Source and Channel Rate Allocation for Channel Codes Satisfying the Gilbert–Varshamov or Tsfasman–Vlăduţ–Zink Bounds

András Méhes and Kenneth Zeger, *Fellow, IEEE*

***Abstract*—We derive bounds for optimal rate allocation between source and channel coding for linear channel codes that meet the Gilbert–Varshamov or Tsfasman–Vlăduţ–Zink bounds. Formulas giving the high resolution vector quantizer distortion of these systems are also derived. In addition, we give bounds on how far below channel capacity the transmission rate should be for a given delay constraint. The bounds obtained depend on the relationship between channel code rate and relative minimum distance guaranteed by the Gilbert–Varshamov bound, and do not require sophisticated decoding beyond the error correction limit. We demonstrate that the end-to-end mean-squared error decays exponentially fast as a function of the overall transmission rate, which need not be the case for certain well-known structured codes such as Hamming codes.**

***Index Terms*—Error-correcting codes, source and channel coding, vector quantization.**

I. Introduction

ONE commonly used approach to transmit 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" technique is to determine the optimal allocation of available transmission rate between source coding and channel coding. Upper [1] and lower [2] distortion bounds on the optimal tradeoff between source and channel coding were previously derived for a binary symmetric channel. They exploit the fact that optimal source coding and optimal channel coding each contribute an exponentially decaying amount to the total distortion (averaged over all index assignments), as a function of the overall transmission rate of the system.

In practice, there is usually a constraint on the overall delay and complexity of such a system. This constraint limits the lengths of source blocks and of channel codewords. 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 amount of available information, cannot be used in practice. Instead, one must often transmit data at a rate substantially below capacity. The amount below capacity was determined in [2] for binary symmetric channels and in [3] for Gaussian channels. However, the results in both [2] and [3] exploit the existence of codes which have exponentially decaying error probabilities achieving the expurgated error exponent. Although such codes are known to exist, no efficiently decodable ones have yet been discovered. Various suboptimal algorithms do 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.

In the present paper we determine bounds on the optimal tradeoff between source and channel coding for classes of channel codes that attain the Gilbert–Varshamov bound. It is known that, asymptotically, a random linear code achieves the Gilbert–Varshamov bound with probability one [4], [5] although most known structured classes of codes fall short of the bound. The existence of certain Goppa codes, alternant codes, self-dual codes, and double-circulant or quasi-cyclic codes, that meet the Gilbert–Varshamov bound has been discussed in [6, p. 557]. A significant breakthrough was achieved by Tsfasman, Vlăduţ, and Zink [7], where sequences of algebraic geometry codes over GF (q) (with $q = p^{2m}$ and p prime) were constructed from reductions of modular curves. These codes exceed the Gilbert–Varshamov bound (in an interval of rates) if $q \geq 49$. Katsman, Tsfasman, and Vlăduţ [8] showed that there is an infinite polynomially constructible family of codes better than the Gilbert–Varshamov bound, although the best presently known (polynomial) algorithms are not yet practical. Another explicit construction of codes above the Gilbert–Varshamov curve was given recently in [9], but a detailed analysis of the algorithmic complexity of the construction is presently lacking. No binary constructions of codes with parameters exceeding the Gilbert–Varshamov bound are known. In fact, it is widely believed that the Gilbert–Varshamov bound is the tightest possible for $q = 2$. The best known binary codes are obtained from good q-ary codes by concatenation. Corresponding bounds are also available, but are generally weaker than the binary version of the Gilbert–Varshamov bound. There are several other bounds for the parameters of both linear and nonlinear, and both binary and nonbinary codes based on algebraic geometry codes. A summary of these bounds is found in [10] and a

Manuscript received October 6, 1997; revised September 12, 1998, November 9, 1999, and May 1, 2000. This work was supported in part by the National Science Foundation. The material in this paper was presented in part at the 31st Annual Conference on Information Sciences and Systems (CISS'97), Baltimore, MD, March 1997.

A. Méhes is with the Royal Institute of Technology, Stockholm, Sweden (e-mai: andras@s3.kth.se).

K. Zeger is with the Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093-0407 USA (e-mail: zeger@ucsd.edu).

Communicated by N. Merhav, Associate Editor for Source Coding and A. Vardy, Editor-in-Chief.

Publisher Item Identifier S 0018-9448(00)06994-7.

standard reference on algebraic-geometry codes is [11]. In [12], variable inner codes and an algebraic-geometry outer code are concatenated to obtain an exponentially decaying probability of error.

To obtain results for families of channel codes attaining the Gilbert–Varshamov or Tsfasman–Vlăduţ–Zink bounds, we only use the property that a positive monotone decreasing function g (in Proposition 2) exists describing the relationship between the channel code rate and the relative minimum distance. Thus the same method of derivation could potentially be used to obtain similar bounds for other classes of asymptotically good channel codes, some of which (e.g. Justesen codes, Blokh–Zyablov codes) are practical. However, it is often difficult to exhibit the function g in an analytically tractable form.

Since our derivation relies only on a standard bound on the probability of error which is valid even when bounded distance decoding is used, we in fact demonstrate that the class of known channel codes for which quantizer distortions decay to zero exponentially fast with increasing transmission rate includes certain suboptimal coding schemes. Note that families of channel codes which are not asymptotically good need not have exponentially decaying distortion as a function of the overall transmission rate. Indeed, repetition codes and other classes of codes with asymptotically vanishing channel code rates can have distortions decreasing to zero but not exponentially fast [13]. As suggested in [2], the distortion decay rates *are faster* when more sophisticated decoding algorithms are used.

The main results of this paper are as follows. In Theorem 1, upper and lower bounds are given for the optimal tradeoff between source and channel coding for channel codes satisfying the Gilbert–Varshamov or Tsfasman–Vlăduţ–Zink inequalities. Theorem 2 extends a result of [2] for the optimal source–channel coding tradeoff over an unrestricted class of channel codes. Theorems 1 and 2 enable a comparison of channel codes that achieve the reliability function of the channel (and in this sense are optimal for the given channel) and certain asymptotically good channel codes that are independent of the underlying channel. Fig. 4 presents an example of the loss in channel code rate due to suboptimality. Note that the bounds compared need not be the tightest possible in all cases. Theorem 3 gives the large-dimension performance of the optimal tradeoff determined in Theorem 1. In [2], the upper and lower bounds on the optimal rate allocation for "optimal" channel codes were shown to coincide for large enough dimensions (dependent on the bit-error probability). Thus we do not derive the large-dimension performance corresponding to Theorem 2, but in the example shown in Fig. 7 we include bounds for both optimal and suboptimal channel codes for comparison.

Throughout this paper we assume a randomized index assignment (i.e., a uniformly random mapping of vector quantizer codevectors to channel codewords). While this assumption is certainly suboptimal from an implementation standpoint, it provides a powerful mathematical tool for obtaining tight performance bounds, analogous in spirit to the classical randomization techniques used to prove Shannon's channel coding theorem. The same index assignment randomization method was used in [1], [2], and [3] as well. Furthermore, it is not presently known if randomization of index assignments is in general asymptotically suboptimal.

We note that at present, implementation of channel codes achieving the Gilbert–Varshamov or Tsfasman–Vlăduţ–Zink bounds is not computationally practical, but we conjecture that future research will yield more efficient codes. Even without such implementations, the present work serves as an improvement in the theoretical understanding of joint source and channel coding.

Section II gives necessary notations, definitions, and lemmas and Section III presents the source and channel coding tradeoff problem. Section IV gives basic results on bounds and error exponents. The main results of the paper are given in Section V and one technically complicated proof is left to the Appendix.

II. Preliminaries

The following notations will be useful in our asymptotic analysis.

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

- $f=O(g)$, if there is a positive real number c, and a positive integer n_0 such that $|f(n)|\leq c|g(n)|$, whenever $n>n_0$;
- $f = o(g)$, 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 positive real numbers c_1 and c_2, and a positive integer n_0, such that $c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|$, for all $n > n_0$.

We obtain bounds on the optimal rate allocation for the cascaded system depicted in Fig. 1. In this model, the source coder is a vector quantizer.

Definition 1: A k-dimensional, M-point *vector quantizer* is a mapping from k-dimensional Euclidean space $\mathbb{R}^k$ to a set of *codevectors* $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_M\} \subset \mathbb{R}^k$. Associated with each codevector $\boldsymbol{y}_i$ is an *encoder region* $\mathcal{R}_i \subset \mathbb{R}^k$, the set of all points in $\mathbb{R}^k$ that are mapped by the quantizer to $\boldsymbol{y}_i$. The set of encoder regions forms a partition of $\mathbb{R}^k$. The *rate* (or resolution) of a vector quantizer is defined as $R_s = (\log_2 M)/k$.

A vector quantizer is commonly decomposed into a quantizer *encoder* and a quantizer *decoder*. For each input vector, the encoder produces the index $i \in \{1, \dots, M\}$ of the encoder region $\mathcal{R}_i$ containing the input vector. For each index i, the decoder outputs the codevector $\boldsymbol{y}_i$.

The *pth-power distortion* of a vector quantizer is

$$D_0 = \sum_{i=1}^{M} \int_{\mathcal{R}_i} \|\boldsymbol{x} - \boldsymbol{y}_i\|^p \, d\mu(\boldsymbol{x}) \tag{1}$$

where $\|\cdot\|$ is the usual Euclidean norm, and μ is the probability distribution of a k-dimensional source vector. The subscript 0 is used to distinguish the distortion on an error-free channel from the distortion due to a noisy channel (to be discussed later). The high-resolution (i.e., large R_s) behavior of D_0 can be described by Zador's formula.

Lemma 1 (Zador [14]): The minimum pth-power distortion of a rate R_s vector quantizer is asymptotically given by

$$D_0 = 2^{-pR_s+O(1)}.$$

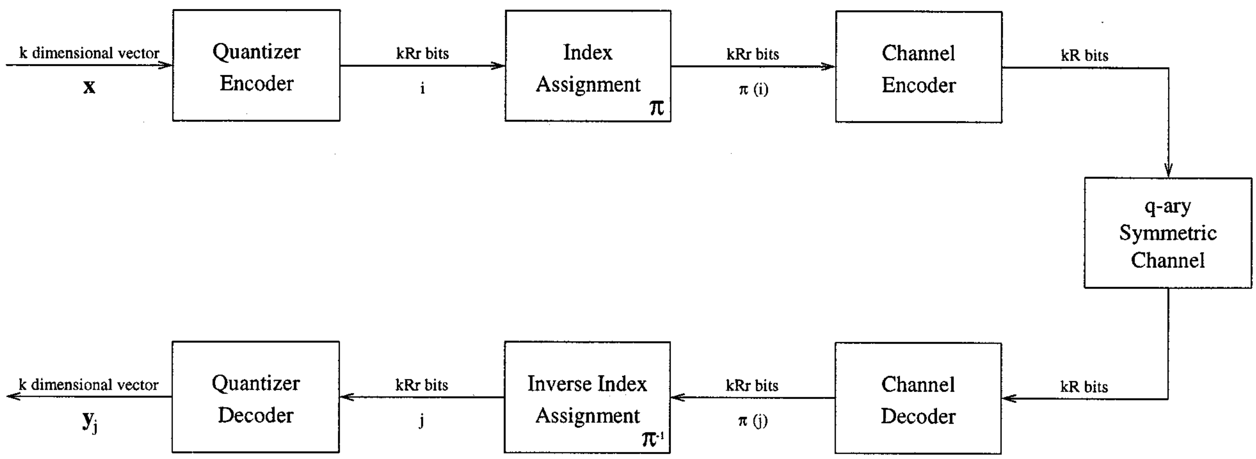


Fig. 1. Cascaded vector quantizer and channel coder system.

This is often referred to as the "6 dB/bit/component rule" for $p = 2$, since

$$10\log_{10}(2^{-pR_s}/2^{-p(R_s+1)}) \approx 3p.$$

In addition to the minimum distortion achieved by optimal quantizers, the asymptotic distortions of several other classes of vector quantizers, including uniform quantizers and other lattice-based quantizers, have the same high-resolution decay rate.

Definition 2: We call a vector quantizer that achieves the asymptotic distortion of Lemma 1 a *good vector quantizer*.

Motivated by the (nonbinary) alphabet size requirements of algebraic-geometry codes, we consider channel codes over GF(q) and use a q-ary symmetric channel in our system model shown in Fig. 1. The following two definitions formally introduce q-ary symmetric channels and q-ary linear block channel codes.

Definition 3: A *discrete memoryless channel* is a probabilistic mapping from an input alphabet $\mathcal{A}$ to an output alphabet $\mathcal{B}$ characterized by *channel transition probabilities* $P(b|a)$, i.e., the probability that the channel maps an input symbol $a \in \mathcal{A}$ to the output symbol $b \in \mathcal{B}$. A *q-ary symmetric channel* with symbol error probability $\epsilon \in [0, 1-q^{-1}]$ is a discrete memoryless channel having $\mathcal{A} = \mathcal{B} = \{0, \dots, q-1\}$ and channel transition probabilities

$$P(a|b) = I_{\{a=b\}}(1-\epsilon) + I_{\{a\neq b\}}\frac{\epsilon}{q-1}, \qquad a, b \in \{0, \dots, q-1\} \quad (2)$$

where I denotes the indicator function.

