SRC TR 85-33

Optimal Quantizer Design For Noisy Channels: An Approach To Combined Source-Channel Coding

by

N. Farvardin

and

V. Vaishampayan

OPTIMAL QUANTIZER DESIGN FOR NOISY CHANNELS: AN APPROACH TO COMBINED SOURCE-CHANNEL CODING+

by

N. Farvardin and V. Vaishampayan

Electrical Engineering Department
and
Systems Research Center
University of Maryland
College Park, Maryland 20742

ABSTRACT

In this paper, we present an analysis of the zero-memory quantization of memoryless sources when the quantizer output is to be encoded and transmitted across a noisy channel. Necessary conditions for the joint optimization of the quantizer and the encoder/decoder pair are presented and a recursive algorithm for obtaining a locally optimum system is developed. The performance of this locally optimal system, obtained for the class of generalized Gaussian distributions and the Binary Symmetric Channel is compared against the Optimum Performance Theoretically Attainable (using Rate-Distortion theoretic arguments), as well as against the performance of Lloyd-Max quantizers encoded using Natural Binary Codes. It is shown that this optimal design could result in substantial performance improvements. The performance improvements are more noticeable at high bit rates and for more broad-tailed densities.

+This work was supported in part by a grant from Martin Marietta Laboratories and in part by the National Science Foundation under grant NSFD CDR-85-00108.

I. Introduction:

In a typical communication system, we wish to transmit information about a source $\{X_n\}$ with entropy rate (possibly infinite) $H(X)$ across a channel with capacity C (measured in bits/source symbol) in such a way that the resulting reproduction of $\{X_n\}$ in the receiver, say, $\{X_n\}$, is as close a replica of the original source as possible. Shannon's channel coding arguments [1],[3] prove that when $H(X) > C$, it is impossible to obtain an exact replica of $\{X_n\}$ in the receiver. In this case, rate-distortion theoretic arguments [2]-[4] suggest that a source encoder can be used whose effect, essentially, is that of mapping the source output $\{X_n\}$ into an approximation of itself, say, $\{\tilde{X}_n\}$, whose entropy rate $H(\tilde{X})$ satisfies $H(\tilde{X}) < C$. Then, again through channel coding arguments [1],[3],[4], it is well-known that $\{\tilde{X}_n\}$ can be encoded and transmitted through the channel with perfect reliability, i.e., the only distortion between $\{X_n\}$ and the reconstructed sequence in the receiver $\{\hat{X}_n\}$ is that produced by the source encoder. Thus, the best source encoder is that which maps $\{X_n\}$ into $\{\tilde{X}_n\}$ satisfying $H(\tilde{X}) < C$ in such a way that $\{\tilde{X}_n\}$ is as good an approximation of $\{X_n\}$ as possible.

The above argument implies a certain demarcation in the encoding process. Specifically, it suggests that the source and the channel encoder can be separated in such a way that the entropy rate reduction takes place in the source encoder and the protection against channel errors in the channel encoder. Indeed, one of the important results of Shannon's celebrated papers [1],[2] was his demonstration that the source and the channel coding functions are fundamentally separable. However, as indicated clearly in [3], the assumption that the source and the channel coders are separable is justifiable only in the limit of arbitrarily complex encoders and decoders. Of course, in practical situations, there are limitations on the system's complexity which make this separation questionable. It is our intent in this paper to study the interrelationship

between the source and channel coders for a practical system with limited complexity.

The effects of the channel errors on the performance of source coders, as well as the interaction between the source and channel coders, have been studied by several researchers. Among others, Spilker [5] has noted that when the output of a zero-memory quantizer is coded and transmitted via a very noisy channel, quantizers with small number of levels (i.e., large quantization distortion) yield better performance than those with larger number of levels. Kurtenbach and Wintz [6] studied the problem of optimum quantizer design when the quantizer's output is transmitted over a noisy channel. However, in [6], the issue of code assignment is not addressed. On the other hand, Rydbeck and Sundberg [7], without considering the quantizer design problem, have shown that the code assignment plays an important role in determining the system's performance.

On the other hand, a few researchers in the channel coding area have used the abstract Fourier analysis of groups to design an optimum mean-square-error code for discrete-alphabet uniformly distributed sources [8]-[10].

On the more practical side, combined source/channel coding schemes are studied for image coding situations in [11],[12]. Also, similar schemes have been reported for speech coding in [13],[14].

In this paper, we present a rigorous treatment of the joint source-channel optimization for a specific case in which the source is a discrete-time memoryless source, the source coder is a zero-memory quantizer, and the channel coder is a block coder. Here, we make use of the results in [15] which were developed for a more general scheme, to develop necessary conditions for the joint optimality of the source and the channel coders. Moreover, we develop an algorithmic procedure which is sure to result in a locally optimal system. We will demonstrate through our numerical results that, in certain cases, optimal design

of source-channel codes could result in substantial performance improvements.

The organization of the paper is as follows. In Section II, we present a description of the overall system and the issues involved in system optimization. In Section III, the necessary conditions for optimality are derived under the squared-error distortion criterion. This is followed by the description of the algorithm for system design in Section IV. In Section V, we have discussed the issue of separability of the source and the channel distortions for this particular problem. Section VI is devoted to the presentation of the numerical results, and, finally, in Section VII, a summary and suggestions for future research are included.

II. Problem Statement and Preliminary Analysis:

In this paper, our analysis is based on the assumption that the source to be encoded $\{X_n\}$ can be modeled as a discrete-time zero-mean memoryless stationary process with probability density function (p.d.f.) $p(x)$ and variance $\sigma_X^2 = E\{X_n^2\}$.

The block diagram of the system we wish to analyze is illustrated in Figure 1. In the first stage of the encoding process, the source output, which belongs to a continuous alphabet, is discretized by means of a zero-memory quantizer. An N-level zero-memory quantizer is a function $q(.)$ which maps the source output $x \epsilon R$ into one of N values $Q_1, Q_2, \ldots, Q_N$ each in R. This operation can be described in terms of the threshold levels $T_1, T_2, \ldots, T_{N-1}$, partitioning the support of p into N disjoint and exhaustive regions described by $I_\ell = (T_{\ell-1}, T_\ell]$, $\ell = 1,2,\ldots,N$, with $(T_0, T_N] = (\alpha,\beta]$ describing the support of p, such that $x \epsilon R$ is mapped to $y = q(x) \epsilon A_Y \overset{\Delta}{=} \{Q_1, Q_2, \ldots, Q_N\}$ according to

$$q(x) = Q_\ell \ , \quad x \epsilon I_\ell \ , \quad \ell = 1,2,\ldots,N. \tag{II.1}$$

Here, the T_ℓ are called the quantization thresholds and the Q_ℓ are called the quantization (representation) levels.

As indicated in Fig. 1, subsequent to quantization (source encoding), a channel encoder described by the mapping $f: A_Y \to A_U = \{0,1,\ldots,M-1\}$, with $M = 2^r > N$, operates upon the quantizer output Y_n to generate a sequence $U_n = f(Y_n)$ of integers. The binary representation of these integers will then constitute r-bit codes which will be delivered to a binary symmetric channel (BSC) with crossover probability ε. The channel operates on these r-bit long codes on a bit-by-bit basis. Let us suppose that $V_n \varepsilon A_U$ denotes the decimal representation of the r-bit binary code at the channel output. Obviously, for $\varepsilon \neq 0$, there is a nonzero probability that V_n does not equal U_n. From this point on, for the sake of simplicity, we consider our channel as a channel with input U and output V both in A_U. Note that computing the probability matrix of this channel is straightforward.

Finally, to complete the description of the encoding/decoding operation, we should mention the last stage of the operation which consists of taking the channel output V_n and decoding it into $\hat{X}_n = g(V_n)$ that can take on one of the M possible values, called the decoder <u>reconstruction levels</u>, from the set $A_{\hat{X}} = \{R_1, R_2, \ldots, R_M\}$. It is important to note that, in general, M need not be equal to N. Furthermore, even if M = N, the R_ℓ's are not necessarily equal to the Q_ℓ's. We shall elaborate on this issue shortly.

Our goal in designing the above encoding/decoding scheme is to minimize the average distortion between X and $\hat{X}$ for a fixed number of transmitted bits per source symbol, say, r. In the sequel, we will focus attention on the squared-error distortion criterion. Thus, the mean squared-error (MSE) incurred in this operation, which depends upon the choice of the quantizer q, the channel encoder f, and the decoder g, is given by[+]

[+] It must be noted that the quantity D(q,f,g) is also a function of the channel crossover probability ε. This dependency is not reflected in our notation for the sake of brevity.

$$D(q,f,g) = E\{(X - \hat{X})^2\} . \tag{II.2}$$

Let us introduce the following notation

$$u_k = v_k = k - 1 , \quad k = 1,2,\ldots,M , \tag{II.3}$$

and define the channel transition probability from U to V by

$$p_{k|m} = \Pr\{V = v_k | U = u_m\} , \quad m,k = 1,2,\ldots,M . \tag{II.4}$$

Then, (II.2) can be expanded in the following form:

$$D(q,f,g) = \sum_{k=1}^{M} \sum_{\ell=1}^{N} \Pr\{V = v_k | U = f(Q_\ell)\} \int_{T_{\ell-1}}^{T_\ell} (x - R_k)^2 p(x)dx , \tag{II.5}$$

in which $R_k = g(v_k)$. Equation (II.5) can be further simplified to

$$D(q,f,g) = \sum_{k=1}^{M} \sum_{\ell=1}^{N} p_{k|\hat{m}(\ell)} \int_{T_{\ell-1}}^{T_\ell} (x - R_k)^2 p(x)dx , \tag{II.6}$$

where $\hat{m}(.)$ is defined by

$$f(Q_\ell) = u_{\hat{m}(\ell)} , \quad \ell = 1,2,\ldots,N . \tag{II.7}$$

We can now state our objective more formally. We wish to minimize the overall MSE given by (II.6) by appropriate choices of q, f, and g. It is evident from (II.6) that the Q_ℓ's do not play any role in determining the overall MSE. This makes sense because the Q_ℓ's are only intermediate entities in our system whose indices are used to determine the quantization interval to which the source output has fallen, and their actual values are of no significance in determining the overall MSE. With this in mind, the MSE is only a function of the quantization thresholds $\underline{T} = (T_1,T_2,\ldots,T_{N-1})$, the channel encoder f(.), and the reconstruction levels $\underline{R} = (R_1,R_2,\ldots,R_M)$.

