Asymptotic Bounds on Optimal Noisy Channel Quantization Via Random Coding

Kenneth Zeger and Vic Manzella

Abstract—**Asymptotically optimal zero-delay vector quantization in the presence of channel noise is studied using random coding techniques. First, an upper bound is derived for the average rth-power distortion of channel optimized k-dimensional vector quantization at transmission rate R on a binary symmetric channel with bit error probability ϵ. The upper bound asymptotically equals $2^{-rRg(\epsilon,k,r)}$, where $k/(k+r)[1-\log_2(1+2\sqrt{\epsilon(1-\epsilon)})] \leq g(\epsilon,k,r) \leq 1$ for all $\epsilon \geq 0$, $\lim_{\epsilon \to 0} g(\epsilon,k,r)=1$, and $\lim_{k \to \infty} g(\epsilon,k,r)=1$. Numerical computations of $g(\epsilon,k,r)$ are also given. This result is analogous to Zador's asymptotic distortion rate of 2^{-rR} for quantization on noiseless channels. Next, using a random coding argument on nonredundant index assignments, a useful upper bound is derived in terms of point density functions, on the minimum mean squared error of high resolution, regular, vector quantizers in the presence of channel noise. The formula provides an accurate approximation to the distortion of a noisy channel quantizer whose codebook is arbitrarily ordered. Finally, it is shown that the minimum mean squared distortion of a regular, noisy channel VQ with a randomized nonredundant index assignment, is, in probability, asymptotically bounded away from zero.**

Index Terms—**Asymptotic vector quantization, noisy channel, joint source-channel coding, index assignment.**

I. Introduction

INTEREST in combined source/channel coding for bandlimited radio channels has motivated research toward quantifying the effects of channel noise on quantization systems. Dunham and Gray [1] and Kumazawa *et al.* [2] derived necessary conditions for optimal vector quantization in the presence of discrete memoryless channel noise. Their results show that optimal quantizers have encoders and decoders that satisfy generalizations of Lloyd's well-known nearest neighbor and centroid conditions. In [3], [4] algorithms were introduced for finding locally optimal codevector index assignments (or labelings), so as to minimize the average distortion resulting from a particular assignment. These papers showed the importance of choosing a good index assignment in terms of the overall signal-to-noise ratio. Using the assumption of a "greedy" index assignment, some numerical high-resolution bounds for noisy channel vector quantization were presented in [5] in terms of integrals of point density functions.

Consider a vector quantizer source coding system with a channel coder that transmits data across a binary symmetric channel. Let R_s denote the *source rate*, i.e., the number of bits per input vector component used for vector quantization. Let R denote the *channel usage* rate, i.e., the number of binary channel uses per input vector component. Let $R_c = R_s/R$ be the *channel code rate*, i.e., the fraction of transmitted bits that are used as information bits for source coding. The vector quantizer has a codebook of size 2^{kR_s} and the overall transmission rate of the system is R, where in general $R_s \leq R$. The difference $R - R_s$ is the number of redundancy bits per vector component used for error correction coding. If a redundancy free channel code is used then $R_s = R$, such as when the channel is noiseless.

Zador [6], [8] showed that, assuming a noiseless channel, the asymptotic mean rth-power distortion, $E\|X - Q(X)\|^r$, of an optimal rate R, k-dimensional vector quantizer decays as 2^{-rR} (our Lemma 1). No such explicit formula, however, has yet been displayed for quantizers in the presence of channel noise.

Asymptotic vector quantization theory assumes the rate of data transmission R grows without bound. This assumption implies that as R grows, there exists an ever increasing channel bandwidth available for transmission. Equivalently, one can assume that kR binary channel uses are made for each k-dimensional input vector. Throughout this paper it will be assumed that all data is transmitted across a binary symmetric channel.

In a quantization system that allows delay, the output binary data from a quantized sequence of input vectors can be blocked together and sent over the channel in the form of long channel codewords. Shannon's channel coding theorem guarantees that up to CkR bits per source vector can be reliably conveyed in this way, hence achieving distortions down to $D(CR)$, where $D(\cdot)$ is the distortion-rate function of the source and $C \in [0, 1]$ is the channel capacity. This is one typical conclusion about separating the source and channel coder components in a communication system.

Manuscript received June 9, 1992; revised September 12, 1994. This work was supported in part by the National Science Foundation under Grants NCR-90-09766 and NCR-91-57770, and by the Hewlett-Packard Co. This paper was presented in part at the International Symposium on Information Theory and Its Applications, Ibusuki, Japan, Dec. 1991.

K. Zeger is with the Coordinated Science Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, IL 61801.

V. Manzella is with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI 02881.

IEEE Log Number 9406320.

However, the same conclusion does not follow for source coding systems that do not allow delay; that is, systems that require that the kR bits corresponding to an input vector be transmitted before the next input vector is encoded. We call these *zero delay vector quantizers* and assume throughout this paper that all quantizers mentioned are zero delay quantizers. Without arbitrarily long blocking of input symbols there exists a nonzero probability of incorrectly decoding a channel codeword. This in turn induces an extra component of "analog" distortion between the source vector and the final reproduction vector.

For a given vector dimension and transmission rate, the only "blocking" advantage that can be exploited for source or channel coding in a zero delay system is that of the kR bits simultaneously transmitted per source vector. On one hand, Shannon's channel coding theorem guarantees that one can achieve a bit error probability at least as small as $2^{-(kR)E_{\max}(R_c)}$, where $E_{\max}$ is the maximum of the error exponent function E_r and the expurgated error exponent function E_{ex} [7]. However, minimizing the end to end analog distortion does not necessarily imply one should strive to reliably convey the maximal amount of binary data across the channel. In fact, it might be desirable to tolerate some amount of bit errors in order to increase the quantization resolution.

It can be shown that the minimal average rth-power distortion of nonzero delay vector quantizers on a noisy channel decays to zero as the number of transmitted bits per sample grows, provided one is willing to block together multiple input samples before transmission and thus incur delay. It has been an open problem, however, to find the rate of decay of the minimum distortion for zero delay quantizers. Part of the difficulty in determining this lies in the complexity of mathematically analyzing the index assignment problem, since for an N-level quantizer there are $N!$ possible index assignments.

In this paper we present several results to help answer these questions for zero delay quantizers. First (Theorem 1), we show that for high resolution vector quantization on a binary symmetric channel (BSC) with bit error probability ϵ, the minimum mean square-error (MSE)-decays to zero at least exponentially fast in R.

An MSE upper bound is given that approaches Zador's 2^{-rR} optimal decay rate (Zador's rate holds for $\epsilon = 0$ and for all k) either in the limit as $\epsilon \to 0$ or in the limit as $k \to \infty$. In general, the decay rate of the bound can exactly be computed by simple numerical means. Also, an analytic upper bound to the decay rate is derived which is accurate over a certain range of values of r and k. As the vector dimension k grows, the decay rate of the noisy channel bound approaches that of the noiseless channel bound and is thus asymptotically (in vector dimension) tight. This result however, implicitly assumes that a block channel coder follows an ordinary vector quantizer; however, this channel coder is obtained using the existence proof in Shannon's channel coding theorem. Even if an explicit cascade of a quantizer and channel coder were not used, a channel coder would in essence be implicitly embedded in a quantizer of higher rate. For practical coders, this is not generally feasible due to complexity; in such cases the result provides a theoretically achievable quantizer performance level.

On the other hand, one can consider quantization systems for noisy channels that have *nonredundant* channel coders, that is, channel coders that merely determine the assignment of binary channel words to codevectors and do not add redundancy bits (i.e., $R_s = R$). This is the focus of Section III. To derive a useful description of optimal noisy channel quantization, we introduce a random coding technique to analyze the MSE of noisy channel vector quantizers (VQ's), by averaging the MSE of a given VQ over all possible nonredundant index assignments. In particular, attention is focused on regular VQ's [a quantizer is *regular* if each encoding cell R_i is convex and contains the codevector $y_i = Q(R_i)$]. Quantizers with convex encoding cells that satisfy the centroid condition are regular [8], as are lattice quantizers. An *expected* MSE is derived (Theorem 2), giving an asymptotic upper bound on the MSE of any VQ having a *better than average* index assignment. It thus provides a mathematical tool analogous to Zador's formula for analytically describing the MSE. Corollary 1 gives a useful formula for the minimum MSE of a vector quantizer that transmits across a noisy channel with randomized index assignments and provides an upper bound on the MSE over all possible index assignments. For certain ranges of transmission rates this expected MSE is shown (Corollary 3) to have a unique minimum as a function of the transmission rate, an interesting "diminishing returns" result. That is, for zero delay regular VQ with a randomized index assignment, there is no benefit to increasing the transmission rate beyond a certain point.

It is also shown (Theorem 3) that, under a regularity constraint, the average (over randomized nonredundant index assignments) mean squared quantization distortion with a BSC is, in probability, asymptotically bounded away from zero. That is, the MSE of optimal zero-delay quantization using random codeword labeling does not typically decay to zero. This contrasts with Theorem 1 where the MSE decays to zero. An explanation for this lies in the fact that Theorem 1 imposes a regularity constraint (or equivalently only allows nonredundant channel coding) on the quantizers in order to reduce computational complexity.

Section II establishes the asymptotic MSE upper bound for VQ distortion on a binary symmetric channel without any regularity assumption. Section III introduces the technique of computing mean squared quantization error for a noisy channel by randomizing over all possible nonredundant index assignments, with a regularity constraint. Here, an MSE formula is given in terms of point density functions. Appendix A contains the statements and proofs of some technical lemmas used in the paper. Lemma 2 is particularly useful and is used to prove most of the other lemmas. Appendix A is dedicated to proving Proposition 2, which requires examination of many separate cases.