Definition 4: An (n, k) *block channel code* is a set of length n strings of q-ary symbols, called *codewords*. A *linear q-ary $[n, k, d]_q$ block channel code* is a linear subspace of $[\mathrm{GF}(q)]^n$, containing $M = q^k$ codewords, each (except the all-zero codeword) with at least d nonzero components. The number $r = \frac{k}{n} \in (0, 1]$ is the *channel code rate*.

Associated with a channel code is a *channel encoder* and a *channel decoder*. The channel encoder is a one-to-one mapping of messages (e.g., quantizer indices) to channel codewords for transmission. The channel decoder, on the other hand, is a many-to-one mapping. It maps received sequences of channel symbols (not necessarily codewords) to messages. Denoting the channel codeword corresponding to m by $c^{(m)}$, and the set of length n sequences decoded into l by S_l, the transition probabilities of the coded channel are

$$\beta_{l|m} = \sum_{u\in S_l}\prod_{i=1}^{n} P(u_i|c_i^{(m)})$$

where $u_i, c_i^{(m)} \in \{0, \dots, q-1\}$ are the ith symbols of u and $c^{(m)}$, respectively. The *average probability of decoding error* (for a uniform source) is

$$P_e = \frac{1}{M}\sum_{l=1}^{M}(1-\beta_{l|l}). \quad (3)$$

Although we never assume a uniform source, this definition of P_e is notationally convenient in what follows. The following two lemmas state classical asymptotic upper and lower bounds on P_e.

Lemma 2 [15, pp. 140, 153]: For every $r < C$, there exist sequences of (n, rn) channel codes such that

$$P_e < e^{-nE_{\max}(r)+o(n)}$$

where C denotes the capacity of the channel, and

$$E_{\max}(r) = \max(E_{\mathrm{rc}}(r), E_{\mathrm{ex}}(r))$$

is the maximum of the "random coding" and the "expurgated" error exponents.[1]

Lemma 2 characterizes the class of channel codes considered in [2]. For easier reference, we introduce the following terminology.

Definition 5: We call a block channel code that achieves the asymptotic error exponent in Lemma 2 an *efficient channel code*.

Lemma 3 [15, p. 157]: Any sequence of (n, rn) channel codes on a discrete memoryless channel must satisfy

$$P_e > e^{-nE_{\mathrm{sp}}(r)+o(n)}$$

where $E_{\mathrm{sp}}(r)$ is the "sphere packing" error exponent.

[1]The notation E_{rc} is used instead of the usual E_r to avoid confusing the subscript and the rate r.

While Lemma 2 is an existence result, Lemma 3 holds for all channel codes. The error exponent functions depend on the channel statistics. Definitions of $E_{\rm rc}$, $E_{\rm ex}$, and $E_{\rm sp}$ in terms of the transition probabilities of a discrete memoryless channel, and a derivation of closed-form expressions for q-ary symmetric channels are given in the Appendix. All three of these error exponent functions are known to be positive and convex in the range $0 < r < C$.

Another element of our system model shown in Fig. 1 is an index assignment.

Definition 6: An *index assignment* π is a permutation of the index set $\{1, \ldots, M\}$.

The purpose of an index assignment is to match a vector quantizer and a channel coder in a cascaded system in order to minimize the end-to-end distortion. Distance properties of channel codewords and quantizer codevectors should be aligned, so that on average a likely channel error (small Hamming distance) results in a tolerable quantization error (small Euclidean distance).

III. Problem Formulation

Consider a k-dimensional vector quantizer cascaded with a channel coder operating over a q-ary symmetric channel with a fixed overall transmission rate R measured in bits per vector component, as shown in Fig. 1. For each k-dimensional input vector, a channel codeword consisting of n q-ary symbols is transmitted across the channel to the receiver. The transmission rate is $R = (n \log_2 q)/k$. Let $r \in [0, 1]$ denote the rate of a q-ary $[n, rn, d]_q$ linear block channel code, where d is the minimum distance of the code (in q-ary symbols). The source coding rate and the overall transmission rate are related by $R_s = Rr$. Let M denote the number of quantizer codevectors (equivalently, the number of channel codewords). Then, $M = 2^{kR_s} = 2^{kRr} = q^{rn}$. For each input vector $\boldsymbol{x} \in \mathbb{R}^k$, the quantizer encoder produces an integer index $i \in \{1, \ldots, M\}$, which in turn is mapped to another index $\pi(i)$ by an index assignment. The channel encoder transmits the $\pi(i)$th-channel codeword through a q-ary symmetric channel (n q-ary symbols corresponding to kR bits). At the receiver, the channel decoder reconstructs an index $\pi(j)$ from the (possibly corrupted) n q-ary symbols received from the channel. Then the inverse index assignment is performed and the quantizer codevector $\boldsymbol{y}_j \in \mathbb{R}^k$ corresponding to the resulting index j is presented at the output.

For a given index assignment π, the average pth-power distortion can be expressed as

$$D(\pi) = \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{\pi(j)|\pi(i)} \int_{\mathcal{R}_i} ||\boldsymbol{x} - \boldsymbol{y}_j||^p \, d\mu(\boldsymbol{x}). \tag{4}$$

There are no known general techniques for analytically determining $\min_\pi D(\pi)$. As an alternative, we randomize the choice of index assignment. This technique serves as a tool in obtaining an existence theorem, and also models the choice of index assignment in systems where index design is ignored. Hence, we examine the following distortion:

$$D = \frac{1}{M!} \sum_{\pi} D(\pi) = \sum_{i=1}^{M} \sum_{j=1}^{M} \left[\frac{1}{M!} \sum_{\pi} \beta_{\pi(j)|\pi(i)} \right] \cdot \int_{\mathcal{R}_i} ||\boldsymbol{x} - \boldsymbol{y}_j||^p \, d\mu(\boldsymbol{x}) \tag{5}$$

where the sums over π are taken over all $M!$ permutations of the integers $\{1, \ldots, M\}$. The averaging effectively replaces the original q-ary symmetric channel by a "new" M-ary symmetric channel whose symbol error probability equals the average probability of channel decoding error P_e of the underlying channel. We have

$$\begin{aligned}
\frac{1}{M!} &\sum_{\pi} \beta_{\pi(j)|\pi(i)} \\
&= \frac{1}{M!} \sum_{\pi} \sum_{l=1}^{M} \sum_{m=1}^{M} \beta_{m|l} I_{\{\pi(i)=l, \pi(j)=m\}} \\
&= I_{\{i=j\}} \sum_{l=1}^{M} \beta_{l|l} \frac{1}{M!} \sum_{\pi} I_{\{\pi(i)=l\}} \\
&\quad + I_{\{i \neq j\}} \sum_{l=1}^{M} \sum_{\substack{m=1 \\ m \neq l}}^{M} \beta_{m|l} \frac{1}{M!} \sum_{\pi} I_{\{\pi(i)=l, \pi(j)=m\}} \\
&= I_{\{i=j\}} \sum_{l=1}^{M} \beta_{l|l} \frac{(M-1)!}{M!} \\
&\quad + I_{\{i \neq j\}} \sum_{l=1}^{M} (1 - \beta_{l|l}) \frac{(M-2)!}{M!} \\
&= I_{\{i=j\}} (1 - P_e) + I_{\{i \neq j\}} \frac{P_e}{M-1}.
\end{aligned} \tag{6}$$

Substituting (6) into (5) yields

$$\begin{aligned}
D = (1 - P_e) &\sum_{i=1}^{M} \int_{\mathcal{R}_i} ||\boldsymbol{x} - \boldsymbol{y}_i||^p \, d\mu(\boldsymbol{x}) \\
&+ \frac{P_e}{M-1} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \int_{\mathcal{R}_i} ||\boldsymbol{x} - \boldsymbol{y}_j||^p \, d\mu(\boldsymbol{x}).
\end{aligned} \tag{7}$$

The sum in the first term of (7) is the distortion for a noiseless channel. We assume that the source has compact support, in which case

$$\begin{aligned}
\frac{1}{M-1} &\sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \int_{\mathcal{R}_i} ||\boldsymbol{x} - \boldsymbol{y}_j||^p \, d\mu(\boldsymbol{x}) \\
&\le \frac{1}{M-1} \sum_{i=1}^{M} (M-1) \max_{\boldsymbol{x} \in \mathcal{R}_i, j \neq i} ||\boldsymbol{x} - \boldsymbol{y}_j||^p \int_{\mathcal{R}_i} d\mu(\boldsymbol{x}) \\
&\le \operatorname{diam}(\mu)
\end{aligned}$$

where $\operatorname{diam}(\mu)$ is the diameter of the support region. Unless the source is deterministic, a nonzero lower bound on the same double sum can be obtained using the pth-moment type quantity

$$\nu_p(\mu) = \min_{\boldsymbol{y}} \int_{\mathbb{R}^k} ||\boldsymbol{x} - \boldsymbol{y}||^p \, d\mu(\boldsymbol{x}).$$

Namely,

$$\begin{aligned}\frac{1}{M-1}\sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M}\int_{\mathcal{R}_i}\|\boldsymbol{x}-\boldsymbol{y}_j\|^p\,d\mu(\boldsymbol{x})\\ =\frac{1}{M-1}\left(\sum_{j=1}^{M}\sum_{i=1}^{M}\int_{\mathcal{R}_i}\|\boldsymbol{x}-\boldsymbol{y}_j\|^p\,d\mu(\boldsymbol{x})\right.\\ \left.-\sum_{i=1}^{M}\int_{\mathcal{R}_i}\|\boldsymbol{x}-\boldsymbol{y}_i\|^p\,d\mu(\boldsymbol{x})\right)\\ \geq\frac{M\nu_p(\mu)-D_0}{M-1}\geq\nu_p(\mu).\end{aligned}$$

Note that both the upper and lower bounds above depend solely on the source and not on the channel. Thus returning to (7) we have

$$D=(1-P_e)D_0+P_e\Theta(1). \tag{8}$$

We assume a good vector quantizer and an efficient channel code. Then, using Lemma 1 to bound D_0, and Lemmas 3 and 2 to bound P_e, the average pth-power distortion D of a cascaded source coder and rate r channel coder, with transmission rate R, can asymptotically (as $R\to\infty$) be bounded as

$$2^{-pRr+O(1)}+2^{-kRE_{\rm sp}(r)+o(R)}\leq D\leq 2^{-pRr+O(1)}+2^{-kRE_{\max}(r)+o(R)} \tag{9}$$

where the error exponents have been scaled by a factor of $\ln q$ as compared to Lemmas 2 and 3, in order to change the unit of block length from symbols to bits. The minimum value of the right side of (9) over all $r\in[0,1]$ is an asymptotically achievable (as $R\to\infty$) distortion D, and the minimum value of the left side of (9) is a lower bound on D for any choice of r. Let $r_{\max}$ and $r_{\rm sp}$, respectively, denote the values of r which minimize (asymptotically) the right and left sides of (9). Then $r_{\max}\leq r^*\leq r_{\rm sp}$, where r^* is the optimal rate allocation. It can be seen that to minimize the bounds in (9), the exponents of the two decaying exponentials in each bound have to be balanced, so that

$$E_X(r_X)=\frac{p}{k}r_X+o(1) \tag{10}$$

where formally $X\in\{\mathrm{sp},\max\}$ and $o(1)\to0$ as $R\to\infty$. The distortion achieved with a channel code rate r^* in this case is

$$D=2^{-pRr^*+O(1)}.$$

The values of $r_{\max}$ and $r_{\rm sp}$ were determined in [2] for efficient binary channel codes. We investigate the problem of optimal rate allocation for channel codes that attain the Gilbert-Varshamov bound and/or the Tsfasman–Vlăduţ–Zink bound (or "basic algebraic-geometry bound"). Such codes are in general weaker than those in [2], but are potentially less algorithmically complex. Our results also generalize those in [2] to q-ary channels.

IV. Error Exponents

In this section, we present the classical channel coding error exponents $E_{\rm rc}$, $E_{\rm ex}$, and $E_{\rm sp}$ specialized to a q-ary symmetric channel, and derive two new q-ary error exponents $E_{\rm GV}$ and $E_{\rm TVZ}$ for channel codes that satisfy the Gilbert–Varshamov and Tsfasman–Vlăduţ–Zink inequalities, respectively. All five of these error exponents can be concisely written using q-ary versions of the entropy, the relative entropy, and Rényi's entropy of order $1/2$. We start with the general definitions of these information measures.

Definition 7: Let P and $\hat{P}$ be probability distributions on a finite set.

The *entropy* of P is

$$H(P)=-\sum_x P(x)\log_2 P(x). \tag{11}$$

The *relative entropy* between P and $\hat{P}$ is

$$D(P\|\hat{P})=\sum_x P(x)\log_2(P(x)/\hat{P}(x)). \tag{12}$$

The *Rényi entropy of order* α of P is

$$H_\alpha(P)=\frac{1}{1-\alpha}\log_2\sum_x[P(x)]^\alpha \tag{13}$$

for $\alpha>0$, $\alpha\neq1$. Jensen's inequality implies $H_\alpha(P)\geq H(P)$ for $\alpha\in(0,1)$, and $H_\alpha(P)\leq H(P)$ for $\alpha>1$. Details of Rényi's information measures are given in [16].