The problem of obtaining the optimum thresholds $\underline{T}^*$, the optimum reconstruction levels $\underline{R}^*$, and the optimum encoding function f* will be addressed in the

next section. Notice that f is a function from a finite alphabet A_Y to a finite alphabet A_U, and hence there is only a finite number of possibilities for f. In fact, it is simple to show that the total number of distinct possibilities for f is $M!/(M-N)!$. Various researchers have considered the problem of system design under fixed encoder choices. Among these, Kurtenbach and Wintz [6] consider the problem of quantization for noisy channels. They restricted attention to the natural binary codes (NBC) and the Gray code (GC). Also, Rydbeck and Sundberg [7] consider a similar problem with other code assignments, such as folded binary codes (FBC) and minimum distance codes (MDC).

In any case, upon fixing the encoder structure f, the problem is that of determining the vectors $\underline{T}^*$ and $\underline{R}^*$. For a fixed f and $\underline{R}$ the best set of the threshold levels $\underline{T}$ is determined by minimizing $D(q,f,g)$ with respect to $\underline{T}$ subject to the constraint that $T_\ell < T_{\ell+1}$, $\ell = 0,1,\ldots N-1$. This constraint will be referred to as the <u>realizability constraint</u>.

Let us for the moment drop the realizability constraint and attempt to solve the unconstrained minimization problem. The best set of the threshold levels $\underline{T}$ is determined by setting the partial deviatives of (II.6) with respect to the T_ℓ's equal to zero, i.e.,

$$\frac{\partial}{\partial T_\ell} D(q,f,g) = \sum_{k=1}^{M} \left[P_{k|\hat{m}(\ell)} - P_{k|\hat{m}(\ell+1)}\right](T_\ell - R_k)^2 p(T_\ell) = 0 \ , \qquad \text{(II.8)}$$

which, in turn, implies

$$T_\ell = \frac{1}{2} \frac{\sum_{k=1}^{M} \left[P_{k|\hat{m}(\ell)} - P_{k|\hat{m}(\ell+1)}\right] R_k^2}{\sum_{k=1}^{M} \left[P_{k|\hat{m}(\ell)} - P_{k|\hat{m}(\ell+1)}\right] R_k} , \quad \ell = 1,2,\ldots,N \ . \qquad \text{(II.9)}$$

If this unconstrained minimization results in a solution $\underline{T}$ that satisfies the realizability constraint, then this $\underline{T}$ is also the solution to the constrained problem.

On the other hand, fixing $\underline{T}$ and f and setting partial derivatives of (II.6) with respect to the R_k's equal to zero, yields

$$\frac{\partial}{\partial R_k} D(q,f,g) = - \sum_{\ell=1}^{N} P_{k|\hat{m}(\ell)} \int_{T_{\ell-1}}^{T_\ell} (x - R_k)p(x)dx = 0 , \tag{II.10}$$

which, in turn, implies

$$R_k = \frac{\sum_{\ell=1}^{N} P_{k|\hat{m}(\ell)} \int_{T_{\ell-1}}^{T_\ell} xp(x)dx}{\sum_{\ell=1}^{N} P_{k|\hat{m}(\ell)} \int_{T_{\ell-1}}^{T_\ell} p(x)dx} , \quad k = 1,2,\ldots,M . \tag{II.11}$$

From an estimation theory point of view, for fixed $\underline{T}$ and f, the reconstruction levels should be chosen to minimize the reconstruction error $E\{(X - \hat{X})^2\}$ upon observing V. Therefore, from well-known results of estimation theory [16], the best R_k is given by the conditional expectation of X given $V = v_k$, i.e.,

$$R_k = E\{X|V = v_k\} , \tag{II.12}$$

which is just a concise form of (II.11).

Equations (II.9) and (II.11) constitute a set of M + N nonlinear simultaneous equations whose solution (if it exists) results in vectors $\underline{T}^*$ and $\underline{R}^*$ that satisfy the necessary conditions for optimality for a fixed encoding rule f. It is important here to note that (II.9) and (II.11) are simply the generalized versions of the similar set of necessary conditions obtained by Kurtenbach and Wintz [6] for the case in which M = N.

Several comments about the above necessary conditions for optimality are in order. These comments are especially important as they provide motivation for this work and underscore the contribution of this paper. Kurtenbach and Wintz [6] proposed an iterative algorithm for solving equations (II.9) and (II.11). This algorithm, which is a straightforward extension of Lloyd's

algorithm (1st method) for minimum distortion quantizer design [7], suffers from some difficulties which are not mentioned in [6]. Most important of all, it can be proven that Lloyd's iterative algorithm for quantizer design actually converges. This is easily shown by proving that the average distortion is a non-increasing function of the number of iterations. Here, however, the same result does not hold. Specifically, it is straightforward to prove that for a fixed set of threshold levels $\underline{T}$, updating the reconstruction levels $\underline{R}$ according to (II.11) cannot increase the overall MSE. But, unfortunately, if we fix $\underline{R}$ and update $\underline{T}$ according to (II.9), it does not immediately imply that the distortion is not increased. In fact, the matrix of second derivatives of D(q,f,g) with respect to T_ℓ's given by (II.9) is described by

$$\frac{\partial^2}{\partial T_m \partial T_\ell} D(q,f,g) = \begin{cases} 0 \quad ; & m \neq \ell \\ 2p(T_\ell) \sum_{k=1}^{M} [P_{k|\hat{m}(\ell+1)} - P_{k|\hat{m}(\ell)}] R_k ; & m = \ell \end{cases} , \qquad \text{(II.13)}$$

which implies that the second derivative matrix is a diagonal matrix whose entries are described by (II.13). To ensure that the solution to (II.9) does not increase the average distortion, we need to show that this matrix is non-negative definite, or,

$$\sum_{k=1}^{M} [P_{k|\hat{m}(\ell+1)} - P_{k|\hat{m}(\ell)}] R_k = E\{\hat{X}|Y = Q_{\ell+1}\} - E\{\hat{X}|Y = Q_\ell\} \geq 0, \ \ell = 1,2,\ldots,(N-1). \qquad \text{(II.14)}$$

Therefore, a sufficient condition for the convergence of the algorithm is that (II.14) is satisfied at every step of the iteration process. The second important point is the realizability constraint. This needs to be stated explicitly because there is no guarantee that the unconstrained minimization will result in a solution that satisfies the realizability constraint. This is in marked contrast to the noiseless channel case (Lloyd's Aglorithm) where the realizability constraint is

always satisfied and hence is not stated explicitly.

In summary, the algorithm in [6] suffers from two major problems: (i) lack of proof of convergence, and (ii) lack of guarantee for realizability. It can further be shown (Sec. IV) that if the unconstrained minimization results in a realizable solution then the matrix of 2nd derivatives is non-negative definite and hence the average distortion is not increased when we update $\underline{T}$.

In what follows, we present an algorithm in which both of the above problems are resolved and a locally optimal system is obtained in which the quantizer, the encoder, and the decoder satisfy the necessary conditions for optimality all at the same time.

III. Necessary Conditions for Optimal System Design:

We approach the problem of optimal quantizer design from a slightly different perspective now. Our goal is to come up with a set of necessary conditions for the optimality of the quantizer, the channel encoder, and the decoder. Furthermore, we wish to develop an algorithm which does not suffer from the realizability and convergence problems of the Kurtenbach-Wintz algorithm [6].

Here, instead of separating the quantizer (source encoder) and the channel encoder, we concentrate on designing an encoder whose input is the source output $\{X_n\}$ and whose output is the channel input $\{U_n\}$. In essence, we search for an optimal mapping $\gamma: R \to A_U$ which is described by the composite function $\gamma(x) \equiv f[q(x)]$.

The approach consists of two stages. First, for a fixed decoder g, we develop necessary conditions for optimality of the encoder function γ. Then, for a fixed γ, we develop necessary conditions for optimality of g. The pair of conditions given in these two stages can then be used to establish a set of conditions for the optimality of the entire system. The same approach has been used by Fine [15] for optimization of a more general system. Certain extensions

of Fine's work can be found in [18].

Let us now proceed by describing these necessary conditions for optimality.

We assume first that the decoder g is known and fixed. We would like to determine the best encoder structure. Recall that our objective is to minimize the MSE given by

$$D(\gamma,g) \equiv D(q,f,g) = \int_{-\infty}^{\infty} p(x)E\{(X - \hat{X})^2 | X = x\}dx \ . \qquad \text{(III.1)}$$

Since p(x) is a non-negative quantity, to minimize $D(\gamma,g)$ it suffices to minimize $E\{(X - \hat{X})^2 | X = x\}$. But,

$$E\{(X - \hat{X})^2 | X = x\} = \sum_{i=1}^{M} E\{(X - \hat{X})^2 | X = x,\ U = u_i\}Pr\{U = u_i | X = x\}. \qquad \text{(III.2)}$$

Notice that our encoder is a deterministic mapping, and hence it maps a given x to some $u = \gamma(x) \in A_U$. This implies that for this u,

$$Pr\{U = u | X = x\} = 1 \ , \qquad \text{(III.3)}$$

and, hence,

$$E\{(X - \hat{X})^2 | X = x\} = E\{(X - \hat{X})^2 | X = x,\ U = u\}$$

$$= E\{(x - \hat{X})^2 | U = u\} \ . \qquad \text{(III.4)}$$

Therefore, to minimize $D(\gamma,g)$, we have to obtain a mapping γ that minimizes (III.4) for every value of x. In other words, upon defining the set $A_i(g)$ as the collection of all values of x that should be encoded to the ith channel input, i.e., u_i, we must satisfy

$$A_i(g) = \{x : E\{(x - \hat{X})^2 | U = u_i\} < E\{(x - \hat{X})^2 | U = u_j \text{ for all } j \neq i\}\}, \qquad \text{(III.5)}$$

which can be written as

$$A_i(g) = \bigcap_{\substack{j=1 \\ j\neq i}}^{M} A_{ij}(g) \ , \tag{III.6}$$

where $A_{ij}(g)$ is defined as

$$A_{ij}(g) = \{x: 2x[E\{\hat{X}|U = u_j\} - E\{\hat{X}|U = u_i\}] < E\{\hat{X}^2|U = u_j\} - E\{\hat{X}^2|U = u_i\}\}. \tag{III.7}$$

Here, $A_{ij}(g)$ specifies a set on the real line whose members, if mapped to u_i instead of u_j, will result in a lower MSE.

