Tradeoff Between Source and Channel Coding for Codes Satisfying the Gilbert-Varshamov Bound

András Méhes and Kenneth Zeger
Department of Electrical and Computer Engineering,
University of California, San Diego, CA 92093-0407
{andras,zeger}@code.ucsd.edu

Abstract— **We derive bounds for optimal rate allocation between source and channel coding for linear codes that meet the Gilbert-Varshamov bound. Analogous bounds based on Shannon's channel coding theorem and Zador's high-resolution quantization formula have previously been given.**

I. Introduction

A theoretical study of the fundamental tradeoff of transmission rate between source and channel coding was given in [1]. They studied a cascaded vector quantizer and channel coder system operating on a binary symmetric channel, and derived upper and lower bounds on optimal rate allocation between the two subsystems. Their results rely on the fact that both subsystems contribute an exponentially decaying term to the total distortion (averaged over all index assignments), as a function of the overall transmission rate of the system.

Various suboptimal algorithms exist for vector quantizer design for noisy channels, but their implementation and design complexities generally grow exponentially fast as a function of the transmission rate of the system. A commonly used approach to transmitted source information across a noisy channel is to cascade a vector quantizer designed for a noiseless channel, and a block channel coder designed independently of the source coder.

A fundamental question for this traditional "separation" approach is to determine the optimal allocation of available transmission rate between source coding and channel coding. In practice, there is usually a constraint on the overall delay and complexity of such a system. This constraint generally limits the length of source block lengths and of channel codeword block lengths. As a result, the classical approach of Shannon, to transmit channel information at a rate close to the channel's capacity and to encode the source with the corresponding available amount of information, cannot be used in practice. In reality, one must often transmit data at a rate substantially below capacity. The amount below capacity that one must transmit was determined in [1]. However, the results in [1] exploit the availability of codes which achieve the reliability function of the channel.

In the present paper we determine bounds on the optimal tradeoff between source and channel coding, for classes of linear channel codes achieving the Gilbert-Varshamov bound. Examples include certain Goppa codes, alternant codes, self-dual codes, and double circulant or quasi-cyclic codes [2, p. 557].

The research was supported in part by the National Science Foundation.

II. Preliminaries

We consider a k-dimensional vector quantizer cascaded with a channel coder operating on a binary symmetric channel with a fixed overall transmission rate per vector component R, as shown in Figure 1. Let $r \in [0,1]$ denote the channel code rate. Then, for each input vector $\mathbf{x} \in \mathbb{R}^k$ the quantizer encoder produces a kRr-bit index i, which is passed through an index assignment π, and then encoded into kR bits before entering a binary symmetric channel. The index assignment is a permutation of the set of all possible kRr-bit indices. At the receiver, the channel decoder reconstructs a kRr-bit word $\pi(j)$ from the (possibly corrupted) kR bits received from the channel. Then the inverse of the index assignment is performed, and one of the 2^{kRr} quantizer codepoints $\mathbf{y}_j \in \mathbb{R}^k$ corresponding to the resulting index j is presented at the output.

Using the p^{th}-power distortion as a figure of merit, for a given index assignment π the performance of this system can be expressed as

$$D_\pi = \sum_{i=0}^{2^{kRr}-1} \sum_{j=0}^{2^{kRr}-1} Q_{\pi(j)|\pi(i)} \int_{R_i} \|\mathbf{x} - \mathbf{y}_j\|^p d\mu(\mathbf{x}), \tag{1}$$

where $\|\cdot\|$ is the usual Euclidean distance, μ is the probability distribution of the input, $Q_{l|m}$ is the probability that the channel decoder outputs l given that the input to the channel encoder was m, and R_i is the region of the encoder partition corresponding to codepoint $\mathbf{y}_i$. There are no known general techniques for analyzing the performance of such a system for an arbitrary index assignment. Instead we randomize the choice of index assignment. This technique serves as a tool in obtaining an existence theorem and at the same time accurately models the choice of index assignment in systems where active index design is ignored. Hence, we examine the distortion averaged over all index assignments,

$$D = \frac{1}{(2^{kRr})!} \sum_{\pi \in S_{2^{kRr}}} D_\pi,$$

where $S_{2^{kRr}}$ is the set of all possible permutations of all kRr-bit indices. The distortion can be decomposed into two components, one due to source coding and one due to

