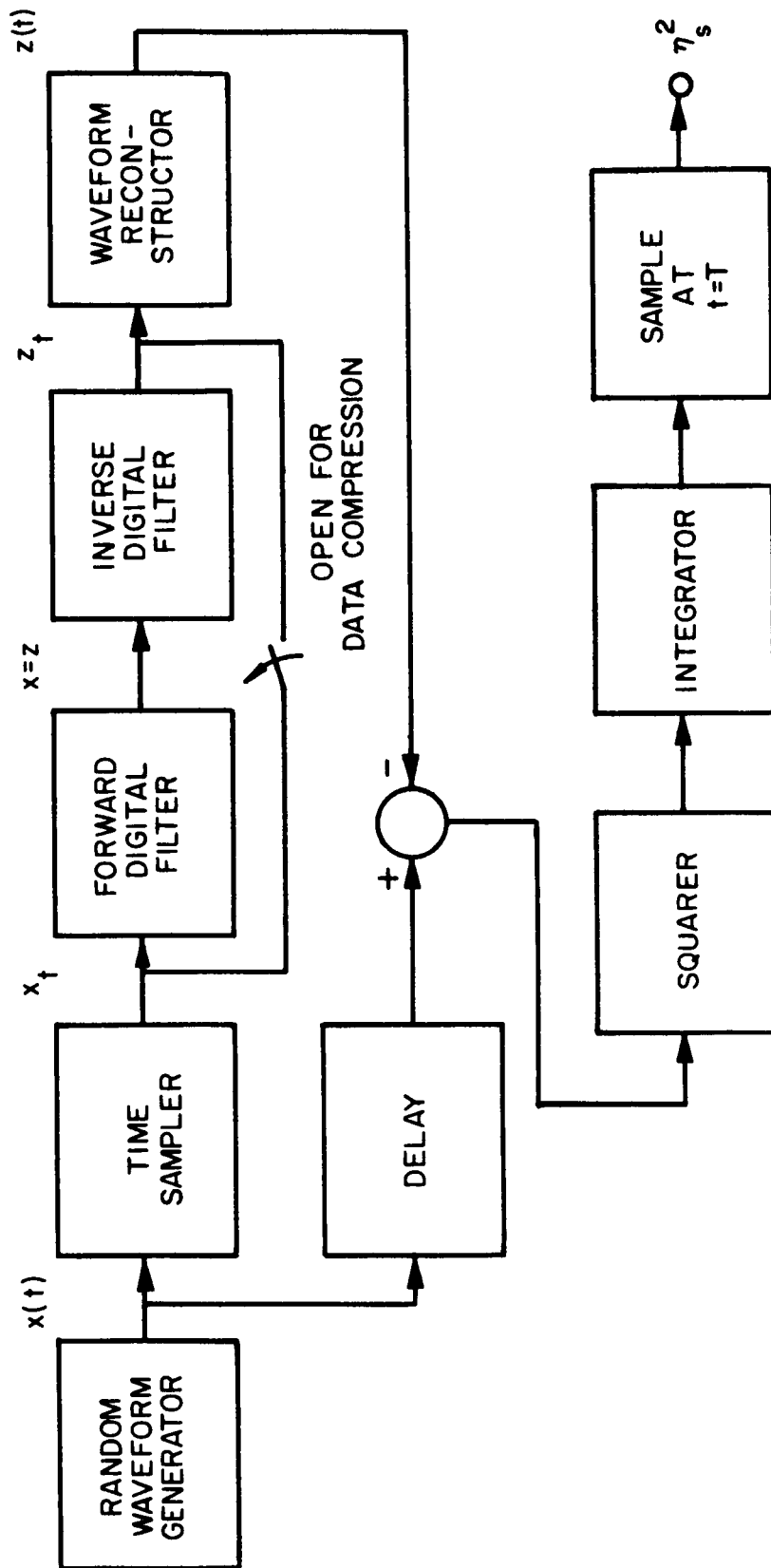


Figure B.1 Block Diagram of Simulated Time Sampler and Data Compression System

$$\eta_f^2 = 2 \int_{W_f}^{\infty} S(f) df \quad (B.2)$$

For the process of (B.1)

$$S(f) = \frac{2}{1 + (2\pi f)^2} \quad (B.3)$$

Substituting this $S(f)$ into (B.2) and letting $N = 300$, $T = 1$, and therefore $W_f = 150$, we obtain $\eta_f^2 = 0.0007$ which is insignificant compared to the approximation error plotted in Figure 2.2.

The dense array of properly correlated, simulated time samples results from correlating independent, zero mean, pseudo-random, Gaussian variables[†]. By applying the Markov property to the conditional density function for the j^{th} correlated variable q_j we have

$$p(q_j | q_1, \dots, q_{j-1}) = p(q_j | q_{j-1}) \quad (B.4)$$

Then applying Bayes rule to the Gaussian density we have

$$\begin{aligned} p(q_j | q_{j-1}) &= \frac{p(q_j, q_{j-1})}{p(q_{j-1})} \\ &= \frac{\frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\{-\frac{1}{2} \frac{q_{j-1}^2 - 2\rho q_{j-1}q_j + q_j^2}{(1-\rho^2)}\}}{\frac{1}{\sqrt{(2\pi)}} \exp\{-\frac{1}{2} q_{j-1}^2\}} \\ &= \frac{1}{\sqrt{(2\pi)} \sqrt{(1-\rho^2)}} \exp\{-\frac{1}{2} \frac{(q_j - \rho q_{j-1})^2}{(1-\rho^2)}\} \end{aligned} \quad (B.5)$$

where

$$\rho = R(T/N) = E(q_{j-1}q_j), \quad j = 2, \dots, N \quad (B.6)$$

Let r_j be distributed $N(0,1)$ and consider the transformation

[†]For a description of the computer routine used to generate the pseudo-random variables, see the Purdue University Computer Science Center Library Routine No. G5.01.4.

$$r_j = \frac{q_j - \rho q_{j-1}}{\sqrt{(1-\rho^2)}} \quad (\text{B.7})$$

Solving B.7 for q_j we have

$$q_j = (1-\rho^2)^{1/2} r_j + \rho q_{j-1} \quad (\text{B.8})$$

Our procedure now is to call a $N(0,1)$ pseudo-random variable, label it r_1 , and set $q_1 = r_1$; then call a second, label it r_2 , and compute q_2 from (B.8). We continue this procedure until the array of q_j ($j = 1, \dots, N$) is completed. Call this array q .

The time sampler in Figure B.1 is simulated by selecting n_1 of the q_j uniformly spaced in the index j . Call this array x_t . It represents the vector obtained by uniformly time sampling an input function $x(t)$. For data compression these n_1 samples are transformed into n samples by programming the matrix equation (2.20) and the n samples back to n_1 samples by (2.24). This is the array, call it z_t , from which the output time function $z(t)$ is constructed. For no data compression we have $n = n_1$.

Three methods of waveform reconstruction were simulated. The first, called zero order hold, simply holds the value of the i^{th} member of the n_1 dimensional array z_t for T/n_1 seconds, then the $i+1^{\text{st}}$, etc. The second, called linear interpolation merely connects adjacent values of the n_1 dimensional array z_t . Again only the points corresponding to the original N dimensional array of q_j 's are computed. Call this vector q^* . It represents N time samples taken on the reconstructed waveform $z(t)$. We approximate the integral

$$\eta_s^2 = \int_0^T [x(t) - z(t)]^2 dt \quad (\text{B.9})$$

using the vectors q and q^* and the trapezoidal approximation to integration. The random variable η_s^2 is then averaged by simply repeating the experiment and taking the average of all values taken on by η_s^2 . The third method of waveform reconstruction utilizes $\sin(\pi x) / \pi x$ functions to obtain the vector q^* . The elements of this vector are computed by

$$q_j^* = \frac{\sum_{i=1}^{n_1} z_{ti} \sin\left\{\pi \frac{N}{T} \left(j \frac{T}{N} - i \frac{T}{n_1}\right)\right\}}{\pi \frac{N}{T} \left(j \frac{T}{N} - i \frac{T}{n_1}\right)} \quad (B.10)$$

where the z_{ti} are the elements of the n_1 dimensional array presented to the reconstructor.

When using the Monte Carlo approach to estimate the mean value of a random variable we must have a measure of the necessary number of experiments to perform in order to have confidence in the estimate obtained. An estimate of the standard deviation of the sample mean was used as this measure. We now derive this estimate.

Define the following estimators:

$$\text{Sample mean} = \bar{x} \triangleq \frac{1}{L} \sum_{i=1}^L x_i \quad (B.11)$$

$$\text{Sample variance} = s^2 \triangleq \frac{1}{L-1} \sum_{i=1}^L (x_i - \bar{x})^2 \quad (B.12)$$

Let

$$E(x_i) < \infty \quad (B.13)$$

$$\mu = E(x_i)$$

$$\sigma^2 = \text{var}(x_i) = E(x_i - \mu)^2 \quad (B.14)$$

Then $\bar{x}$ and s^2 are unbiased estimators of μ and σ^2 .

The variance of the sample mean is by definition

$$\text{var}(\bar{x}) = E(\bar{x} - E(\bar{x}))^2 \quad (\text{B.15})$$

When the x_i result from independent trials, which we assume, it is a straight forward matter to show that

$$\text{var}(\bar{x}) = \frac{\sigma^2}{L} \quad (\text{B.16})$$

and hence the standard deviation

$$\text{S.D.}(\bar{x}) = \frac{\sigma}{\sqrt{L}} \quad (\text{B.17})$$

Substituting the sample variance for the true variance we have as an estimate of the S.D.

$$\hat{\text{S.D.}} = \left\{ \frac{1}{L} \frac{1}{L-1} \sum_{i=1}^L (x_i - \frac{1}{L} \sum_{i=1}^L x_i)^2 \right\}^{1/2} \quad (\text{B.18})$$

By replacing x_i by $\eta_{s_i}^2$ and letting L be the number of waveforms simulated, (B.18) is an estimate of convergence of the error estimate to the true value. For all of the simulation results used in Figure 2.2, L was taken sufficiently large so that

$$\text{S.D.} < 0.05 \hat{\epsilon}_a^2$$

and

$$\hat{\epsilon}_a^2 = \frac{1}{L} \sum_{i=1}^L \eta_{s_i}^2 \text{ was used as the approximation error.}$$

APPENDIX C

DERIVATION OF THE QUANTIZER DENSITY FUNCTIONS

The density functions needed to evaluate the expectation operation in (3.9) are derived in this appendix. First note that x_i is a continuous random variable, while y_i and z_i are discrete random variables. By allowing Dirac delta functions, we are able to obtain the probability density $p_{x_i, z_i}(X_i, Z_i)$. First consider the probability

$$\text{Prob}[x_i \leq X_i, y_i = v_k, z_i = v_\ell]$$

Using Bayes rule, we have

$$\begin{aligned} & \text{Prob}[x_i \leq X_i, y_i = v_k, z_i = v_\ell] \\ &= \text{Prob}[z_i = v_\ell | x_i \leq X_i, y_i = v_k] \text{Prob}[x_i \leq X_i | y_i = v_k] \\ & \quad \cdot \text{Prob}[y_i = v_k] \end{aligned} \tag{C.1}$$

The first term on the right-hand side (RHS) of (C.1) is

$$\begin{aligned} & \text{Prob}[z_i = v_\ell | x_i \leq X_i, y_i = v_k] \\ &= \begin{cases} \text{Prob}[z_i = v_\ell | y_i = v_k] \triangleq P_{k\ell}, & u_k < X_i \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{C.2}$$