Let us now look more closely at (III.7) as it becomes important in our algorithm. Upon defining,+

$$\alpha_{ij} \triangleq E\{\hat{X}^2|U = u_j\} - E\{\hat{X}^2|U = u_i\} \ , \tag{III.8a}$$

$$\beta_{ij} \triangleq E\{\hat{X}|U = u_j\} - E\{\hat{X}|U = u_i\} \ , \tag{III.8b}$$

and

$$t_{ij} \triangleq \frac{\alpha_{ij}}{2\beta_{ij}} \ , \qquad \beta_{ij} \neq 0 \ , \tag{III.8c}$$

we have

$$A_{ij}(g) = \begin{cases} (-\infty, t_{ij}] \ , & \beta_{ij} > 0 \ , \\ [t_{ij}, \infty) \ , & \beta_{ij} < 0 \ , \\ (-\infty, \infty) \ , & \beta_{ij} = 0 \ , \quad \alpha_{ij} > 0 \ , \\ \phi \ , & \beta_{ij} = 0 \ , \quad \alpha_{ij} < 0 \ . \end{cases} \tag{III.9}$$

Therefore, $A_{ij}(g)$ is an <u>interval</u> and hence $A_i(g)$ is also an interval as it is a finite intersection of intervals. If we define the upper and lower endpoints of $A_i(g)$ by

$$t_i^u = \min_{j:\beta_{ij}>0} t_{ij} \ , \tag{III.10a}$$

+Notice that α_{ij}, β_{ij}, and t_{ij} all depend upon the decoder structure g. But, here, to avoid complicating notation, this direct dependence is not reflected in our notation.

and

$$t_i^{\ell} = \max_{j:\beta_{ij}<0} t_{ij} \quad , \tag{III.10b}$$

respectively, we can characterize $A_i(g)$ by

$$A_i(g) = \begin{cases} [t_i^{\ell}, t_i^{u}] \ , \\ \phi \quad , \text{ if } \beta_{ij} = 0 \text{ and } \alpha_{ij} < 0 \text{ for some } j. \end{cases} \tag{III.11}$$

Therefore, the optimum encoder mapping γ for a fixed decoder g is given by

$$\gamma(x) = u_i \quad , \quad x \in A_i(g) \quad , \quad i = 1,2,\ldots,M \ , \tag{III.12}$$

where $A_i(g)$ can be explicitly determined by (III.11). Notice that (III.12) satisfies the <u>necessary</u> and <u>sufficient</u> conditions for the optimality of γ, given a fixed g.

On the other hand, for a fixed encoder γ, the optimal decoder is described, as mentioned in the previous section, by the conditional expectation of the input given the channel output. Specifically,

$$R_i(\gamma) = g(v_i) = E\{X|V = v_i\} \quad , \quad i = 1,2,\ldots,M \ . \tag{III.13}$$

It should be mentioned that (III.13) is also the <u>necessary</u> and <u>sufficient</u> condition for the optimality of g, given γ.

Before we proceed to give an elaborate description of the algorithm for system design, there are several comments concerning the conditions described by (III.12) and (III.13). First, it is of fundamental importance to note that in (III.11) we could encounter a situation in which $t_i^{\ell} > t_i^{u}$ even though $\beta_{ij} \neq 0$ for all $j \neq i$. This, in fact, implies that no value of x should be encoded as u_i, or, equivalently, $A_i(g) = \phi$. As we will describe in the numerical results section, this phenomenon, which occurs when the channel is highly noisy, indicates that in certain cases the total number of quantization regions should be smaller than the

total number of available codes M. Secondly, an iterative algorithm, which works based on successive application of (III.12) and (III.13), guarantees convergence. This is because the sequence of values of the MSE obtained in these iterations is a nonincreasing sequence of nonnegative numbers, and hence it converges. Furthermore, despite the fact that (III.12) and (III.13) are both necessary and sufficient conditions when g and γ are fixed, respectively, the final solution obtained by the iterative algorithm need not satisfy the sufficient conditions for the system's optimality. That is, the final solution obtained by this algorithm is only a locally optimum+ solution.

In the following section, we present a precise description of the algorithm used for the optimal encoder/decoder design.

IV. Algorithm.

Our objective is to develop an algorithm based on the results of the previous section that, for a fixed value of ε, generates a locally optimum encoder/decoder pair. We would like to obtain these results for a range of different values of the channel crossover probability ε, say, $0 \leq \varepsilon \leq \varepsilon_{max}$.

The following is a brief description of the algorithm that is used to compute the minimum MSE associated with the locally optimum system.

(a) Set $\varepsilon = \varepsilon_{max}$. Set $\underline{R} = \underline{R}_0$, the initial set of reconstruction levels.

(b.1) Set $i = 0$ (i is the iteration index). Choose an initial set of reconstruction levels $\underline{R}^{(0)} = \underline{R}_0$. Set $D^{(0)} = \infty$.

(b.2) Use (III.12) to obtain the best encoder $\gamma^{(i)}$ for the fixed $\underline{R}^{(i)}$.

(b.3) Set $i = i + 1$. Use (III.13) to obtain the best set of reconstruction levels $\underline{R}^{(i)}$ for the fixed $\gamma^{(i-1)}$.

+An example of the situation with multiple locally optimum solutions is the case of $\varepsilon = 0$. In this case, our problem reduces to that of minimum MSE quantizer design and our algorithm is that of Lloyd's 1st method [17]. It is well-known in this case that the algorithm may not converge to the globally optimum solution [20].

(b.4) Compute the MSE $D^{(i)}$ associated with $\underline{R}^{(i)}$ and $\gamma^{(i-1)}$. If $(D^{(i-1)}-D^{(i)})/D^{(i)} < \delta$, where $\delta > 0$, go to (c). Otherwise, go to (b.2).

(c) Set $\varepsilon = \varepsilon - \Delta\varepsilon$ ($\Delta\varepsilon > 0$ is the increment by which ε is changed). If $\varepsilon > 0$, go to (b.1), otherwise, stop.

Clearly, in steps (b.1)-(b.4), the locally optimum system for a fixed value of ε is computed. In the actual implementation of the algorithm, to insure that a <u>good</u> locally optimum system is chosen, we have considered the following variation of the algorithm. At the end of step (c), when $\varepsilon = 0$, we begin to increase ε again by increments of $\Delta\varepsilon$. We continue this until $\varepsilon = \varepsilon_{max}$. By means of this process, we obtain a curve of the MSE as a function of ε which may or may not coincide with the previous one. If it coincides for all values of ε, we stop. Otherwise, we start decreasing ε from ε_{max} down to $\varepsilon = 0$. We will continue this process back and forth until some two MSE curves coincide, in which case we stop. Then, for each value of ε, we take the smallest MSE obtained in these iterations.

Before we close this section, we should mention a modification used in (b.2) which helps reduce the complexity of the algorithm.

Careful examination of (III.12) reveals that in order to determine $A_i(g)$ at each step of the iteration, we need to make M computations. The following procedure will result in a noticeable amount of saving in the number of computations, and hence the complexity of the algorithm.

After computing $E\{\hat{X}|U = u_i\}$, $i = 1,2,\ldots,M$, we arrange them in the increasing order, i.e., we reshuffle the codes and the corresponding reconstruction levels in such a way that

$$E\{\hat{X}|U = u_1'\} < E\{\hat{X}|U = u_2'\} < \ldots < E\{\hat{X}|U = u_M'\}, \qquad \text{(IV.1)}$$

in which $u_i' \varepsilon A_U$ is such that $E\{\hat{X}|U = u_i'\}$ is the ith smallest element of the set $\{E\{\hat{X}|U = u_i\}, i = 1,2,\ldots,M\}$. Note that this reshuffling of the codes does not affect our analysis of section III at all. However, it introduces substantial simplification in the implementation of the algorithm. The following theorem shows that the above reordering of the codes results in an encoder which partitions the real line in essentially the same order.

Theorem 1: For a given g, or equivalently R, if the codes are reshuffled to satisfy (IV.1), and if $i < j$, it is guaranteed that $A_i(g)$ lies to the left of $A_j(g)$ provided that $A_i(g)$ and $A_j(g)$ are both well-defined, i.e., $t_k^u > t_k^\ell$, $k = i,j$. The converse is also true, i.e., if we obtain an encoder for which $A_i(g)$ lies to the left of $A_j(g)$, then $E\{\hat{X}|U = u_i'\} < E\{\hat{X}|U = u_j'\}$.

Proof: Let us use the same notation as in section III for the rehsuffled code. Then, it is easy to see that+

$$t_i^u = \min_{k>i} \{t_{ik}\} \quad , \quad i = 1,2,\ldots,M-1 \ , \tag{IV.2.a}$$

and, similarly,

$$t_i^\ell = \max_{k<i} \{t_{ik}\} \quad , \quad i = 2,3,\ldots,M \ . \tag{IV.2.b}$$

Let us suppose that $t_k^u > t_k^\ell$, $k = i,j$. Then, to prove that $A_i(g)$ lies to the left of $A_j(g)$ for $i < j$, it suffices to show that $t_i^u < t_j^\ell$. But

$$t_i^u = \min_{k>i} \{t_{ik}\} < t_{ij} \quad , \tag{IV.3.a}$$

and

$$t_j^\ell = \max_{k<j} \{t_{jk}\} > t_{ji} \ . \tag{IV.3.b}$$

+Here, $t_1^\ell = \alpha$ and $t_M^u = \beta$. Also, in the sequel without loss of generality, we will assume $\alpha = -\infty$ and $\beta = \infty$.

But, by definition, $t_{ij} = t_{ji}$, which implies $t_i^u < t_j^\ell$.

To prove the converse, we proceed as follows: Since $A_i(g)$ lies to the left of $A_j(g)$, we have

$$E\{(\hat{X}-x)^2|U = u_i'\} < E\{(\hat{X} - x)^2|U = u_j'\}, \text{ for } x < t_{ij} \quad . \tag{IV.4.a}$$

Also, from the definition of t_{ij}, we have

$$E\{(\hat{X} - t_{ij})^2|U = u_i'\} = E\{(\hat{X} - t_{ij})^2|U = u_j'\}. \tag{IV.4.b}$$

Subtracting (IV.4.b) from (IV.4.a) and further simplification yields

$$(t_{ij} - x)E\{\hat{X}|U = u_i'\} < (t_{ij} - x)E\{\hat{X}|U = u_j'\}, \text{ for } x < t_{ij}, \tag{IV.5}$$

which, in turn, implies

$$E\{\hat{X}|U = u_i'\} < E\{\hat{X}|U = u_j'\}. \tag{IV.6}$$

The following corollary shows that intervals with neighboring indices are actually adjacent.