II. Upper Bound on Asymptotic Channel Optimized Vector Quantizer Distortion

A rate R, k-dimensional *vector quantizer* is a mapping $Q: \mathscr{R}^k \to \{y_0, \cdots, y_{2^{kR}-1}\} \subset \mathscr{R}^k$ where $y_0, \cdots, y_{2^{kR}-1}$ are called *codevectors*. Q is also said to be a 2^{kR}-*level* vector quantizer. The *encoder* and *decoder* of Q are, respectively, functions $Q_E: \mathscr{R}^k \to \{0,1\}^{kR}$ and $Q_D: \{0,1\}^{kR} \to \{y_0, \cdots, y_{2^{kR}-1}\}$ such that $Q_D(i) = y_i$ for all i and $Q = Q_D \circ Q_E$. For each i, $R_i \stackrel{\text{def}}{=} Q_E^{-1}\{i\}$ is the *cell* associated with y_i. Here the notation for the binary string $i \in \{0,1\}^{kR}$ is used to represent its numerical value.

A *noisy channel vector quantizer*, of rate R with respect to a binary channel having conditional probabilities $P_{j|i} = Pr[j \text{ received}|i \text{ sent}]$ is the composition $\hat{Q} = Q_D \circ \eta \circ Q_E$ where the random permutation $\eta: \{0,1\}^{kR} \to \{0,1\}^{kR}$ is defined by $Pr[\eta(i) = j] = P_{j|i}$. Here, Rk uses of a binary channel are made. We will frequently refer to a noisy channel vector quantizer as a vector quantizer which transmits across a noisy channel.

It is well known that with a noiseless channel, an optimal quantizer satisfies a *nearest neighbor condition* and a *centroid condition*. These two conditions, respectively, assert that every cell contain all the points closer to its codevector than any other codevector and that every codevector is the conditional mean of its cell. For noisy channel quantizers, there are also well-known generalizations of these conditions.

Zador's formula provides a useful rule of thumb of "$3r$ dB/bit" increase in SNR for each bit added to a high resolution scalar quantizer using an rth power distortion measure. This is reasonably accurate for many low-resolution cases as well. Below, it is shown that on a noisy channel an optimal quantizer's average rth power distortion decreases asymptotically at least as fast as "$3rg(\epsilon, k, r)$ dB/bit," where $g(\epsilon, k, r) \in (0,1\,]$ for all $\epsilon \geq 0$. To the best of our knowledge, this is the only known upper bound on the average distortion of a quantizer with a noisy channel. This result does not assume that the centroid condition is necessarily satisfied.

Lemma 1: (Zador [6]) Let $X \in \mathscr{R}^k$ be a random vector having a density f. The minimum average rth-power distortion of a rate R vector quantizer is asymptotically equal to

$$b_{r,k} \|f\|_{k/(k+r)} 2^{-rR} \tag{1}$$

where $\|f\|_p \stackrel{\text{def}}{=} (\int_{\mathscr{R}^k} |f|^p)^{1/p}$ and $b_{r,k}$ is a constant independent of f and R.

Theorem 1: Let $X \in \mathscr{R}^k$ be a random vector having a density f with compact support. The minimum average rth-power distortion of a rate R noisy channel vector quantizer on a binary symmetric channel with crossover probability ϵ is asymptotically bounded above by $2^{-rRg(\epsilon,k,r)}$, where $\lim_{\epsilon \to 0} g(\epsilon, k, r) = 1$ and $k/(k+r)[1 - \log_2(1 + 2\sqrt{\epsilon(1-\epsilon)})] \leq g(\epsilon, k, r) \leq 1$.

Proof: Many of the facts and terminologies used in this proof can be found in [7]. It suffices to exhibit any noisy channel quantizer that satisfies the bound for the given f. Assume for convenience that R and R_s are positive integers, and consider a rate R, k-dimensional noisy channel vector quantizer $\hat{Q}$ that is composed of an optimal rate R_s noiseless channel quantizer Q and a (kR, kR_s) channel coder $\psi: \{0,1\}^{kR_s} \to \{0,1\}^{kR}$ with channel decoder $\phi: \{0,1\}^{kR} \to \{0,1\}^{kR_s}$. That is, if the encoder of Q is $Q_E: \mathscr{R}^k \to \{0,1\}^{kR_s}$ and the decoder of Q is $Q_D: \{0,1\}^{kR_s} \to \mathscr{R}^k$, and if we define a new vector quantizer of rate $R \geq R_s$ with encoder $\hat{Q}_E = \psi \circ Q_E$ and decoder $\hat{Q}_D = Q_D \circ \phi$ then the noisy channel vector quantizer is given by $\hat{Q} = \hat{Q}_E \circ \eta \circ \hat{Q}_D$, where η is the random permutation $\eta: \{0,1\}^{kR} \to \{0,1\}^{kR}$ defined by $Pr[\eta(i) = j] = P_{j|i}$.

The capacity of a BSC with crossover probability ϵ is $C = 1 - H(\epsilon)$ in bits per channel use, where the *binary entropy function* $H(x)$ is defined as

$$H(x) = -x \log_2 x - (1-x) \log_2 (1-x). \tag{2}$$

If $R_c < C$, then Shannon's channel coding theorem guarantees the existence of a function ψ such that kR_s information bits can be reliably transmitted for every block of kR total bits sent, as R becomes asymptotically large. More precisely, kR_s bits can be conveyed with a probability of error

$$P_e \leq 2^{-(kR)E_{\max}(R_c)} \tag{3}$$

where $E_{\max}(R_c) = \max\{E_r(R_c), E_{\text{ex}}(R_c)\}$, E_r is the *error exponent* function, and E_{ex} is the *expurgated error component* function. Let us define the linear function

$$E_{\text{lin}}(R_c) \stackrel{\text{def}}{=} 1 - \log_2\left(1 + 2\sqrt{\epsilon(1-\epsilon)}\right) - R_c. \tag{4}$$

On a BSC, the function $E_r(R_c)$ is known to be positive and convex for all channel code rates less than capacity C, and equal to $E_{\text{lin}}(R_c)$ in the range $0 < R_c \leq R_2$, where

$$R_2 = 1 - H\left(\frac{\sqrt{\epsilon}}{\sqrt{\epsilon} + \sqrt{1-\epsilon}}\right). \tag{5}$$

The convexity of E_r implies that $E_r(R_c) \geq E_{\text{lin}}(R_c)$ for all $R \in [0, C]$. E_r can be computed numerically in the nonlinear region using

$$E_r(R_c) = \max_{0 \leq \rho \leq 1} \Big[\rho(1 - R_c) - (\rho + 1) \cdot \log_2 [\epsilon^{1/(1+\rho)} + (1-\epsilon)^{1/(1+\rho)}] \Big] \quad R_2 \leq R_c \leq C. \tag{6}$$

For R_c greater than R_2, the error exponent $E_r(R_c)$ decreases slower than linearly until it reaches zero at $R_c = C$.

The function $E_{\text{ex}}(R_c)$ is strictly greater than $E_r(R_c)$ for all R_c in the range $0 < R_c < R_1 < R_2$ where

$$R_1 = 1 - \log_2\left(1 + 2\sqrt{\epsilon(1-\epsilon)}\right) + \left(\frac{2\sqrt{\epsilon(1-\epsilon)}}{1 + 2\sqrt{\epsilon(1-\epsilon)}}\right) \log_2\left(2\sqrt{\epsilon(1-\epsilon)}\right) \tag{7}$$

and $E_{ex}(R_c) = E_{lin}(R_c)$ in the range $R_1 < R_c < C$. Hence $E_{ex}(R_c) \geq E_{lin}(R_c)$ for all $R_c \in [0, C]$. The function E_{ex} can be computed numerically in the nonlinear region using (for $0 < R_c \leq R_1$).

$$E_{ex}(R_c) = \sup_{\rho \geq 1} \cdot \left\{ \rho \left[(1 - R_c) - \log_2 \left[1 + \left(2\sqrt{\epsilon(1-\epsilon)} \right)^{1/\rho} \right] \right] \right\}. \quad (8)$$

The definition of E_r and its properties can be found in [7] and the formula for E_{ex} on a BSC in (8) follows from a result in [15]. In summary, we have

$$E_{max}(R_c) = \begin{cases} E_{ex}(R_c) & 0 < R_c < R_1 \\ E_{lin}(R_c) & R_1 \leq R_c \leq R_2 \\ E_r(R_c) & R_2 < R_c < C. \end{cases} \quad (9)$$

Let I and J, respectively, be the indices transmitted and received across a BSC by a vector quantizer. The mean rth-power vector error of such a noisy channel vector quantizer, averaged over both the source and channel statistics, is given by

$$D = E\|X - y_J\|^r. \quad (10)$$

By conditioning the expectation over the events that either a channel error does or does not occur, the average distortion can be bounded as

$$\begin{aligned} D &= E[\|X - y_J\|^r | J = I](1 - P_e) + E[\|X - y_J\|^r | J \neq I] P_e \\ &\leq E\|X - y_I\|^r + E[\|X - y_J\|^r | J \neq I] P_e \\ &\leq G_1 2^{-rRR_c} + G_2 2^{-kRE_{max}(R_c)} \\ &\stackrel{\text{def}}{=} D_u(R_c) \; \forall R_c \in (0, C) \end{aligned} \quad (11)$$

where

$$G_1 \stackrel{\text{def}}{=} b_{r,k} \|f\|_{k/(k+r)}$$
$$G_2 \stackrel{\text{def}}{=} \text{diam}\,[\text{supp}\,(f)]^r \quad (12)$$

and diam[supp(f)] is the diameter of the support of f. The last inequality follows from Lemma 1 (large R) and Shannon's channel coding theorem.

The overall transmission rate R of the system is fixed and we wish to minimize the bound D_u over all source rates R_s, or equivalently over channel code rates R_c. For convenience we use the notation $D_u(R_c)$ to indicate the dependence on R_c. The choice of R_s trades off between source and channel coding.