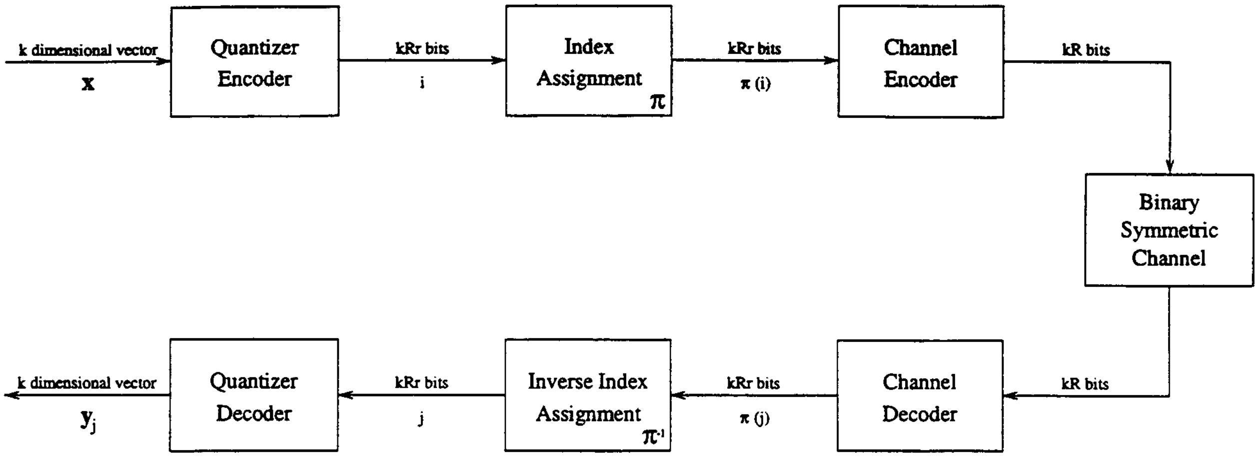

Fig. 1. Cascaded vector quantizer and channel coder system

channel coding. The behavior given that no uncorrected channel error occurred (equivalent to noiseless quantization) is described by Zador's formula; whereas the asymptotics (in R) of the component corresponding to an uncorrected channel error are governed by the probability of channel decoding error.

Let $f(n)$ and $g(n)$ be real-valued sequences. Then, we write

- $f = O(g)$, if there is a real number $c > 0$, and a positive integer n_0 such that $|f(n)| \leq c|g(n)|$, for all $n > n_0$;
- $f = o(g)$, if g has only a finite number of zeros, and $f(n)/g(n) \to 0$ as $n \to \infty$;
- $f = \Theta(g)$, if there are real numbers $c_1, c_2 > 0$, and a positive integer n_0 such that $c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|$, for all $n > n_0$.

In [1], it is shown that the average p^{th}-power distortion D of a cascaded system, with total available transmission rate R, consisting of a k-dimensional vector quantizer and a rate r channel coder can be bounded as

$$\begin{array}{c} 2^{-pRr+\Theta(1)} + 2^{-kRE_{\rm L}(r)+o(R)} \\ \leq \; D \; \leq \\ 2^{-pRr+\Theta(1)} + 2^{-kRE_{\rm U}(r)+o(R)} \end{array} \qquad (2)$$

where $E_{\rm L}$ and $E_{\rm U}$ are the corresponding error exponents in the lower and upper bound for the probability of decoding error. Thus, by solving the equations

$$E_{\rm L}(r_1) = \frac{p}{k}r_1 + o(1), \qquad (3)$$

and

$$E_{\rm U}(r_2) = \frac{p}{k}r_2 + o(1), \qquad (4)$$

one obtains the following bounds on the optimal rate r:

$$r_2 \leq r \leq r_1, \qquad (5)$$

and the corresponding distortion satisfies

$$D = 2^{-pRr+\Theta(1)}.$$

For a fixed transmission rate R, the performance of the system varies depending upon the rate allocation between source and channel coding. In [1], [3], bounds on the optimal channel coding rate r are given assuming that the channel code performs as predicted by Shannon's random coding argument. We investigate the problem of optimal rate allocation for channel codes satisfying the Gilbert-Varshamov bound.