<u>Corollary 1</u>: If $t_k^u > t_k^\ell$, $k = i, i + 1$, then $t_i^u = t_{i+1}^\ell$, $i = 1,2,\ldots,M-1$.

<u>Proof</u>: We know from Theorem 1 that $t_i^u < t_{i+1}^\ell$. Let us assume that $t_i^u < t_{i+1}^\ell$. Since $t_{i+1}^\ell < t_k^\ell$, $k = i+2,\ldots, M$, the interval (t_i^u, t_{i+1}^ℓ) will be mapped to no code. This is impossible because for every $x \epsilon R$, there must exist a code which minimizes the distortion. Recall that condition (III.5) is a necessary and sufficient condition.

The only issue which remains to be resolved is that of determining those codes for which the corresponding upper endpoint is strictly smaller than the lower endpoint. Let us denote the set of all i's satisfying this condition by I^*, given by

$$I^* = \{i : t_i^u \leq t_i^\ell\} . \tag{IV.7}$$

Essentially, $i \in I^*$ implies that for every $x \in R$ there is always a code $j \notin I^*$ that performs better than (or equally well as) the ith code. Therefore, I^* is the set of all "useless" codes which should be discarded.

The following theorem provides a simple way of simultaneous identification of the best partition and the codes in I^*.

Theorem 2: Let us suppose that the codes are reshuffled to satisfy (IV.1). Let us suppose that the ith interval is well-defined, i.e., $t_i^\ell < t_i^u$, and suppose $t_i^u = t_{ij}$ for some $j = i+1, i+2, \ldots, M$. Then,

(a) all the codes with indices $k = i+1, i+2, \ldots, j-1$ belong to I^*, and

(b) the jth code does not belong to I^*, i.e., $t_j^\ell \leq t_j^u$.

Proof: As we have assumed,

$$t_{ij} \leq t_{ik} \quad , \quad k = i+1, i+2, \ldots, M. \tag{IV.8}$$

We will first show that

$$t_{kj} \leq t_{ij} \quad , \quad k = i+1, i+2, \ldots, j-1. \tag{IV.9}$$

Consider $x \in (t_{ij}, t_{ik})$. Then, for encoding this x, the jth code is preferable to the ith code, and the ith code is preferable to the kth code, and hence the jth code is preferable to the kth code, i.e., x must satisfy $x > t_{kj}$. Since this must hold for all x in (t_{ij}, t_{ik}), we must have $t_{kj} \leq t_{ij}$. Notice that in arriving at this result, the fact that $i < k < j$ plays an important role.

Using (IV.9), we have

$$t_k^u = \min_{m>k} \{t_{km}\} \leq t_{kj} \leq t_{ij} = t_i^u \ , \ k = i+1, i+2, \ldots, j-1 \ , \tag{IV.10}$$

which together with the fact that $t_k^\ell > t_i^u$, $k = i+1, i+2, \ldots, j-1$, (see Corollary 1) implies that $t_k^u < t_k^\ell$, $k = i+1, i+2, \ldots, j-1$, which proves (a).

Now, we proceed to show that $t_j^u > t_j^\ell$. We have

$$t_j^u = \min_{k>j} \{t_{jk}\} > \min_{k>j} \{t_{ik}\} > t_{ij} , \tag{IV.11.a}$$

$$t_j^\ell = \max_{k<j} \{t_{jk}\} < t_{ij} = t_i^u , \tag{IV.12.b}$$

which implies that $t_j^u > t_j^\ell$. Therefore, the jth code is well-defined. Also, according to Corollary 1, the jth interval lies to the right of the ith interval and hence $t_j^\ell = t_i^u$, for otherwise there will be a gap between the ith interval and the jth interval to which no code is assigned.

Combining the results of Theorems 1 and 2 and Corollary 1 yields the following very simple algorithm for determining the best partition used in step (b.2) of our algorithm.

(b.2.1) Set $i = 1$.

(b.2.2) Compute t_i^u. Let $j(i)$ be such that $t_i^u = t_{ij(i)}$. Eliminate all codes with indices $i+1, i+2, \ldots, j(i)-1$. Set $t_{j(i)}^\ell = t_i^u$.

(b.2.3) Set $i = j(i)$. If $i < M$, go to (b.2.2). Otherwise, set $t_i^u = \infty$, stop.

It is important to note, as we will emphasize in the numerical results section, that for highly noisy channels, it turns out that a large number of quantization intervals belong to I^*. In this case, the algorithm for generating the best partition results in a quick way of identifying these "bad" codes and hence is much more efficient than ad hoc algorithms, which simply compute all the thresholds and eliminate the "bad" codes.

Before leaving this section, let us prove our assertion in Section II that if, while updating the thresholds, a solution to the unconstrained minimization problem

results in a realizable quantizer, then the MSE will not have increased. Recall that the solution to the unconstrained problem is given by (II.9). If (II.9) should result in a solution for which $T_1 < T_2 < \ldots < T_{N-1}$ holds, then Theorem 1 implies that $E\{\hat{X}|U = u_i\} < E\{\hat{X}|U = u_{i+1}\}$, $i = 0,1,\ldots,N-1$, which, in turn, implies that the matrix of second derivatives given by (II.13) is non-negative definite. Hence, when we update $\underline{T}$ by (II.9), the MSE will not increase if the realizability condition is satisfied.

We are now in a position to present the results of this algorithm as applied to the encoding of specific memoryless sources. However, before presenting the results, in the next section we examine the interaction between the quantization error effects and the channel error effects.

V. Separability of the Overall MSE.

In this section, we will study the relationship between the overall MSE and the MSE due to the quantization noise and the channel noise separately. This separability issue (or the lack thereof) has been addressed in various places such as [5]-[7], and [19]. It would be convenient, as it is sometimes erroneously assumed, if the overall MSE could be written as the sum of the quantization noise and the channel noise. Indeed, Totty and Clark [19] have shown that this is the case when the decoder reconstruction levels are chosen to be the <u>same</u> as the encoder quantization levels (i.e., with our notation, $N = M$ and $R_\ell = Q_\ell$, $\ell = 1,2,\ldots,M$), and when the Q_ℓ's are chosen to be the center of probability mass of their corresponding intervals.

We would like to prove the same type of separability in our scheme which is more general than that in [19]. In our scheme, in which the encoder is described in terms of a partitioning of the real line followed by a mapping to the code alphabet, the first outstanding issue is that of "defining" the quantization noise. This is because, in the scheme described in section III, there is no mention of the quantization levels in the encoder.

In what follows, we will show that if the quantization noise is defined as the smallest possible squared-error distortion attainable by a quantizer whose threshold levels are the same as the threshold levels of the encoder, then certain nice properties hold, and the overall MSE can be decomposed into the sum of the quantization noise and the channel noise.

Theorem 3: Let us suppose that in the encoder/decoder pair described in section III, for a fixed encoder, the optimum decoder reconstruction levels $\underline{R}$ is chosen. We define

$$Q_\ell = \frac{\int_{T_{\ell-1}}^{T_\ell} xp(x)dx}{\int_{T_{\ell-1}}^{T_\ell} p(x)dx}, \quad \ell = 1,2,\ldots,N, \tag{V.1}$$

in which the T_ℓ's are used to denote the boundaries of the encoder partitions. Then, the following hold:

(a) The best reconstruction levels R_k are given by

$$R_k = E\{X|V = v_k\} = E\{Y|V = v_k\}, \quad k = 1,2,\ldots,M, \tag{V.2}$$

(b) $$D = E\{(X - \hat{X})^2\} = E\{X - Y)^2\} + E\{(Y - \hat{X})^2\}, \tag{V.3}$$

i.e., the overall MSE is decomposable, and

(c) $$D = \sigma_X^2 - \sigma_{\hat{X}}^2 \tag{V.4}$$

where $\sigma_{\hat{X}}^2$ is the variance of $\{\hat{X}_n\}$.

Proof:

(a) We first note that

$$E\{X|V = v_k\} = \int_{-\infty}^{\infty} xp(x|V = v_k)dx \quad ,$$

$$= \sum_{\ell=1}^{N} \frac{Pr(Y = Q_\ell, V = v_k)}{Pr(Y = Q_\ell)Pr(V = v_k)} \int_{T_{\ell-1}}^{T_\ell} xp(x)dx \ . \tag{V.5}$$

Now, upon substitution of (V.1) in (V.5), we will obtain

$$E\{X|V = v_k\} = \sum_{\ell=1}^{N} \frac{Pr(Y = Q_\ell, V = v_k)}{Pr(V = v_k)} Q_\ell = E\{Y|V = v_k\} \ . \tag{V.6}$$

(b) We have

$$D = E\{(X - \hat{X})^2\} = E\{(X - Y)^2\} + E\{(Y - \hat{X})^2\} + 2E\{(X - Y)(Y - \hat{X})\}. \tag{V.7}$$

An immediate conclusion of (V.1) is that

$$E\{(X - Y)Y\} = 0 \ . \tag{V.8}$$

Furthermore,

$$E\{X\hat{X}\} = \sum_{k=1}^{M} E\{X\hat{X}|\hat{X} = R_k\}Pr(\hat{X} = R_k)$$

$$= \sum_{k=1}^{M} R_k E\{X|V = v_k\}Pr(V = v_k) \ . \tag{V.9}$$

Now, if we use the result of part (a), we have

$$E\{X\hat{X}\} = \sum_{k=1}^{M} R_k E\{Y|V = v_k\}Pr(V = v_k) = E\{Y\hat{X}\} \ , \tag{V.10}$$

or, equivalently,

$$E\{(X - Y)\hat{X}\} = 0 \ . \tag{V.11}$$

Combining (V.8), (V.11), and (V.7) yields the desired result in (V.3).

(c) It is easy to show that the quantization noise can be written as

$$E\{(X - Y)^2\} = \sigma_X^2 - \sigma_Y^2 \quad ; \tag{V.12}$$

similarly, as a result of (V.2), we have

$$E\{(Y - \hat{X})^2\} = \sigma_Y^2 - \sigma_{\hat{X}}^2 \ . \tag{V.13}$$

Therefore, combining (V.12), (V.13), and (V.3) yields (V.4).