Next, we introduce the various q-ary entropy functions defined for one-parameter distributions related to the transition probabilities of a q-ary symmetric channel.

Definition 8: Let $\epsilon,\delta\in[0,1-q^{-1}]$, and let $\mathcal{P}_\epsilon$ and $\mathcal{P}_\delta$ be probability distributions on $\{0,\ldots,q-1\}$ with respective probabilities $(1-\epsilon,\frac{\epsilon}{q-1},\ldots,\frac{\epsilon}{q-1})$, and $(1-\delta,\frac{\delta}{q-1},\ldots,\frac{\delta}{q-1})$.

The *q-ary entropy function* is defined as

$$\begin{aligned}\mathcal{H}_q(\epsilon)\triangleq H(\mathcal{P}_\epsilon)/\log_2 q=\epsilon\log_q(q-1)\\ -\epsilon\log_q\epsilon-(1-\epsilon)\log_q(1-\epsilon).\end{aligned} \tag{14}$$

For $q=2$ this gives the *binary entropy function*

$$h(\epsilon)=-\epsilon\log_2\epsilon-(1-\epsilon)\log_2(1-\epsilon).$$

The derivative of $\mathcal{H}_q$ with respect to ϵ is

$$\mathcal{H}_q'(\epsilon)=\log_q(q-1)-\log_q\epsilon+\log_q(1-\epsilon) \tag{15}$$

and the second derivative is

$$\mathcal{H}_q'(\epsilon)=-\frac{\log_q e}{\epsilon(1-\epsilon)}. \tag{16}$$

Thus $\mathcal{H}_q(\epsilon)$ is concave, strictly increasing on $[0,1-q^{-1}]$, and achieves its maximum $\mathcal{H}_q(1-q^{-1})=1$ and its minimum $\mathcal{H}_q(0)=0$. The notation $\mathcal{H}_q^{-1}$ denotes the inverse of $\mathcal{H}_q\colon[0,1-q^{-1}]\to[0,1]$. Clearly, $\mathcal{H}_q^{-1}$ is convex, from (16). The *capacity* of a q-ary symmetric channel with symbol error probability $\epsilon\in[0,1-q^{-1}]$ expressed in q-ary symbols is

$$C_q=1-\mathcal{H}_q(\epsilon). \tag{17}$$

The *(q-ary) relative entropy (information divergence) function* is defined as

$$\begin{aligned}\mathcal{D}_q(\delta\|\epsilon)&=D(\mathcal{P}_\delta\|\mathcal{P}_\epsilon)/\log_2 q\\ &=\delta\log_q\frac{\delta}{\epsilon}+(1-\delta)\log_q\frac{1-\delta}{1-\epsilon}\end{aligned} \tag{18}$$

which can also be expressed in terms of the q-ary entropy function as

$$\mathcal{D}_q(\delta\|\epsilon) = \mathcal{H}_q(\epsilon) + (\delta-\epsilon)\mathcal{H}_q'(\epsilon) - \mathcal{H}_q(\delta). \tag{19}$$

For $|\delta-\epsilon|$ small, a Taylor series approximation of $\mathcal{H}_q(\delta)$ around ϵ gives

$$\mathcal{D}_q(\delta\|\epsilon) = -\frac{1}{2}(\delta-\epsilon)^2\mathcal{H}_q''(\epsilon) + O(|\delta-\epsilon|^3). \tag{20}$$

We restrict attention to Rényi's entropy of order $1/2$, and the corresponding channel capacity of order $1/2$ for a q-ary symmetric channel. The *q-ary entropy function of order* $1/2$ is defined as

$$\mathcal{H}_q^{(1/2)}(\epsilon) \triangleq \mathcal{H}_{\frac{1}{2}}(\mathcal{P}_\epsilon)/\log_2 q = 2\log_q(\sqrt{1-\epsilon} + \sqrt{\epsilon(q-1)}). \tag{21}$$

The capacity of order $1/2$ of a q-ary symmetric channel with symbol error probability $\epsilon \in [0, 1-q^{-1}]$ expressed in q-ary symbols is

$$C_q^{(1/2)} = 1 - \mathcal{H}_q^{(1/2)}(\epsilon) \tag{22}$$

which Csiszár [16] showed to equal the "cutoff rate" of the channel.

The error exponents of Lemmas 2 and 3 can be specialized to a q-ary symmetric channel as follows (the proof of Proposition 1 is given in the Appendix).

Proposition 1:

$$E_{\rm sp}(r) = \mathcal{D}_q(\mathcal{H}_q^{-1}(1-r)\|\epsilon) \qquad r \in (0, C_q) \tag{23}$$

$$E_{\rm rc}(r) = \begin{cases} C_q^{(1/2)} - r & r \in (0, r_2] \\ \mathcal{D}_q(\mathcal{H}_q^{-1}(1-r)\|\epsilon) & r \in [r_2, C_q) \end{cases} \tag{24}$$

$$E_{\rm ex}(r) = \begin{cases} \mathcal{H}_q^{-1}(1-r)\log_q \frac{q-1}{q^{\mathcal{H}_q^{(1/2)}(\epsilon)}-1} & r \in (0, r_1] \\ C_q^{(1/2)} - r & r \in [r_1, C_q^{(1/2)}) \end{cases} \tag{25}$$

where

$$r_1 = 1 - \mathcal{H}_q(1 - q^{-\mathcal{H}_q^{(1/2)}(\epsilon)})$$

and

$$r_2 = 1 - \mathcal{H}_q\left(\frac{\sqrt{(q-1)\epsilon}}{\sqrt{1-\epsilon}+\sqrt{(q-1)\epsilon}}\right).$$

Also, since $r_2 \le C_q^{(1/2)} \le C_q$, and $r_1 \le r_2$ for $\epsilon < 1-q^{-1}$, we have

$$E_{\max}(r) = \begin{cases} \mathcal{H}_q^{-1}(1-r)\log_q \frac{q-1}{q^{\mathcal{H}_q^{(1/2)}(\epsilon)}-1} & r \in (0, r_1] \\ C_q^{(1/2)} - r & r \in [r_1, r_2] \\ \mathcal{D}_q(\mathcal{H}_q^{-1}(1-r)\|\epsilon) & r \in [r_2, C_q). \end{cases} \tag{26}$$

The lower bound on P_e given in Lemma 3 holds for an arbitrary code. The upper bound of Lemma 2, however, is an existence result. Analogous upper bounds, and corresponding error exponents can be obtained for "asymptotically good" families of codes.

For a sequence of (n, nr, d) codes to be asymptotically good, both the rate r and the relative minimum distance d/n must be bounded away from zero as the block length n increases. Usually, bounds are given in the form $d \ge ng(r)$ or $r \ge g^{-1}(d/n)$, for some monotonic decreasing function g. In this paper we consider two of the best known such bounds, the Gilbert–Varshamov bound and the Tsfasman–Vlăduţ–Zink bound (see [11, p. 609] for a summary of these and several related bounds).

Definition 9: An $[n, nr, d]_q$ code is said to satisfy the

- Gilbert–Varshamov bound, if
$$r \ge 1 - \mathcal{H}_q(d/n);$$
- Tsfasman–Vlăduţ–Zink bound, if
$$r \ge 1 - d/n - (\sqrt{q}-1)^{-1}.$$

The following lemma provides a bound on the tail of a binomial distribution.

Lemma 4 [15, p. 531]: For $\delta > \epsilon \ge 0$

$$\sum_{i=n\delta}^{n} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \le 2^{-n\mathcal{D}_2(\delta\|\epsilon)}.$$

Proposition 2: If an $[n, rn, d]_q$ linear block channel code has minimum distance $d \ge ng(r)$ for some positive monotone decreasing function g, then the average probability of decoding error on a q-ary symmetric channel with symbol error probability ϵ satisfies

$$P_e \le 2^{-n\mathcal{D}_2(\frac{1}{2}g(r)\|\epsilon)}, \qquad r \in (0, g^{-1}(2\epsilon)).$$

Proof: Since a code with minimum distance d can correct at least $\lfloor \frac{d-1}{2} \rfloor$ errors

$$P_e \le \sum_{i=\lfloor \frac{d-1}{2}\rfloor+1}^{n} \binom{n}{i} (q-1)^i (1-\epsilon)^{n-i} \left(\frac{\epsilon}{q-1}\right)^i \tag{27}$$

$$\le \sum_{i=n\frac{1}{2}g(r)}^{n} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \tag{28}$$

$$\le 2^{-n\mathcal{D}_2(\frac{1}{2}g(r)\|\epsilon)} \tag{29}$$

where inequality (28) follows from

$$\left\lfloor \frac{d-1}{2} \right\rfloor + 1 \ge d/2 \ge ng(r)/2$$

and inequality (29) from Lemma 4. □

The bound on P_e in (27) used to obtain Proposition 2 holds even when bounded distance decoding is used in the channel decoder. While tighter bounds on P_e would also improve the rate allocation bounds derived later in the paper, we opted for the "standard" bound (inequality (27)) for two main reasons. First, rate allocation bounds are already available for efficient channel codes assuming optimal decoding [2]. Our goal is to show that certain suboptimal coding schemes also achieve a high-resolution distortion which decays to zero exponentially fast with increasing transmission rate. Note that this need not be the case in general [13]. Second, (27) depends only on the minimum distance, which enables us to directly apply the function g relating the rate and the relative minimum distance, without any further assumptions on the structure of the channel codes. The upper bound on P_e given in Proposition 2 only depends on the code parameters n and r, the symbol error probability ϵ, and the function g. The following two corollaries follow immediately from Proposition 2 and will also be useful in what follows.

Corollary 1: Consider the cascade of a good k-dimensional vector quantizer, a q-ary linear block channel coder that achieves the Gilbert–Varshamov bound, and a q-ary symmetric channel with symbol error probability ϵ and overall transmission rate R. For every $r < 1 - \mathcal{H}_q(2\epsilon) = C_q(2\epsilon)$, the average probability of channel decoding error satisfies

$$P_e \le 2^{-kRE_{\mathrm{GV}}(r)}$$

where

$$E_{\mathrm{GV}}(r) = \mathcal{D}_q\left(\frac{1}{2}\mathcal{H}_q^{-1}(1-r)\|\epsilon\right) \tag{30}$$

is the Gilbert–Varshamov error exponent.

Corollary 2: Consider the cascade of a good k-dimensional vector quantizer, a q-ary linear block channel coder that achieves the Tsfasman–Vlăduţ–Zink bound, and a q-ary symmetric channel with symbol error probability ϵ and overall transmission rate R. For every $r < 1 - (\sqrt{q}-1)^{-1} - 2\epsilon$, the average probability of channel decoding error satisfies

$$P_e \le 2^{-kRE_{\mathrm{TVZ}}(r)}$$

where

$$E_{\mathrm{TVZ}}(r) = \mathcal{D}_q\left(\frac{1}{2}(1-r-(\sqrt{q}-1)^{-1})\middle\|\,\epsilon\right) \tag{31}$$

is the Tsfasman–Vlăduţ–Zink error exponent.

Analogously to $E_{\max}(r) = \max(E_{\mathrm{rc}}(r), E_{\mathrm{ex}}(r))$, we define

$$E'_{\max}(r) = \max(E_{\mathrm{GV}}(r), E_{\mathrm{TVZ}}(r)).$$

For $q \le 49$, $E'_{\max}(r) = E_{\mathrm{GV}}(r)$. For $q > 49$

$$E'_{\max}(r) = \begin{cases} \mathcal{D}_q(\frac{1}{2}(1-r-(\sqrt{q}-1)^{-1})\|\epsilon), \\ \qquad\qquad r \in [r'_1, r'_2] \\ \mathcal{D}_q(\frac{1}{2}\mathcal{H}_q^{-1}(1-r)\|\epsilon), \\ \qquad\qquad r \in (0, r'_1] \bigcup [r'_2, C_q(2\epsilon)) \end{cases}$$

where $r'_1 < r'_2$ are roots of $\mathcal{H}_q^{-1}(1-r) = 1-r+(\sqrt{q}-1)^{-1}$. (In [7] this equation is shown to have two distinct roots for $q > 49$.)

V. Optimal Rate Allocation

The bounds we obtain on the optimal rate allocation in a cascaded vector quantizer and channel coder system are functions of the vector dimension k, the channel symbol error probability ϵ, and the parameter p in the distortion criterion. These bounds do not depend, however, on the source statistics. We obtain analytic bounds on the optimal rate allocation for two important cases of interest: a large vector dimension k, and a small symbol error probability ϵ. In each case, the remaining parameters are assumed fixed but arbitrary. To obtain these bounds, we analyze the error exponents E_{sp}, $E_{\max}$, and $E'_{\max}$.

First, we note that on the interval $[r_1, r_2]$, the function $E_{\max}(r) = C_q^{(1/2)} - r$ is linear. Let r_{lin} be a solution of (10) (for $X = \max$) such that $r_{\mathrm{lin}} \in [r_1, r_2]$, whenever such a solution exists. Then

$$C_q^{(1/2)} - r_{\mathrm{lin}} = \frac{p}{k} r_{\mathrm{lin}}$$

or, equivalently,

$$r_{\mathrm{lin}} = C_q^{(1/2)}/(1+(p/k)).$$

If k is fixed, and ϵ approaches zero, then $r_{\mathrm{lin}} \to (1+(p/k))^{-1}$ and $r_1 \to 1$. Hence, $r_{\mathrm{lin}} < r_1$ for ϵ sufficiently small. Thus for ϵ sufficiently small, $r_{\max} < r_1$ and it, therefore, suffices to restrict attention to E_{ex} instead of $E_{\max}$ (see Fig. 2(a)).