As in [1], we take $E_{\rm L}$ to be the sphere-packing exponent, which holds for any code. Following [4, p. 165], this can be expressed as

$$E_{\rm sp}(r) = \mathcal{D}\left(\mathcal{H}^{-1}(1-r)\| \, \epsilon\right), \qquad (6)$$

or in parametric form as

$$\begin{aligned} r &= 1 - \mathcal{H}(\delta) & (7) \\ E_{\rm sp}(r) &= \mathcal{D}(\delta \| \, \epsilon), & (8) \end{aligned}$$

where $\mathcal{H}(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta)$ is the binary entropy, $\mathcal{D}(\delta\| \, \epsilon) = \delta \log \frac{\delta}{\epsilon} + (1-\delta)\log\frac{1-\delta}{1-\epsilon}$ is the information divergence between two Bernoulli distributions with parameters δ and ϵ, and all logarithms are base 2.

A similar expression can be obtained for $E_{\rm U}$ for sequences of codes that meet the Gilbert-Varshamov bound. First, consider the following bounds for the tail of a binomial distribution.

Lemma 1 ([4, p. 531]) For $\mu > p$,

$$\sum_{i=n\mu}^{n} \binom{n}{i} p^i (1-p)^{n-i} \leq 2^{-n\mathcal{D}(\mu\|p)}. \qquad (9)$$

Proposition 1: For (n, m) linear block channel codes with rate $r = m/n$ that meet the Gilbert-Varshamov bound, the probability of decoding error over a binary symmetric channel with bit error probability ϵ can be bounded as

$$P_e \leq 2^{-n\mathcal{D}\left(\frac{1}{2}\mathcal{H}^{-1}(1-r)\|\epsilon\right)} \qquad r \in (0, 1-\mathcal{H}(2\epsilon)). \qquad (10)$$

Proof

For an (n, nr, d) linear block channel code that meets the Gilbert-Varshamov bound we have

$$\mathcal{H}^{-1}(1-r) < \frac{d}{n} \leq \frac{1}{2} \qquad (11)$$

(see [4, p. 537] for details).

Using a standard bound on the probability of decoding error (e.g. [2, p. 19]) we have

$$\begin{aligned} P_e &\leq \sum_{i=\lfloor\frac{d-1}{2}\rfloor+1}^{n} \binom{n}{i}\epsilon^i(1-\epsilon)^{n-i} \\ &\leq \sum_{i=n\frac{1}{2}\mathcal{H}^{-1}(1-r)}^{n} \binom{n}{i}\epsilon^i(1-\epsilon)^{n-i} \\ &\leq 2^{-n\mathcal{D}(\frac{1}{2}\mathcal{H}^{-1}(1-r)\|\epsilon)} \end{aligned}$$

where the second inequality follows from (11), and the last one from Lemma 1. ∎

Thus, for "good" linear codes an error exponent can be written as

$$E_{\mathrm{GV}}(r) = \mathcal{D}\left(\frac{1}{2}\mathcal{H}^{-1}(1-r)\middle\|\,\epsilon\right), \qquad (12)$$

or in parametric form

$$\begin{aligned} r &= 1-\mathcal{H}(\delta) & (13) \\ E_{\mathrm{GV}}(r) &= \mathcal{D}(\delta/2\|\,\epsilon)\,. & (14) \end{aligned}$$

As stated in Proposition 1, this exponent is only valid for $r < 1-\mathcal{H}(2\epsilon)$. For $r \geq 1-\mathcal{H}(2\epsilon)$, the trivial bound of $P_e \leq 1$ can be used, which corresponds to the error exponent being zero.