The operation of the quantizer determines the last term on the RHS of (C.1) as

$$\text{Prob}[y_i = v_k] = \int_{u_k}^{u_{k+1}} p_{x_i}(X_i) dX_i \triangleq a_k \tag{C.3}$$

The second term on the RHS of (C.1) is now written as

$$\text{Prob}[x_i \leq X_i | y_i = v_k] = \frac{\text{Prob}[x_i \leq X_i, y_i = v_k]}{\text{Prob}[y_i = v_k]} \quad (\text{C.4})$$

where

$$\text{Prob}[x_i \leq X_i, y_i = v_k] = \begin{cases} 0 & , \quad X_i \leq u_k \\ \int_{u_k}^{X_i} p_{x_i}(X_i) dX_i & , \quad u_k < X_i \leq u_{k+1} \\ \int_{u_k}^{u_{k+1}} p_{x_i}(X_i) dX_i & , \quad u_{k+1} < X_i \end{cases} \quad (\text{C.5})$$

After substituting (C.2), (C.4), and (C.5) into (C.1) we have

$$\text{Prob}[x_i \leq X_i, y_i = v_k, z_i = v_\ell] = \begin{cases} 0 & , \quad X_i \leq u_k \\ P_{k\ell} \int_{u_k}^{X_i} p_{x_i}(X_i) dX_i & , \quad u_k < X_i \leq u_{k+1} \\ P_{k\ell} a_k & , \quad u_{k+1} < X_i \end{cases} \quad (\text{C.6})$$

Therefore the probability density $p_{x_i, y_i, z_i}(X_i, Y_i, Z_i)$ is

$$\begin{aligned} & p_{x_i, y_i, z_i}(X_i, Y_i, Z_i) \\ &= \sum_{\ell=1}^M \sum_{k=1}^M P_{k\ell} p_{x_i}(X_i) [1(X_i - u_{k+1}) - 1(X_i - u_k)] \\ & \cdot \delta(Y_i - v_k) \delta(Z_i - v_\ell) \end{aligned} \quad (\text{C.7})$$

where $1(X_i - b)$ is the unit step function and $\delta(Y_i - b)$ is Dirac's delta function. Integrating with respect to Y_i , we have

$$\begin{aligned}
 p_{x_i, z_i}(x_i, z_i) &= \sum_{k=1}^M \sum_{\ell=1}^M p_{k\ell} p_{x_i}(x_i) \\
 &\cdot [1(x_i - u_{k+1}) - 1(x_i - u_k)] \delta(z_i - v_\ell)
 \end{aligned}
 \tag{C.8}$$

and integrating further with respect to x_i , we have

$$\begin{aligned}
 p_{z_i}(z_i) &= \sum_{k=1}^M \sum_{\ell=1}^M p_{k\ell} \int_{u_k}^{u_{k+1}} p_{x_i}(x_i) dx_i \delta(z_i - v_\ell) \\
 &= \sum_{k=1}^M \sum_{\ell=1}^M p_{k\ell} a_k \delta(z_i - v_\ell)
 \end{aligned}
 \tag{C.9}$$

APPENDIX D

DERIVATION OF THE QUANTIZER STRUCTURES

D.1 The Optimum Structure

Necessary conditions for the m.s.e. in (3.10) to be a minimum are that the v_j ($j = 1, \dots, M$) and u_j ($j = 2, \dots, M$) satisfy the equations

$$\frac{\partial E[(x_i - z_i)^2]}{\partial v_j} = -2 \sum_{k=1}^M P_{kj} \int_{u_k}^{u_{k+1}} x_i p_{x_i}(X_i) dX_i \quad (D.1)$$

$$+ 2v_j \sum_{k=1}^M P_{kj} \int_{u_k}^{u_{k+1}} p_{x_i}(X_i) dX_i = 0, \quad j = 1, \dots, M$$

$$\frac{\partial E[(x_i - z_i)^2]}{\partial u_j} = -2u_j \left[\sum_{\ell=1}^M v_\ell (P_{j-1\ell} - P_{j\ell}) \right] p_{x_i}(u_j) \quad (D.2)$$

$$+ \left[\sum_{\ell=1}^M v_\ell (P_{j-1\ell} - P_{j\ell}) \right] p_{x_i}(u_j) = 0, \quad j = 2, \dots, M$$

These equations define the stationary points of the function

$$\epsilon_i^*(u, v) = E(x_i - z_i)^2 \quad (D.3)$$

for a given channel matrix, $[P_{k\ell}]$, input density function, $p_{x_i}(X_i)$, and number of quantization values, M . On inspection these equations exhibit certain isolation and coupling properties which are more evident in another form. To see this we solve the sets of equations (D.1) and (D.2) for the quantization values v_j ($j = 1, \dots, M$) and

transition levels u_j ($j = 2, \dots, M$) respectively and obtain

$$v_j = \frac{\sum_{k=1}^M P_{kj} \int_{u_k}^{u_{k+1}} x_i p_{x_i}(x_i) dx_i}{\sum_{k=1}^M P_{kj} \int_{u_k}^{u_{k+1}} p_{x_i}(x_i) dx_i}, \quad j = 1, \dots, M \quad (D.4)$$

$$u_j = \frac{\frac{1}{2} \sum_{\ell=1}^M v_\ell^2 (P_{j\ell} - P_{j-1\ell})}{\sum_{\ell=1}^M v_\ell (P_{j\ell} - P_{j-1\ell})}, \quad j = 2, \dots, M \quad (D.5)$$

We recall that (D.4) and (D.5) only are necessary conditions for the m.s.e. in (3.10) to be minimum. Sufficient conditions have not been obtained but it is reasonable to expect that they might follow in a manner similar to the case of the quantizer error only as obtained by Fleischer [56].

D.2 The Optimum Uniform Structure

In practice for ease of implementation, it is often desirable to require that the quantization values and transition levels be uniformly dispersed. When this constraint given by (3.15) is added the expression for m.s.e. (3.10) becomes

$$\begin{aligned} E[(x_i - z_i)^2] &= E(x_i^2) + \sum_{k=2}^{M-1} \left\{ -2 \left[\sum_{\ell=1}^M \left(\ell - \frac{M+1}{2} \right) r P_{k\ell} \right] \right. \\ &\quad \cdot \int_{(k-1-\frac{M}{2})r}^{(k-\frac{M}{2})r} x_i p_{x_i}(x_i) dx_i + \left. \left[\sum_{\ell=1}^M \left(\ell - \frac{M+1}{2} \right)^2 r^2 P_{k\ell} \right] \right\} \\ &\quad \cdot \int_{(k-1-\frac{M}{2})r}^{(k-\frac{M}{2})r} p_{x_i}(x_i) dx_i \} - 2 \left[\sum_{\ell=1}^M \left(\ell - \frac{M+1}{2} \right) r (P_{M\ell} - P_{1\ell}) \right] \end{aligned} \quad (D.6)$$

Equation (D.6) cont'd.

$$\cdot \int_{(M/2-1)r}^{u_{M+1}} X_i p_{X_i}(X_i) + \left[\sum_{\ell=1}^M \left(\ell - \frac{M+1}{2} \right)^2 r^2 (P_{M\ell} + P_{1\ell}) \right]$$

$$\cdot \int_{(M/2-1)r}^{u_{M+1}} p_{X_i}(X_i) dX_i$$

For the perfect channel (D.6) reduces to

$$E[(x_i - z_i)^2] = \sum_{k=2}^{M-1} \int_{(k-1-\frac{M}{2})r}^{(k-\frac{M}{2})r} [X_i - (k - \frac{M+1}{2})r]^2 p_{X_i}(X_i) dX_i$$

$$+ 2 \int_{(M/2-1)r}^{u_{M+1}} (X_i - \frac{M-1}{2}r)^2 p_{X_i}(X_i) dX_i \quad (D.7)$$

Using the following definitions[†]

$$f_M(r) = -f_1(r) \triangleq \int_{(M/2-1)r}^{u_{M+1}} X_i p_{X_i}(X_i) dX_i \quad (D.8)$$

$$g_M(r) = g_1(r) \triangleq \int_{(\frac{M}{2}-1)r}^{u_{M+1}} p_{X_i}(X_i) dX_i \quad (D.9)$$

$$f_k(r) \triangleq \int_{(k-1-\frac{M}{2})r}^{(k-\frac{M}{2})r} X_i p_{X_i}(X_i) dX_i \quad k = 2, \dots, M-1 \quad (D.10)$$

$$g_k(r) \triangleq \int_{(k-1-\frac{M}{2})r}^{(k-\frac{M}{2})r} p_{X_i}(X_i) dX_i \quad k = 2, \dots, M-1 \quad (D.11)$$

$$d_k \triangleq \sum_{\ell=1}^M \left(\ell - \frac{M+1}{2} \right) P_{k\ell} \quad k = 1, \dots, M \quad (D.12)$$

$$e_k \triangleq \sum_{\ell=1}^M \left(\ell - \frac{M+1}{2} \right)^2 P_{k\ell} \quad k = 1, \dots, M \quad (D.13)$$

[†]These definitions are for a symmetric density, $p_{X_i}(X_i)$.

we can write (D.6) and (D.7) more compactly as

$$E(x_i - z_i)^2 = E(x_i^2) + \sum_{k=2}^{M-1} \{-2r d_k f_k(r) + r^2 e_k g_k(r)\} \\ - 2r(d_M - d_1) f_M(r) + r^2(e_1 + e_M) g_M(r) \quad (D.14)$$

and for the perfect channel

$$E(x_i - z_i)^2 = E(x_i - y_i)^2 = E(x_i^2) \\ + \sum_{k=2}^{M-1} \{-2r(k - \frac{M+1}{2}) f_k(r) + r^2(k - \frac{M+1}{2})^2 g_k(r)\} \\ - 2(M-1)rf_M(r) + \frac{1}{2}(M-1)^2 r^2 g_M(r) \quad (D.15)$$

An implicit equation for r is obtained by equating the derivative of (D.14) with respect to r to zero. Using the definitions above plus the following two.

$$h_k(r) \triangleq p_{x_i}[(k-1-\frac{M}{2})r], \quad k = 1, \dots, M \quad (D.16)$$

and

$$w_k \triangleq (k-1-\frac{M}{2}), \quad k = 2, \dots, M \quad (D.17)$$

we have

$$\frac{dE(x_i - z_i)^2}{dr} = \sum_{k=2}^{M-1} \{-2d_k[r^2(w_{k+1}^2 h_{k+1}(r) - w_k^2 h_k(r)) + f_k(r)] \\ + e_k[r^2(w_{k+1} h_{k+1}(r) - w_k h_k(r)) + 2r g_k(r)]\} \\ + 2(d_M - d_1)[r^2 w_M^2 h_M(r) - f_M(r)] \\ - (e_1 + e_M)[r^2 w_M h_M(r) - 2r g_M(r)] \\ = 0 \quad (D.18)$$

For the case of the perfect channel (D.18) reduces to

$$\sum_{k=1}^M \left\{ -\left(k - \frac{M+1}{2}\right) f_M(r) + r \left(k - \frac{M+1}{2}\right)^2 g_k \right\} = 0 \quad (\text{D.19})$$

which was previously obtained by Max. Obtaining (D.19) by the method suggested above requires a bit of algebra; the same result can be obtained more readily by simply differentiating (D.7) with respect to r .