If ϵ is fixed and k increases, then $r_{\mathrm{lin}} \to C_q^{(1/2)}$. Hence, $r_{\mathrm{lin}} > r_2$ for k sufficiently large. Thus for k sufficiently large, $r_{\max} > r_2$ and thus it suffices to restrict attention to E_{rc} instead of $E_{\max}$ (see Fig. 2(b)). Also note that $E_{\mathrm{rc}}(r) = E_{\mathrm{sp}}(r)$ for all $r \in [r_2, C_q)$. Thus the upper and lower bounds coincide as k increases, and hence, it suffices to consider E_{sp}.

Next, we examine $E'_{\max}$. For $q < 49$, $E'_{\max} = E_{\mathrm{GV}}$ for all r. For $q \ge 49$, note that r'_2 is independent of both k and ϵ and depends only on q. Thus for k fixed, and ϵ decreasing, $\frac{p}{k}r'_2$ (the right-hand side of (10)) is constant, whereas

$$\mathcal{D}_q(\mathcal{H}_q^{-1}(1-r'_2)/2\|\epsilon)$$

(the left-hand side of (10)) increases without bound. Hence, for ϵ small enough

$$\mathcal{D}_q(\mathcal{H}_q^{-1}(1-r'_2)/2\|\epsilon) > \frac{p}{k}r'_2. \tag{32}$$

Since $\frac{p}{k}r$ is a monotone increasing function of r, and $E'_{\max}(r)$ is monotone decreasing in r, (32) implies that if

$$E'_{\max}(r'_{\max}) = \frac{p}{k}r'_{\max}$$

then $r'_{\max} > r'_2$ (see Fig. 3(a)). For ϵ fixed and k increasing, $\frac{p}{k}r'_2$ (the right-hand side of (10)) is decreasing, while

$$\mathcal{D}_q(\mathcal{H}_q^{-1}(1-r'_2)/2\|\epsilon)$$

(the left-hand side of (10)) is constant. Hence, for k large enough, (32) holds, and by the same monotonicity argument used above, $r'_{\max} > r'_2$ (see Fig. 3(b)). Consequently, it suffices to work with E_{GV} instead of $E'_{\max}$. Thus we henceforth omit E_{TVZ} from our analysis.

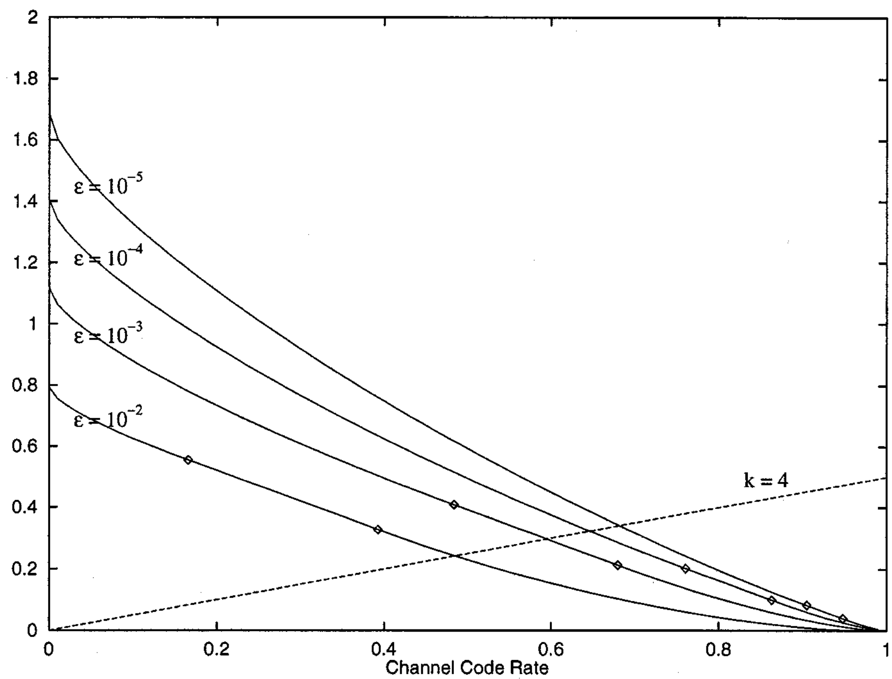
We note that a slightly more complicated differentiable bound relating d/n and r is also known. This bound, called "Vlăduts bound" [sic] in [17], effectively "smoothes the edges" of the maximum of the Gilbert–Varshamov and Tsfasman–Vlăduţ–Zink bounds. Applying Proposition 2, a "Vlăduţ error exponent" could also be obtained, but there exists a rate, analogous to r'_2 (independent from ϵ and k), beyond which the Vlăduţ and Gilbert–Varshamov error exponents coincide. Hence, by the same argument given above, it suffices to restrict attention to E_{GV} instead of the Vlăduţ error exponent.

A. Small Bit-Error Probability

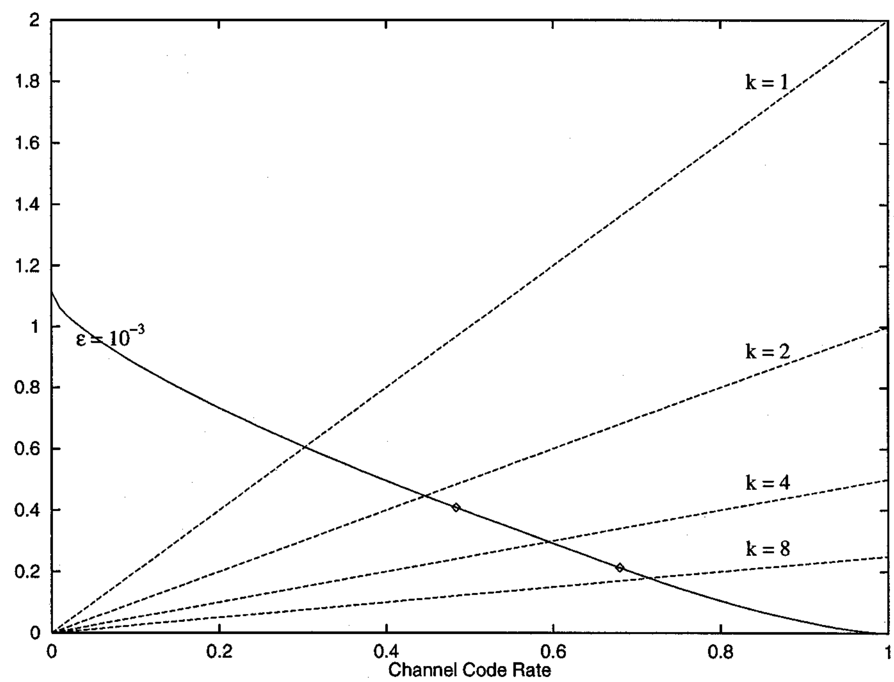
In this section, we determine the behavior of the solution to (10) for small ϵ, and fixed k and p. First, we set $\delta = \mathcal{H}_q^{-1}(1-r)$ and rewrite the error exponents as

$$E_{\mathrm{sp}}(\delta) = \mathcal{D}_q(\delta\|\epsilon), \qquad \delta \in (\epsilon, 1-q^{-1}) \tag{33}$$

$$E_{\mathrm{ex}}(\delta) = \delta \log_q \frac{q-1}{q^{\mathcal{H}_q^{(1/2)}(\epsilon)} - 1}, \qquad \delta \in [1-q^{-\mathcal{H}_q^{(1/2)}(\epsilon)}, 1-q^{-1}) \tag{34}$$



(a)



(b)

Fig. 2. A graphical solution to (10) for $E_{\max}$ ($p = 2$, $q = 64$). The solid curves show $E_{\max}(r)$ for different values of ϵ, and the dashed lines have slope p/k. The two dots on each error exponent curve correspond to r_1 and r_2. (a) Small bit-error probability. (b) Large vector dimension.

$$E_{\rm GV}(\delta) = \mathcal{D}_q(\delta/2 \| \epsilon), \qquad \delta \in (2\epsilon, 1 - q^{-1}). \tag{35}$$

Next we find a real number δ_X that satisfies

$$E_X(\delta_X) = c r(\delta_X) \tag{36}$$

where formally $X \in \{\mathrm{sp}, \mathrm{ex}, \mathrm{GV}\}$

$$r(\delta) = 1 - \mathcal{H}_q(\delta) \tag{37}$$

and $c = p/k$. Then, we obtain the solution to (10) by setting

$$r_X^* = r(\delta_X) + o(1) \tag{38}$$

where the $o(1)$ term vanishes as $R \to \infty$.

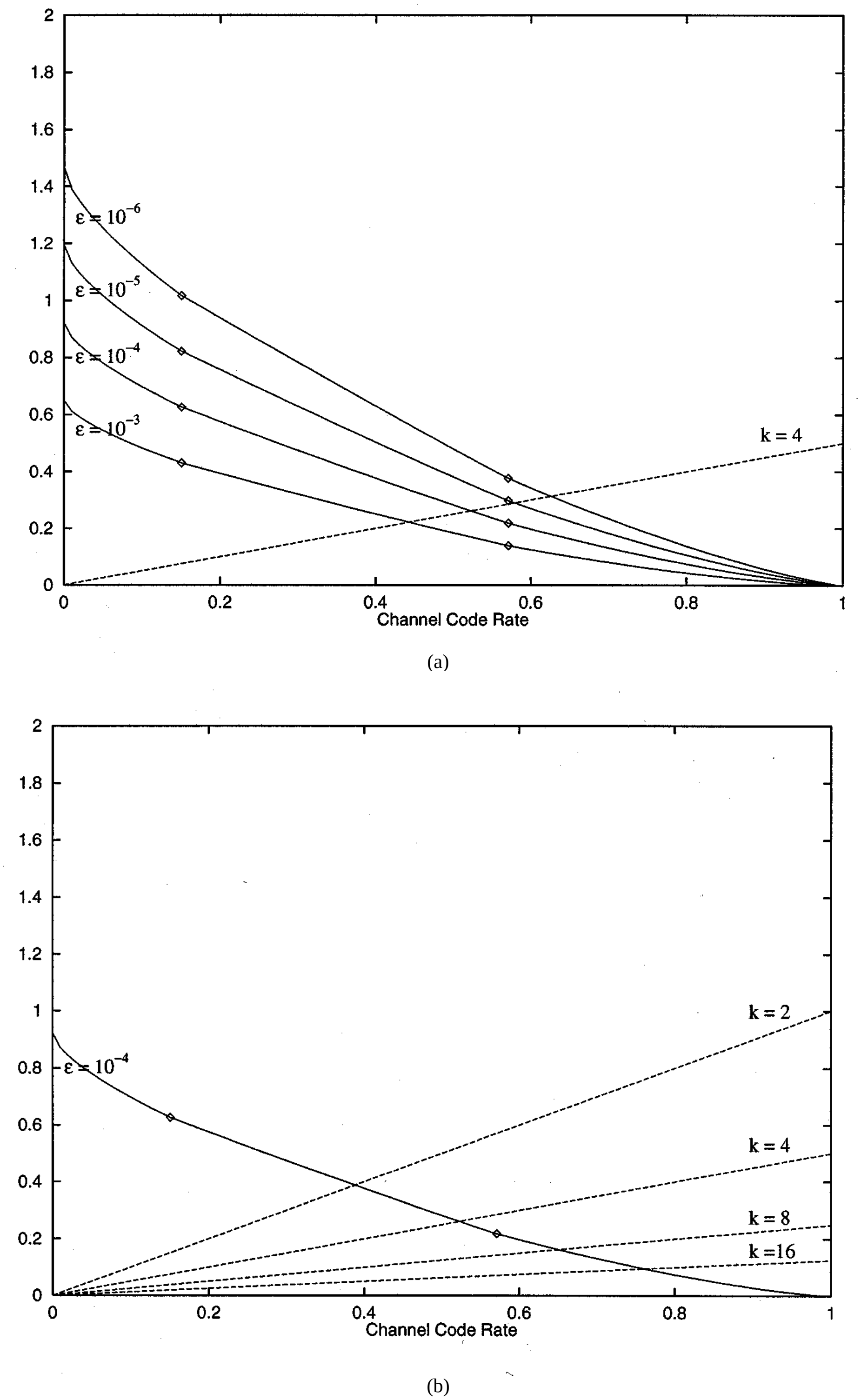


Fig. 3. A graphical solution to (10) for $E'_{\max}$ ($p = 2$, $q = 64$). The solid curves show $E'_{\max}(r)$ for different values of ϵ, and the dashed lines have slope p/k. The two dots on each error exponent curve correspond to r'_1 and r'_2. (a) Small bit-error probability. (b) Large vector dimension.