Equations (6) through (8), and (12) through (14) differ only in a factor of 2. Also, equations (3) and (4) have essentially the same form. To reduce the redundancy in solving two nearly identical problems, we take a unified approach, and rewrite those expressions to obtain for $i = 1, 2$,

$$E_i(r_i) = \mathcal{D}\left(\frac{1}{i}\mathcal{H}^{-1}(1-r_i)\middle\|\,\epsilon\right). \qquad (15)$$

Alternatively,

$$\begin{aligned} r_i &= 1-\mathcal{H}(\delta_i), & (16) \\ E_i(r_i) &= \mathcal{D}(\delta_i/i\|\,\epsilon)\,, & (17) \end{aligned}$$

and the equations for balancing the exponents are

$$E_i(r_i) = \frac{p}{k}r_i + o(1), \qquad (18)$$

where $E_1 \equiv E_{\mathrm{L}} \equiv E_{\mathrm{sp}}$, and $E_2 \equiv E_{\mathrm{U}} \equiv E_{\mathrm{GV}}$.

III. Results

The bounds on the optimal rate allocation in a cascaded vector quantizer and channel coder system are functions of the vector dimension k, the channel bit error probability ϵ, and the parameter p in the distortion criterion. They do not depend, however, on the statistics of the source. We analytically characterize the optimal rate allocation for two important cases of interest: a large vector dimension k, and a small bit error probability ϵ. In each case the remaining parameters are assumed fixed but arbitrary.

A. Small Bit Error Probability

Lemma 2: For any p, k, and ϵ sufficiently small, let $\alpha_i \in \left(\frac{i}{2}\frac{p}{k}, 2i\frac{p}{k}\right)$ (*for* $i = 1,2$) satisfy

$$\begin{aligned} \alpha_i =& i\frac{p}{k} + \frac{\alpha_i}{\log 1/\epsilon}\left(1 - i\frac{p}{k}\right)(\log\log 1/\epsilon + \log e - \log\alpha_i) \\ &+ \frac{\alpha_i \log i}{\log 1/\epsilon} - \frac{\log e}{2i}\left(1-i^2\frac{p}{k}\right)\left(\frac{\alpha_i}{\log 1/\epsilon}\right)^2. \end{aligned} \qquad (19)$$

Then, the channel code rate minimizing the bounds in (2) is

$$r_i^* = 1 - \mathcal{H}\left(\frac{\alpha_i}{\log 1/\epsilon}\right) + O\left(\frac{1}{\log^3 1/\epsilon}\right) + o(1), \qquad (20)$$

where the $O\left(\frac{1}{\log^3 1/\epsilon}\right)$ term goes to zero as $\epsilon \to 0$ for any R, and the $o(1)$ term goes to zero as $R \to \infty$.

Proof

Using an intermediate value theorem argument (at $\alpha_i = \frac{i}{2}\frac{p}{k}$ and $\alpha_i = 2i\frac{p}{k}$) shows that an α_i satisfying (19) must exist. Thus, since the sequence of α_i is bounded, we have $\alpha_i \to i\frac{p}{k}$ as $\epsilon \to 0$.

After substituting $\delta_i \triangleq \frac{\alpha_i}{\log 1/\epsilon}$, and applying standard power series expansions, (16) becomes

$$\begin{aligned} r_i &= 1 + \frac{\alpha_i}{\log 1/\epsilon}\log\frac{\alpha_i}{\log 1/\epsilon} \\ &\quad + \left(1-\frac{\alpha_i}{\log 1/\epsilon}\right)\log\left(1-\frac{\alpha_i}{\log 1/\epsilon}\right) \\ &= 1 - \frac{\alpha_i}{\log 1/\epsilon}(\log\log 1/\epsilon + \log e - \log\alpha_i) \\ &\quad + \frac{\log e}{2}\left(\frac{\alpha_i}{\log 1/\epsilon}\right)^2 + O\left(\frac{1}{\log^3 1/\epsilon}\right), \end{aligned}$$