The above theorem has enabled us to define the quantization noise in such a way that a separation between the quantization and the channel noise has become possible. In the next section, we will make use of this result in studying the interaction between these two sources of distortion.

VI. Numerical Results:

In this section, we provide numerical results determining the performance of the system described in section III operating on various source distributions. We make comparisons against the performance of the Lloyd-Max quantization scheme encoded by natural binary codes as well as the optimum performance theoretically attainable (OPTA) obtained through rate-distortion theoretic arguments.

A. Source Description:

In what follows, we assume that the p.d.f. of the assumed memoryless source is chosen from the class of generalized Gaussian distributions described by

$$p(x) = \left[\frac{\alpha\eta(\alpha,\beta)}{2\Gamma(1/\alpha)}\right] \exp\left\{-\left[\eta(\alpha,\beta)|x|\right]^{\alpha}\right\} \ , \ -\infty < x < \infty \ , \tag{VI.1.a}$$

where

$$\eta(\alpha,\beta) \triangleq \beta^{-1}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2} \ , \tag{VI.1.b}$$

with $\alpha > 0$ describing the exponential rate of decay, β a positive quantity representing a scale parameter, and $\Gamma(.)$ is the gamma function. The variance of the associated random variable is given by $\sigma_X^2 = \beta^2$. For $\alpha = 2$, we have the

Gaussian distribution, while for $\alpha = 1$ we have the Laplacian distribution. The generalized Gaussian distribution with values of α in the range $0.1 < \alpha < 1.0$ provides a useful model for broad-tailed processes [21]. It is also useful to note that for large values of α, the distribution tends to a uniform distribution. Typical behavior of p.d.f.'s as a function of normalized input are illustrated in Fig. 2 for selected parameter choices.

B. Rate-Distortion Derived Bounds:

It is well-known from the results of source coding subject to a fidelity criterion that for a channel with capacity C bits/channel use, and a source with distortion-rate function D(R), where R is measured in bits/source symbol, the smallest attainable average distortion is given by [2]-[4],

$$D_{min} = D(rC) \ , \tag{VI.2}$$

in which r is the number of channel uses for each source symbol.

In our problem, the channel is a binary symmetric channel with crossover probability ε. It is easily shown [3] that, in this case,

$$C = 1 + \varepsilon \log_2 \varepsilon + (1 - \varepsilon) \log_2 (1 - \varepsilon), \text{ bits/channel use.} \tag{VI.3}$$

In general, closed form expressions for D(R) do not exist; however, Blahut's algorithm [22] can be used for numerical computation of D(R).

C. Lloyd-Max Quantizer Performance on Noisy Channels:

We compare our results against the performance of the Lloyd-Max quantizer optimized for the same source, when the quantizer output is encoded by natural binary codes and transmitted via a binary symmetric channel. In effect, we compare the performance of our system designed by the noisy channel against a system whose designer has been completely ignorant about the channel noise but has made every effort to minimize the quantization noise. We will

denote the performance of this system by $D_{LM}(R)$. This quantity can be easily computed from (II.6). It might be interesting to note at this point that for small values of ε ($\varepsilon \approx 0$), $D_{LM}(R)$ grows approximately linearly with respect to ε.

D. Performance Results:

For the selected values of α used in Fig. 2, we have obtained the performance of the locally optimum encoder/decoder pair of Section III for different values of ε in the range [0,0.1]. The resulting signal-to-noise ratios are illustrated in Figs. 3-6. In these figures, we have also included the optimum attainable performance based on (VI.2) and the performance of the Lloyd-Max quantizer transmitted using the natural binary code. Similar results are tabulated in Tables 1-4 for r = 1,2,3, and 4 bits/sample, and ε = 0.005, 0.01, 0.05, and 0.1.

It can be concluded from these results that the optimized system offers performance improvements over the Lloyd-Max quantizer. Let us emphasize here that these performance improvements are quite substantial; for example, for the Laplacian source which closely approximates the distribution of speech signals, at ε = 0.01, a noticeable 3.38 dB SNR improvement is obtained for r = 4 bits/sample (Table 2). The general trend that can be observed from the results in Figs. 3-6 or Tables 1-4 is that this performance improvement becomes more noticeable for larger r, larger ε, and more broad-tailed densities. Indeed, for r = 4 bits/sample, ε = 0.1 and α = 0.5 (Table 1), the SNR improvement is more than 11.5 dB.

One could conclude from the above discussion that the optimal design of the system in general results in performance improvements that cannot be neglected, especially if the channel is very noisy.

We have used (V.3) to write the overall MSE in terms of the quantization MSE and the channel MSE, separately. It is observed that the quantization MSE increases as a function of ε. In other words, the system design is such that the quantizer's struc-

ture is changed, as ϵ is increased, in such a way that more distortion will be incurred in the quantization portion (source coding) so that the overall MSE is minimized. This type of exchange of MSE between the source and channel encoder, of course, does not exist in Shannon's coding arguments.

In Fig. 8, we have illustrated the behavior of the quantization thresholds in an 8-level optimum scheme as a function of ϵ. It is interesting to observe that, while the quantizer starts off with 8 levels at $\epsilon = 0$, for some value of $\epsilon > 0$ it switches to a 4-level quantizer (i.e., higher quantization noise). Indeed, our experimental results have shown that for some value of $\epsilon > 0.1$, this quantizer will switch again to a 2-level quantizer. In other words, consistent with our claims in Sections III and IV, some of the codes become "useless" (those denoted by I^* in Section IV) as the channel becomes noisier, and, therefore, they will not be used in transmission of any values of the source output.

VII. Summary and Conclusions:

We have studied the problem of optimal quantization and coding when the quantizer outputs are to be transmitted via a noisy channel. Only memoryless sources, zero-memory quantization, and memoryless channels have been considered. An iterative algorithm has been developed for obtaining a locally optimal quantizer and coder. On the basis of the numerical results obtained for a wide variety of sources, it can be concluded that this design technique offers substantial improvements over Lloyd-Max quantization followed by natural binary codes. The improvement is more noticeable for more noisy channels, for more broad-tailed densities, and at higher bit rates.

Having defined the quantization error properly, we have shown that under certain conditions which are satisfied in our problem, the overall MSE can be decomposed into the sum of the quantization noise and the channel noise. It is shown through numerical results that in the optimal system, there is close

interaction between the quantization noise and the channel noise to minimize the overall MSE.

An open problem which remains unresolved is that of obtaining the globally optimum system. At this point, the locally optimum system to which our algorithm converges depends upon the initial point, and we know of no way of choosing the initial point to ensure convergence to the globally optimum system.

A second interesting problem worthwhile studying is the effect of channel noise in block transform encoding of sources with memory. In this context, we must mention that the optimal design of the system not only affects the structure of the quantizer and the encoder, but it could affect the bit assignment too [23]. The study of this problem in block transform coding of Gauss-Markov sources is currently underway.

Finally, in most practical cases, the exact value of the channel crossover probability is not known. An interesting issue is that of designing a robust system where only a partial knowledge of the channel's characteristics is available.

REFERENCES

1. C. E. Shannon, "A Mathematical Theory of Communication," Bell Syst. Tech. J., Vol. 27, pp. 379-423 and 623-656, 1948.

2. C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," IRE Nat. Conv. Rec., pp. 142-163, March 1959.

3. A. J. Viterbi and J. K. Omura, Principles of Digital Communications and Coding, McGraw-Hill, New York, 1979.

4. T. Berger, Rate-Distortion Theory: A Mathematical Basis for Data Compression, Prentice-Hall, Englewood Cliffs, NJ, 1971.

5. J. J. Spilker, Jr., Digital Communications by Satellite, Prentice-Hall, Englewood Cliffs, NJ, 1977.

6. A. Kurtenbach and P. Wintz, "Quantizing for Noisy Channels," IEEE Trans. Commun. Technology, Vol. COM-17, pp. 291-302, April 1969.

7. N. Rydbeck and C. W. Sundberg, "Analysis of Digital Errors in Nonlinear PCM Systems," IEEE Trans. Commun., Vol. COM-24, pp. 59-65, January 1976.

8. T. R. Crimmins, H. M. Horwitz, C. J. Palermo and R. V. Palermo, "Minimization of Mean-Square Error for Data Transmitted Via Group Codes," IEEE Trans. Inform. Theory, Vol. IT-15, pp. 72-78, January 1969.

9. G. A. Wolf and R. Redinbo, "The Optimum Mean-Square Estimate for Decoding Binary Block Codes," IEEE Trans. Inform. Theory, Vol. IT-20, pp. 344-351, May 1974.

10. G. R. Redinbo, "On the Design of Mean-Square Error Channel Coding Systems Using Cyclic Codes," IEEE Trans. Inform. Theory, Vol. IT-28, pp. 406-413, May 1982.

11. J. W. Modestino and D. G. Daut, "Combined Source-Channel Coding of Images," IEEE Trans. Commun., Vol. COM-27, pp. 1644-1659, November 1979.

12. J. W. Modestino, D. G. Daut and A. L. Vickers, "Combined Source-Channel Coding of Images Using the Block Cosine Transform," IEEE Trans. Commun., Vol. COM-29, pp. 1261-1274, September 1981.

13. P. Noll, "Effects of Channel Errors on the Signal-to-Noise Performance of Speech Encoding Systems," Bell Syst. Tech. J., Vol. 54, pp. 1615-1636, November 1975.

14. R. Zelinski, "Effects of Transmission Errors on the Mean-Squared Error Performance of Transform Coding Systems," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, pp. 531-537, October 1979.

15. T. Fine, "Properties of an Optimum Digital System and Applications," IEEE Trans. Inform. Theory, Vol. IT-10, pp. 287-296, October 1964.

16. A. P. Sage and J. L. Melsa, Estimation Theory with Applications to Communications and Control, McGraw-Hill, New York, 1971.

17. S. P. Lloyd, "Least Squares Quantization in PCM," IEEE Trans. Inform. Theory, Vol. IT-28, pp. 129-137, March 1982.

18. J. D. Gibson and T. S. Fischer, "Alphabet-Constrained Data Compression," IEEE Trans. Inform. Theory, Vol. IT-28, pp. 443-457, May 1982.

19. R. E. Totty and G. C. Clark, "Reconstruction Error in Waveform Transmission," IEEE Trans. Inform. Theory, Vol. IT-13, pp. 336-338, April 1967.

20. E. Abaya and G. L. Wise, "Some Notes on Optimal Quantization," in Proc. IEEE International Conf. Commun., Denver, CO, pp. 30.7.1-30.7.5, June 1981.