Observe that the sphere packing and Gilbert–Varshamov exponents can both be written as

$$\begin{aligned} E_X(\delta) &= \mathcal{D}_q(\delta/i\|\epsilon) \\ &= \frac{\delta}{i}\log_q\frac{\delta}{i} + \frac{\delta}{i}\log_q 1/\epsilon + \left(1-\frac{\delta}{i}\right)\log_q\left(1-\frac{\delta}{i}\right) \\ &\quad - \left(1-\frac{\delta}{i}\right)\log_q(1-\epsilon) \end{aligned} \tag{39}$$

where $i = 1$ when $X = \mathrm{sp}$, and $i = 2$ when $X = \mathrm{GV}$. Using (72) and (70), the expurgated exponent can be rewritten as

$$\begin{aligned} E_{\mathrm{ex}}(\delta) &= -\delta\log_q\left(2\sqrt{(1-\epsilon)\frac{\epsilon}{q-1}} + (q-2)\frac{\epsilon}{q-1}\right) \\ &= \delta\left[\frac{1}{2}\log_q 1/\epsilon - \log_q\left(2\sqrt{\frac{1-\epsilon}{q-1}} + (q-2)\frac{\sqrt{\epsilon}}{q-1}\right)\right]. \end{aligned} \tag{40}$$

Since δ is bounded, the dominant term on the left-hand side of (36) (as given in (39) and (40)) equals $(\delta/i)\log_q 1/\epsilon$ in all three cases, while the right-hand side is bounded between 0 and c, independent of ϵ. Hence, as ϵ approaches zero, for equality to hold in (36), δ has to approach zero at least as fast as $(\log_q 1/\epsilon)^{-1}$. On the other hand, the right-hand side of (36) approaches the finite constant c if $\delta \to 0$. Thus δ cannot converge to zero faster than $(\log_q 1/\epsilon)^{-1}$ for the left-hand side to stay bounded away from zero. We therefore conclude that the solution to (36) must be of the form

$$\delta_X = \frac{ic+\alpha_X}{\log_q 1/\epsilon} \tag{41}$$

where $\alpha_X \to 0$ as $\epsilon \to 0$, and $i=1$ when $X=\mathrm{sp}$, and $i=2$ when $X \in \{\mathrm{GV}, \mathrm{ex}\}$. To characterize δ_X more precisely, α_X has to be determined. In what follows, all $O(\cdot)$ terms go to zero as $\epsilon \to 0$.

Substituting (41) in (39), and applying power series expansions yields

$$\begin{aligned} E_X(\alpha_X) &= \frac{ic+\alpha_X}{i\log_q 1/\epsilon}\left[\log_q \frac{ic+\alpha_x}{i\log_q 1/\epsilon} + \log_q 1/\epsilon\right] \\ &\quad + \left(1-\frac{ic+\alpha_X}{i\log_q 1/\epsilon}\right) \\ &\quad \cdot \left[\log_q\left(1-\frac{ic+\alpha_X}{i\log_q 1/\epsilon}\right)+\log_q(1-\epsilon)\right] \\ &= c+(\alpha_x/i) - \frac{ic+\alpha_X}{i\log_q 1/\epsilon} \\ &\quad \cdot [\log_q\log_q 1/\epsilon - \log_q(c+(\alpha_x/i))] \\ &\quad - \left(1-\frac{ic+\alpha_X}{i\log_q 1/\epsilon}\right) \\ &\quad \cdot \left[\frac{ic+\alpha_X}{i\log_q 1/\epsilon}\log_q e + O\left(\frac{1}{\log_q^2 1/\epsilon}\right)+O(\epsilon)\right] \\ &= c+(\alpha_X/i) \\ &\quad - \frac{ic+\alpha_X}{i\log_q 1/\epsilon}[\log_q\log_q 1/\epsilon \\ &\quad - \log_q(c+(\alpha_X/i))+\log_q e] + O\left(\frac{1}{\log_q^2 1/\epsilon}\right) \end{aligned} \tag{42}$$

where $i=1$ for $X=\mathrm{sp}$, and $i=2$ for $X=\mathrm{GV}$.

The same steps applied to (40) result in

$$\begin{aligned} E_{\mathrm{ex}}(\alpha_{\mathrm{ex}}) &= \frac{2c+\alpha_{\mathrm{ex}}}{\log_q 1/\epsilon}\left[\frac{1}{2}\log_q 1/\epsilon\right. \\ &\quad \left. - \log_q\left(2\sqrt{\frac{1-\epsilon}{q-1}}+(q-2)\frac{\sqrt{\epsilon}}{q-1}\right)\right] \\ &= c+(\alpha_{\mathrm{ex}}/2) - \frac{2c+\alpha_{\mathrm{ex}}}{\log_q 1/\epsilon} \\ &\quad \cdot \left[\frac{1}{2}\log_q(q-1)-\log_q 2+O(\sqrt{\epsilon})\right] \\ &= c+(\alpha_{\mathrm{ex}}/2) - \frac{2c+\alpha_{\mathrm{ex}}}{2\log_q 1/\epsilon} \\ &\quad \cdot [\log_q(q-1)-\log_q 4] + O\left(\frac{1}{\log_q^2 1/\epsilon}\right). \end{aligned} \tag{43}$$

To obtain the right-hand side of (36) as a function of α_X, we write

$$\begin{aligned} r_X(\alpha_X) &= 1-\frac{ic+\alpha_X}{\log_q 1/\epsilon}\left(\log_q(q-1)-\log_q\frac{ic+\alpha_X}{\log_q 1/\epsilon}\right) \\ &\quad + \left(1-\frac{ic+\alpha_X}{\log_q 1/\epsilon}\right)\log_q\left(1-\frac{ic+\alpha_X}{\log_q 1/\epsilon}\right) \\ &= 1-\frac{ic+\alpha_X}{\log_q 1/\epsilon}(\log_q(q-1) \\ &\quad -\log_q(ic+\alpha_X)+\log_q\log_q 1/\epsilon) \\ &\quad -\left(1-\frac{ic+\alpha_X}{\log_q 1/\epsilon}\right)\frac{ic+\alpha_X}{\log_q 1/\epsilon}\log_q e+O\left(\frac{1}{\log_q^2 1/\epsilon}\right) \\ &= 1-\frac{ic+\alpha_X}{\log_q 1/\epsilon}(\log_q\log_q 1/\epsilon+\log_q(q-1) \\ &\quad -\log_q(ic+\alpha_X)+\log_q e)+O\left(\frac{1}{\log_q^2 1/\epsilon}\right). \end{aligned} \tag{44}$$

Next, we proceed to solve (36) for α_X. Comparing (36), (42)–(44), we conclude that

$$\alpha_X = O\left(\frac{\log_q\log_q 1/\epsilon}{\log_q 1/\epsilon}\right).$$

Based on this observation, the $\log_q(c+(\alpha_X/i))$ terms in (42) and (44) can be further expanded to obtain

$$\begin{aligned} E_X(\alpha_X) &= c+(\alpha_X/i) \\ &\quad -\frac{ic+\alpha_X}{i\log_q 1/\epsilon}\left[\log_q\log_q 1/\epsilon-\log_q c-\frac{\alpha_X\log_q e}{ic}\right. \\ &\quad \left. +O(\alpha_X^2)+\log_q e\right]+O\left(\frac{1}{\log_q^2 1/\epsilon}\right) \\ &= c+(\alpha_X/i) \\ &\quad -\frac{c}{\log_q 1/\epsilon}[\log_q\log_q 1/\epsilon-\log_q c+\log_q e] \\ &\quad -\frac{\alpha_X}{i\log_q 1/\epsilon}[\log_q\log_q 1/\epsilon-\log_q c]+O\left(\frac{1}{\log_q^2 1/\epsilon}\right) \\ &= c-\frac{c}{\log_q 1/\epsilon}[\log_q\log_q 1/\epsilon-\log_q c+\log_q e] \\ &\quad +(\alpha_X/i)\left(1-\frac{\log_q\log_q 1/\epsilon-\log_q c}{\log_q 1/\epsilon}\right) \\ &\quad +O\left(\frac{1}{\log_q^2 1/\epsilon}\right) \end{aligned} \tag{45}$$

and

$$\begin{aligned} r_X(\alpha_X) &= 1-\frac{ic+\alpha_X}{\log_q 1/\epsilon}\left(\log_q\log_q 1/\epsilon-\log_q(ic)\right. \\ &\quad \left. -\frac{\alpha_X\log_q e}{ic}+O(\alpha_X^2)+\log_q e(q-1)\right) \\ &\quad +O\left(\frac{1}{\log_q^2 1/\epsilon}\right) \\ &= 1-\frac{ic}{\log_q 1/\epsilon}(\log_q\log_q 1/\epsilon-\log_q(ic) \\ &\quad +\log_q e(q-1)) \end{aligned}$$

$$- \frac{\alpha_X}{\log_q 1/\epsilon}(\log_q \log_q 1/\epsilon + \log_q(q-1) - \log_q(ic)) + O\left(\frac{1}{\log_q^2 1/\epsilon}\right). \quad (46)$$

1) Sphere Packing and Gilbert-Varshamov Exponents: Substituting (45) and (46) into (36) and rearranging terms yields

$$\begin{aligned}0 &= i(E_X(\alpha_X) - cr_X(aX)) \\ &= \alpha_X[1 - ((1-ic)(\log_q \log_q 1/\epsilon - \log_q c) \\ &\quad - ic(\log_q(q-1) - \log_q i))(\log_q 1/\epsilon)^{-1}] \\ &\quad - \frac{ic}{\log_q 1/\epsilon}[(1-ic)(\log_q \log_q 1/\epsilon - \log_q c + \log_q e) \\ &\quad - ic(\log_q(q-1) - \log_q i)] + O\left(\frac{1}{\log_q^2 1/\epsilon}\right).\end{aligned}$$

Thus

$$\begin{aligned}\alpha_X &= [ic(1-ic)(\log_q \log_q 1/\epsilon - \log_q c + \log_q e) \\ &\quad - (ic)^2(\log_q(q-1) - \log_q i)] \\ &\quad \cdot [\log_q 1/\epsilon - (1-ic)(\log_q \log_q 1/\epsilon - \log_q c) \\ &\quad + ic(\log_q(q-1) - \log_q i)]^{-1} + O\left(\frac{1}{\log_q^2 1/\epsilon}\right).\end{aligned}$$

Substituting this in (46), gives

$$\begin{aligned}r_X &= 1 - [ic(\log_q \log_q 1/\epsilon - \log_q ic + \log_q e(q-1))] \\ &\quad \cdot [\log_q 1/\epsilon - (1-ic)(\log_q \log_q 1/\epsilon - \log_q c) \\ &\quad + ic(\log_q(q-1) - \log_q i)]^{-1} + O\left(\frac{1}{\log_q^2 1/\epsilon}\right).\end{aligned}$$

Now, using (38) the bounds on the optimal rate are summarized in the following two lemmas.

Lemma 5: For any p and k, and sufficiently small $\epsilon > 0$

$$\begin{aligned}r^*_{\text{sp}} &= 1 - \left[\frac{p}{k}\left(\log_q \log_q 1/\epsilon - \log_q \frac{p}{k} + \log_q e(q-1)\right)\right] \\ &\quad \cdot \left[\log_q 1/\epsilon - (1-\frac{p}{k})(\log_q \log_q 1/\epsilon - \log_q \frac{p}{k}) \right. \\ &\quad \left. + \frac{p}{k}\log_q(q-1)\right]^{-1} + O\left(\frac{1}{\log_q^2 1/\epsilon}\right) + o(1)\end{aligned}$$

where the $O(\frac{1}{\log_q^2 1/\epsilon})$ term goes to zero as $\epsilon \to 0$ for any R, and the $o(1)$ term goes to zero as $R \to \infty$.

Lemma 6: For any p and k, and sufficiently small $\epsilon > 0$

$$\begin{aligned}r^*_{\text{GV}} &= 1 - \left[\frac{2p}{k}\left(\log_q \log_q 1/\epsilon - \log_q \frac{2p}{k} + \log_q e(q-1)\right)\right] \\ &\quad \cdot \left[\log_q 1/\epsilon - (1-\frac{2p}{k})(\log_q \log_q 1/\epsilon - \log_q \frac{p}{k}) \right. \\ &\quad \left. + \frac{2p}{k}(\log_q(q-1) - \log_q 2)\right]^{-1} + O\left(\frac{1}{\log_q^2 1/\epsilon}\right) + o(1)\end{aligned}$$

where the $O(\frac{1}{\log_q^2 1/\epsilon})$ term goes to zero as $\epsilon \to 0$ for any R, and the $o(1)$ term goes to zero as $R \to \infty$.

Combining Lemmas 5 and 6 gives the desired bounds for optimal rate allocation for codes attaining the Gilbert–Varshamov bound, as summarized in the following theorem.