APPENDIX E

GENERATION OF CHANNEL MATRICES FOR USE IN ERROR ANALYSIS

When communication channels are analyzed under error criterion other than minimizing channel bit error, it is frequently necessary to have the channel matrix in hand. This need has been pointed out in recent papers [64], [65], and [66]. Consider, for example, the mean square error (m.s.e.) criterion applied to the telemetry of numerical data over a noisy channel. Let the source vocabulary consist of M words representing quantized values of a continuous variable. If we constrain the output vocabulary to the same M words labeled $v_1, \dots, v_M$, then the channel transition probabilities are

$$P_{k\ell} = \text{Prob}(z = v_\ell / y = v_k) \quad k, \ell = 1, \dots, M \quad (\text{E.1})$$

where z is the output variable and y the input variable. The entries $P_{k\ell}$ ($k, \ell = 1, \dots, M$) constitute the channel matrix. And the m.s.e. is

$$\epsilon_c^2 = E(x-y)^2 = \sum_{k=1}^M \sum_{\ell=1}^M (v_\ell - v_k)^2 P_{k\ell} a_k \quad (\text{E.2})$$

where

$$a_k = \text{Prob}(y = v_k) \quad (\text{E.3})$$

In order to evaluate ϵ_c^2 it is necessary to have the channel matrix

$P = [P_{k\ell}]$ in hand. For M large, say $M = 2^m$ with $m = 9$, which corresponds to using 9 information bits per channel word, the matrix P is a 512 x 512

matrix with $262,144$ entries.

Let us restrict the vocabulary size to an integral power of two and source encoding according to the natural code, vis.

v_1 is coded as 00 ... 000

v_2 is coded as 00 ... 001

v_3 is coded as 00 ... 010

etc.

Then transmitting over the m^{th} extension of the BSC (memoryless), we obtain the channel matrix

$$P^{(m)} = [P_{k\ell}^{(m)}] \quad (\text{E.4})$$

where

$$P_{k\ell}^{(m)} = \text{Prob}[z = v_\ell | y = v_k \text{ and } m \text{ channel bits per source word}] \quad (\text{E.5})$$

The matrix $P^{(m)}$ has nice properties as pointed out by Abramson [58] and Bellman [67], vis. that

$$P^{(1)} = \begin{bmatrix} q & p \\ p & q \end{bmatrix}, \quad q = 1 - p$$

and

$$P^{(m)} = \begin{bmatrix} q P^{(m-1)} & p P^{(m-1)} \\ p P^{(m-1)} & q P^{(m-1)} \end{bmatrix}, \quad m = 2, 3, \dots \quad (\text{E.6})$$

For this class of channel matrices (E.2) becomes

$$\epsilon_c^2 = \sum_{k=1}^{2^m} \sum_{\ell=1}^{2^m} (v_k - v_\ell)^2 P_{k\ell}^{(m)} a_k \quad (\text{E.7})$$

In practice it is desirable to compute this error for various values of the probability of bit error, p . To facilitate the computations it was noted that the entries in the channel matrix $P^{(m)}$ are all of the form

$$P_{k\ell}^{(m)} = p^w q^{m-w}, \quad w = 0, \dots, m \quad (E.8)$$

The r.v. w is the weight of the channel error pattern necessary to permute v_k into v_ℓ , i.e. it is the Hamming distance (the weight of the binary vector) between the representations of v_k and v_ℓ . Call the matrix of these weights $W^{(m)}$, then

$$W^{(1)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and

$$W^{(m)} = \begin{bmatrix} W^{(m-1)} & 1+W^{(m-1)} \\ 1+W^{(m-1)} & W^{(m-1)} \end{bmatrix}, \quad m = 2, 3, \dots \quad (E.9)$$

where 1 is a matrix of the appropriate size with all entries one. Hence in performing the computations one need only generate and store the weight matrix $W^{(m)}$. Since w is generally a small integer, $W^{(m)}$ can be stored using only a few bits per entry. Furthermore, it is clear that for $m > m_1$, the mapping (E.9) is such that each element $w_{k\ell}^{(m)}$ has a unique inverse image in the array $W^{(m_1)}$. Therefore, it is a straightforward matter to determine the element $w_{k\ell}^{(m)}$ given the array $W^{(m_1)}$.

For example, if $m_1 = 4$ and $m = 6$ then $w_{12,60}^{(6)} = 2 + w_{12,12}^{(4)}$. So that one can store some basic weight matrix $W^{(m_1)}$ and compute the entries in

higher order channel matrices when needed in the analysis.

We might ask if such a reproducing channel matrix is limited to source encoding according to the natural code. It turns out that the channel matrix obtained for the Gray code can be generated in a similar manner.

For the Gray code with $m = 1$ we have

v_1 coded as 0

v_2 coded as 1

$$P^{(1)} = \begin{bmatrix} q & p \\ p & q \end{bmatrix}, \quad W^{(1)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

For $m = 2$ we have

v_1 coded as 00

v_2 coded as 01

v_3 coded as 11

v_4 coded as 10

with

$$P^{(2)} = \begin{bmatrix} q^2 & pq & p^2 & pq \\ pq & q^2 & pq & p^2 \\ p^2 & pq & q^2 & pq \\ pq & p^2 & pq & q^2 \end{bmatrix}$$

$$= \begin{bmatrix} qP^{(1)} & pP_R^{(1)} \\ pP_R^{(1)} & qP^{(1)} \end{bmatrix}$$

and

$$W^{(2)} = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} W^{(1)} & 1+W_R^{(1)} \\ 1+W_R^{(1)} & W^{(1)} \end{bmatrix}$$

where

$$P_R^{(1)} = \begin{bmatrix} p & q \\ q & p \end{bmatrix} \quad \text{and} \quad W_R^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In general, the m^{th} extension of the channel matrix (BSC - Gray code) is

$$P^{(m)} = \begin{bmatrix} qP^{(m-1)} & pP_R^{(m-1)} \\ pP_R^{(m-1)} & qP^{(m-1)} \end{bmatrix}, \quad m = 2, 3, \dots \quad (\text{E.10})$$

and the weight or power matrix is

$$W^{(m)} = \begin{bmatrix} W^{(m-1)} & 1+W_R^{(m-1)} \\ 1+W_R^{(m-1)} & W^{(m-1)} \end{bmatrix}, \quad m = 2, 3, \dots \quad (\text{E.11})$$

where $P_R^{(m)}$ and $W_R^{(m)}$ are the reflections of $P^{(m)}$ and $W^{(m)}$, respectively,

about either their horizontal or vertical axis.

The Kronecher power of a matrix is not restricted to the case of a 2x2 initial matrix but is valid for any NxN initial matrix. This corresponds to an N-ary channel signal alphabet. Consider first the ternary channel, $N = 3$; then the channel may be represented by the matrix

$$A^{(1)} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$A^{(2)} = A^{(1)} \times A^{(1)} =$$

$$\begin{bmatrix} a_{11}A^{(1)} & a_{12}A^{(1)} & a_{13}A^{(1)} \\ a_{21}A^{(1)} & a_{22}A^{(1)} & a_{23}A^{(1)} \\ a_{31}A^{(1)} & a_{32}A^{(1)} & a_{33}A^{(1)} \end{bmatrix}$$

and in general

$$A^{(m)} = A^{(1)} \times A^{(m-1)}, \quad m = 2, 3, \dots \quad (E.12)$$

The above generalization is valid for $A^{(1)}$ a symmetric stochastic matrix with the source encoded according to the natural ternary code, e.g. $m = 2$:

$$v_1 : 00$$

$$v_2 : 01$$

$$v_3 : 02$$

$$v_4 : 10$$

$$v_5 : 11$$

$$v_6 : 12$$

$$v_7 : 20$$

$$v_8 : 21$$

$$v_9 : 22$$

By requiring that $a_{11} = a_{22} = a_{33}$, the matrix $A^{(1)}$ becomes

$$A_3^{(1)} = \begin{bmatrix} q & p & p \\ p & q & p \\ p & p & q \end{bmatrix}, \quad q = 1 - 2p \quad (\text{E.13})$$

For m channel symbols per source word this matrix generalizes to

$$A_3^{(m)} = A_3^{(1)} \times A_3^{(m-1)} = \begin{bmatrix} a_{k\ell}^{(m)} \end{bmatrix}, \quad m = 2, 3, \dots \quad (\text{E.14})$$

Again this can be associated with a weight matrix $W^{(m)} = \begin{bmatrix} w_{k\ell}^{(m)} \end{bmatrix}$ with

$$a_{k\ell}^{(m)} = p^{w_{k\ell}^{(m)}} (1-2p)^{m-w_{k\ell}^{(m)}}$$

We have

$$W^{(1)} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$W^{(m)} = \begin{bmatrix} W^{(m-1)} & 1+W^{(m-1)} & 1+W^{(m-1)} \\ 1+W^{(m-1)} & W^{(m-1)} & 1+W^{(m-1)} \\ 1+W^{(m-1)} & 1+W^{(m-1)} & W^{(m-1)} \end{bmatrix}, \quad m = 2, 3, \dots \quad (E.15)$$

where again 1 is a matrix of appropriate size with all entries unity.

The channel matrix in (E.13) represents, in the terminology of Abramson, the r -ary symmetric channel (rSC) with $r = 3$. And hence, (E.14) is m^{th} extension of the 3SC. We conjecture that the procedure is also valid for the m^{th} extension of the rSC. In which case we would have

$$A_r^{(1)} = \begin{bmatrix} q & p & p & \dots & p \\ p & q & p & \dots & p \\ . & . & . & \dots & . \\ p & p & p & \dots & q \end{bmatrix}, \quad q = 1 - (r-1)p \quad (E.16)$$

and

$$A_r^{(m)} = A_r^{(1)} A_r^{(m-1)}, \quad m = 2, 3, \dots \quad (E.17)$$

with the associated weight matrices

$$W_r^{(1)} = \begin{bmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & 1 & \dots & 1 \\ . & . & . & \dots & . \\ 1 & 1 & 1 & \dots & 0 \end{bmatrix} \quad (E.18)$$

and

$$W_r^{(m)} = \begin{bmatrix} W_r^{(m-1)} & 1+W_r^{(m-1)} & \dots & 1+W_r^{(m-1)} \\ 1+W_r^{(m-1)} & W_r^{(m-1)} & \dots & 1+W_r^{(m-1)} \\ \dots & \dots & \dots & \dots \\ 1+W_r^{(m-1)} & 1+W_r^{(m-1)} & \dots & W_r^{(m-1)} \end{bmatrix} \quad (E.19)$$

Table 3.1

Max's Uniform Quantizer Structure for $m = 2, \dots, 9$

M = 2

THE U'S ARE
 C. 0.9816
 THE V'S ARE
 C.4528 1.5100

M = 3

THE U'S ARE
 C. 0.5006 1.0500 1.7480
 THE V'S ARE
 C.2451 0.7560 1.3439 2.1520

M = 4

THE U'S ARE
 C. 0.2582 0.5224 0.7996 1.0993 1.4372 1.8436 2.4010
 THE V'S ARE
 C.1284 0.3881 0.6568 0.9424 1.2563 1.6181 2.0691 2.7328

M = 5

THE U'S ARE
 C. 0.1320 0.2648 0.3991 0.5355 0.6761 0.8210 0.9718
 1.1305 1.2993 1.4816 1.6821 1.9084 2.1738 2.5053 2.9774
 THE V'S ARE
 C.0659 0.1981 0.3314 0.4668 0.6050 0.7473 0.8947 1.0490
 1.2120 1.3866 1.5766 1.7876 2.0292 2.3184 2.6921 3.2626