If one transmits at a rate R_c very close to capacity C, then the number of information bits R_s will be large, and thus the quantization error $E\|X - y_I\|^r$ will be small; however, for large R_c, the probability of an uncorrected channel error cannot be as tightly upper bounded (smaller error exponent), so that the term P_e will contribute more to the overall distortion D. Thus, there is an important trade-off in this case between: i) designating more of the transmitted bits as information bits to reduce quantization error, and ii) devoting more of the transmitted bits toward error control coding to drive the probability of an uncorrected channel error to zero faster.

For each R, one could set the derivative $D_u'(R_c)$ equal to zero to find the rate R_c that optimizes this trade-off in the bound $D_u(R_c)$. To prove the theorem, however, it will turn out to be sufficient to merely choose the fixed rate R_c^* (not varying with R) that forces each of the exponentials in (11) to decay at exactly the same rate. That is, we choose R_c^* to satisfy the equation

$$E_{max}(R_c^*) = (r/k) R_c^*. \quad (13)$$

Exactly one such point exists in the range $(0, C)$, since the left-hand side of (13) is a monotonic decreasing function and the right-hand side is a linearly increasing function.

Equation (13) can be solved exactly by numerical means (see remarks following this proof). If the desired ϵ, r, and k are such that $R_1 \leq R_c^* \leq R_2$, then (13) can be solved analytically. More generally, we can use the linear lower bound on $E_{max}(R_c)$ to obtain

$$\begin{aligned} E_{max}(R_c^*) &\geq E_{lin}(R_c^*) \\ &= 1 - \log_2\left(1 + 2\sqrt{\epsilon(1-\epsilon)}\right) - R_c^*. \end{aligned} \quad (14)$$

Equation (13) and inequality (14) together imply

$$R_c^* \geq \frac{k}{k+r}\left[1 - \log_2\left(1 + 2\sqrt{\epsilon(1-\epsilon)}\right)\right] \quad (15)$$

and hence

$$\begin{aligned} D_u(R_c^*) &= (G_1 + G_2) 2^{-rRR_c^*} \\ &\stackrel{\text{def}}{=} (G_1 + G_2) 2^{-rRg(\epsilon, k, r)} \end{aligned} \quad (16)$$

where we have defined $g(\epsilon, k, r) = R_c^*$, thus

$$g(\epsilon, k, r) \geq \left(\frac{k}{k+r}\right)\left[1 - \log_2\left(1 + 2\sqrt{\epsilon(1-\epsilon)}\right)\right]. \quad (17)$$

This completes the proof since $D \leq D_u(R_c^*)$.

It can be shown that on a BSC, $R_1 \to C$ and $R_2 \to C$ as $\epsilon \to 0$, so that $E_{max}(R_c) = E_{ex}(R_c)$ for R_c arbitrarily close to C, for sufficiently small ϵ. But since $C \to 1$ as $\epsilon \to 0$, we have that $E_{max}(R_c) = E_{ex}(R_c)$ for R_c arbitrarily close to 1 as $\epsilon \to 0$. Also, for any fixed $R_c \in (0, C)$, it is easy to show from the definition of $E_{ex}(R_c)$ that $\lim_{\epsilon \to 0} E_{ex}(R_c) = \infty$. Thus for any $R_c \in (0, C)$, we have $\lim_{\epsilon \to 0} E_{max}(R_c) = \infty$. This in turn implies that for a fixed k and r, the solution, R_c^*, to (13) approaches $R_c^* = 1$ as $\epsilon \to 0$, and thus $\lim_{\epsilon \to 0} g(\epsilon, k, r) = 1$, agreeing with Zador's formula. □

Remarks: The decay rate of our upper bound $2^{-rRg(\epsilon,k,r)}$ approaches the optimal decay rate of 2^{-rR} as either $\epsilon \to 0$ or $k \to \infty$. The analytic lower bound on $g(\epsilon, k, r)$ in (17) also approaches 1 as $k \to \infty$ but unfortunately approaches only $k/(k+r)$ as $\epsilon \to 0$. Thus, from (16) it can be seen that to achieve the distortion upper bound D_u for small ϵ [using the analytic bound on $g(\epsilon, k, r)$], it suffices to convey information across the channel at an approximate rate of $k/(k+r)$ bits per channel. For a scalar quantizer on a BSC with small ϵ, the MSE in this case drops at least as fast as about $2^{-2R/3}$, which is the same rate of decay as an optimal scalar quantizer on a noiseless channel, but using only one-third as many bits. The remaining two-thirds of the bits are for channel error protection.

While for small ϵ the analytic lower bound in (17) does not agree with the optimal decay rate (i.e., does not approach 1), it is still useful for larger ϵ and it will be shown numerically that is differs very little from $g(\epsilon, k, r)$ in some practical cases. In Fig. 1 a comparison is made between $g(\epsilon, k, r)$ and the lower bound in (17) for MSE distortion (i.e., $r = 2$) and several channel bit error probabilities. It can be seen that they are reasonably close to each other over a wide range of ϵ. The curves for $g(\epsilon, k, r)$ give the best known decay rate of asymptotic noisy channel vector quantization.

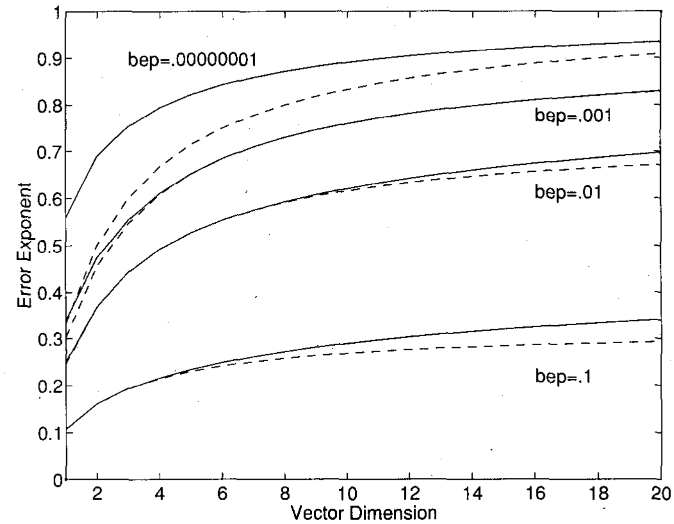

Fig. 1. Plot of the function $g(\epsilon, k, r)$ for $r = 2$ and $\epsilon = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-8}$. The solid lines are the function $g(\epsilon, k, r)$ and the dashed lines immediately below them are the corresponding upper bounds from Eq. (17). For $\epsilon = 10^{-1}$: $E_{\max} = E_{\text{lin}}$ for $k \leq 2$ and $E_{\max} = E_r$ for $k \geq 3$. For $\epsilon = 10^{-2}$: $E_{\max} = E_{\exp}$ for $k = 1$ $E_{\max} = E_{\text{lin}}$ for $2 \leq k \leq 6$ and $E_{\max} = E_r$ for $k \geq 6$. For $\epsilon = 10^{-3}$: $E_{\max} = E_{\exp}$ for $k \leq 5$ $E_{\max} = E_{\text{lin}}$ for $6 \leq k \leq 14$ and $E_{\max} = E_r$ for $k \geq 15$. For $\epsilon = 10^{-8}$: $E_{\max} = E_{\exp}$ for $1 \leq k \leq 20$.

III. Random Coding Theorems for Regular Quantizers on Noisy Channels

In this section, the asymptotic mean squared performance of k-dimensional VQ in the presence of channel noise and under a regularity constraint is examined for *typical* index assignments by modeling the selection of nonredundant index assignments as a random procedure. For an N-level quantizer Q_N define a *randomized index assignment* to be a codevector labeling chosen randomly and uniformly, and independent of the source, from the set S_N of all $N!$ permutations of N elements. It may be argued that this accurately models an arbitrary ordering of the codevectors in a codebook. Arbitrarily ordering the codevectors is inferior to various known index assignment algorithms and is clearly not recommended. However, in practice, some implementations do in fact neglect to choose good index assignments and instead settle for whatever codebook ordering happens to result from a quantizer design algorithm.

While the above argument may provide some mild motivation for studying random index assignments, there is, in fact, a much more significant purpose. Namely, the average distortion using a randomized index assignment gives us an *analytic upper bound* on the performance of the best possible index assignment (and also a lower bound on the worst possible index assignment performance). Until now, there has been no concise formula for predicting or closely bounding noisy channel quantizer distortion. In practice, the performance of randomized index assignments is sometimes close to the best index assignments and sometimes not too close. However, even when the best index assignment is significantly better that the average index assignment, their performances tend to follow the same trends as a function of the channel bit error probability. Thus, the randomized index distortion formula can still provide utility in these cases.

Assume that the source random vector X has a density f over a compact support, $S_f \subset \mathscr{R}^k$. Let y_i denote the ith codevector of Q_N, and R_i the ith partition cell. For notational simplicity we do not explicitly write the size N of the codebook as a subscript of the codevectors and partition regions. The probability of a source vector lying in the ith partition cell is

$$P_i = \int_{R_i} f(x)\, dx \tag{18}$$

and we define

$$\delta_N \stackrel{\text{def}}{=} \max_{i=1,\cdots,N} \operatorname{diam}(R_i \cap S_f) \tag{19}$$

which will serve as an upper bound on intracell Euclidean distances. We will say that a sequence of quantizers *has diminishing cell diameters* if $\lim_{N\to\infty} \delta_N = 0$.

Note that with a noiseless channel any sequence of optimal quantizers Q_N will have diminishing cell diameters if the source density is positive on the support region. To see this, first enclose the support region S_f inside a cube and divide each side of the cube into say M equal length pieces. This partitions the cube into M^k smaller cubes. Now consider those of the cubes that intersect S_f with positive probability. For sufficiently large N each of these cubes will contain at least one codevector, for otherwise the MSE would be bounded away from zero. Hence the diameter of each quantizer cell is upper bounded, for instance, by four times the diagonal length of these cubes, which can be made arbitrarily small by choosing M large. This is a straightforward result, but relies on the compact support and positive density assumptions.