21. V. R. Algazi and R. M. Lerner, "Binary Detection in White Non-Gaussian Noise," Tech. Rept. DS-2138, M.I.T. Lincoln Lab., Lexington, MA, 1965.

22. R. E. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions," IEEE Trans. Inform. Theory, Vol. IT-18, pp. 460-473, July 1972.

23. J. J. T. Huang and P. M. Schultheiss, "Block Quantization of Correlated Gaussian Random Variables," IEEE Trans. Commun. Syst., Vol. CS-11, pp. 298-296, September 1963.

r in bits/sample	$\varepsilon = 0.005$			$\varepsilon = 0.01$			$\varepsilon = 0.05$			$\varepsilon = 0.1$		
	OPT.	LM	OPTA	OPT.	LM	OPTA	OPT.	LM	OPTA	OPT.	LM	OPTA
1	1.51	1.51	8.97	1.48	1.48	8.84	1.21	1.19	7.39	0.93	0.86	6.10
2	5.14	4.58	15.17	4.69	3.80	14.64	2.76	0.24	12.14	1.68	-1.93	9.69
3	7.28	5.37	20.97	6.18	3.13	20.34	3.50	-2.99	16.67	2.32	-5.70	13.20
4	10.01	3.56	26.95	9.00	0.71	26.05	5.13	-5.96	20.92	3.04	-8.69	16.57

Table 1

SNR (in dB) of the Optimum System (OPT.) and Comparisons with the Lloyd-Max Quantizer with NBC (LM), and the Optimum Performance Theoretically Attainable (OPTA) for the Generalized Gaussian Source with $\alpha = 0.5$.

r in bits/sample	$\varepsilon = 0.005$			$\varepsilon = 0.01$			$\varepsilon = 0.05$			$\varepsilon = 0.1$		
	OPT.	LM	OPTA	OPT.	LM	OPTA	OPT.	LM	OPTA	OPT.	LM	OPTA
1	2.93	2.92	6.42	2.84	2.84	6.21	2.25	2.25	4.89	1.67	1.55	3.72
2	6.81	6.75	12.13	6.26	6.08	11.70	4.05	2.84	9.23	2.80	0.73	7.05
3	9.94	9.30	17.87	8.68	7.44	17.23	5.21	1.76	13.52	3.62	-0.93	10.23
4	12.17	9.78	23.65	10.49	7.11	22.77	6.32	0.52	17.81	4.45	-2.29	13.42

Table 2

SNR (in dB) of the Optimum System (OPT.) and Comparisons with the Lloyd-Max Quantizer with NBC (LM), and the Optimum Performance Theoretically Attainable (OPTA) for the Generalized Gaussian Source with $\alpha = 1.0$ (Laplacian).

r in bits/sample	$\varepsilon = 0.005$			$\varepsilon = 0.01$			$\varepsilon = 0.05$			$\varepsilon = 0.1$		
	OPT.	LM	OPTA	OPT.	LM	OPTA	OPT.	LM	OPTA	OPT.	LM	OPTA
1	4.25	4.25	5.75	4.10	4.10	5.54	3.15	3.09	4.31	2.27	2.09	3.22
2	8.50	8.47	11.49	7.88	7.77	11.07	5.20	4.41	8.59	3.63	2.22	6.39
3	11.88	11.61	17.24	10.49	9.85	16.60	6.47	4.28	12.89	4.67	1.56	9.59
4	14.15	12.92	22.99	12.30	10.33	22.13	7.81	3.79	17.18	5.60	0.93	12.97

Table 3

SNR (in dB) of the Optimum System (OPT.) and Comparisons with the Lloyd-Max Quantizer with NBC (LM), and the Optimum Performance Theoretically Attainable (OPTA) for the Generalized Gaussian Source with $\alpha = 2.0$ (Gaussian).

r in bits/sample	$\varepsilon = 0.005$			$\varepsilon = 0.01$			$\varepsilon = 0.05$			$\varepsilon = 0.1$		
	OPT.	LM	OPTA	OPT.	LM	OPTA	OPT.	LM	OPTA	OPT.	LM	OPTA
1	5.77	5.77	6.43	5.53	5.53	6.29	4.06	3.98	4.91	2.84	2.60	3.60
2	10.93	10.90	12.68	10.06	10.04	12.13	6.51	6.02	9.59	4.50	3.59	7.21
3	14.67	14.52	18.60	13.00	12.60	17.97	8.12	6.73	14.05	5.87	3.88	10.68
4	17.01	16.23	24.43	14.87	13.60	23.57	9.74	6.92	18.54	7.03	3.95	13.95

Table 4

SNR (in dB) of the Optimum System (OPT.) and Comparisons with the Lloyd-Max Quantizer with NBC (LM), and the Optimum Performance Theoretically Attainable (OPTA) for the Generalized Gaussian Source with $\alpha = \infty$ (Uniform).

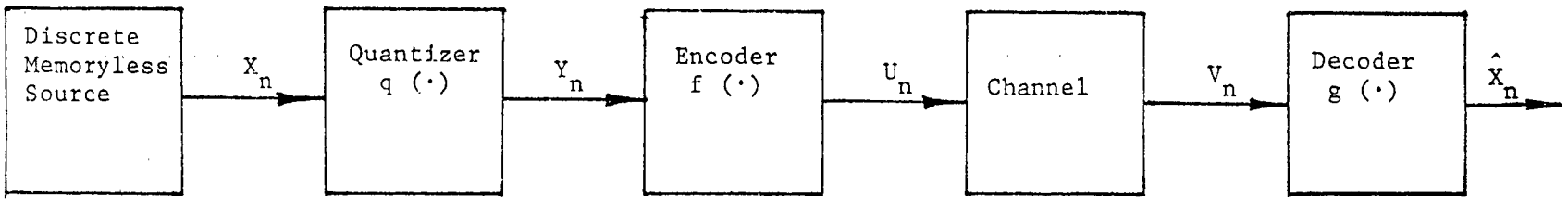

Figure 1

The Block Diagram of an Encoding/Decoding System

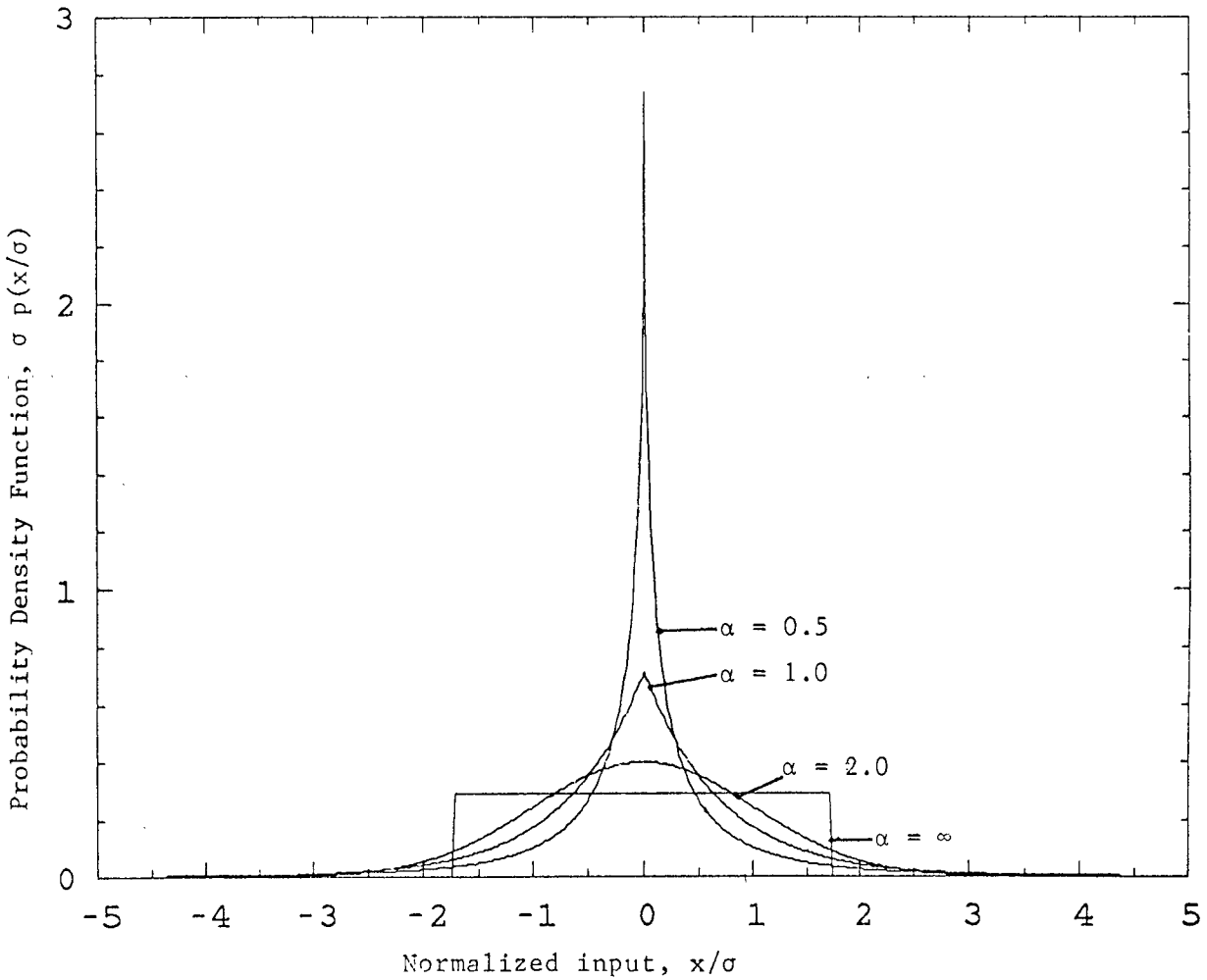

Fig. 2. Probability density function for the generalized Gaussian distribution with selected parameter values.