Theorem 1: Consider the cascade of a good k-dimensional vector quantizer, a q-ary linear block channel coder that meets the Gilbert–Varshamov bound or the Tsfasman–Vlăduţ–Zink bound, and a q-ary symmetric channel with symbol error probability ϵ. The channel code rate r^* that minimizes the pth-power distortion (averaged over all index assignments) satisfies

$$\begin{aligned}&\left[\frac{p}{k}\left(\log_q \log_q 1/\epsilon - \log_q \frac{p}{k} + \log_q e(q-1)\right)\right] \\ &\cdot \left[\log_q 1/\epsilon - (1-\frac{p}{k})(\log_q \log_q 1/\epsilon - \log_q \frac{p}{k}) \right. \\ &\left. + \frac{p}{k}(\log_q(q-1)\right]^{-1} + O\left(\frac{1}{\log_q^2 1/\epsilon}\right) + o(1) \\ &\le 1 - r^* \le \\ &\left[\frac{2p}{k}\left(\log_q \log_q 1/\epsilon - \log_q \frac{2p}{k} + \log_q e(q-1)\right)\right] \\ &\cdot \left[\log_q 1/\epsilon - (1-2\frac{p}{k})(\log_q \log_q 1/\epsilon - \log_q \frac{p}{k} \right. \\ &\left. + \frac{2p}{k}\log_q(q-1) - \log_q 2)\right]^{-1} + O\left(\frac{1}{\log_q^2 1/\epsilon}\right) + o(1)\end{aligned}$$

where the $O(\frac{1}{\log_q^2 1/\epsilon})$ term goes to zero as $\epsilon \to 0$ for any transmission rate R, and the $o(1)$ term goes to zero as $R \to \infty$.

A crude comparison of the upper and lower bounds in Theorem 1 shows a factor of 2 difference in the asymptotically dominant ϵ-dependent term for channel codes that attain the Gilbert–Varshamov bound. The same phenomenon was observed in [2] for efficient binary channel codes. Next, we derive more precise bounds for efficient q-ary channel codes based on the expurgated error exponent, and we present an example comparing the optimal rate-allocation bounds for channel codes that achieve the reliability function of the channel and channel codes that attain the Gilbert–Varshamov bound.

2) Expurgated Exponent: Substituting (43) and (46) into (36) and rearranging terms yields

$$\begin{aligned}0 &= 2(E_{\text{ex}}(\alpha_{\text{ex}}) - cr_{\text{ex}}(\alpha_{\text{ex}})) \\ &= \alpha_{\text{ex}}\left(1 - \frac{1}{\log_q 1/\epsilon}[\log_q(q-1) - \log_q 4 \right. \\ &\quad \left. - 2c(\log_q \log_q 1/\epsilon + \log_q(q-1) - \log_q 2 - \log_q c)]\right) \\ &\quad - \frac{2c}{\log_q 1/\epsilon}\Big[\log_q(q-1) - \log_q 4 \\ &\quad - 2c(\log_q \log_q 1/\epsilon + \log_q(q-1) - \log_q 2 \\ &\quad - \log_q c + \log_q e)\Big] + O\left(\frac{1}{\log_q^2 1/\epsilon}\right).\end{aligned}$$

Thus

$$\begin{aligned}\alpha_{\text{ex}} &= 2c \cdot [\log_q(q-1) - \log_q 4 \\ &\quad - 2c(\log_q \log_q 1/\epsilon + \log_q(q-1) \\ &\quad - \log_q 2 - \log_q c + \log_q e)] \\ &\quad \cdot [\log_q 1/\epsilon - \log_q(q-1) + \log_q 4 \\ &\quad + 2c(\log_q \log_q 1/\epsilon + \log_q(q-1) - \log_q 2 - \log_q c)]^{-1} \\ &\quad + O\left(\frac{1}{\log_q^2 1/\epsilon}\right)\end{aligned}$$

which when substituted into (46), gives

$$r_{\rm ex} = 1 - 2c[(\log_q \log_q 1/\epsilon - \log_q 2c + \log_q e(q-1))] \cdot [\log_q 1/\epsilon + 2c(\log_q \log_q 1/\epsilon - \log_q 2c + \log_q (q-1)) - \log_q(q-1) + \log_q 4]^{-1} + O\left(\frac{1}{\log_q^2 1/\epsilon}\right).$$

Now, using (38), the bound on the optimal rate is summarized in the following lemma.

Lemma 7: For any p and k, and sufficiently small $\epsilon > 0$

$$r_{\rm ex}^* = 1 - \left[\frac{2p}{k}\left(\log_q \log_q 1/\epsilon - \log_q \frac{2p}{k} + \log_q e(q-1)\right)\right] \cdot \left[\log_q 1/\epsilon + \frac{2p}{k}(\log_q \log_q 1/\epsilon - \log_q \frac{2p}{k} + \log_q(q-1)) - \log_q(q-1) + \log_q 4\right]^{-1} + O\left(\frac{1}{\log_q^2 1/\epsilon}\right) + o(1)$$

where the $O(\frac{1}{\log_q^2 1/\epsilon})$ term goes to zero as $\epsilon \to 0$ for any R, and the $o(1)$ term goes to zero as $R \to \infty$.

Combining Lemmas 5 and 7 establishes bounds for the optimal rate allocation based on "random coding." These extend the results of [2] to q-ary channels.

Theorem 2: Consider the cascade of a good k-dimensional vector quantizer, an efficient q-ary linear block channel coder, and a q-ary symmetric channel with symbol error probability ϵ. The channel code rate r^* that minimizes the pth-power distortion (averaged over all index assignments) satisfies

$$\begin{aligned}&\left[\frac{p}{k}\left(\log_q \log_q 1/\epsilon - \log_q \frac{p}{k} + \log_q e(q-1)\right)\right] \\ &\cdot \left[\log_q 1/\epsilon - (1 - \frac{p}{k})(\log_q \log_q 1/\epsilon - \log_q \frac{p}{k}) + \frac{p}{k}\log_q(q-1)\right]^{-1} \\ &+ O\left(\frac{1}{\log_q^2 1/\epsilon}\right) + o(1) \\ &\le 1 - r^* \le \\ &\left[\frac{2p}{k}\left(\log_q \log_q 1/\epsilon - \log_q \frac{2p}{k} + \log_q e(q-1)\right)\right] \\ &\cdot \left[\log_q 1/\epsilon + 2\frac{p}{k}(\log_q \log_q 1/\epsilon - \log_q 2\frac{p}{k} + \log_q(q-1)) - \log_q(q-1) + \log_q 4\right]^{-1} \\ &+ O\left(\frac{1}{\log_q^2 1/\epsilon}\right) + o(1)\end{aligned}$$

where the $O(\frac{1}{\log_q^2 1/\epsilon})$ term goes to zero as $\epsilon \to 0$ for any transmission rate R, and the $o(1)$ term goes to zero as $R \to \infty$.

The dominant ϵ-dependent terms in the upper and lower bounds of Theorem 2 differ only by a factor of 2. This was derived in [2] for efficient binary channel codes and observed in Theorem 1 for general q-ary channel codes that meet the Gilbert–Varshamov bound. The upper bounds of Theorems 1 and 2 are identical, since they both derive from the sphere-packing error exponent. The lower bounds, however, are identical only to the precision afforded in [2]. The main rationale for our more complex formulas is to pinpoint the difference between the two categories of channel codes (i.e., those that achieve the reliability function of the channel and those that attain the Gilbert–Varshamov bound). Fig. 4 compares the rate-allocation bounds of Theorems 1 and 2 for $q = 64$. The choice of alphabet size is motivated by the requirement on algebraic-geometry codes needed in order to lie above the Gilbert–Varshamov bound, but the bounds for other values of q display similar behavior. The solid curves correspond to $r_{\rm sp}$, the upper bound in both Theorems 1 and 2. The dotted curves represent $r_{\max}$, the lower bound in Theorem 2. The dashed curves show $r'_{\max}$, the lower bound in Theorem 1. Thus the dark shaded regions represent the uncertainty of the bounds of Theorem 2 and the corresponding light shaded regions show the discrepancy between the bounds of Theorems 1 and 2. The curves plotted do not omit any $O(\cdot)$ terms.

Note that by substituting

$$\alpha_X = O\left(\frac{\log_q \log_q 1/\epsilon}{\log_q 1/\epsilon}\right)$$

directly into (41), simpler expressions for r_X can be obtained at the expense of precision. Figs. 5 and 6 provide additional motivation for the more intricate analysis. The curves obtained numerically without omitting any $O(\cdot)$ terms are denoted by r_X. The approximations (omitting the $O(\cdot)$ terms) given in [2] are denoted by $r_X^{(HZ)}$, and those given in Lemmas 5 and 7 by $r_X^{(MZ)}$. The expressions we used for $r_X^{(HZ)}$ (see [2, Theorem 1]) are

$$r_{\rm sp}^{(HZ)} = 1 - \frac{p}{k}\,\frac{\log_2 \log_2 1/\epsilon}{\log_2 1/\epsilon}$$

and

$$r_{\rm ex}^{(HZ)} = 1 - \frac{2p}{k}\,\frac{\log_2 \log_2 1/\epsilon}{\log_2 1/\epsilon}.$$

The illustrations in Figs. 5 and 6 are for $q = 2$, $p = 2$, and $k = 8$. We note that the curve for $r_{\rm sp}$ is closely approximated by Lemma 5. The situation is similar for $r_{\rm GV}$ (not plotted here).

B. Large Source Vector Dimension

In this section, we analyze the optimal rate allocation for large source vector dimensions. As noted earlier, it suffices to examine the sphere packing and Gilbert–Varshamov exponents for k sufficiently large. Solutions of (10) based on these two exponents provide upper and lower bounds for the optimal-rate allocation for systems using asymptotically good codes that meet the Gilbert–Varshamov bound. The upper and lower bounds on the optimal rate coincide for efficient channel codes, since $E_{\rm sp}$ and $E_{\rm rc}$ are identical for k large. Hence, an exact asymptotic solution of the rate allocation problem is possible, if the channel codes used obey Lemma 2. Thus in what follows we concentrate on the Gilbert–Varshamov case.

We combine (23) and (30) to obtain

$$E_X(r) = \mathcal{D}_q\left(\frac{1}{i}\mathcal{H}_q^{-1}(1-r)\|\epsilon\right), \qquad r \in (0, C_q(i\epsilon)) \quad (47)$$

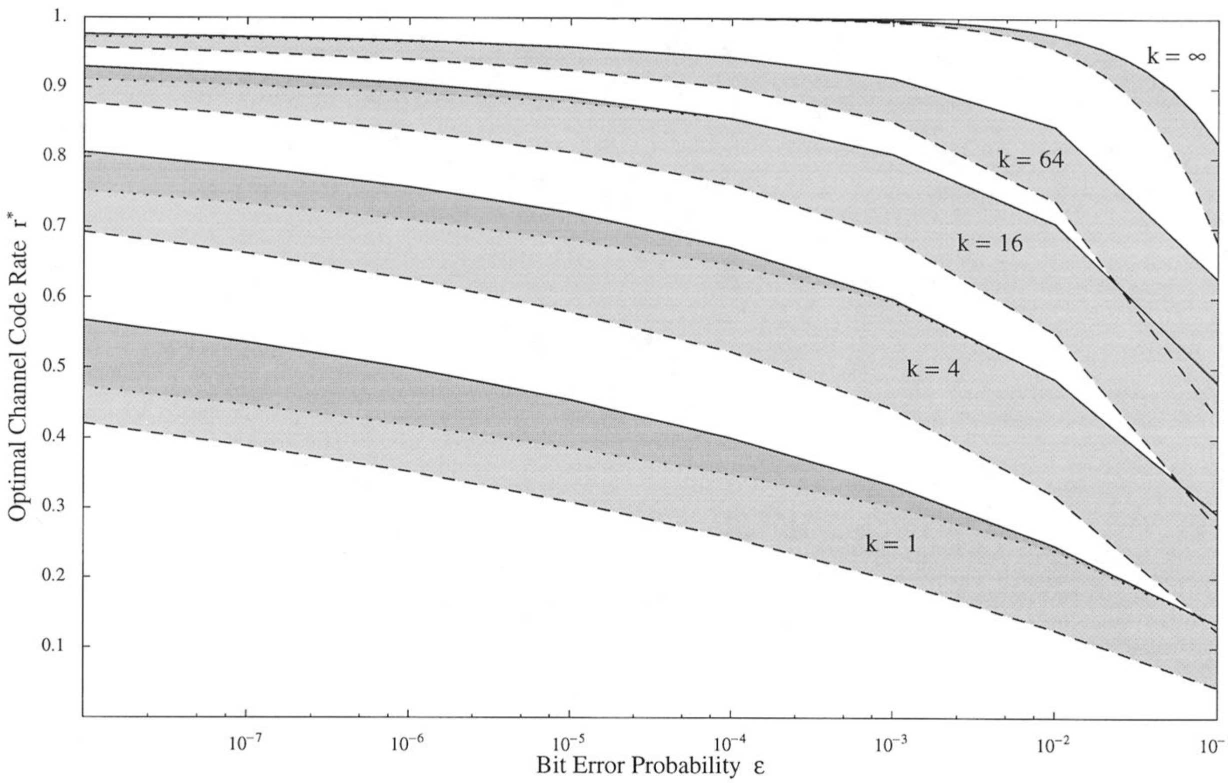


Fig. 4. A comparison of the upper and lower bounds on the optimal channel code rate as given in Theorems 1 and 2 for $p = 2, q = 64$. For $k = 1, 4, 16, 64$, the solid curves show $r_{\rm sp}$, the dotted curves $r_{\max}$, and the dashed curves $r'_{\max}$, respectively. For $k = \infty$, $r_{\rm sp} = r_{\max} = C_q(\epsilon)$ and $r'_{\max} = C_q(2\epsilon)$ are displayed. The corresponding triplets of bounds are indicated by shading. The lightly shaded regions illustrate the rate loss due to suboptimality.

where $C_q(i\epsilon) = 1 - \mathcal{H}_q(i\epsilon)$, and $i = 1$ when $X = \rm sp$, and $i = 2$ when $X = \rm GV$. First, the value of r_X satisfying

$$\mathcal{D}_q\left(\frac{1}{i}\mathcal{H}_q^{-1}(1 - r_X)\bigg\|\,\epsilon\right) = \frac{p}{k}r_X \tag{48}$$

is found, and then a solution to (10) is obtained by setting

$$r_X^* = r_X + o(1) \tag{49}$$

where $o(1) \to 0$ as $R \to \infty$.