Denote by C the random mapping of input indices to output indexes across the channel, and let $\pi \in S_N$ be a permutation of codevector indexes. With this model let I be the source encoder output, a random variable depend-

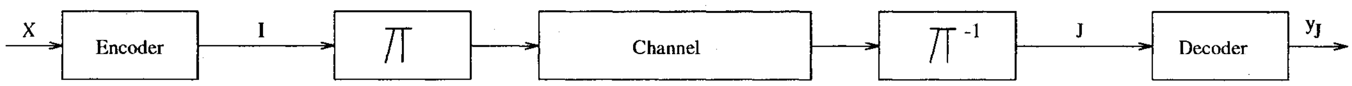

Fig. 2. Block diagram of noisy channel vector quantization system, where π is a nonredundant index assignment function.

ing only on X, and let J be the decoder input, a random variable depending on X and C, as well as the index assignment. A block diagram is shown in Fig. 2. The total distortion $D = D(X, \pi, C) = \|X - y_J\|^2$ is also a random variable, and the quantity of interest is

$$\min_{\pi \in S_N} E_{X,C}[D]. \tag{20}$$

This minimization is in general a quite difficult discrete optimization problem. We introduce in this paper an alternative approach in which the choice of index assignment π is randomized. Letting Π denote an index permutation selected randomly and uniformly from S_N, we emphasize that $D = D(X, \Pi, C)$ now depends on *three* random variables, and D is then averaged over all three random variables, making $E_{X,\Pi,C}[D]$ the new quantity of interest. This average will provide an upper bound on the minimum distortion in (20), and will also be a good estimate of the average distortion for systems that choose the index assignment arbitrarily. We use the subscript notation on expectations in order to make clear which random variables are included in the expectation. If any of the random variables X, Π, or C are omitted in an expectation, this will indicate that the expectation is a conditional expectation (conditioned on the missing random variable).

To obtain the mean squared distortion $E_{X,\Pi,C}[D]$ we decompose the overall distortion into the sum

$$D = D_s + 2D_{sc} + D_c \tag{21}$$

of a source coding distortion $D_s = \|X - y_I\|^2$, a codebook distortion $D_c = \|y_I - y_J\|^2$, and a cross term $D_{sc} = (X - y_I)^t(y_I - y_J)$. D_s depends only on X so that

$$E_{X,\Pi,C}[D] = E_X[D_s] + 2E_{X,\Pi,C}[D_{sc}] + E_{X,\Pi,C}[D_c]. \tag{22}$$

Throughout this paper it will often be convenient to first compute expectations with respect to X and C and then to analyze the expectations over Π separately. This is possible without conditioning since the random index assignment is assumed to be independent of the source and channel random variables. To facilitate this and to emphasize dependence on Π we introduce the additional notation

$$\begin{aligned} D(\Pi) &\overset{\text{def}}{=} E_{X,C}[D] \\ D_c(\Pi) &\overset{\text{def}}{=} E_{X,C}[D_c] \\ D_{sc}(\Pi) &\overset{\text{def}}{=} E_{X,C}[D_{sc}]. \end{aligned} \tag{23}$$

Using this notation we have

$$\begin{aligned} D(\Pi) &= E_X[D_s] + 2D_{sc}(\Pi) + D_c(\Pi) \\ E_{X,\Pi,C}[D] &= E_\Pi[D(\Pi)] \\ &= E_X[D_s] + 2E_\Pi[D_{sc}(\Pi)] + E_\Pi[D_c(\Pi)]. \end{aligned} \tag{24}$$

The following proposition shows that for regular quantizers the expectation of the cross term decays to zero as N increases. If a quantizer is designed optimally for a noiseless channel, then the centroid condition is satisfied and the expectation of the cross term D_{sc} is identically zero for all N. One motivation for stating the following proposition in its generality is so that it applies to the class of lattice quantizers, which in general do not satisfy the centroid condition. This proposition implies that for sequences of regular quantizers with randomized index assignments, the average distortion is asymptotically given by (1), with $r = 2$, plus the term $E_{X,\Pi,C}[D_c]$. We note that the quantity D_{sc} was analyzed in [9] under certain specific assumptions.

Proposition 1: Let $X \in \mathscr{R}^k$ be a random vector having a density with compact support. Then for any sequence of N-level vector quantizers that are regular, have diminishing cell diameters, and that transmit across a binary symmetric channel with bit error probability ϵ, $\lim_{N\to\infty} E_{X,\Pi,C}[D_{sc}] = 0$.

Proof: Since each permutation π is equally likely and is selected independently of the source and channel, the cross-term expectation is

$$E_{X,\Pi,C}[D_{sc}] = \sum_\pi \frac{1}{N!} E_{X,C}[D_{sc}(X, \pi, C)]. \tag{25}$$

For each $\pi \in S_N$,

$$\begin{aligned} &E_{X,C}[D_{sc}(X, \pi, C)] \\ &\quad = \sum_{i=1}^N \sum_{j=1}^N E_X[D_{sc} | I = i,\ J = j] P[I = i,\ J = j] \\ &\quad = \sum_{i=1}^N \sum_{j=1}^N E_X\Big[(X - y_i)^t(y_i - y_j) | X \in R_i\Big] P_{\pi(j)|\pi(i)} P_i \\ &\quad = \sum_{i=1}^N \sum_{j=1}^N (c_i - y_i)^t(y_i - y_j) P_{\pi(j)|\pi(i)} P_i \end{aligned} \tag{26}$$

where $P_{\pi(j)|\pi(i)}$ is the conditional probability that $\pi(j)$ is received given that $\pi(i)$ is transmitted across the channel, and $c_i = E_X[X | X \in R_i]$ is the centroid of cell R_i. To determine the expectation with respect to the randomized permutation selection, note that

$$D_{sc}(\Pi) = \sum_{i=1}^N \sum_{j=1}^N (c_i - y_i)^t(y_i - y_j) P_{\Pi(j)|\Pi(i)} P_i \tag{27}$$

so that

$$\begin{aligned}|E_{X,\Pi,C}[D_{sc}]| &= |E_{\Pi}[D_{sc}(\Pi)]| \\ &= \frac{1}{N!}\left|\sum_{i=1}^{N}\sum_{j=1}^{N}(c_i - y_i)^t(y_i - y_j)P_i\sum_{\pi}P_{\pi(j)|\pi(i)}\right| \\ &= \left|\frac{1-(1-\epsilon)^{\log_2 N}}{N-1}\sum_{i=1}^{N}\sum_{j=1}^{N}P_i(c_i - y_i)^t(y_i - y_j)\right| \quad (28) \\ &\le \frac{1-(1-\epsilon)^{\log_2 N}}{N-1}\sum_{i=1}^{N}\sum_{j=1}^{N}P_i\|c_i - y_i\|\cdot\|y_i - y_j\| \quad (29) \\ &\le [1-(1-\epsilon)^{\log_2 N}]\frac{N}{N-1}\delta_N \operatorname{diam}(S_f) \quad (30) \\ &\to 0 \text{ as } N \to \infty, \text{ since } \delta_N \to 0. \quad (31)\end{aligned}$$

Equation (28) follows from Lemma 3 (see Appendix A), and (29) follows from the Cauchy–Schwarz inequality. The regularity of the quantizer implies that each partition cell is convex and therefore contains both its centroid c_i and the codevector y_i, implying $\|c_i - y_i\| \le \delta_N$ and thus, together with the compact support assumption, implies that (30) follows from (29). □

We next consider the asymptotic behavior of the mean $E_{\Pi}[D(\Pi)]$ and variance $\operatorname{Var}_{\Pi}[D(\Pi)]$ of the distortion random variable $D(\Pi)$. It will be shown in the following propositions that the mean of $D(\Pi)$ can be written in a convenient asymptotic form and that ultimately it is a *positive constant, independent of the quantizer size N.* Also, the variance will be shown to decay to zero as N increases.

In order to examine $E_{\Pi}[D_c(\Pi)]$ the notion of a point density function is used. For any quantizer Q_N with partition cells R_i, having Lebesgue measure μ_i, define a *point density function* [10] by

$$\lambda_N(x) = \sum_{i=1}^{N}\frac{1}{N\mu_i}\chi_{R_i}(x). \quad (32)$$

The characteristic (or selector) function χ is defined for any set S as $\chi_S(x) = 1$ if $x = S$, and $\chi_S(x) = 0$ if $x \notin S$. Assume that for a given sequence of quantizers $\{Q_N\}$ there exists a probability density function λ such that $\lambda_N(x) \to \lambda(x)$ almost everywhere on S_f as $N \to \infty$. The following theorem gives an explicit expression for the asymptotic expectation of D_c in terms of two quantities, one depending on the source density and the other on the quantizer point density.