and (17) becomes

$$\begin{aligned} E_i(r_i) &= \frac{\alpha_i}{i\log 1/\epsilon}\log 1/\epsilon - \left(1-\frac{\alpha_i}{i\log 1/\epsilon}\right)\log(1-\epsilon) \\ &\quad + \frac{\alpha_i}{i\log 1/\epsilon}\log\frac{\alpha_i}{i\log 1/\epsilon} \\ &\quad + \left(1-\frac{\alpha_i}{i\log 1/\epsilon}\right)\log\left(1-\frac{\alpha_i}{i\log 1/\epsilon}\right) \\ &= \frac{p}{k}\left(1 - \frac{\alpha_i}{\log 1/\epsilon}(\log\log 1/\epsilon + \log e - \log\alpha_i)\right. \\ &\quad \left. + \frac{\log e}{2}\left(\frac{\alpha_i}{\log 1/\epsilon}\right)^2\right) + O\left(\frac{1}{\log^3 1/\epsilon}\right) \\ &= \frac{p}{k}r_i + O\left(\frac{1}{\log^3 1/\epsilon}\right) \end{aligned}$$

where (19) was used. Since r_i^* satisfies (18), we have $r_i^* = r_i + O\left(\frac{1}{\log^3 1/\epsilon}\right) + o(1)$. ∎

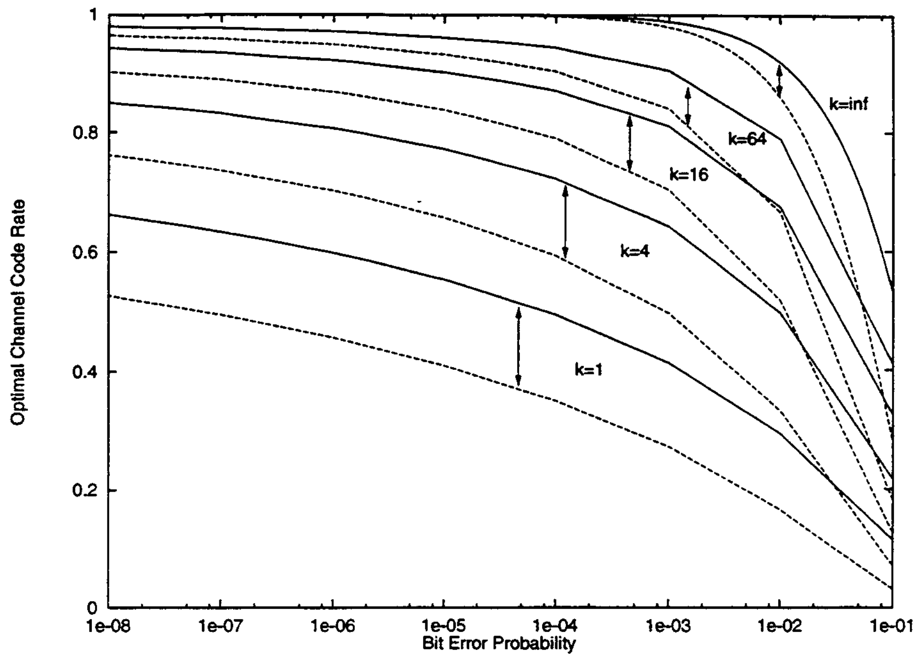

Fig. 2. A set of upper and lower bounds as given in Theorem 1 for mean-squared distortion ($p = 2$). Solid (upper bound) and dashed (lower bound) curves come in pairs for $k = 1, 4, 16, 64, \infty$. (The optimal rate increases as k gets larger.)

Theorem 1: Consider the cascade of a k-dimensional vector quantizer, a channel coder using a linear block channel code that meets the Gilbert-Varshamov bound, and a binary symmetric channel with bit error probability ϵ. Then, the channel code rate r that minimizes the p^{th}-power distortion (averaged over all index assignments) satisfies

$$\begin{aligned} & 1-\left(\frac{2p}{k}\right)\frac{\log\log 1/\epsilon+\log e-\log\frac{2p}{k}}{\log 1/\epsilon} \\ & -\left(\frac{2p}{k}\right)^2\frac{\log e}{2\log^2 1/\epsilon}+O\left(\frac{1}{\log^3 1/\epsilon}\right)+o(1) \\ \leq \quad r \quad \leq \quad & \\ & 1-\left(\frac{p}{k}\right)\frac{\log\log 1/\epsilon+\log e-\log\frac{p}{k}}{\log 1/\epsilon} \\ & -\left(\frac{p}{k}\right)^2\frac{\log e}{2\log^2 1/\epsilon}+O\left(\frac{1}{\log^3 1/\epsilon}\right)+o(1), \end{aligned}$$

where the $O\left(\frac{1}{\log^3 1/\epsilon}\right)$ term goes to zero as $\epsilon \to 0$ for any R, and the $o(1)$ term goes to zero as $R \to \infty$.