M = 6

THE U'S ARE
 C. 0.0668 0.1338 0.2009 0.2684 0.3362 0.4046 0.4736
 C.5434 0.6141 0.6858 0.7587 0.8330 0.9088 0.9864 1.0661
 1.1481 1.2327 1.3204 1.4117 1.5070 1.6072 1.7130 1.8257
 1.9467 2.0781 2.2228 2.3850 2.5716 2.7945 3.0784 3.4914
 THE V'S ARE
 C.0334 0.1003 0.1673 0.2346 0.3022 0.3703 0.4390 0.5083
 C.5785 0.6497 0.7219 0.7955 0.8705 0.9471 1.0257 1.1065
 1.1897 1.2758 1.3651 1.4582 1.5558 1.6585 1.7675 1.8839
 2.0095 2.1467 2.2988 2.4711 2.6720 2.9170 3.2398 3.7429

M = 7

THE U'S ARE
 C. 0.0337 0.0673 0.1010 0.1348 0.1685 0.2024 0.2363
 C.2703 0.3044 0.3387 0.3731 0.4076 0.4423 0.4771 0.5122
 C.5474 0.5829 0.6187 0.6547 0.6910 0.7276 0.7645 0.8018
 0.8394 0.8775 0.9159 0.9549 0.9943 1.0343 1.0748 1.1159
 1.1576 1.2000 1.2432 1.2871 1.3319 1.3776 1.4243 1.4720
 1.5209 1.5710 1.6225 1.6755 1.7300 1.7864 1.8447 1.9051
 1.9680 2.0336 2.1023 2.1744 2.2505 2.3313 2.4174 2.5101
 2.6105 2.7206 2.8431 2.9817 3.1429 3.3375 3.5882 3.9575
 THE V'S ARE
 C.0168 0.0505 0.0842 0.1179 0.1516 0.1854 0.2193 0.2533
 0.2874 0.3215 0.3558 0.3903 0.4249 0.4596 0.4946 0.5297
 0.5651 0.6007 0.6366 0.6727 0.7092 0.7459 0.7830 0.8205
 C.8583 0.8966 0.9353 0.9745 1.0142 1.0544 1.0952 1.1366
 1.1787 1.2214 1.2650 1.3093 1.3545 1.4007 1.4479 1.4962
 1.5457 1.5964 1.6486 1.7023 1.7577 1.8150 1.8743 1.9360
 2.0001 2.0672 2.1374 2.2114 2.2897 2.3729 2.4620 2.5581
 2.6629 2.7784 2.9077 3.0557 3.2300 3.4450 3.7315 4.1835

P= 8

THE U'S ARE

C.	0.0169	0.0338	0.0507	0.0675	0.0844	0.1014	0.1183
C.1352	0.1521	0.1691	0.1861	0.2030	0.2201	0.2371	0.2541
C.2712	0.2883	0.3054	0.3226	0.3398	0.3570	0.3743	0.3916
C.4089	0.4263	0.4437	0.4612	0.4787	0.4963	0.5139	0.5316
C.5493	0.5671	0.5849	0.6028	0.6208	0.6388	0.6569	0.6751
C.6933	0.7117	0.7301	0.7486	0.7671	0.7858	0.8045	0.8234
C.8423	0.8614	0.8805	0.8998	0.9192	0.9386	0.9582	0.9780
C.9978	1.0178	1.0379	1.0582	1.0786	1.0992	1.1199	1.1408
1.1618	1.1830	1.2044	1.2260	1.2478	1.2698	1.2920	1.3144
1.3370	1.3598	1.3829	1.4062	1.4298	1.4537	1.4778	1.5022
1.5269	1.5520	1.5774	1.6031	1.6292	1.6557	1.6825	1.7098
1.7375	1.7657	1.7943	1.8234	1.8531	1.8833	1.9141	1.9455
1.9776	2.0103	2.0439	2.0782	2.1133	2.1494	2.1865	2.2245
2.2637	2.3041	2.3458	2.3888	2.4334	2.4797	2.5278	2.5780
2.6304	2.6854	2.7432	2.8042	2.8689	2.9379	3.0118	3.0915
3.1784	3.2741	3.3810	3.5026	3.6445	3.8164	4.0374	4.3616

THE V'S ARE

C.0084	0.0253	0.0422	0.0591	0.0760	0.0929	0.1098	0.1267
C.1437	0.1606	0.1776	0.1945	0.2115	0.2286	0.2456	0.2627
0.2798	0.2969	0.3140	0.3312	0.3484	0.3656	0.3829	0.4002
C.4176	0.4350	0.4524	0.4699	0.4875	0.5051	0.5227	0.5404
C.5582	0.5760	0.5939	0.6118	0.6298	0.6479	0.6660	0.6842
C.7025	0.7208	0.7393	0.7578	0.7764	0.7951	0.8139	0.8328
C.8518	0.8709	0.8901	0.9094	0.9289	0.9484	0.9681	0.9879
1.0078	1.0278	1.0480	1.0684	1.0888	1.1095	1.1303	1.1512
1.1724	1.1937	1.2152	1.2369	1.2587	1.2808	1.3031	1.3256
1.3483	1.3713	1.3945	1.4179	1.4417	1.4657	1.4899	1.5145
1.5394	1.5646	1.5902	1.6161	1.6423	1.6690	1.6961	1.7235
1.7515	1.7799	1.8087	1.8381	1.8681	1.8985	1.9296	1.9614
1.9938	2.0269	2.0608	2.0955	2.1311	2.1677	2.2052	2.2438
2.2836	2.3246	2.3669	2.4107	2.4561	2.5033	2.5524	2.6036
2.6572	2.7135	2.7728	2.8356	2.9023	2.9735	3.0501	3.1330
3.2238	3.3244	3.4377	3.5676	3.7214	3.9113	4.1636	4.5596

P= 9

THE U'S ARE

C.	0.0084	0.0168	0.0253	0.0337	0.0421	0.0505	0.0589
C.0674	0.0758	0.0842	0.0927	0.1011	0.1095	0.1179	0.1264
0.1348	0.1433	0.1517	0.1601	0.1686	0.1770	0.1855	0.1939
C.0204	0.2109	0.2193	0.2278	0.2363	0.2448	0.2533	0.2618
C.2703	0.2788	0.2873	0.2958	0.3043	0.3128	0.3214	0.3299
C.3385	0.3471	0.3556	0.3642	0.3728	0.3814	0.3900	0.3986
C.4073	0.4159	0.4246	0.4333	0.4419	0.4506	0.4593	0.4681
C.4768	0.4855	0.4943	0.5030	0.5118	0.5206	0.5294	0.5382
C.5470	0.5559	0.5647	0.5736	0.5825	0.5914	0.6003	0.6092
C.6182	0.6271	0.6361	0.6451	0.6541	0.6631	0.6722	0.6812
C.6903	0.6994	0.7085	0.7177	0.7268	0.7360	0.7452	0.7544
C.7637	0.7729	0.7822	0.7915	0.8008	0.8102	0.8196	0.8290
0.8384	0.8478	0.8573	0.8668	0.8763	0.8859	0.8955	0.9051
0.9147	0.9244	0.9341	0.9439	0.9536	0.9634	0.9733	0.9832
C.9931	1.0030	1.0130	1.0230	1.0330	1.0431	1.0532	1.0633
1.0735	1.0838	1.0940	1.1043	1.1146	1.1249	1.1353	1.1457
1.1562	1.1667	1.1772	1.1878	1.1984	1.2091	1.2198	1.2305
1.2413	1.2521	1.2630	1.2739	1.2849	1.2960	1.3070	1.3182
1.3294	1.3407	1.3520	1.3634	1.3749	1.3864	1.3979	1.4096
1.4213	1.4331	1.4449	1.4568	1.4688	1.4809	1.4931	1.5053
1.5176	1.5300	1.5425	1.5550	1.5677	1.5804	1.5932	1.6062
1.6191	1.6322	1.6454	1.6586	1.6720	1.6855	1.6991	1.7127
1.7265	1.7404	1.7544	1.7686	1.7828	1.7973	1.8118	1.8265
1.8413	1.8563	1.8714	1.8866	1.9019	1.9174	1.9330	1.9488
1.9647	1.9808	1.9971	2.0136	2.0303	2.0472	2.0643	2.0816
2.0991	2.1169	2.1349	2.1530	2.1714	2.1901	2.2090	2.2282
2.2477	2.2674	2.2875	2.3078	2.3285	2.3495	2.3708	2.3925
2.4147	2.4372	2.4602	2.4836	2.5075	2.5319	2.5568	2.5823
2.6084	2.6351	2.6624	2.6905	2.7193	2.7488	2.7792	2.8106
2.8430	2.8765	2.9112	2.9471	2.9842	3.0228	3.0629	3.1048

3.1484	3.1942	3.2424	3.2932	3.3465	3.4040	3.4647	3.5300
3.6014	3.6805	3.7687	3.8679	3.9812	4.1131	4.2741	4.4882
THE V'S ARE							
C.0042	0.0126	0.0210	0.0295	0.0379	0.0463	0.0547	0.0632
0.0716	0.0800	0.0884	0.0969	0.1053	0.1137	0.1222	0.1306
C.1390	0.1475	0.1559	0.1644	0.1728	0.1813	0.1897	0.1982
C.2066	0.2151	0.2236	0.2321	0.2405	0.2490	0.2575	0.2660
0.2745	0.2830	0.2915	0.3000	0.3086	0.3171	0.3257	0.3342
0.3428	0.3513	0.3599	0.3685	0.3771	0.3857	0.3943	0.4030
0.4116	0.4203	0.4289	0.4376	0.4463	0.4550	0.4637	0.4724
0.4812	0.4899	0.4987	0.5074	0.5162	0.5250	0.5338	0.5426
0.5515	0.5603	0.5692	0.5780	0.5869	0.5958	0.6048	0.6137
0.6226	0.6316	0.6406	0.6496	0.6586	0.6676	0.6767	0.6858
C.6948	0.7039	0.7131	0.7222	0.7314	0.7406	0.7498	0.7590
C.7683	0.7776	0.7869	0.7962	0.8055	0.8149	0.8243	0.8337
0.8431	0.8526	0.8621	0.8716	0.8811	0.8907	0.9003	0.9099
C.9196	0.9293	0.9390	0.9487	0.9585	0.9684	0.9782	0.9881
C.9980	1.0080	1.0180	1.0280	1.0380	1.0481	1.0583	1.0684
1.0786	1.0889	1.0991	1.1094	1.1198	1.1301	1.1405	1.1510
1.1614	1.1720	1.1825	1.1931	1.2037	1.2144	1.2251	1.2359
1.2467	1.2576	1.2685	1.2794	1.2904	1.3015	1.3126	1.3238
1.3350	1.3463	1.3577	1.3691	1.3806	1.3921	1.4037	1.4154
1.4272	1.4390	1.4508	1.4628	1.4749	1.4870	1.4991	1.5114
1.5238	1.5362	1.5487	1.5613	1.5740	1.5868	1.5997	1.6126
1.6257	1.6388	1.6520	1.6653	1.6787	1.6922	1.7059	1.7196
1.7334	1.7474	1.7615	1.7757	1.7900	1.8045	1.8191	1.8339
1.8488	1.8638	1.8790	1.8942	1.9096	1.9252	1.9409	1.9567
1.9727	1.9890	2.0053	2.0219	2.0387	2.0557	2.0729	2.0903
2.1080	2.1258	2.1439	2.1622	2.1807	2.1995	2.2185	2.2378
2.2575	2.2774	2.2976	2.3181	2.3389	2.3600	2.3816	2.4035
2.4258	2.4486	2.4718	2.4955	2.5196	2.5442	2.5694	2.5952
2.6216	2.6486	2.6763	2.7047	2.7338	2.7638	2.7947	2.8265
2.8595	2.8935	2.9288	2.9653	3.0031	3.0424	3.0834	3.1261
3.1707	3.2177	3.2671	3.3192	3.3746	3.4335	3.4960	3.5641
3.6388	3.7221	3.8153	3.9206	4.0419	4.1844	4.3638	4.6176

Table 3.3

The Optimum Quantizer Structure for PSK Reception and
the Natural Code with $m = 2, \dots, 9$

M=2 BETA=0.01 EPSTARSQ= 0.9519720E 00

THE U'S ARE
0.00 0.005
THE V'S ARE
0.000 0.126

M=3 BETA=5. EPSTARSQ= 0.1762832E 00

THE U'S ARE
0.000 0.145 0.497 1.098
THE V'S ARE
0.037 0.309 0.757 1.501

M=4 BETA=14.3 EPSTARSQ= 0.3216890E-01

THE U'S ARE
0.000 0.087 0.277 0.493 0.702 1.008 1.372 1.917
THE V'S ARE
0.048 0.181 0.381 0.588 0.846 1.167 1.587 2.257

M=5 BETA=28.6 EPSTARSQ= 0.5344853E-02

THE U'S ARE
0.000 0.082 0.183 0.293 0.409 0.529 0.654 0.782
0.909 1.064 1.232 1.418 1.627 1.880 2.201 2.670
THE V'S ARE
0.042 0.133 0.238 0.351 0.468 0.590 0.716 0.843
0.984 1.145 1.320 1.515 1.741 2.019 2.385 2.956

M=6 BETA=40.