Theorem 2: Let $X \in \mathscr{R}^k$ be a random vector having a density f with compact support, mean m_X, and component variances σ_i^2 $(1 \le i \le k)$. Consider a sequence of N-level vector quantizers, that are regular, have diminishing cell diameters, randomized nonredundant index assignments, limiting point density function λ, and whose outputs are transmitted across a binary symmetric channel with bit error probability $\epsilon > 0$. Then $E_{\Pi}[D_c(\Pi)]$ is asymptotically (in N) equal to

$$(1-(1-\epsilon)^{\log_2 N})\left[\sum_{i=1}^{k}\sigma_i^2 + \int_{\mathscr{R}^k}\|x - m_X\|^2\lambda(x)\,dx\right]. \quad (33)$$

Proof: We can write

$$D_c(\Pi) = \sum_{i=1}^{N}\sum_{j=1}^{N}\|y_i - y_j\|^2 P_{\Pi(j)|\Pi(i)}P_i \quad (34)$$

so that the expectation of $D_c(\Pi)$ may be computed using Lemma 3:

$$\begin{aligned}E_{\Pi}[D_c(\Pi)] &= \frac{1}{N!}\sum_{i=1}^{N}\sum_{j=1}^{N}\|y_i - y_j\|^2 P_i\sum_{\pi}P_{\pi(j)|\pi(i)} \\ &= \frac{1-(1-\epsilon)^{\log_2 N}}{N-1}\sum_{i=1}^{N}\sum_{j=1}^{N}P_i\|y_i - y_j\|^2. \quad (35)\end{aligned}$$

To evaluate the double summation, define

$$g(N) \stackrel{\text{def}}{=} \sum_{i=1}^{N}\sum_{j=1}^{N}P_i\frac{1}{N}\|y_i - y_j\|^2 \quad (36)$$

and observe that g can be written as $g(N) = E\|W_N - V_N\|^2$, where W_N and V_N are independent discrete random vectors for each N. V_N is distributed uniformly over the finite codevector set $\{y_1, y_2, \cdots, y_N\}$, while W_N is distributed over the same set of codevectors, but with $\Pr[W_N = y_i] = P_i$. For each N define $\overline{W}_N = E[W_N]$ and write

$$\begin{aligned}g(N) &= E\|W_N - V_N\|^2 \\ &= E\|W_N - \overline{W}_N\|^2 \\ &\quad + 2E\left[(W_N - \overline{W}_N)^t(\overline{W}_N - V_N)\right] \\ &\quad + E\|V_N - \overline{W}_N\|^2. \quad (37)\end{aligned}$$

Note that $\lim_N \overline{W}_N = m_X$ since

$$\begin{aligned}\|\overline{W}_N - m_X\| &= \left\|\sum_{i=1}^{N}y_iP_i - \int_{\mathscr{R}^k}xf(x)\,dx\right\| \\ &\le \sum_{i=1}^{N}\int_{R_i}\|y_i - x\|f(x)\,dx \\ &\le \delta_N \to 0 \text{ as } N \to \infty. \quad (38)\end{aligned}$$

The middle term in (37) is zero by the independence of W_N and V_N. To compute the first term in (37) for large N, note that

$$\begin{aligned}\left|E\|W_N - \overline{W}_N\|^2 - \sum_{i=1}^{k} \sigma_i^2\right| &= \left|\sum_{i=1}^{N} \|y_i - \overline{W}_N\|^2 P_i - \int_{\mathscr{R}^k} \|x - m_X\|^2 f(x)\, dx\right| \\ &\le \sum_{i=1}^{N} \int_{R_i} \left| \|y_i - \overline{W}_N\|^2 - \|x - m_X\|^2 \right| f(x)\, dx \\ &\le \sum_{i=1}^{N} \int_{R_i} \left(\|y_i - \overline{W}_N\| + \|x - m_X\|\right) \\ &\quad \cdot \|y_i - x + m_X - \overline{W}_N\| f(x)\, dx \\ &\le 2\,\mathrm{diam}\,(S_f)\left(\delta_N + \|m_X - \overline{W}_N\|\right) \to 0 \text{ as } N \to \infty. \end{aligned} \tag{39}$$

where (38) was used. Each of the three above inequalities follows from the triangle inequality. So, $E\|W_N - \overline{W}_N\|^2 \to \sum_{i=1}^{k} \sigma_i^2$ as $N \to \infty$. Next, since $\lambda_N(y_i)\mu_i = 1/N$, we have

$$\begin{aligned} E\|V_N - \overline{W}_N\|^2 &= \sum_{i=1}^{N} \|y_i - \overline{W}_N\|^2 \frac{1}{N} \\ &= \sum_{i=1}^{N} \|y_i - \overline{W}_N\|^2 \lambda_N(y_i)\mu_i \end{aligned} \tag{40}$$

and then by adding and subtracting the quantity $\lambda(x)\|y_i - \overline{W}_N\|^2$ one obtains

$$\begin{aligned}\left|E\|V_N - \overline{W}_N\|^2 - \int_{\mathscr{R}^k} \|x - m_X\|^2 \lambda(x)\, dx\right| &\le \sum_{i=1}^{N} \int_{R_i} \left| \|y_i - \overline{W}_N\|^2 \lambda_N(y_i) - \|x - m_X\|^2 \lambda(x) \right| dx \\ &\le \sum_{i=1}^{N} \int_{R_i} \Big(\|y_i - \overline{W}_N\|^2 |\lambda_N(x) - \lambda(x)| \\ &\quad + \lambda(x)\left(\|y_i - \overline{W}_N\| + \|x - m_X\|\right) \\ &\quad \cdot \left(\|y_i - x\| + \|m_X - \overline{W}_N\|\right)\Big)\, dx \\ &\le \mathrm{diam}^2\,(S_f) \int_{S_f} |\lambda_N(x) - \lambda(x)|\, dx + 2\,\mathrm{diam}\,(S_f) \\ &\quad \cdot \left(\delta_N + \|m_X - \overline{W}_N\|\right) \end{aligned} \tag{41}$$

where we have used the fact that $\int_{\mathscr{R}^k} \lambda(x)\, dx = 1$, and the inequality

$$\left| \|a - b\| - \|c - d\| \right| \le \|a - c\| + \|d - b\|. \tag{42}$$

Scheffé's theorem [11] implies that $\lim_N \int_{\mathscr{R}^k} |\lambda_N(x) - \lambda(x)|\, dx = 0$, so the quantity in (41) tends to zero as $N \to \infty$, completing the proof. □

Corollary 1 below follows from (24), the centroid condition, Lemma 1, and Theorem 2. One consequence of Corollary 1 (and equivalently Theorem 2) is that an existence result can be inferred. A positive coding theorem can be stated that says there exists a sequence of index assignments $\{\pi_N\}$ that asymptotically achieve MSE's at least as small as in (43). Corollaries 2 and 3 follow directly from Theorem 2.

Corollary 1: Let $X \in \mathscr{R}^k$ be a random vector having a positive density f with compact support, mean m_X, and component variances σ_i^2 $(1 \le i \le k)$. Consider a sequence of N-level vector quantizers that are optimized for a noiseless channel and have randomized nonredundant index assignments. If the output indices are transmitted across a binary symmetric channel with bit error probability $\epsilon > 0$, then the mean squared error, $E_{X,\Pi,C}[D]$, is asymptotically (in N) equal to

$$b_{2,k}\|f\|_{k/(k+2)} N^{-2/k} + \left(1 - (1 - \epsilon)^{\log_2 N}\right) \cdot \left[\sum_{i=1}^{k} \sigma_i^2 + \|f\|_{k/(k+2)}^{-k/(k+2)} \cdot \int_{\mathscr{R}^k} \|x - m_X\|^2 f^{k/(k+2)}(x)\, dx\right]. \tag{43}$$

Corollary 2: The assumptions of Theorem 2 imply that $\lim_{N \to \infty} E_{X,\Pi,C}[D] > 0$.

Corollary 3: The assumptions of Theorem 2 imply that the rate that minimizes the mean squared error is $R_{\mathrm{opt}} = (1/2)\log_2(2\alpha/(k\beta\epsilon))$, and at this rate, $E_{X,\Pi,C}[D] \approx (\epsilon\beta k/2)\ln(2\alpha e/(k\beta\epsilon))$, where $\alpha = b_{2,k}\|f\|_{k/(k+2)}$ and $\beta = (1/\ln 2)(\sum_{i=1}^{k}\sigma_i^2 + \int_{\mathscr{R}^k}\|x - m_X\|^2\lambda(x)\, dx)$.

Proof: For a given ϵ, the MSE, written as a function of the number of quantizer codevectors N, is given asymptotically by

$$\begin{aligned} D(N) &\overset{\mathrm{def}}{=} E_{X,\Pi,C}[D] \\ &= \alpha N^{-2/k} + \left(1 - (1 - \epsilon)^{\log_2 N}\right)\beta \ln 2 \\ &\approx \alpha N^{-2/k} + \epsilon\beta \ln N. \end{aligned} \tag{44}$$

Setting $dD(N)/dN = 0$ then yields $N_{\mathrm{opt}} = (2\alpha/(k\beta\epsilon))^{k/2}$, and then we use $R_{\mathrm{opt}} = (1/k)\log_2 N_{\mathrm{opt}}$. □

Equation (33) can easily be computed for many sources and together with (1) provides a useful formula for estimating the MSE of a vector quantizer in the presence of channel noise for a randomized index assignment. The minimum mean squared error in this case provides an upper bound to the minimum mean squared error of the quantizer using the best possible index assignment and this performance can be guaranteed at the optimal rate. Perhaps future research will gain intuition from this analysis in order to obtain the performance formulas and best operating rates for quantizers having optimal index assignments.

The MSE formula given in Theorem 2 is for a randomized index assignment and can be useful for several purposes. First, it gives an analytical upper bound on the MSE performance of the best possible index assignment, as well as a lower bound on the MSE performance of the worst possible index assignment. No other analytical bounds are presently known. Some experimental results are given in [14] and [4], where algorithms for finding good index assignments are presented. In [14 (Figs. 3–5)] the randomized index assignment performance is often close to the "worst case" performance while in [4 (Figs. 3–5)] the performance of the index assignment resulting from the generalized Lloyd algorithm is often close to the best possible index assignment. Together these experimental results suggest that the randomized index assignment performance formula provides an upper bound on the MSE of the index assignment generated by the generalized Lloyd algorithm.

An example for the Gaussian source is given below. While our results have been proven only for sources with bounded support, we feel the Gaussian source is very illustrative and indeed shows the utility of the formulas for more general sources. We have found experimentally that truncating the support region of a Gaussian density to several standard deviations yields nearly identical results.