The error exponents are decreasing functions of the rate and vanish for rates above $C_q(i\epsilon)$. As k increases, the right-hand side of (48) decreases, so $r_X \to C_q(i\epsilon)$ as $k \to \infty$. Thus for large k, $\mathcal{H}_q^{-1}(1 - r_X)$ can be approximated by its Taylor series around $1 - C_q(i\epsilon)$ as

$$\mathcal{H}_q^{-1}(1 - r_X) = \mathcal{H}_q^{-1}(1 - C_q(i\epsilon)) + \frac{C_q(i\epsilon) - r_X}{\mathcal{H}'_q(\mathcal{H}_q^{-1}(1 - C_q(i\epsilon)))} + O((C_q(i\epsilon) - r_X)^2) \tag{50}$$

where $\mathcal{H}'_q$ is the first derivative of the q-ary entropy function, given in (15).

Let $(x_i)_k \triangleq C_q(i\epsilon) - r_X$ (note that r_X depends on k via (48)). Then $(x_i)_k \ge 0$, and $(x_i)_k \to 0$ as $k \to \infty$. Thus using

$$\mathcal{H}_q^{-1}(1 - C_q(i\epsilon)) = \mathcal{H}_q^{-1}(\mathcal{H}_q(i\epsilon)) = i\epsilon$$

we can rewrite (50) as

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

which approaches ϵ, as $k \to \infty$. Applying (20), gives

$$\begin{aligned}\mathcal{D}_q&\left(\frac{1}{i}\mathcal{H}_q^{-1}(1 - r_X)\bigg\|\,\epsilon\right)\\ &= -\frac{\mathcal{H}''_q(\epsilon)}{2}\left(\frac{(x_i)_k}{i\mathcal{H}'_q(i\epsilon)} + O((x_i)_k^2)\right)^2 + O((x_i)_k^3)\\ &= -\frac{\mathcal{H}''_q(\epsilon)(x_i)_k^2}{2i^2[\mathcal{H}'_q(i\epsilon)]^2} + O((x_i)_k^3)\end{aligned} \tag{51}$$

where $\mathcal{H}''_q$ is the second derivative of the q-ary entropy function, given in (16). Let

$$\gamma_i \triangleq \frac{2p[i\mathcal{H}'_q(i\epsilon)]^2}{-\mathcal{H}''_q(\epsilon)}.$$

Substituting (51) and $r_X = C_q(i\epsilon) - (x_i)_k$ in (48), yields

$$(\boldsymbol{x}_i)_k^2 = \frac{\gamma_i}{k}(C_q(i\epsilon) - (x_i)_k) + O((x_i)_k^3). \tag{52}$$

The nonnegative root of the quadratic in (52) is

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

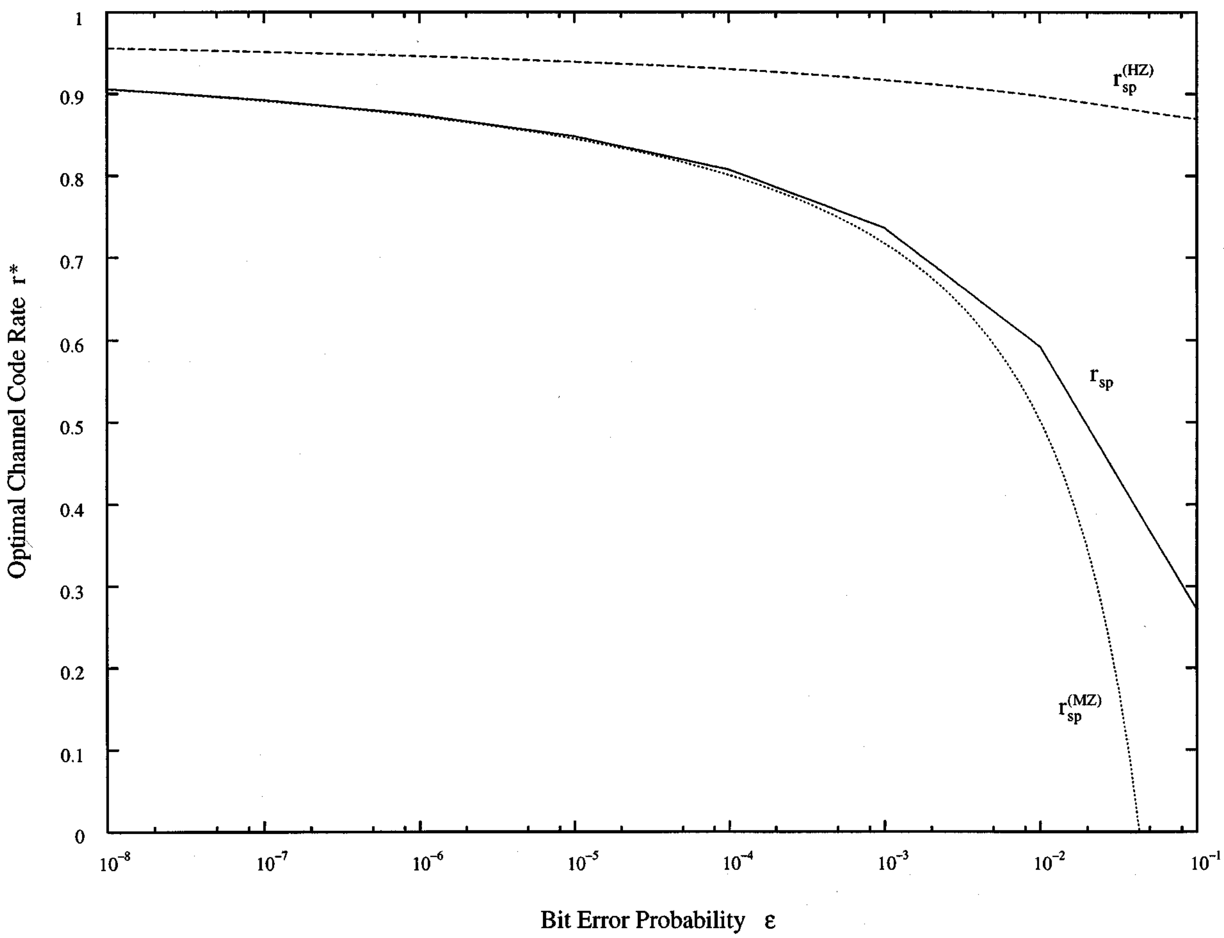


Fig. 5. Approximations to $r_{\rm sp}$ given in [2, Theorem 1] (dashed curve $r_{\rm sp}^{(HZ)}$), and by our Lemma 5 (dotted curve $r_{\rm sp}^{(MZ)}$). The solid curve $r_{\rm sp}$ was obtained by numerical solution of (10) for $q = 2, p = 2$, and $k = 8$.

Then, by (49) we have

$$r_X^* = C_q(i\epsilon) - \frac{i\mathcal{H}_q(i\epsilon)'}{\sqrt{k}} \left(\frac{2pC_q(i\epsilon)}{-\mathcal{H}_q''(\epsilon)}\right)^{\frac{1}{2}} + O\left(\frac{1}{k}\right) + o(1)$$

and we can state the following theorem.

Theorem 3: Consider the cascade of a good k-dimensional vector quantizer, a q-ary linear block channel code that attains the Gilbert–Varshamov bound or the Tsfasman–Vlăduţ–Zink bound, and a q-ary symmetric channel with symbol error probability ϵ. The channel code rate r^* that minimizes the pth-power distortion (averaged over all index assignments) satisfies

$$\begin{aligned} C_q(2\epsilon) - \frac{2\mathcal{H}_q'(2\epsilon)}{\sqrt{k}} \left(\frac{2pC_q(2\epsilon)}{-\mathcal{H}_q''(\epsilon)}\right)^{\frac{1}{2}} + O\left(\frac{1}{k}\right) + o(1) \\ \le r^* \le C_q(\epsilon) - \frac{\mathcal{H}_q'(\epsilon)}{\sqrt{k}} \left(\frac{2pC_q(\epsilon)}{-\mathcal{H}_q''(\epsilon)}\right)^{\frac{1}{2}} + O\left(\frac{1}{k}\right) + o(1) \end{aligned}$$

where the $O(\frac{1}{k})$ terms approach zero as $k \to \infty$ for any transmission rate R, the $o(1)$ terms approach zero as $R \to \infty$, $\mathcal{H}_q$ is the q-ary entropy function, and $C_q(x)$ is the capacity of a q-ary symmetric channel with symbol error probability x.

Fig. 7 illustrates the upper and lower bounds of Theorem 3. As before, the solid curves represent the upper bound $r_{\rm sp}$, and the dashed curves show the lower bound $r'_{\rm max}$. For the sake of comparison we also plotted the corresponding lower bounds for efficient channel codes using dotted curves. As shown in [2] for the binary case, the dotted and solid curves converge for k sufficiently large. The light shading illustrates how the bounds for codes attaining the Gilbert–Varshamov bound compare to the bounds obtained for efficient codes. The curves plotted do not omit any $O(\cdot)$ terms.

VI. Conclusion

To determine the optimal tradeoff between source and channel coding for certain structured linear block channel codes, we have derived upper and lower bounds on the channel code rate that minimizes the pth-power distortion of a k-dimensional vector quantizer cascaded with a linear block channel coder on a q-ary symmetric channel. We have presented bounds based on the Gilbert–Varshamov and Tsfasman–Vlăduţ–Zink bounds as well as random coding arguments for q-ary alphabets. Comparisons of the two types of results were also given.

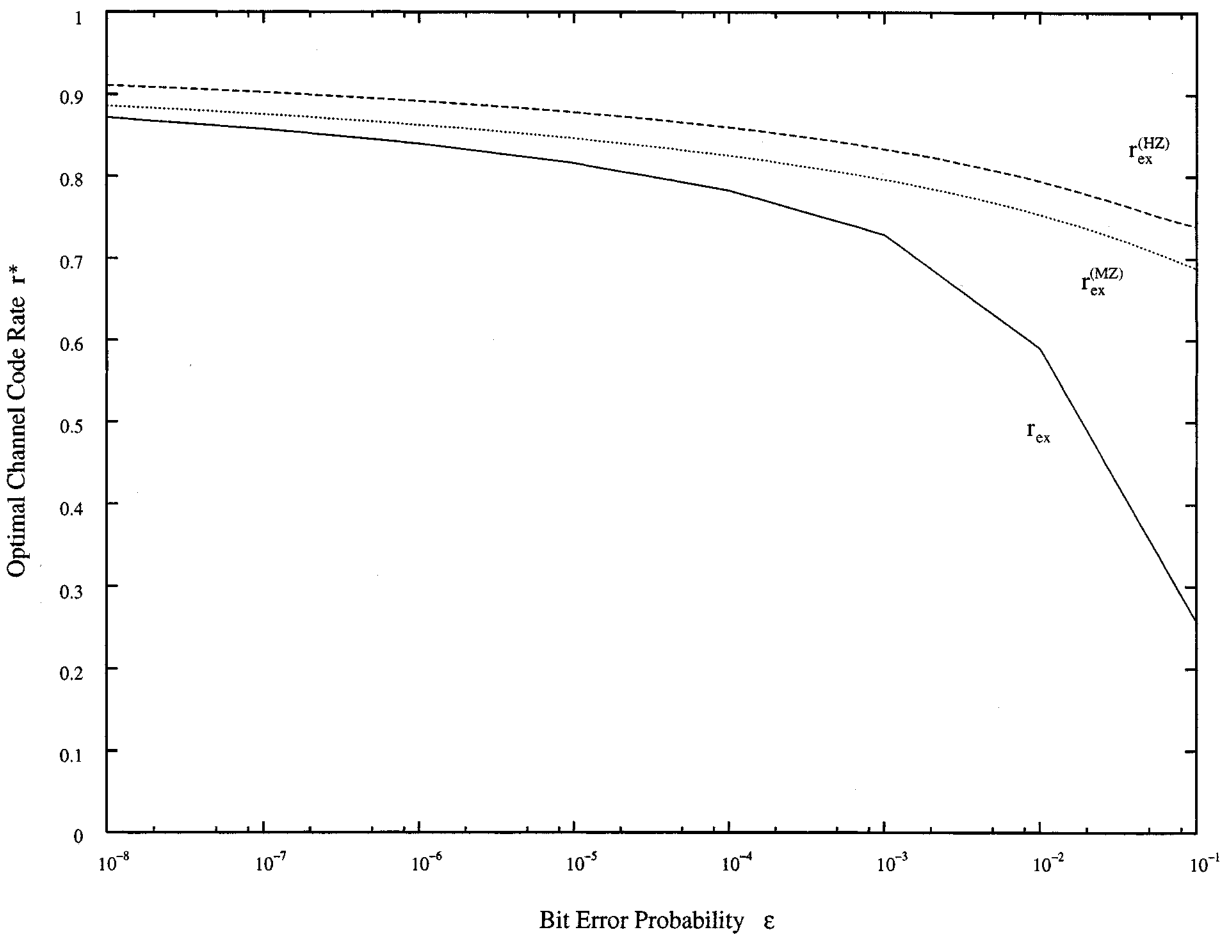


Fig. 6. Approximations to $r_{\rm ex}$ given in [2, Theorem 1] (dashed curve $r_{\rm ex}^{(HZ)}$), and by our Lemma 7 (dotted curve $r_{\rm ex}^{(MZ)}$). The solid curve $r_{\rm ex}$ was obtained by numerical solution of (10) for $q = 2$, $p = 2$, and $k = 8$.