EPSTARSQ= 0.1839630E-02

THE U'S ARE

0.000	0.051	0.113	0.180	0.249	0.318	0.389	0.459
0.530	0.602	0.674	0.747	0.821	0.895	0.970	1.043
1.112	1.197	1.283	1.372	1.464	1.558	1.656	1.758
1.862	1.978	2.104	2.244	2.404	2.600	2.854	3.241

THE V'S ARE

0.026	0.082	0.147	0.215	0.284	0.354	0.424	0.495
0.566	0.638	0.711	0.784	0.858	0.932	1.006	1.077
1.154	1.240	1.327	1.417	1.510	1.606	1.706	1.809
1.918	2.039	2.170	2.318	2.492	2.708	3.001	3.481

M=7 BETA=52.

EPSTARSQ= 0.7194850E-03

THE U'S ARE

0.000	0.021	0.053	0.088	0.123	0.159	0.194	0.229
0.264	0.299	0.334	0.369	0.404	0.439	0.474	0.509
0.545	0.581	0.617	0.653	0.689	0.726	0.763	0.801
0.839	0.877	0.916	0.955	0.995	1.034	1.072	1.108
1.138	1.185	1.232	1.278	1.324	1.370	1.417	1.465
1.513	1.563	1.614	1.667	1.720	1.776	1.833	1.890
1.947	2.012	2.080	2.150	2.223	2.300	2.381	2.466
2.555	2.652	2.756	2.870	2.999	3.157	3.368	3.700

THE V'S ARE

0.011	0.037	0.071	0.106	0.142	0.177	0.212	0.247
0.282	0.317	0.352	0.387	0.422	0.457	0.492	0.527
0.563	0.599	0.635	0.671	0.707	0.744	0.782	0.819
0.858	0.896	0.935	0.975	1.014	1.053	1.090	1.123
1.162	1.209	1.255	1.301	1.347	1.394	1.441	1.489
1.538	1.588	1.640	1.693	1.748	1.804	1.861	1.918
1.979	2.046	2.114	2.186	2.261	2.340	2.422	2.509
2.602	2.702	2.810	2.930	3.070	3.245	3.490	3.911

M=8 BETA=65.

EPSTARSQ= 0.3354251E-03

THE U'S ARE

0.000	0.014	0.032	0.049	0.066	0.083	0.101	0.118
0.135	0.152	0.169	0.186	0.203	0.220	0.237	0.254
0.271	0.288	0.305	0.322	0.340	0.357	0.374	0.391
0.409	0.426	0.444	0.461	0.479	0.496	0.514	0.531
0.549	0.567	0.585	0.603	0.621	0.639	0.657	0.675
0.693	0.712	0.730	0.749	0.767	0.786	0.805	0.823
0.842	0.861	0.881	0.900	0.919	0.939	0.958	0.978
0.998	1.018	1.038	1.059	1.079	1.100	1.121	1.141
1.154	1.182	1.204	1.226	1.248	1.270	1.292	1.314
1.337	1.360	1.383	1.406	1.430	1.453	1.478	1.502
1.527	1.552	1.577	1.603	1.629	1.655	1.682	1.709
1.737	1.765	1.794	1.823	1.853	1.883	1.914	1.945
1.974	2.009	2.043	2.077	2.113	2.149	2.186	2.224
2.263	2.303	2.345	2.388	2.432	2.478	2.526	2.576
2.628	2.683	2.741	2.801	2.865	2.934	3.007	3.086
3.171	3.265	3.370	3.487	3.621	3.779	3.968	4.219

THE V'S ARE

0.008	0.024	0.042	0.059	0.076	0.093	0.110	0.127
0.144	0.160	0.177	0.194	0.211	0.228	0.246	0.263
0.280	0.297	0.314	0.331	0.348	0.366	0.383	0.400
0.418	0.435	0.452	0.470	0.487	0.505	0.523	0.540
0.558	0.576	0.594	0.612	0.630	0.648	0.666	0.684
0.702	0.721	0.739	0.758	0.776	0.795	0.814	0.833
0.852	0.871	0.890	0.909	0.929	0.948	0.968	0.988
1.008	1.028	1.048	1.068	1.089	1.110	1.131	1.150
1.171	1.193	1.215	1.237	1.259	1.281	1.303	1.325
1.348	1.371	1.394	1.418	1.441	1.465	1.490	1.514
1.539	1.564	1.590	1.616	1.642	1.669	1.696	1.723
1.751	1.779	1.808	1.838	1.868	1.898	1.929	1.960
1.993	2.026	2.060	2.095	2.130	2.167	2.204	2.243
2.283	2.323	2.366	2.409	2.455	2.502	2.551	2.602
2.655	2.711	2.770	2.832	2.899	2.969	3.045	3.127
3.216	3.315	3.425	3.549	3.694	3.865	4.071	4.368

M=9 BETA=90.

EPSTARSO= 0.5368143E-04

THE U'S ARE

0.000	0.008	0.016	0.025	0.033	0.042	0.050	0.059
0.067	0.076	0.084	0.093	0.101	0.109	0.118	0.126
0.135	0.143	0.152	0.160	0.169	0.177	0.185	0.194
0.202	0.211	0.219	0.228	0.236	0.245	0.253	0.262
0.270	0.279	0.287	0.296	0.304	0.313	0.321	0.330
0.338	0.347	0.356	0.364	0.373	0.381	0.390	0.399
0.407	0.416	0.425	0.433	0.442	0.451	0.459	0.468
0.477	0.486	0.494	0.503	0.512	0.521	0.529	0.538
0.547	0.556	0.565	0.574	0.582	0.591	0.600	0.609
0.618	0.627	0.636	0.645	0.654	0.663	0.672	0.681
0.690	0.699	0.709	0.718	0.727	0.736	0.745	0.754
0.764	0.773	0.782	0.792	0.801	0.810	0.820	0.829
0.838	0.848	0.857	0.867	0.876	0.886	0.895	0.905
0.915	0.924	0.934	0.944	0.954	0.963	0.973	0.983
0.993	1.003	1.013	1.023	1.033	1.043	1.053	1.063
1.074	1.084	1.094	1.104	1.115	1.125	1.136	1.146
1.154	1.167	1.177	1.188	1.198	1.209	1.220	1.230
1.241	1.252	1.263	1.274	1.285	1.296	1.307	1.318
1.329	1.341	1.352	1.363	1.375	1.386	1.398	1.410
1.421	1.433	1.445	1.457	1.469	1.481	1.493	1.505
1.518	1.530	1.542	1.555	1.568	1.580	1.593	1.606
1.619	1.632	1.645	1.659	1.672	1.685	1.699	1.713
1.726	1.740	1.754	1.769	1.783	1.797	1.812	1.826
1.841	1.856	1.871	1.887	1.902	1.917	1.933	1.949
1.964	1.981	1.997	2.014	2.030	2.047	2.064	2.082
2.099	2.117	2.135	2.153	2.171	2.190	2.209	2.228
2.248	2.267	2.287	2.308	2.328	2.349	2.371	2.392
2.415	2.437	2.460	2.483	2.507	2.532	2.557	2.582
2.608	2.635	2.662	2.690	2.719	2.749	2.779	2.810
2.843	2.876	2.911	2.947	2.984	3.022	3.062	3.104
3.148	3.193	3.241	3.292	3.346	3.403	3.463	3.528
3.600	3.678	3.765	3.863	3.974	4.104	4.274	4.542

THE V'S ARE

0.004	0.012	0.021	0.029	0.038	0.046	0.055	0.063
0.072	0.080	0.088	0.097	0.105	0.114	0.122	0.131
0.139	0.147	0.156	0.164	0.173	0.181	0.190	0.198
0.207	0.215	0.224	0.232	0.241	0.249	0.258	0.266
0.275	0.283	0.292	0.300	0.309	0.317	0.326	0.334
0.343	0.351	0.360	0.369	0.377	0.386	0.394	0.403
0.412	0.420	0.429	0.438	0.446	0.455	0.464	0.472
0.481	0.490	0.499	0.507	0.516	0.525	0.534	0.543
0.551	0.560	0.569	0.578	0.587	0.596	0.605	0.614
0.623	0.632	0.641	0.650	0.659	0.668	0.677	0.686
0.695	0.704	0.713	0.722	0.731	0.741	0.750	0.759
0.768	0.778	0.787	0.796	0.806	0.815	0.824	0.834
0.843	0.853	0.862	0.872	0.881	0.891	0.900	0.910
0.920	0.929	0.939	0.949	0.959	0.968	0.978	0.988
0.998	1.008	1.018	1.028	1.038	1.048	1.058	1.068
1.079	1.089	1.099	1.109	1.120	1.130	1.141	1.151
1.161	1.172	1.182	1.193	1.204	1.214	1.225	1.236
1.247	1.258	1.268	1.279	1.290	1.301	1.313	1.324
1.335	1.346	1.358	1.369	1.381	1.392	1.404	1.415
1.427	1.439	1.451	1.463	1.475	1.487	1.499	1.511
1.524	1.536	1.549	1.561	1.574	1.587	1.600	1.613
1.626	1.639	1.652	1.665	1.679	1.692	1.706	1.720
1.733	1.747	1.761	1.776	1.790	1.804	1.819	1.834
1.849	1.864	1.879	1.894	1.910	1.925	1.941	1.957
1.973	1.989	2.005	2.022	2.039	2.056	2.073	2.090
2.108	2.126	2.144	2.162	2.181	2.199	2.218	2.238
2.257	2.277	2.297	2.318	2.339	2.360	2.381	2.403
2.426	2.448	2.472	2.495	2.519	2.544	2.569	2.595
2.621	2.648	2.676	2.704	2.734	2.764	2.794	2.826
2.859	2.893	2.928	2.965	3.003	3.042	3.083	3.125
3.170	3.217	3.266	3.318	3.373	3.432	3.494	3.562
3.637	3.719	3.811	3.914	4.033	4.176	4.372	4.713

Table 3.4

The Quantizing Interval r of Max's Uniform Quantizer for
 $m = 1, \dots, 9$

m	r	m	r
1	1.59577	6	0.10406
2	0.99570	7	0.05687
3	0.58602	8	0.03076
4	0.33520	9	0.01650
5	0.18814		

Table 3.5

The Optimum Quantizing Interval r of the Optimum Uniform
 Quantizer for PSK Reception and the Natural Code with $m = 5, 6, 7, 8$

β		$m=5$		$m=6$		$m=7$		$m=8$	
in db	in Units	r	ϵ_{*1}^2	r	ϵ_{*1}^2	r	ϵ_{*1}^2	r	ϵ_{*1}^2
20.0	100.0			.1041	.001041	.05686	.0003059	.03064	.0009475
19.212	83.4	.1881	.003496	.1041	.001042	.05675	.0003144	.02990	.0001387
18.543	71.5	.1881	.003496	.1040	.001048	.05620	.0003592	.02830	.0003098
17.701	58.9	.1881	.003503	.1035	.001110	.05374	.0006444		
16.990	50.00	.1879	.003542	.1016	.001362	.05003	.001489	.02336	.0002915
16.021	40.00	.1861	.003874	.09496	.002776	.04439	.004853		
15.224	33.3	.1808	.005016	.08704	.006142	.04004	.01117		
15.031	25.0	.1631	.01149	.07458	.01926	.03417	.03132		
13.010	20.0	.1461	.02405	.06612	.03917	.03039	.05789	.01418	.07813
12.227	16.7	.1331	.04093	.06023	.06258	.02779	.08668		
10.00	10.0	.1036	.1252	.04754	.1629				

APPENDIX G

COMPUTER PROGRAMS USED FOR COMPUTING QUANTIZER STRUCTURES

These programs are all for the input data distributed $N(0,1)$. By translating and spreading, the results may be applied to Gaussian data with any mean or variance. The channel matrix used is that for the additive white Gaussian channel with no bandwidth restrictions. Both PSK and FSK reception were considered.