Gaussian Source: Let $k = 1$ and consider an independent and identically distributed (i.i.d.) Gaussian source with density function $f(x) = N(0, \sigma^2)$. Some arithmetic yields $\|f\|_{1/3} = 6\pi\sqrt{3}\,\sigma^2$, $b_{2,1} = 1/12$, $\alpha = \sqrt{3}/2\pi\sigma^2$, and $\beta = 4\sigma^2/\ln 2$. The point density $\lambda(x)$ is chosen to be the optimal point density for the high-resolution compandor model, $\lambda(x) = f^{1/3}/\int_{\mathscr{R}} f^{1/3} = N(0, 3\sigma^2)$. Thus,

$$N_{\text{opt}} = \frac{\epsilon^{-\frac{1}{2}}}{2}\sqrt{\pi\sqrt{3}\ln 2} \approx \frac{.971}{\sqrt{\epsilon}} \tag{45}$$

$$R_{\text{opt}} \approx 1.66 \cdot \log_{10}\frac{1}{\epsilon} \text{ (bits)} \tag{46}$$

$$D(N_{\text{opt}}) = 2\sigma^2\epsilon \log_2\left(\frac{\pi e\sqrt{3}\ln 2}{4\epsilon}\right) \approx \sigma^2\epsilon\left(2.716 + 6.644 \cdot \log_{10}\frac{1}{\epsilon}\right). \tag{47}$$

The distortion in (44) is plotted in Fig. 3 as a function of the rate for several different values of ϵ for this Gaussian source. The minimum value of each curve corresponds to R_{opt}. It can be seen in the figure and in (46) that if the bit error probability decreases by a factor of 10 in this case then the optimal rate increases by about 1.66 bits. An intuitive explanation for this is that as the rate increases, the resulting longer transmitted codevector indices are more exposed to damaging channel errors. Thus, on channels with smaller ϵ, one can transmit longer indices to achieve the minimum MSE.

Fig. 4 shows the performance of scalar quantization in terms of SNR versus the channel's bit error probability ϵ, again for the same Gaussian source. Three experimental curves are shown corresponding respectively to the best index assignment, worst index assignment, and average over all index assignments from among 5000 randomly chosen index assignments. The average index assignment curve is accurately modeled by the curve labeled "Random Coding Formula," the asymptotic formula from (43).

It was observed experimentally in [12] that for "channel optimized" quantizers, as the transmission rate increased, the number of quantization levels would at some point not increase any more. Instead the additional rate available was better used in channel coding. Adding more quantization levels would in fact have a detrimental effect in such cases. This observation is supported by Corollary 3, which shows that the MSE of quantization can actually increase in the presence of channel noise as the number of levels is increased. The average distortion $D(N)$ in (44) is strictly increasing as a function of N for $N > N_{\text{opt}}$; increasing N beyond N_{opt} only reduces the performance. It should be noted however, that Corollary 3 is based on the assumption that the quantizers are regular, a property that channel optimized quantizers might not always have for certain ϵ, though insight can still be gained from this reasoning.

The formula in (33) provides an asymptotic approximation for the MSE due to channel errors when a *typical index assignment* is used. Thus it serves as an asymptotic upper bound on the channel term D_c of the MSE for all *better than average* index assignments. For realistic large values of N, one might anticipate that this bound would be most useful for quantizers satisfying the centroid condition, as in this case the cross term $E_{X,\Pi,C}[D_{sc}]$ is guaranteed to be zero for all N.

An interesting feature of Corollary 2 is that for any source the expected distortion $E_{X,\Pi,C}[D]$, tends asymptotically to a strictly positive value. The regularity assumption played a key role in this conclusion since it guaranteed the decay to zero of the cross term $E_{\Pi}[D_{sc}(\Pi)]$. For sequences of quantizers satisfying the centroid condition even more can be said. For such sequences the cross term $E_{\Pi}[D_{sc}(\Pi)]$ is identically zero for all N, so that *for each* N the expected distortion is strictly positive.

It is important to consider the rate of convergence for Corollary 2. Though the first term $1 - (1 - \epsilon)^{\log_2 N}$ does eventually converge to 1, N must be of the order $2^{1/\epsilon}$ to approximate this asymptote (in which case the MSE is larger than the source variance), whereas for smaller N the first term in (33) may be approximated as $\epsilon \log_2 N$ for small values of ϵ.

In what follows the asymptotic behavior of the variance is examined. Recall that

$$\begin{aligned}\text{Var}_{\Pi}[D(\Pi)] &= \text{Var}_{\Pi}[E_X[D_s] + 2D_{sc}(\Pi) + D_c(\Pi)] \\ &= \text{Var}_{\Pi}[2D_{sc}(\Pi) + D_c(\Pi)] \\ &= E_{\Pi}\left[(2D_{sc}(\Pi) + D_c(\Pi))^2\right] \\ &\quad - E_{\Pi}^2[2D_{sc}(\Pi) + D_c(\Pi)].\end{aligned} \tag{48}$$

The second term in (48) has already been examined:

$$E_{\Pi}^2[2D_{sc}(\Pi) + D_c(\Pi)] = 4E_{\Pi}^2[D_{sc}(\Pi)] + 4E_{\Pi}[D_{sc}(\Pi)]E_{\Pi}[D_c(\Pi)] + E_{\Pi}^2[D_c(\Pi)]. \tag{49}$$

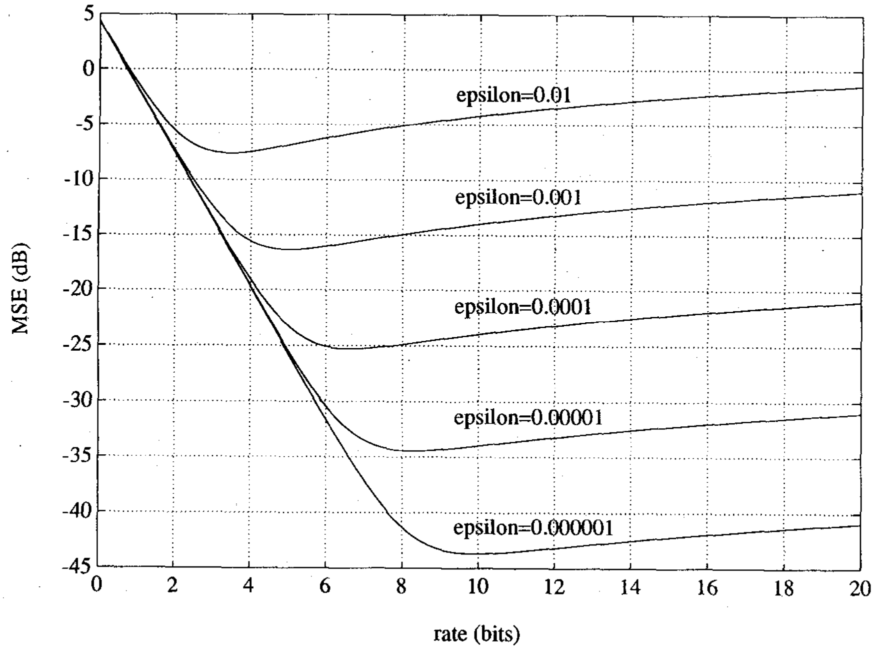

Fig. 3. Plot of noisy channel scalar quantization system's overall mean squared error, $10\log_{10} D(N)$, for a zero-mean, unit-variance, Gaussian i.i.d. source, and for different values of a binary symmetric channel's bit error probability ϵ.

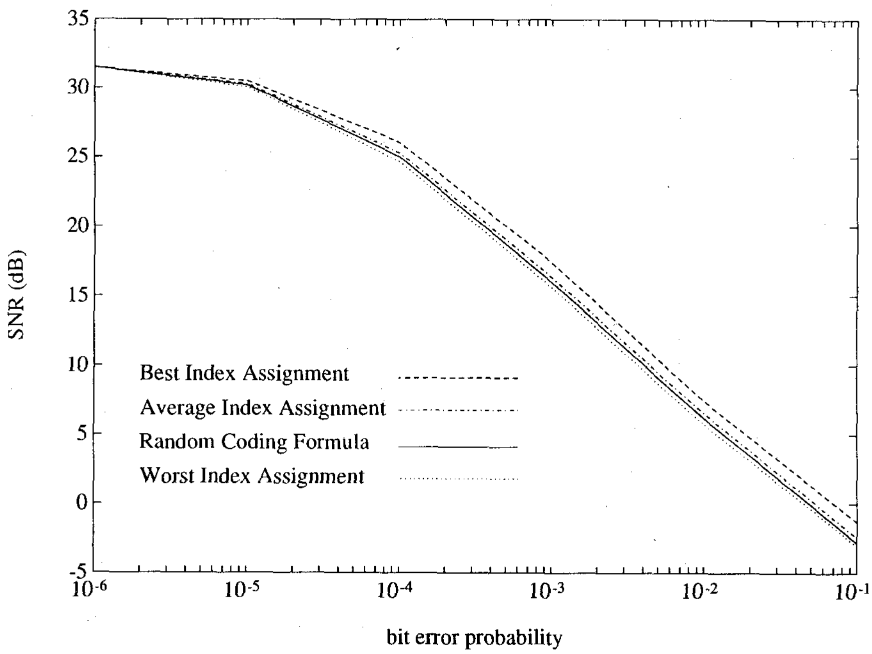

Fig. 4. Overall SNR versus bit error probability on a binary symmetric channel for noisy channel scalar quantization system. Plots are given for the best, worst, and average index assignments, and for the theoretically predicted average index assignment performance based on the random coding formula derived in Theorem 2. The source is i.i.d. zero-mean, unit-variance, Gaussian, at a rate of 6 b/sample.

By Proposition 1, $E_{\Pi}[D_{sc}(\Pi)] \to 0$ as $N \to \infty$, so the term $E_{\Pi}^2[2D_{sc}(\Pi) + D_c(\Pi)]$ is given asymptotically by $(\lim_{N\to\infty} E_{\Pi}[D_c(\Pi)])^2$, whose existence was established by Theorem 2. The following proposition is proven in Appendix B.

Proposition 2: The assumptions of Theorem 2 imply that $\lim_{N\to\infty}\mathrm{Var}_{\Pi}[D(\Pi)] = 0$.

Combining the results of Corollary 2 and Proposition 2 and applying Chebychev's inequality gives the theorem below. In it is stated that the mean squared error of Q_N

is, in probability asymptotically bounded away from zero. The precise meaning of this statement is that

There exists $\epsilon > 0$ such that

$$\lim_{N\to\infty} \Pr\left[E\|X - Q_N(X)\|^2 < \epsilon\right] = 0 \quad (50)$$

where the probability is with respect to the randomized index assignment. Informally, this says that if index assignments are chosen arbitrarily then the MSE will likely not decay to zero as the resolution of the quantizer increases.