Figure 2 illustrates the upper and lower bounds of Theorem 1. The curves plotted are generated numerically without omitting any $O(\cdot)$ terms.

B. Large Source Vector Dimension

For $i = 1, 2$, let $C_i = 1 - \mathcal{H}(i\epsilon)$. Then, only the case of $r_i \in (0, C_i)$ corresponds to a nonzero error exponent. And, since the error exponents $E_i(r_i)$ are decreasing functions of the rate r_i, and as k increases the right hand side of (18) decreases, for R sufficiently large we have $r_i \to C_i$ as $k \to \infty$. Thus, for large k, $\mathcal{H}^{-1}(1-r_i)$ can be approximated by its Taylor series around $1 - C_i$

$$\begin{aligned} \mathcal{H}^{-1}(1-r_i) &= \mathcal{H}^{-1}(1-C_i)+\frac{C_i-r_i}{\mathcal{H}'(\mathcal{H}^{-1}(1-C_i))} \\ & \quad +O((C_i-r_i)^2) \end{aligned} \tag{21}$$

where $\mathcal{H}'(x) = \log\frac{1-x}{x}$ is the first derivative of the binary entropy $\mathcal{H}$.

We have $\mathcal{H}^{-1}(1-C_i) = \mathcal{H}^{-1}(\mathcal{H}(i\epsilon)) = i\epsilon$. Also, let us define $(x_i)_k \triangleq C_i - r_i$ Then $(x_i)_k \geq 0$, and $(x_i)_k \to 0$ as $k \to \infty$. Thus, rewriting (21) in a form suitable for substitution in (15) we obtain

$$\frac{1}{i}\mathcal{H}^{-1}(1-r_i) = \epsilon + \frac{(x_i)_k}{i\mathcal{H}'(i\epsilon)} + O((x_i)_k^2) \tag{22}$$

which approaches ϵ as k increases.

The information divergence can be expressed in terms of entropies as

$$\mathcal{D}\left(a\|\, b\right) = \mathcal{H}(b) + (a-b)\mathcal{H}'(b) - \mathcal{H}(a). \tag{23}$$

Thus,

$$\mathcal{D}\left(\frac{1}{i}\mathcal{H}^{-1}(1-r_i)\middle\|\,\epsilon\right) = -\frac{\mathcal{H}''(\epsilon)(x_i)_k^2}{2i^2\left[\mathcal{H}'(i\epsilon)\right]^2} + O((x_i)_k^3)$$

where $\mathcal{H}''(x) = -\frac{\log e}{x(1-x)}$ is the second derivative of the binary entropy $\mathcal{H}$. Let $\gamma_i \triangleq 2p\left[i\mathcal{H}(i\epsilon)\right]^2\left[-\mathcal{H}''(\epsilon)\right]^{-1}$. Using $r_i = C_i - (x_i)_k$, and substituting the above results in the equations for balancing the exponents yields

$$(x_i)_k^2 = \frac{\gamma_i}{k}(C_i - (x_i)_k) + O((x_i)_k^3) + o(1). \tag{24}$$