G.1 Determining Max's Uniform Quantizer

Subroutine UANDV. This calculates

$$V_k \quad k = 2, \dots, \text{NEND} \quad (*)$$

$$U_k \quad k = 2, \dots, \text{NEND} \quad (**)$$

where [†]

$$V_k = \frac{\varphi(U_k) - \varphi(U_{k-1})}{\frac{1}{2}(\text{erf}(\frac{U_{k-1}}{\sqrt{2}}) - \text{erf}(\frac{U_k}{\sqrt{2}}))}$$

$$U_k = \frac{V_k + V_{k-1}}{2}$$

Method: Regula falsi is used several times to get a good estimate; then Newton's method is employed.

Calling Sequence: Call UANDV(U, V, 2, NEND, EPS)

$$^{\dagger} \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

U = one dimensional array where U(1) has been set to 0.0

V = one dimensional array where some V(1) is input into the subroutine. (i.e., V(1) must be assigned a value before entering the subroutine)

NEND = upper limit of indices; see (*) and (**) above; in my programs $NEND = 2^{m-1}$

EPS = some small number to be used in convergence criterion; if accuracy corresponding to this epsilon cannot be obtained, a message will be printed out and epsilon will be multiplied by 10. The new value of epsilon will also be printed. In all my programs, $EPS = .00001$

Note: Actually the subroutine was designed for a more general case where

V_i and U_i are set by the calling program and the user desires to obtain:

$$V_{i+1}, V_{i+2}, \dots, V_r$$

$$U_{i+1}, U_{i+2}, \dots, U_r$$

in which case the call would appear:

Call UANDV(U, V, I+1, R, EPS)

(the 3rd and 4th arguments are integers where)
 $R \geq I+1$

Search program. This program uses the UANDV subroutine to determine a V_1 such that

$$\frac{2\phi(U_{2^{m-1}})}{1 - \operatorname{erf}\left(\frac{U_{2^{m-1}}}{\sqrt{2}}\right)} - V_{2^{m-1}} = 0$$

Method: In general, the expression of the left above is considered a function of V(1):

Part 1 $F(V(1) = \frac{1}{2^{m-1}})$ is considered to be a positive number, i.e.

$V(1) = \frac{1}{2^{m-1}}$ is to the left of the root we search for. (If this is not so, the program stops.) Using $\delta = \frac{1}{2^{2(m-1)}}$ (selected rather arbitrarily)

a value to the right of the root is found by adding (and subtracting) integral multiples of δ .

Part 2 Once the root has been bracketed by two numbers (called $V_{\phi NE1}$ and $V_{\phi NE2}$) Regula Falsi is used.

Punch Program. This reads in a value of $V(1)$ (previously determined from the Search program) and a value of M . Using the UANDV subroutine it punches out the

$$U(1), \dots, U(2^{m-1}) \text{ and} \\ V(1), \dots, V(2^{m-1})$$

and an initial label card with the value of M .

Note: Several cards with values of $V(1)$ and M may be supplied.

Subroutine BITDIF (Map Program). This calculates the "bit difference" between binary representations of two integers. The "bit difference" is defined to be the number of binary positions in which the representations differ. Example: 1001_2 and 0110_2 have B.D. of 4.

Method: Using ERA (exclusive "or") between the two numbers gives a third number, in which the number of 1's is the bit difference. This third number is successively divided by 2 and the remainder checked to count the number of 1's in it.

Call Sequence: Call BITDIF(I, J, K)

$$\left. \begin{array}{l} I = \\ J = \end{array} \right\} \text{Input integers}$$

$K =$ Bit difference of I, J

Quantizer and Channel Error Program - ϵ_{*1}^2

This program calculates ϵ_{*1}^2 of (3.22) using the BITDIF subroutine. The program is designed to read in output of the Punch program, and decreasingly ordered values of $\beta = ST/N_0 n$ supplied in the data deck. The data deck set-up is described in the initial comments. An example is given in Figure G.1.

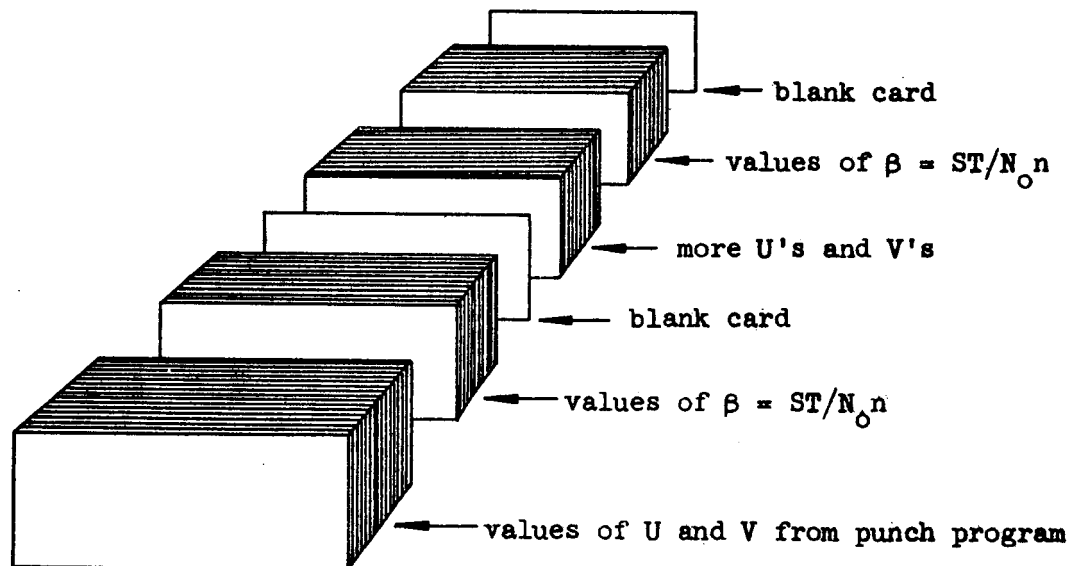


Figure G.1 The Data Deck Set-Up

G.2 Determining the Optimum Quantizer

G.2.1 Newton's Method. Newton's iteration method is used to solve the system of equations

$$f(u, v) = 0$$

where the first $2^{m-1} - 1$ equations of the system are

$$f_i(u, v) = \sum_{\ell=1}^{2^{m-1}} v_{\ell}^2 Q_{j\ell} + 2u_j v_{\ell} R_{j\ell} = 0, \quad j = 2, \dots, 2^{m-1}$$

and the remaining 2^{m-1} equations are

$$f_i(u, v) = \sum_{k=1}^{2^{m-1}} \int_{u_k}^{u_{k+1}} [(P_{kj} - P_{k(2^m+1-j)})X - (P_{kj} + P_{k(2^m+1-j)})v_j] \cdot p_X(X) dX = 0, \quad j = 1, \dots, 2^{m-1}$$

with

$$Q_{j\ell} = P_{j-1\ell} - P_{j\ell} + P_{(2^m+2-j)\ell} - P_{(2^m+1-j)\ell}$$

$$R_{j\ell} = P_{j\ell} - P_{j-1\ell} + P_{(2^m+2-j)\ell} - P_{(2^m+1-j)\ell}$$

The u and v vectors are iterated as follows[†]

$$(u, v)_{n+1} = (u, v)_n + \bar{\delta}$$

where $\bar{\delta}$ is the solution of

$$[J_n] \bar{\delta} = -f(u, v)_n$$

and J_n is the Jacobian of f at $(u, v)_n$. The flow diagram of the program for computing the optimum u & v vectors as a function of the channel is given in Figure G.2.

[†]In this case the couple notation $(u, v)_n$ simply means the combined array of the u and v vectors. The n stands for number of iterations already performed.

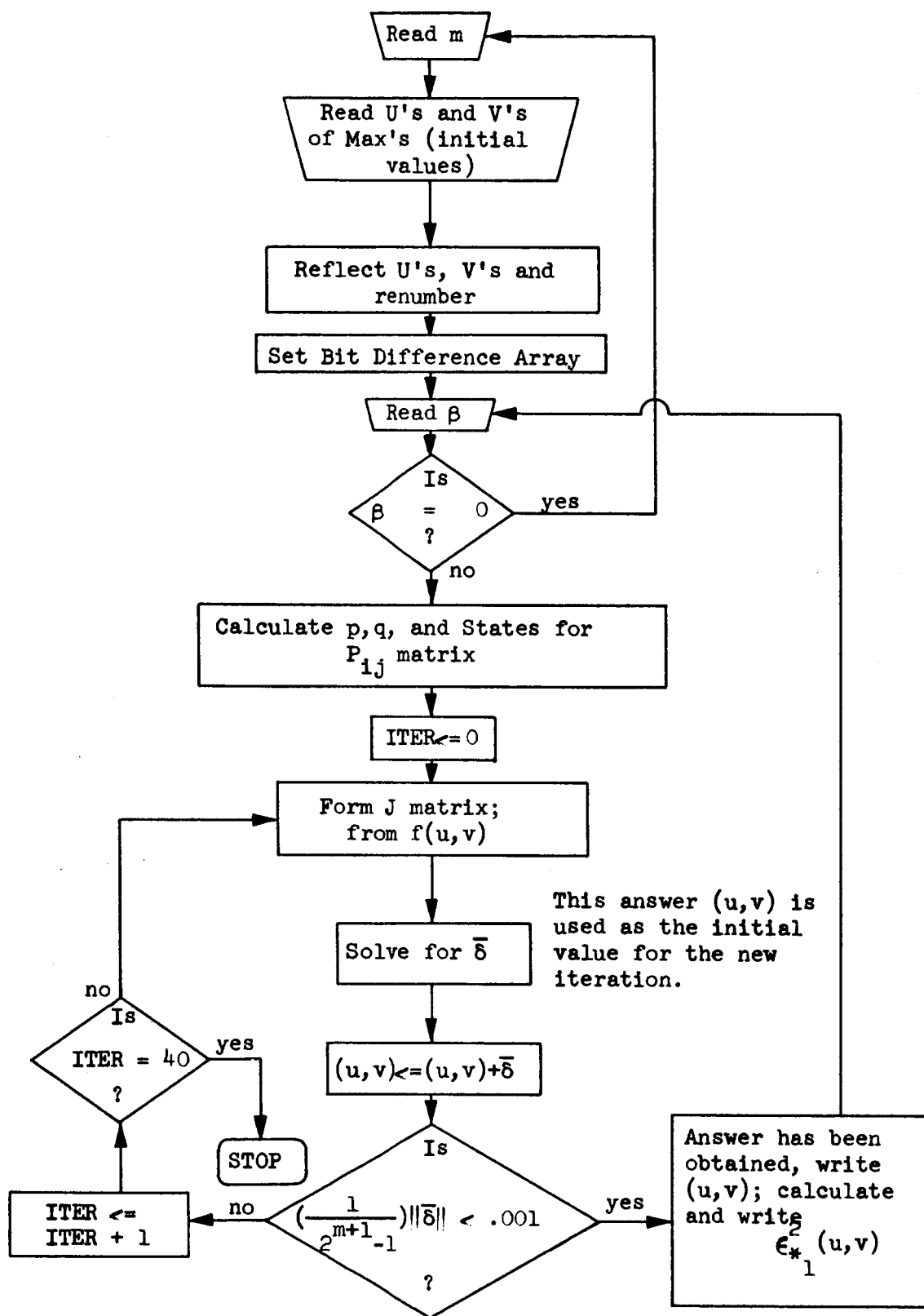


Figure G.2 Flow Diagram for Computing the Optimum Quantizer Noisy Newton's Method