Theorem 3: Let $X \in \mathscr{R}^k$ be a random vector having a density f with compact support. Consider a sequence of N-level vector quantizers Q_N that are regular, have diminishing cell diameters, randomized nonredundant index assignments, and whose outputs are transmitted across a binary symmetric channel with bit error probability $\epsilon > 0$. Then, in probability, the mean squared error (taken over source and channel statistics) of Q_N is asymptotically bounded away from zero.

Without channel noise, the MSE of an optimal N-level vector quantizer Q_N decays to zero. If channel noise is added to the output of Q_N, then according to the above theorem, the probability is asymptotically zero that the resulting MSE decays to zero.

IV. Conclusion

Random coding techniques are introduced to study the performance of asymptotically optimal zero-delay vector quantizers in the presence of channel noise. An upper bound, related to Zador's formula, is given for the minimum mean rth power distortion. For a noisy channel vector quantizer with randomized (or arbitrary) nonredundant index assignments a useful asymptotic formula for the MSE is given. A future research direction would be to mathematically investigate the minimum MSE of noisy channel quantizers for the best index assignments, which would improve upon the results for the average index assignments given in this paper. Along these lines one might study specific structured classes of index assignments rather than aggregate averages. Also, most of the results presented in this paper assume the source has a density with compact support. To remove this assumption would be an interesting challenge. Extending these results to sources and channels with memory is another obvious path to be followed.

Appendix A

Statements and Proofs of Lemmas

Lemma 2: Let $K < N$, $f: \{1, 2, \cdots, N\}^K \to \mathscr{R}$, and suppose that $i_1, i_2, \cdots, i_K$ are distinct elements of $\{1, 2, \cdots, N\}$. Then

$$\sum_{\pi\in S_N} f(\pi(i_1), \pi(i_2), \cdots, \pi(i_K)) = (N-K)! \sum_{\substack{m_1=1 \\ m_1, m_2, \cdots, m_k \text{ distinct}}}^{N} \sum_{m_2=1}^{N} \cdots \sum_{m_K=1}^{N} f(m_1, m_2, \cdots, m_K). \quad (51)$$

Proof: For every K-tuple $(m_1, m_2, \cdots, m_K)$ with $m_1, m_2, \cdots, m_K$ distinct, there exist exactly $(N-K)!$ permutations π in S_N that satisfy $\pi(i_1) = m_1, \pi(i_2) = m_2, \cdots, \pi(i_K) = m_K$. The result is immediate. □

Lemma 3: If $N \geq 2$, then for every pair (i, j) with $i \neq j$

$$\sum_{\pi\in S_N} P_{\pi(j)|\pi(i)} = [1 - (1-\epsilon)^{\log_2 N}]N(N-2)!. \quad (52)$$

Proof: By Lemma 2

$$\begin{aligned}\sum_{\pi\in S_N} P_{\pi(j)|\pi(i)} &= (N-2)! \sum_{k=1}^{N} \sum_{\substack{l=1 \\ l\neq k}}^{N} P_{k|l} \\ &= (N-2)! \sum_{k=1}^{N} \left(\sum_{l=1}^{N} P_{k|l} - P_{k|k} \right) \\ &= (N-2)! \left(\sum_{l=1}^{N} \sum_{k=1}^{N} P_{k|l} - \sum_{k=1}^{N} P_{k|k} \right) \\ &= [1 - (1-\epsilon)^{\log_2 N}]N(N-2)!. \quad (53)\end{aligned}$$

□

Lemma 4: If $N \geq 2$, then for every pair (i, j) with $i \neq j$

$$\sum_{\pi\in S_N} P^2_{\pi(j)|\pi(i)} \leq N(N-2)!. \quad (54)$$

Proof: By Lemma 2,

$$\begin{aligned}\sum_{\pi\in S_N} P^2_{\pi(j)|\pi(i)} &= (N-2)! \sum_{n=1}^{N} \sum_{\substack{m=1 \\ m\neq n}}^{N} P^2_{m|n} \\ &\leq (N-2)! \sum_{n=1}^{N} \sum_{m=1}^{N} P_{m|n} \\ &= N(N-2)!. \quad (55)\end{aligned}$$

□

Lemma 5: If $N \geq 3$, then for every 3-tuple (i, j, k) with i, j, k distinct

$$\sum_{\pi\in S_N} P_{\pi(i)|\pi(j)} P_{\pi(k)|\pi(j)} \leq N(N-3)!. \quad (56)$$

Proof: By Lemma 2,

$$\begin{aligned}\sum_{\pi\in S_N} P_{\pi(i)|\pi(j)} P_{\pi(k)|\pi(j)} &= (N-3)! \sum_{\substack{m=1 \\ m,n,r \text{ distinct}}}^{N} \sum_{n=1}^{N} \sum_{r=1}^{N} P_{m|n} P_{r|n} \\ &\leq (N-3)! \sum_{n=1}^{N} \sum_{m=1}^{N} P_{m|n} \sum_{r=1}^{N} P_{r|n} \\ &= N(N-3)!. \quad (57)\end{aligned}$$

□

Lemma 6: If $N \geq 4$, then for every 4-tuple (i, j, k, l) with i, j, k, l distinct

$$\sum_{\pi\in S_N} P_{\pi(j)|\pi(i)} P_{\pi(l)|\pi(k)} \leq N^2(N-4)!. \quad (58)$$

Proof: By Lemma 2

$$\sum_{\pi \in S_N} P_{\pi(j)|\pi(i)} P_{\pi(l)|\pi(k)}$$

$$= (N-4)! \sum_{\substack{m=1 \\ m,n,r,s \text{ distinct}}}^{N} \sum_{n=1}^{N} \sum_{r=1}^{N} \sum_{s=1}^{N} P_{m|n} P_{r|s}$$

$$\leq (N-4)! \left(\sum_{n=1}^{N} \sum_{m=1}^{N} P_{m|n} \right)^2 = N^2 (N-4)!. \quad (59)$$

□

Lemma 7: Let P_i be the probability of the ith cell in an optimal N-level vector quantizer. Then

$$\lim_N \sum_{i=1}^{N} P_i^2 = 0. \quad (60)$$

Proof:

$$\sum_{i=1}^{N} P_i^2 \leq \max_i \{P_i\} \sum_{i=1}^{N} P_i = \max_i \{P_i\}. \quad (61)$$

For optimal quantization, the MSE decreases to zero so that $\lim_N \max_i \{P_i\} = 0$. □

Appendix B

Proof of Proposition 2

Proof: To show that $\lim_N \mathrm{Var}_\Pi [D(\Pi)] = 0$ it suffices to show

$$\limsup_N E_\Pi \left[(2D_{sc}(\Pi) + D_c(\Pi))^2 \right] \leq \left(\lim_{N\to\infty} E_\Pi [D_c(\Pi)] \right)^2. \quad (62)$$

Consider each of three terms in the equation

$$E_\Pi \left[(2D_{sc}(\Pi) + D_c(\Pi))^2 \right] = 4E_\Pi [D_{sc}^2(\Pi)] + 4E_\Pi [D_{sc}(\Pi) D_c(\Pi)] + E_\Pi [D_c^2(\Pi)]. \quad (63)$$

We first show that $\limsup_N E_\Pi [D_c^2(\Pi)] \leq (\lim_{N\to\infty} E_\Pi [D_c(\Pi)])^2$ and then show that each of the remaining two terms decays to zero, completing the proof. Begin by noting that

$$E_\Pi [D_c^2(\Pi)] = \frac{1}{N!} \sum_{i_1=1}^{N} \sum_{j_1=1}^{N} P_{i_1} \|y_{i_1} - y_{j_1}\|^2 \cdot \sum_{i_2=1}^{N} \sum_{j_2=1}^{N} P_{i_2} \|y_{i_2} - y_{j_2}\|^2 \sum_\pi P_{\pi(j_1)|\pi(i_1)} P_{\pi(j_2)|\pi(i_2)}. \quad (64)$$

An upper bound for this expression can be found by splitting it into a sum of seven terms, $E_1, \cdots, E_7$ corresponding to seven mutually disjoint and exhaustive cases for the indices i_1, j_1, i_2, j_2.

Case 1 (i_1, j_1, i_2, j_2 all distinct): By Lemma 6, $\sum_\pi P_{\pi(j_1)|\pi(i_1)} P_{\pi(j_2)|\pi(i_2)} \leq N^2 (N-4)!$.

$$E_1 \leq \frac{N}{(N-1)(N-2)(N-3)} \left[\sum_{i_1=1}^{N} P_{i_1} \sum_{j_1=1}^{N} \|y_{i_1} - y_{j_1}\|^2 \right]^2$$

$$= \frac{N^3}{(N-1)(N-2)(N-3)} \cdot \left[\sum_{i_1=1}^{N} \sum_{j_1=1}^{N} P_{i_1} \frac{1}{N} \|y_{i_1} - y_{j_1}\|^2 \right]^2$$

$$\to \left(\lim_{N\to\infty} E_\Pi [D_c(\Pi)] \right)^2 \text{ as } N \to \infty \quad (65)$$

as shown in Theorem 2.

Case 2 ($i_2 = i_1, j_2 = j_1$): By Lemma 4, $\sum_\pi P_{\pi(j_1)|\pi(i_1)} \cdot P_{\pi(j_2)|\pi(i_2)} \sum_\pi P^2_{\pi(j)|\pi(i)} \leq N(N-2)!$.

$$E_2 \leq \frac{1}{N-1} \sum_{i_1=1}^{N} P_{i_1}^2 \sum_{j_1=1}^{N} \|y_{i_1} - y_{j_1}\|^4$$

$$\leq \frac{N}{(N-1)} [\mathrm{diam}\,(S_f)]^4 \sum_{i_1=1}^{N} P_{i_1}^2$$

$$\to 0 \text{ as } N \to \infty. \quad (66)$$

The sum in the last inequality above tends to zero by Lemma 7.