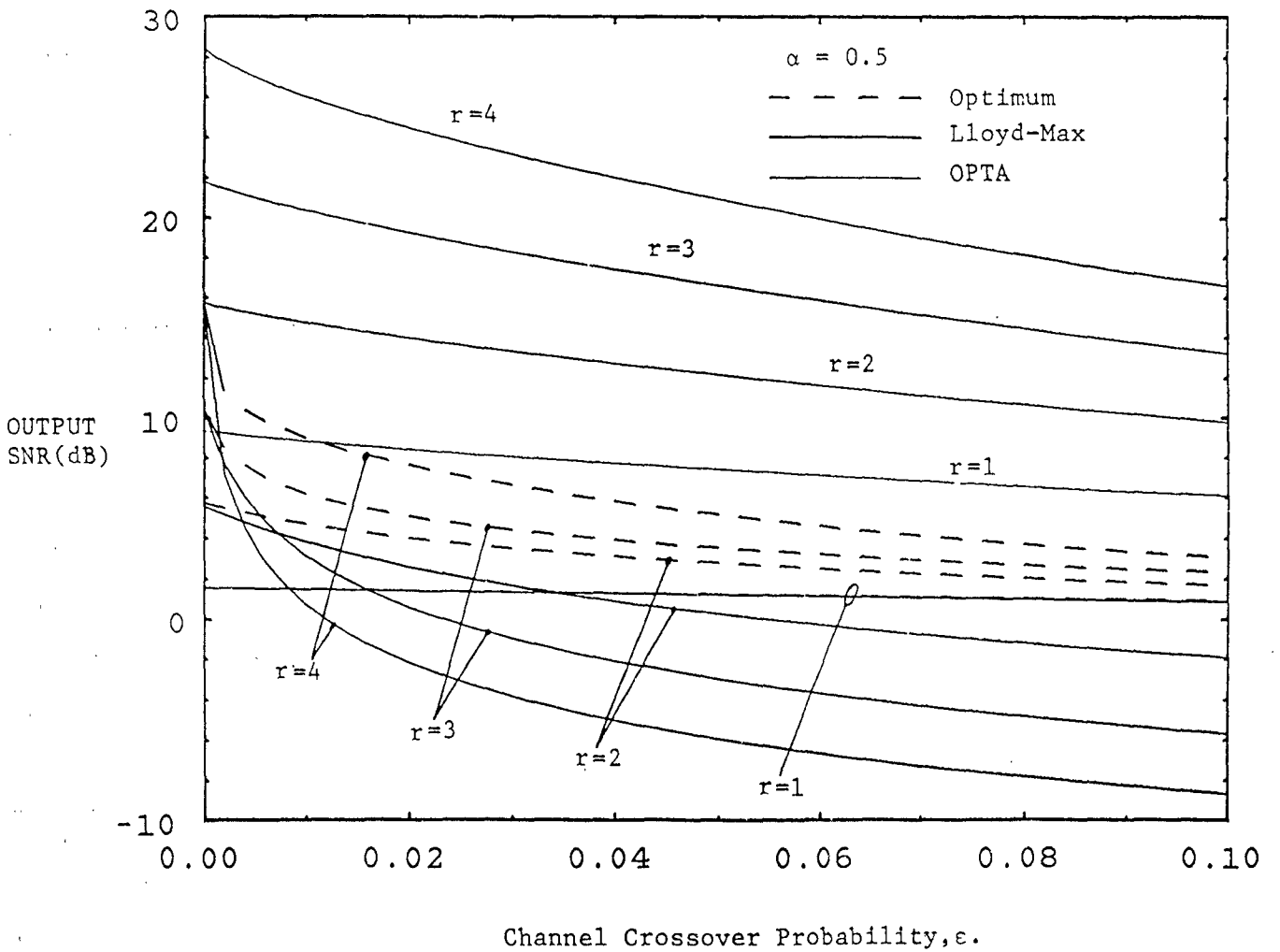

Fig. 3. Performance of the Optimum Quantizer, Lloyd-Max Quantizer and the OPTA; $\alpha = 0.5$.

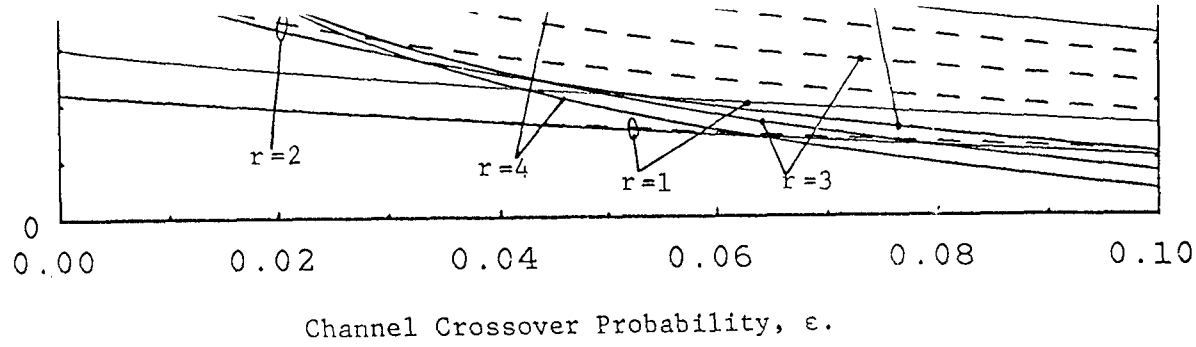

Fig. 5. Performance of the Optimum Quantizer, Lloyd-Max Quantizer and the OPTA; $\alpha = 2.0$.

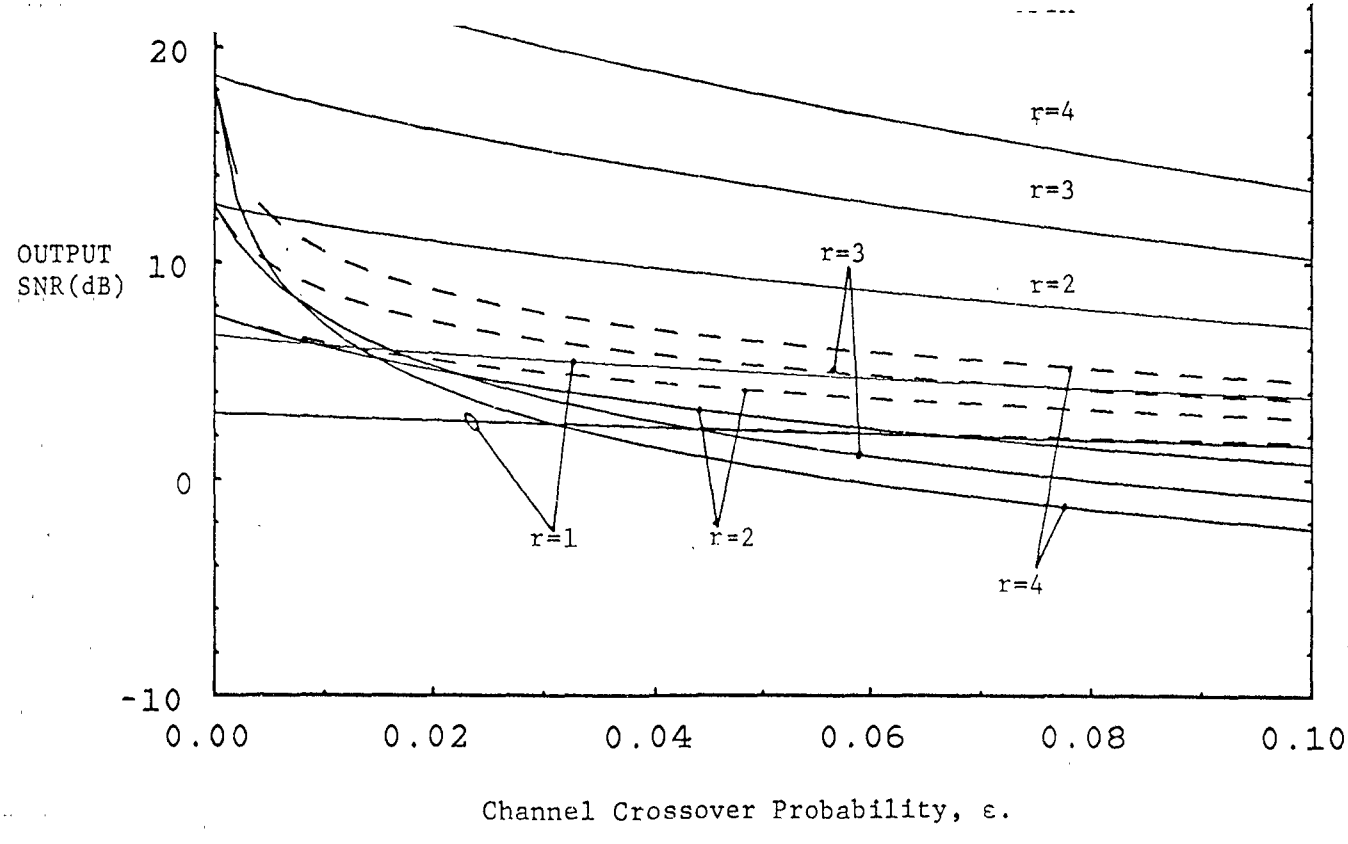

Fig. 4. Performance of the Optimum Quantizer, Lloyd-Max Quantizer and the OPTA; $\alpha = 1.0$.