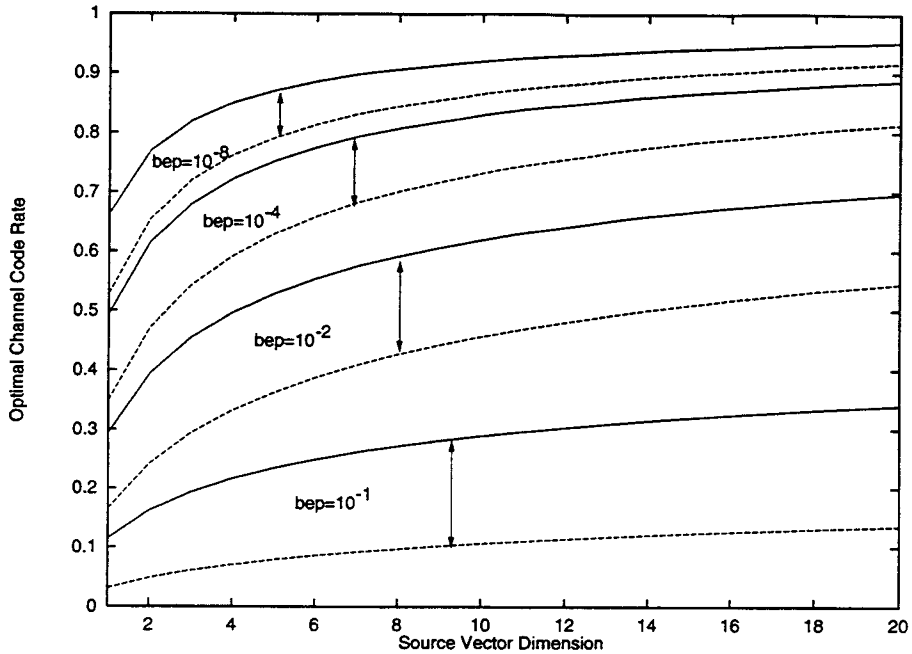

Fig. 3. Upper and lower bounds as given in Theorem 2 for mean-squared distortion ($p = 2$). Solid (upper bound) and dashed (lower bound) curves come in pairs for $\epsilon = 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}$. (The optimal rate increases as ϵ gets smaller.)

The nonnegative root of the quadratic is

$$\begin{aligned}(x_i)_k &= \sqrt{\frac{\gamma_i^2}{4k^2} + \frac{\gamma_i}{k}C_i + O((x_i)_k^3) + o(1)} - \frac{\gamma_i}{2k} \\ &= \sqrt{\frac{\gamma_i C_i}{k}} + O\left(\frac{1}{k}\right) + o(1).\end{aligned}$$

Thus, for $i = 1, 2$,

$$r_i = C_i - \frac{i\mathcal{H}'(i\epsilon)}{\sqrt{k}}\left(\frac{2pC_i}{-\mathcal{H}''(\epsilon)}\right)^{\frac{1}{2}} + O\left(\frac{1}{k}\right) + o(1),$$

and we can state the following.

Theorem 2: Consider the cascade of a k-dimensional vector quantizer, a channel coder using a linear block channel code that meets the Gilbert-Varshamov bound, and a binary symmetric channel with bit error probability ϵ. Then, the channel code rate r that minimizes the p^{th}-power distortion (averaged over all index assignments) satisfies

$$\begin{aligned}&C' - \frac{2\mathcal{H}'(2\epsilon)}{\sqrt{k}}\left(\frac{2pC'}{-\mathcal{H}''(\epsilon)}\right)^{\frac{1}{2}} + O\left(\frac{1}{k}\right) + o(1) \\ \leq\ r\ \leq\ &C - \frac{\mathcal{H}'(\epsilon)}{\sqrt{k}}\left(\frac{2pC}{-\mathcal{H}''(\epsilon)}\right)^{\frac{1}{2}} + O\left(\frac{1}{k}\right) + o(1),\end{aligned}$$

where $C = 1 - \mathcal{H}(\epsilon)$ is the channel capacity, and $C' = 1 - \mathcal{H}(2\epsilon)$; the $O\left(\frac{1}{k}\right)$ term goes to zero as $k \to \infty$ for any R, and the $o(1)$ term goes to zero as $R \to \infty$.

References

[1] B. Hochwald and K. Zeger, "Tradeoff between Source and Channel Coding." to appear in *IEEE Trans. Info. Theory*.

[2] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977–1993.

[3] K. Zeger and V. Manzella, "Asymptotic Bounds on Optimal Noisy Channel Quantization Via Random Coding," *IEEE Trans. Info. Theory*, vol. IT-40, pp. 1926–1938, November 1994.

[4] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.