G.2.2 Alternate Method. This method takes advantage of the coupling of the u and v vectors as shown in (3.23) and (3.24). The iteration technique employed is to initially take β large and using Max's u and v vectors alternately apply (3.23) and (3.24) until (3.23) and (3.24) are solved. In this case

$$(u,v)_{n+1} = F(u,v)_n$$

where it is known that

$$(u,v) = F(u,v)$$

and the functional relationship $F(u,v)$ is given by (3.23) and (3.24). A flow diagram of the program used for this computation is presented in Figure G.3.

G.3 Determining the Uniform Quantizer

We first determine the optimum interval for the noiseless channel which we call Max's r . This is done by solving the equation

$$\text{FUNCT}_m(r) = 0$$

where $\text{FUNCT}_m(r)$ is given by (D.19). The program used is described in Figure G.4.

We also determine the optimum interval for progressively more noisy channels. This is done by solving (D.18) for r . A flow diagram of the program used for this computation is shown in Figure G.5. The function $f(r)$ is that given in (D.18), i.e.

$$f(r) = 0$$

The equation for $\epsilon_{*1}^2(r)$ is (D.6).

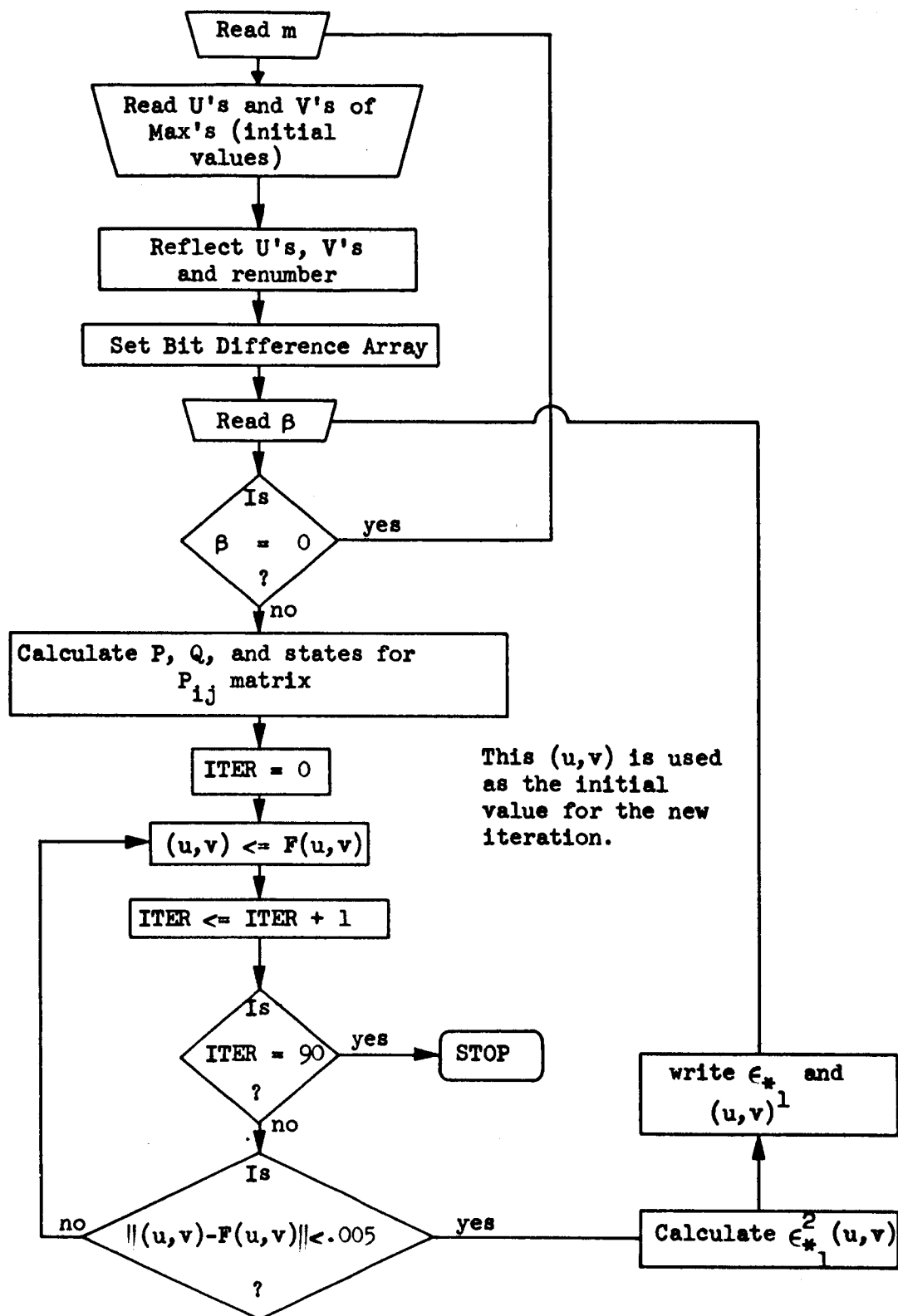


Figure G.3 Flow Diagram for Computing the Optimum Quantizer Using the Alternate Method

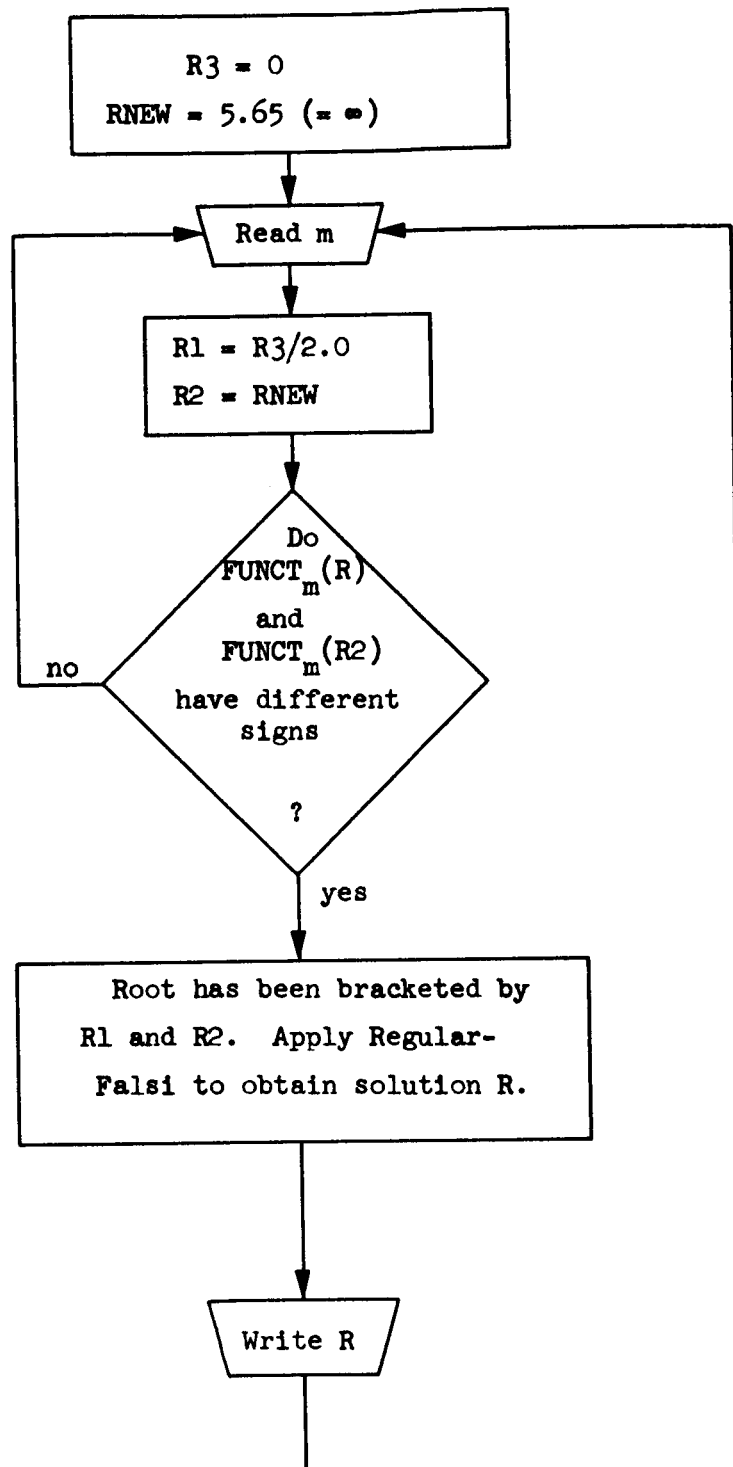
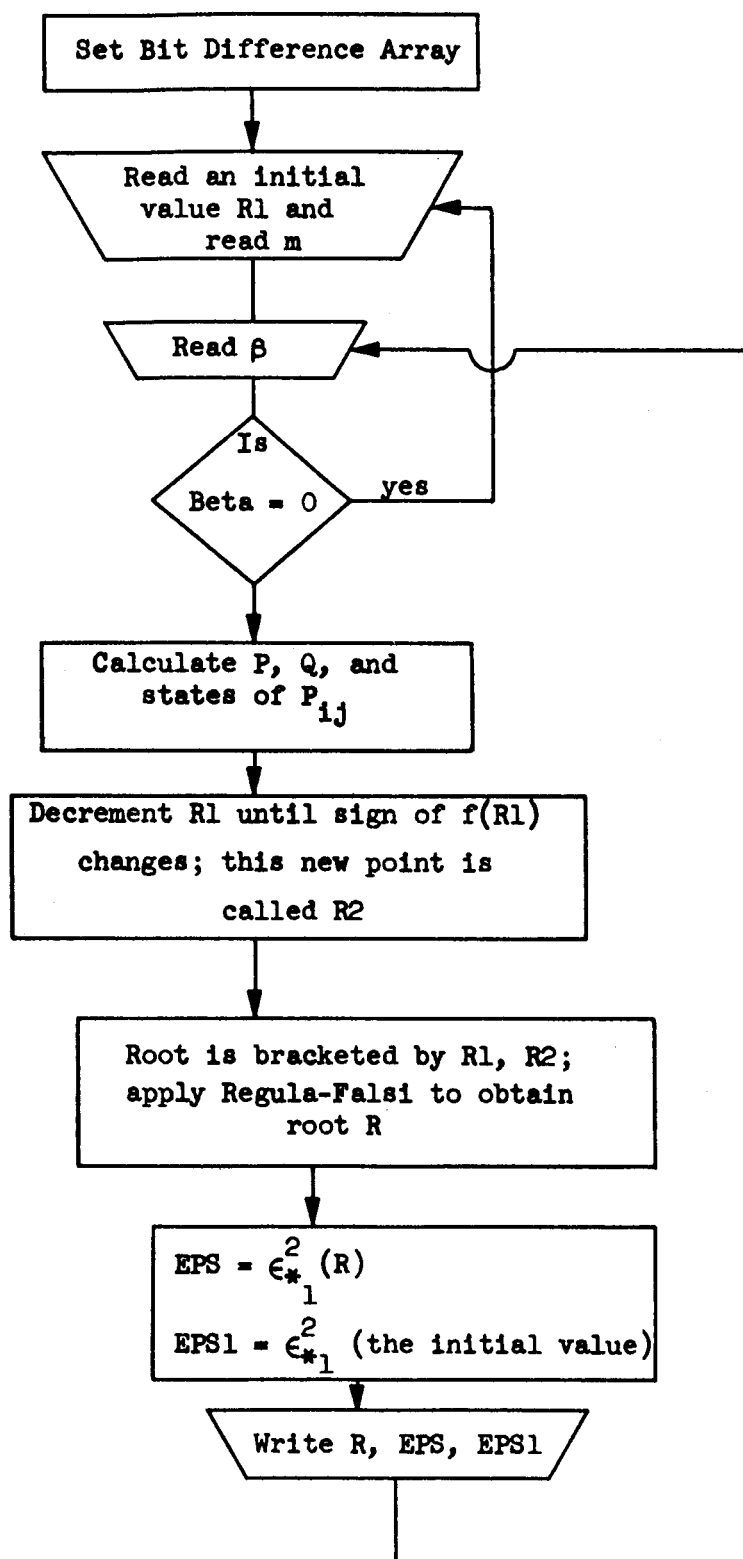