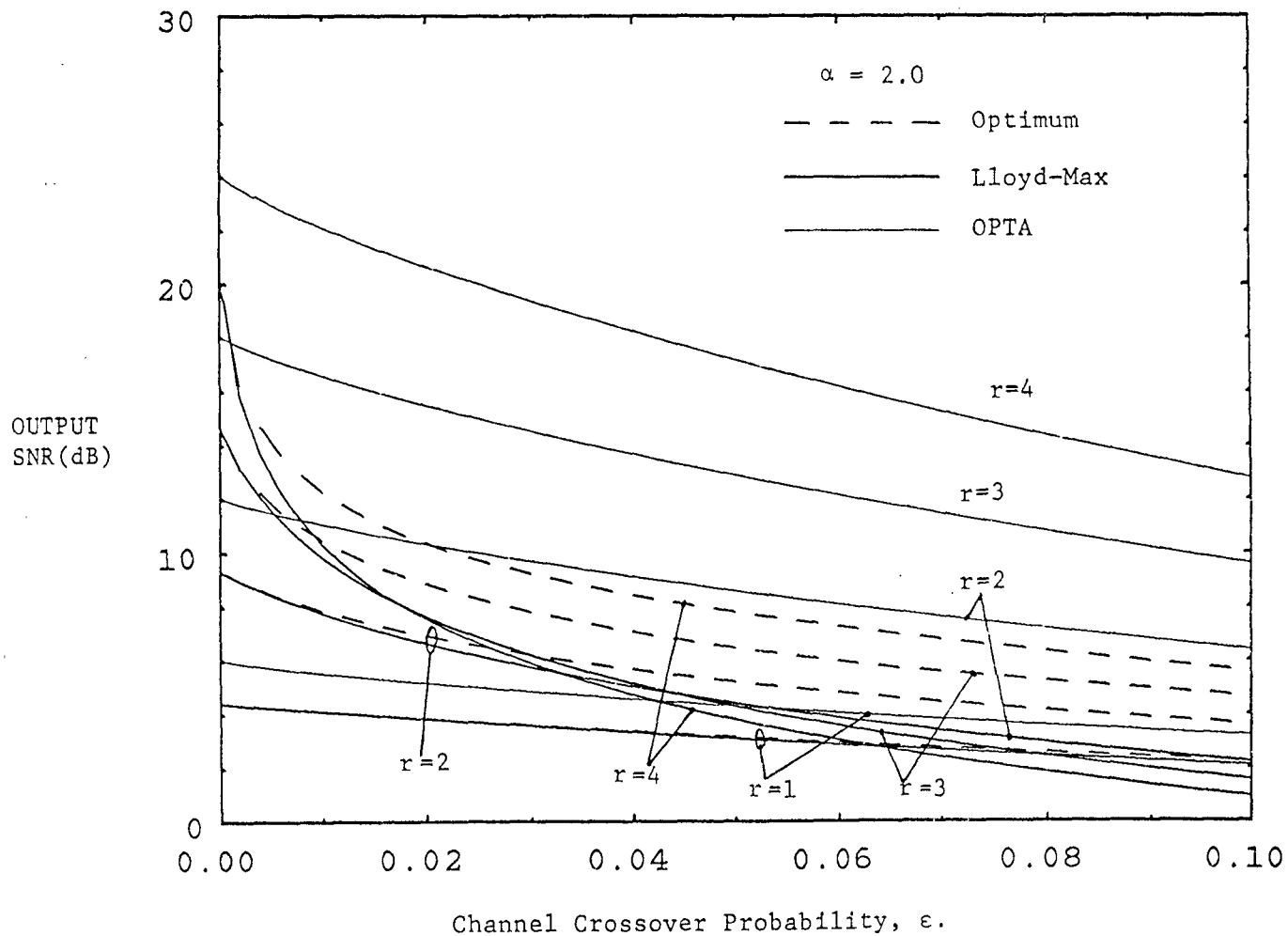

Fig. 5. Performance of the Optimum Quantizer, Lloyd-Max Quantizer and the OPTA; $\alpha = 2.0$.

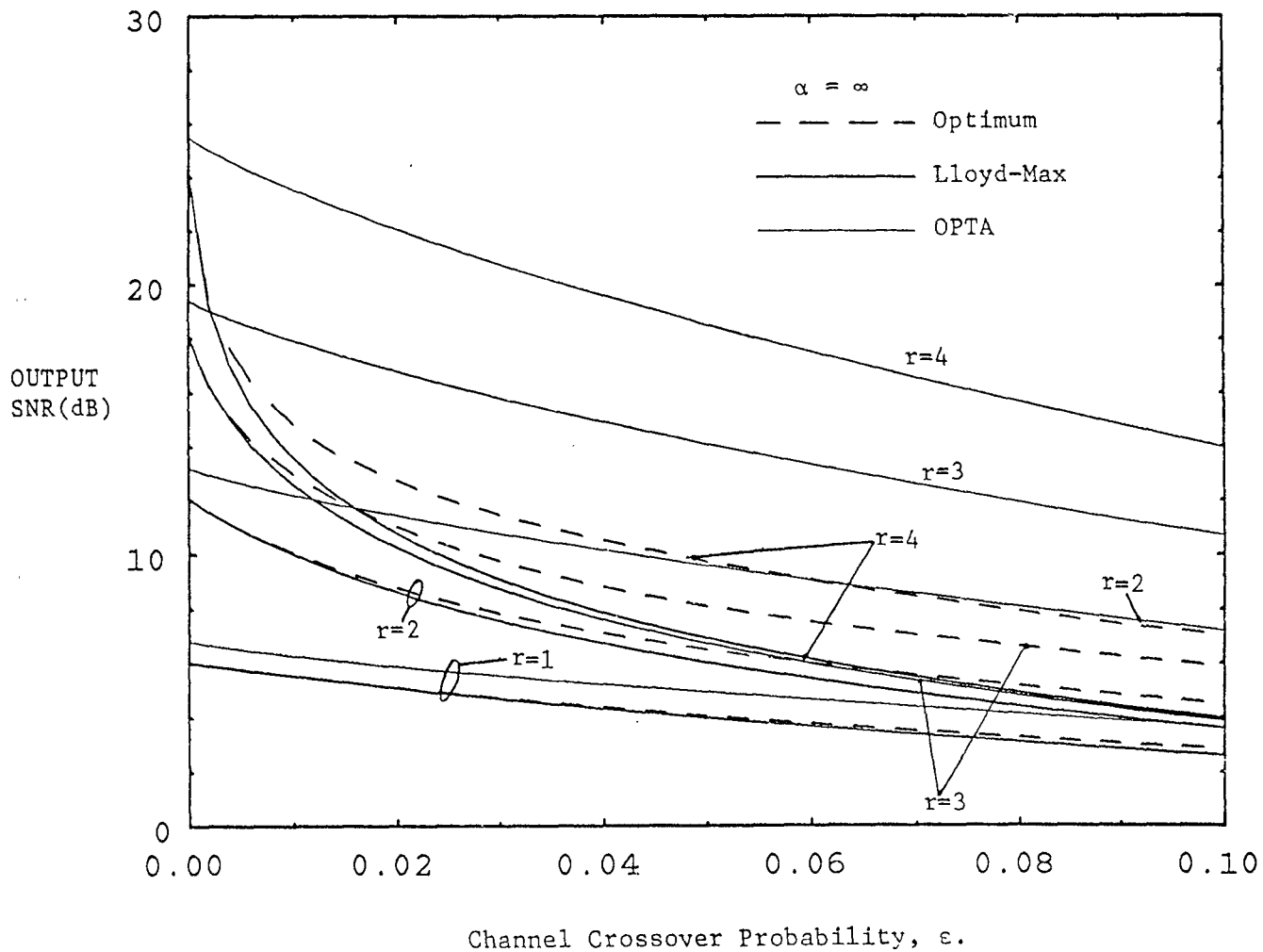

Fig. 6. Performance of the Optimum Quantizer, Lloyd-Max Quantizer, and the OPTA; $\alpha = \infty$.

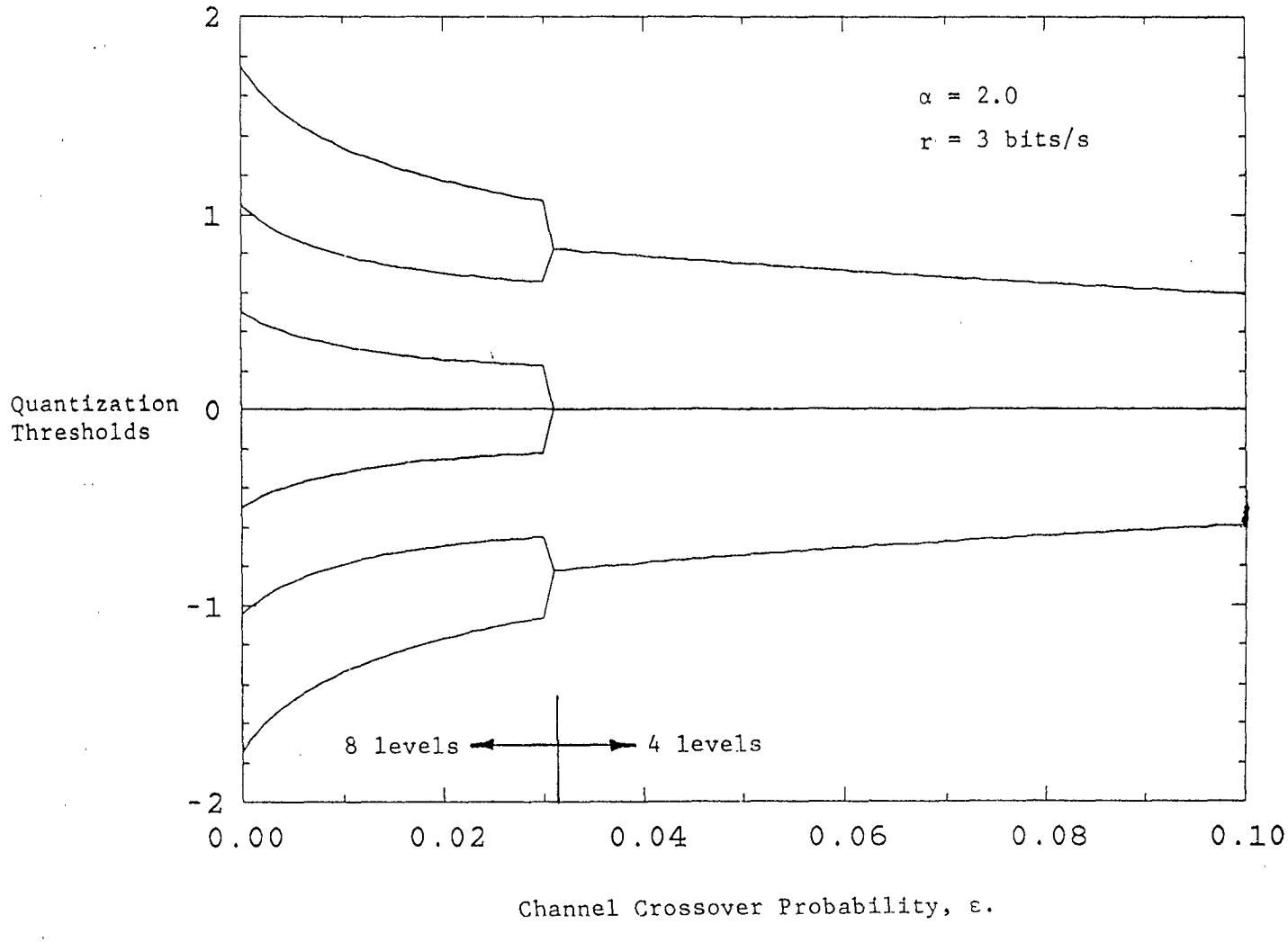

Fig. 7. Optimal quantization thresholds as a function of channel crossover probability; $\alpha = 2.0$; $r = 3$ bits/s.