Case 3 ($i_2 = j_1, j_2 = i_1$): By Lemma 4, $\sum_\pi P_{\pi(j_1)|\pi(i_1)} \cdot P_{\pi(j_2)|\pi(i_2)} \leq N(N-2)!$.

$$E_3 \leq \frac{1}{N-1} \sum_{i_1=1}^{N} P_{i_1} \sum_{j_1=1}^{N} P_{j_1} \|y_{i_1} - y_{j_1}\|^4$$

$$\leq \frac{1}{N-1} [\mathrm{diam}\,(S_f)]^4 \sum_{i_1=1}^{N} P_{i_1} \sum_{j_1=1}^{N} P_{j_1}$$

$$= \frac{1}{N-1} [\mathrm{diam}\,(S_f)]^4$$

$$\to 0 \text{ as } N \to \infty. \quad (67)$$

Case 4 ($i_2 = i_1, j_2 \neq j_1$): By Lemma 5, $\sum_\pi P_{\pi(j_1)|\pi(i_1)} \cdot P_{\pi(j_2)|\pi(i_2)} \leq N(N-3)!$.

$$E_4 \leq \frac{1}{(N-1)(N-2)} \sum_{i_1=1}^{N} P_{i_1}^2 \cdot \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} \|y_{i_1} - y_{j_1}\|^2 \|y_{i_1} - y_{j_2}\|^2$$

$$\leq \frac{N^2}{(N-1)(N-2)} [\mathrm{diam}\,(S_f)]^4 \sum_{i_1=1}^{N} P_{i_1}^2$$

$$\to 0 \text{ as } N \to \infty. \quad (68)$$

The sum in the last inequality above tends to zero by Lemma 7.

Case 5 ($i_2 \neq i_1, j_2 = j_1$): By Lemma 5, $\sum_\pi P_{\pi(j_1)|\pi(i_1)} \cdot P_{\pi(j_2)|\pi(i_2)} \leq N(N-3)!$.

$$E_5 \leq \frac{1}{(N-1)(N-2)} \sum_{i_1=1}^{N} P_{i_1} \sum_{i_2=1}^{N} P_{i_2} \cdot \sum_{j_1=1}^{N} \|y_{i_1} - y_{j_1}\|^2 \|y_{i_2} - y_{j_1}\|^2$$

$$\leq \frac{N}{(N-1)(N-2)} [\mathrm{diam}\,(S_f)]^4 \sum_{i_1=1}^{N} P_{i_1} \sum_{i_2=1}^{N} P_{i_2}$$

$$= \frac{N}{(N-1)(N-2)} [\mathrm{diam}\,(S_f)]^4$$

$$\to 0 \text{ as } N \to \infty. \quad (69)$$

Case 6 $(i_2 = j_1, j_2 \neq i_1)$: By Lemma 5, $\sum_\pi P_{\pi(j_1)|\pi(i_1)} \cdot P_{\pi(j_2)|\pi(i_2)} \leq N(N-3)!$.

$$\begin{aligned} E_6 &\leq \frac{1}{(N-1)(N-2)} \sum_{i_1=1}^{N} P_{i_1} \sum_{j_1=1}^{N} P_{j_1} \cdot \sum_{j_2=1}^{N} \|y_{i_1} - y_{j_1}\|^2 \|y_{j_1} - y_{j_2}\|^2 \\ &\leq \frac{N}{(N-1)(N-2)} [\operatorname{diam}(S_f)]^4 \sum_{i_1=1}^{N} P_{i_1} \sum_{j_1=1}^{N} P_{j_1} \\ &= \frac{N}{(N-1)(N-2)} [\operatorname{diam}(S_f)]^4 \\ &\to 0 \text{ as } N \to \infty. \end{aligned} \tag{70}$$

Case 7 $(i_2 \neq j_1, j_2 = i_1)$: By Lemma 5, $\sum_\pi P_{\pi(j_1)|\pi(i_1)} \cdot P_{\pi(j_2)|\pi(i_2)} \leq N(N-3)!$.

$$\begin{aligned} E_7 &\leq \frac{1}{(N-1)(N-2)} \sum_{i_1=1}^{N} P_{i_1} \sum_{i_2=1}^{N} P_{i_2} \cdot \sum_{j_1=1}^{N} \|y_{i_1} - y_{j_1}\|^2 \|y_{i_2} - y_{i_1}\|^2 \\ &\leq \frac{N}{(N-1)(N-2)} [\operatorname{diam}(S_f)]^4 \sum_{i_1=1}^{N} P_{i_1} \sum_{i_2=1}^{N} P_{i_2} \\ &= \frac{N}{(N-1)(N-2)} [\operatorname{diam}(S_f)]^4 \\ &\to 0 \text{ as } N \to \infty. \end{aligned} \tag{71}$$

We have thus established that $\limsup E_\Pi[D_c^2(\Pi)] \leq (\lim_{N\to\infty} E_\Pi[D_c(\Pi)])^2$. Next, applying the triangle and Cauchy–Schwarz inequalities gives

$$\begin{aligned} &E_\Pi[D_{sc}^2(\Pi)] \\ &= \left| \frac{1}{N!} \sum_{i_1=1}^{N} \sum_{j_1=1}^{N} P_{i_1}(c_{i_1} - y_{i_1})^t (y_{i_1} - y_{j_1}) \cdot \sum_{i_2=1}^{N} \sum_{j_2=1}^{N} P_{i_2}(c_{i_2} - y_{i_2})^t (y_{i_2} - y_{j_2}) \cdot \sum_\pi P_{\pi(j_1)|\pi(i_1)} P_{\pi(j_2)|\pi(i_2)} \right| \\ &\leq \frac{1}{N!} \sum_{i_1=1}^{N} \sum_{j_1=1}^{N} P_{i_1} \|c_{i_1} - y_{i_1}\| \cdot \|y_{i_1} - y_{j_1}\| \cdot \sum_{i_2=1}^{N} \sum_{j_2=1}^{N} P_{i_2} \|c_{i_2} - y_{i_2}\| \cdot \|y_{i_2} - y_{j_2}\| \cdot \sum_\pi P_{\pi(j_1)|\pi(i_1)} P_{\pi(j_2)|\pi(i_2)} \\ &\leq \delta_N^2 \operatorname{diam}(S_f)^2 \sum_{i_1=1}^{N} P_{i_1} \sum_{i_2=1}^{N} P_{i_2} \cdot \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} \frac{1}{N!} \sum_\pi P_{\pi(j_1)|\pi(i_1)} P_{\pi(j_2)|\pi(i_2)}. \end{aligned} \tag{72}$$

Noting that $\delta_N \to 0$ as $N \to \infty$, this quantity can be bounded by a sum of seven terms (using seven cases as above), each of which decays to zero asymptotically. Finally we note that

$$\begin{aligned} &|E_\Pi[D_{sc}(\Pi) D_c(\Pi)]| \\ &= \left| \frac{1}{N!} \sum_{i_1=1}^{N} \sum_{j_1=1}^{N} P_{i_1}(c_{i_1} - y_{i_1})^t (y_{i_1} - y_{j_1}) \sum_{i_2=1}^{N} \sum_{j_2=1}^{N} \cdot P_{i_2} \|y_{i_2} - y_{j_2}\|^2 \sum_\pi P_{\pi(j_1)|\pi(i_1)} P_{\pi(j_2)|\pi(i_2)} \right| \\ &\leq \frac{1}{N!} \sum_{i_1=1}^{N} \sum_{j_1=1}^{N} P_{i_1} \|c_{i_1} - y_{i_1}\| \cdot \|y_{i_1} - y_{j_1}\| \cdot \sum_{i_2=1}^{N} \sum_{j_2=1}^{N} P_{i_2} \|y_{i_2} - y_{j_2}\|^2 \sum_\pi P_{\pi(j_1)|\pi(i_1)} P_{\pi(j_2)|\pi(i_2)} \\ &\leq \delta_N \operatorname{diam}(S_f)^3 \sum_{i_1=1}^{N} P_{i_1} \sum_{i_2=1}^{N} P_{i_2} \cdot \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} \frac{1}{N!} \sum_\pi P_{\pi(j_1)|\pi(i_1)} P_{\pi(j_2)|\pi(i_2)}. \end{aligned} \tag{73}$$

Again noting that $\delta_N \to 0$ as $N \to \infty$, this quantity can also be bounded by a sum of seven terms each of which decays to zero asymptotically. □

Acknowledgments

The authors would like to thank T. Linder and G. Lugosi for helpful technical discussions.

References

[1] J. Dunham and R. M. Gray, "Joint source and noisy channel trellis encoding," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 516–519, July 1981.
[2] H. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electron. Eng. Japan*, vol. 67-B, no. 4, pp. 39–47, 1984.
[3] J. De Marca and N. Jayant, "An algorithm for assigning binary indices to the codevectors of a multi-dimensional quantizer," in *Proc. IEEE Int. Conf. Commun.*, Seattle, WA, June 1987.
[4] K. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Trans. Commun.*, vol. 38, pp. 2147–2158, Dec. 1990.
[5] S. McLaughlin and D. Neuhoff, "Asymptotic bounds in source-channel coding," in *Proc. IEEE Int. Symp. Info. Theory (ISIT)*, Budapest, Hungary, June 1991.
[6] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 139–149, Mar. 1982.
[7] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[8] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer Academic, 1992.
[9] W. M. Lam and A. Reibman, "Asymptotic quantization for signal estimation with noisy channels," preprint.
[10] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.
[11] P. Billingsley, *Convergence of Probability Measures*. New York: Wiley, 1968.
[12] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 827–838, Nov. 1987.
[13] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[14] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 799–809, July 1990.
[15] F. Jelinck, "Evaluation of expurgated bound exponents," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 501–505, May 1968.