Figure G.4 Flow Diagram for Computing Max's r .



The R obtained is used for initial value (R1) in new search.

Figure G.5 Flow Diagram for Computing the Optimum r

APPENDIX H

FORMULATION OF THE TRADE-OFF PROBLEM AS AN
INTEGER NONLINEAR PROGRAMMING PROBLEM

The problem is to minimize the sum

$$\epsilon_*^2 = \sum_{i=1}^n \sigma_i^2 \epsilon_{*i}^2(m_i) \quad (\text{H.1})$$

where n is an integer and the $\sigma_i^2 (i = 1, \dots, n)$ are positive constants (the eigenvalues from (2.8)) ordered such that

$$\sigma_1^2 \geq \dots \geq \sigma_n^2 \quad (\text{H.2})$$

The $\epsilon_{*i}^2(m_i) (i=1, \dots, n)$ are those connected by solid lines in Figure 3.12 for a fixed value of $\Delta = \Gamma/m_b$. Hence the $\epsilon_{*i}^2(m_i)$ are constrained to the set

$$\epsilon_{*i}^2(m_i) \in \{e(1), e(2), \dots, e(\hat{m})\} \quad (\text{H.3})$$

where

$$e(1) \geq e(2) \geq \dots \geq e(\hat{m}) \quad (\text{H.4})$$

And $\hat{m}$ is the largest member of the set of integers which the $m_i (i = 1, \dots, n)$ take on, i.e.

$$m_i \in \{1, 2, \dots, \hat{m}\} \quad (\text{H.5})$$

The $m_i (i = 1, \dots, n)$ are constrained by the requirement

$$m_b = \sum_{i=1}^n m_i \quad (\text{H.6})$$

where $m_b \leq n\hat{m}$ is also an integer. Using (H.1) through (H.6) we also have that

$$\begin{aligned} m_1 &\geq m_2 \\ m_2 &\geq m_3 \\ &\dots\dots\dots \\ m_{n-1} &\geq m_n \\ m_n &\geq 1 \end{aligned} \tag{H.7}$$

For assume not, then for some j and $j+1$ we have from (H.4) that

$$m_{j+1} \geq m_j \Rightarrow e(m_j) \geq e(m_{j+1}) \tag{H.8}$$

and from (H.1) that

$$\epsilon_*^2 = \sum_{\substack{i=1 \\ i \neq j, j+1}}^n \sigma_i^2 \epsilon_{*i}^2(m_i) + \sigma_j^2 e(m_j) + \sigma_{j+1}^2 e(m_{j+1}) \tag{H.9}$$

where from (H.2)

$$\sigma_j^2 \geq \sigma_{j+1}^2 \tag{H.10}$$

Now, using (H.8) we obtain

$$\sigma_j^2 [e(m_j) - e(m_{j+1})] \geq \sigma_{j+1}^2 [e(m_j) - e(m_{j+1})] \tag{H.11}$$

and rearranging the terms in (H.11) we have

$$\sigma_j^2 e(m_j) + \sigma_{j+1}^2 e(m_{j+1}) \geq \sigma_j^2 e(m_{j+1}) + \sigma_j^2 e(m_j) \tag{H.12}$$

Therefore, ϵ_*^2 given by (H.9) is not minimum. Hence, we have the linear constraints given by (H.7). It also follows immediately from (H.6) and (H.7) that

$$m_1 \geq m_b/n \tag{H.13}$$

According to the terminology of Hadley [60], this is a problem in integer nonlinear programming with linear constraints and a separable objective function.

(7.4)

$$x_j^m \leq x_j^m$$

$$x_j^m \leq x_j^m$$

$$x_j^m \leq x_j^m$$

$$x_j^m \leq x_j^m$$

For some j , from (7.4) and (7.5) we have from (7.4) that

(8.4)

$$x_j^m \leq x_j^m \leq x_j^m$$

and from (8.4) that

(9.4)

$$x_j^m \leq x_j^m \leq x_j^m$$

where from (9.4)

(10.4)

$$x_j^m \leq x_j^m$$

Now, using (10.4) we obtain

(11.4)

$$x_j^m \leq x_j^m \leq x_j^m$$

and respectively the same in (11.4) we have

(12.4)

$$x_j^m \leq x_j^m \leq x_j^m$$

Therefore, x_j^m given by (12.4) is not minimum. Hence, we have the linear

constraint given by (12.4). It also follows immediately from (12.4) and

(12.4) that

(13.4)

$$x_j^m \leq x_j^m$$

**RECENT RESEARCH PUBLICATIONS
SCHOOL OF ELECTRICAL ENGINEERING
PURDUE UNIVERSITY**

- TR-EE67-1** **PARAMETER ESTIMATION WITH UNKNOWN SYMBOL SYNCHRONIZATION**
J. C. Hancock and T. L. Stewart, NSF Grant GP-2898, PRF 3955-50-285, January 1967
- TR-EE67-2** **NONLINEAR OSCILLATION OF A GYROSCOPE**
Chikaro Sato, PRF 4891, National Science Foundation Grant No. GK-01235
- TR-EE67-3** **ON THE DESIGN OF SPECIFIC OPTIMAL CONTROLLERS**
V. B. Haas and S. Murtuza, NASA, NGR 15-005-021
- TR-EE67-4** **ON THE ANALYSIS AND SYNTHESIS OF CONTROL SYSTEMS USING A WORST CASE DISTURBANCE APPROACH**
V. B. Haas and A. S. Morse, NASA, NGR 15-005-021
- TR-EE67-5** **ANALYSIS AND SYNTHESIS OF MULTI-THRESHOLD THRESHOLD LOGIC**
K. S. Fu and W. C. W. Mow, Contract ONR-N00016-66-C0076, A04 Joint Services Electronic Program
- TR-EE67-6** **ESTIMATION OF SONAR TARGET PARAMETERS**
G. R. Cooper and J. U. Kincaid, National Science Foundation Grant GK-189, PRF 4243
- TR-EE67-7** **ON GENERALIZATIONS OF ADAPTIVE ALGORITHMS AND APPLICATION OF THE FUZZY SETS CONCEPTS TO PATTERN CLASSIFICATION**
K. S. Fu and W. G. Wee, National Science Foundation Grant GK-696, PRF 4502
- TR-EE67-8** **COMPUTER AIDED ANALYSIS AND SYNTHESIS OF MULTIVALUED MEMORYLESS NETWORKS**
L. O. Chua and W. H. Stelhorn, National Science Foundation Grant GK-01235, Joint Services Electronics Program Contract ONR-N00016-66-C0076-A04, PRF 4711
- TR-EE67-9** **OPTIMUM FINITE SEQUENTIAL PATTERN RECOGNITION**
K. S. Fu and G. P. Cardillo, National Science Foundation Grant GK-696, PRF 4502
- TR-EE67-10** **MINIMUM SENSITIVITY OPTIMAL CONTROL FOR NONLINEAR SYSTEMS**
V. Haas and A. Steinberg, NASA NGR 15-005-021, PRF 4558
- TR-EE67-11** **RANDOM SIGNAL RADAR (Final Report)**
G. R. Cooper and C. D. McGillem, NASA, NSG-543
- TR-EE67-12** **INTRODUCTION TO THE PERFORMANCE OF DISTRIBUTION FREE, MINIMUM CONDITIONAL RISK LEARNING SYSTEMS**
E. A. Patrick and F. P. Fisher II, Naval Ship Systems Command Contract N00024-67-C-1162 PRF No. 4925-53-2859
- TR-EE67-13** **P-I-N THERMO-PHOTO-VOLTAIC DIODE**
R. J. Schwartz and C. W. Kim, Joint Services Electronic Program Contract ONR-N00014-67 A00226-004
- TR-EE67-14** **ESTIMATION OF PROBABILITY DENSITY AND DISTRIBUTION FUNCTIONS**
R. L. Kashyap and C. C. Blaydon, National Science Foundation Grant GK-1970, PRF 5181, and Joint Services Electronics Program Contract N00014-67-A0226-0004, PRF 4711.
- TR-EE67-15** **OPTIMIZATION OF STOCHASTIC CONTROL PROCESSES WITH RESPECT TO PROBABILITY OF ENTERING A TARGET MANIFOLD**
J. Y. S. Luh and G. E. O'Connor, Jr., Jet Propulsion Laboratory Contract 950670, PRF 3807
- TR-EE67-16** **CODES FOR UNSUPERVISED LEARNING OF SOURCE AND BINARY CHANNEL PROBABILITIES**
E. A. Patrick and G. Carayannopoulos, Air Force Avionics Laboratory (AVWC), Wright-Patterson AFB, Dayton, Ohio, AF 33(615)3768, PRF 4575
- TR-EE67-17** **DIGITAL COMMUNICATION SYSTEMS SUBJECT TO FREQUENCY SELECTIVE FADING**
J. C. Lindenlaub and C. C. Bailey, National Aeronautics and Space Administration, NSG-553, PRF 5120
- TR-EE67-18** **DEMO I-A SUPERVISED OR UNSUPERVISED LEARNING COMPUTER DESIGNED FOR DEMONSTRATION**
E. A. Patrick, F. C. Monds, G. L. Carayannopoulos, J. P. Costello, T. A. Martin, Air Force Avionics Laboratory (AVWC), Wright-Patterson Air Force Base, Dayton, Ohio, AF33(615)3768, PRF 4575