APPENDIX
PROOF OF PROPOSITION 1

We use definitions and notation from [15], but we scale the error exponents by a factor of $\ln q$. (Note: There have been recent improvements in error exponent bounds [18] but these are not needed in our analysis). We denote by $\boldsymbol{Q} = (Q(0), Q(1), \ldots, Q(q-1))$ the input distribution, and by $P(i|j)$, $i, j \in \{0, 1, \ldots, q-1\}$ the transition probabilities of a q-ary channel.

Random Coding and Sphere Packing Exponents

Let us slightly reformulate some definitions from [15, pp. 144, 157]

$$F_0(\rho, \boldsymbol{Q}) \triangleq \sum_{i=0}^{q-1} \left(\sum_{j=0}^{q-1} Q(j) P(i|j)^{1/(1+\rho)} \right)^{1+\rho} \tag{53}$$

$$E_0(\rho, \boldsymbol{Q}) \triangleq -\log_q F_0(\rho, \boldsymbol{Q}) \tag{54}$$

$$\hat{E}_0(\rho) \triangleq -\rho r + \max_{\boldsymbol{Q}} E_0(\rho, \boldsymbol{Q}) \tag{55}$$

$$E_{\rm rc}(r) \triangleq \max_{0 \le \rho \le 1} \hat{E}_0(\rho) \tag{56}$$

$$E_{\rm sp}(r) \triangleq \sup_{\rho > 0} \hat{E}_0(\rho). \tag{57}$$

Since the expressions for $E_{\rm sp}$ and $E_{\rm rc}$ differ only in the range of ρ, much of our forthcoming derivation is common to both. Clearly,

$$\operatorname{argmax}_{\boldsymbol{Q}} E_0(\rho, \boldsymbol{Q}) = \operatorname{argmin}_{\boldsymbol{Q}} F_0(\rho, \boldsymbol{Q}).$$

For a q-ary symmetric channel (see (2))

$$F_0(\rho, \boldsymbol{Q}) = \sum_{i=0}^{q-1} \Bigg[Q(i)(1-\epsilon)^{1/(1+\rho)} + (1 - Q(i)) \left(\frac{\epsilon}{q-1} \right)^{1/(1+\rho)} \Bigg]^{1+\rho}.$$

Thus the Jacobian is a diagonal matrix

$$\left[\frac{\partial^2}{\partial Q(i) \partial Q(j)} F_0(\rho, \boldsymbol{Q}) \right] = \operatorname{diag} \Bigg\{ \rho(1+\rho) \Bigg[Q(i))(1-\epsilon)^{1/(1+\rho)}$$

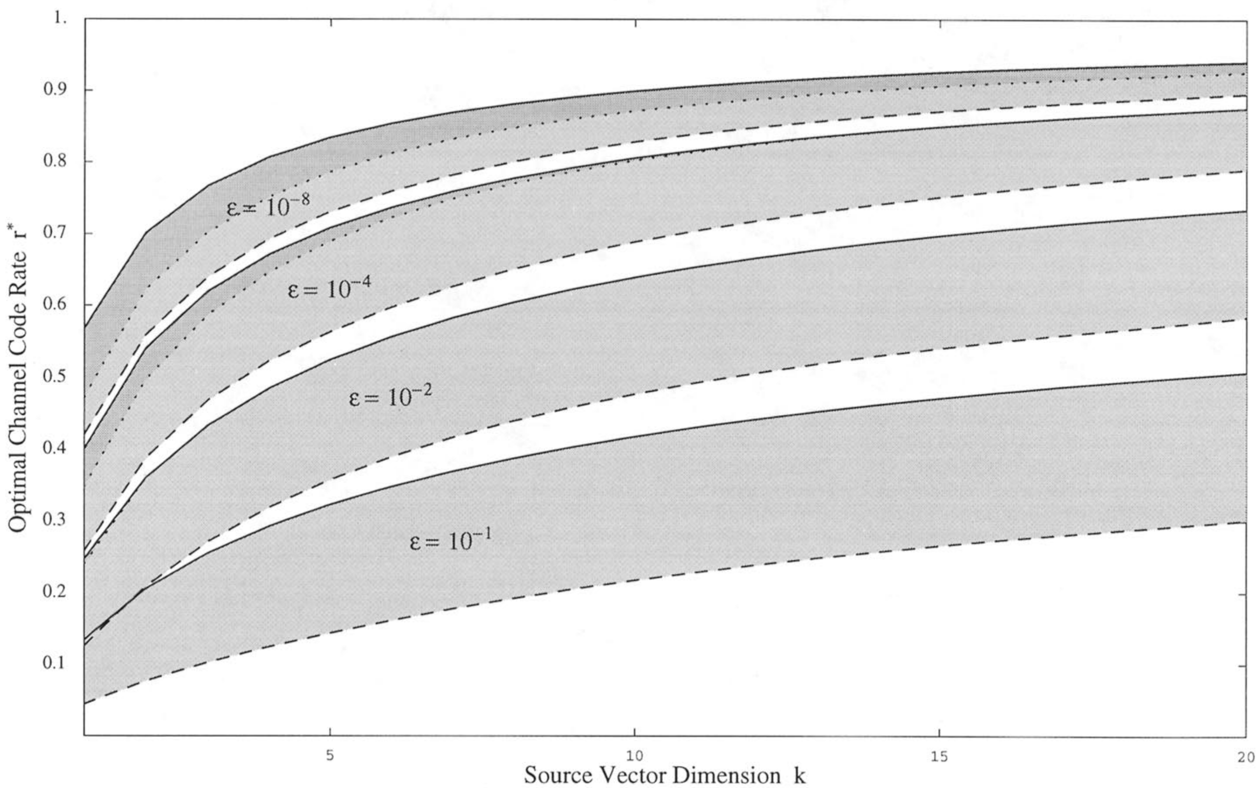


Fig. 7. Upper and lower bounds on the optimal channel code rate as given in Theorem 3 for $p = 2$, $q = 64$ are shown by solid (upper bound, $r_{\rm sp}$) and dashed (lower bound, $r'_{\max}$) curves for $\epsilon = 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}$. For comparison, lower bounds ($r_{\max}$) corresponding to efficient channel codes are plotted by dotted curves. The matching triplets of bounds for the same value of ϵ are indicated by shading. The lightly shaded regions illustrate the rate loss due to suboptimality.

$$+ (1 - Q(i)) \left(\frac{\epsilon}{q-1}\right)^{1/(1+\rho)} \Bigg]^{\rho-1}$$
$$\times \left((1-\epsilon)^{1/(1+\rho)} - \left(\frac{\epsilon}{q-1}\right)^{1/(1+\rho)} \right)^2 \Bigg\}$$

which is positive-definite for $\epsilon \in (0, 1 - q^{-1})$. Hence, setting

$$\frac{\partial}{\partial Q(j)} (F_0(\rho, \boldsymbol{Q}) - \lambda \sum_{i=0}^{q-1} Q(i)) = 0$$

yields a minimum. By symmetry, the minimizing distribution is uniform (this is also easily verified using [15, Theorem 5.6.5]. Then

$$\begin{aligned}\hat{E}_0(\rho) &= -\rho r - \Bigg\{ \log_q q \\ &\quad \cdot \left[\frac{(1-\epsilon)^{1/(1+\rho)} + (q-1)(\epsilon/((q-1))^{1/(1+\rho)}}{q} \right]^{1+\rho} \Bigg\} \\ &= \rho(1-r) - (1+\rho) \\ &\quad \cdot \log_q [(1-\epsilon)^{1/(1+\rho)} + (q-1)(\epsilon/(q-1)^{1/(1+\rho)}]\end{aligned} \tag{58}$$

and thus

$$\begin{aligned}\frac{d}{d\rho}\hat{E}_0(\rho) &= 1 - r - \log_q \Big[(1-\epsilon)^{1/(1+\rho)} \\ &\quad + (q-1)(\epsilon/(q-1))^{1/(1+\rho)} \Big] \\ &\quad + \Big[(1-\epsilon)^{1/(1+\rho)} \log_q (1-\epsilon)^{1/(1+\rho)} \\ &\quad + (q-1)\left(\frac{\epsilon}{q-1}\right)^{1/(1+\rho)} \log_q \left(\frac{\epsilon}{q-1}\right)^{1/(1+\rho)} \Big] \\ &\quad \cdot \Big[(1-\epsilon)^{1/(1+\rho)} + (q-1)(\epsilon/(q-1))^{1/(1+\rho)} \Big]^{-1} \\ &= 1 - r - \mathcal{H}_q(\delta)\end{aligned} \tag{59}$$

where

$$\delta = \frac{(q-1)(\frac{\epsilon}{q-1})^{1/(1+\rho)}}{(1-\epsilon)^{1/(1+\rho)} + (q-1)(\epsilon/(q-1))^{1/(1+\rho)}}. \tag{60}$$

Then, since

$$\begin{aligned}\frac{\partial \delta}{\partial \rho} &= -\frac{1}{(1+\rho)^2} \Bigg\{ (q-1)\left(\frac{\epsilon}{q-1}\right)^{1/(1+\rho)} \ln\left(\frac{\epsilon}{q-1}\right) \\ &\quad \cdot [(1-\epsilon)^{1/(1+\rho)} + (q-1)(\epsilon/(q-1))^{1/(1+\rho)}] \\ &\quad - (q-1)\left(\frac{\epsilon}{q-1}\right)^{1/(1+\rho)} \Bigg[(1-\epsilon)^{1/(1+\rho)} \ln(1-\epsilon) \\ &\quad + (q-1)\left(\frac{\epsilon}{q-1}\right)^{1/(1+\rho)} \ln\left(\frac{\epsilon}{q-1}\right) \Bigg] \Bigg\} \\ &\quad \times [(1-\epsilon)^{1/(1+\rho)} + (q-1)(\epsilon/(q-1)^{1/(1+\rho)}]^{-2} \\ &= \frac{\delta(1-\delta)}{1+\rho} (\log_q(q-1) - \log_q \delta + \log_q(1-\delta)) \ln q \\ &= \frac{\delta(1-\delta)}{1+\rho} \mathcal{H}'_q(\delta) \ln q\end{aligned}$$

we obtain

$$\begin{aligned}\frac{d^2}{d\rho^2}\hat{E}_0(\rho) &= -\mathcal{H}_q'(\delta)\frac{\partial\delta}{\partial\rho}\\ &= -\frac{\delta(1-\delta)}{1+\rho}(\mathcal{H}_q'(\delta))^2\ln q\\ &< 0\end{aligned}$$

for all values of δ corresponding to $\rho \geq 0$. Thus the stationary point ρ^*, found by setting $d\hat{E}_0/d\rho = 0$, is a maximum. Instead of solving explicitly for ρ^*, we obtain a parametric expression for the error exponents in terms of $\delta^* = \delta|_{\rho=\rho^*}$. From (59), we have

$$r = 1 - \mathcal{H}_q(\delta^*) \tag{61}$$

which when substituted into (58) gives

$$\begin{aligned}\hat{E}_0(\rho^*) &= \rho^*\mathcal{H}_q(\delta^*) - (1+\rho^*)\\ &\quad\cdot \log_q[(1-\epsilon)^{1/(1+\rho^*)} + (q-1)(\epsilon/(q-1))^{1/(1+\rho^*)}]\\ &= -\mathcal{H}_q(\delta^*) + (1+\rho^*)\\ &\quad\cdot\{\delta^*\log_q(q-1) - \delta^*\log_q\delta^* - (1-\delta^*)\log_q(1-\delta^*)\\ &\quad -\log_q[(1-\epsilon)^{1/(1+\rho^*)} + (q-1)(\epsilon/(q-1))^{1/(1+\rho^*)}]\}\\ &= -\mathcal{H}_q(\delta^*) - \delta^*\log_q\left(\frac{\delta^*}{q-1}\Big[(1-\epsilon)^{1/(1+\rho^*)}\right.\\ &\qquad\left. + (q-1)(\epsilon/(q-1))^{1/(1+\rho^*)}\Big]\right)^{1+\rho^*}\\ &\quad -(1-\delta^*)\log_q\left((1-\delta^*)\Big[(1-\epsilon)^{1/(1+\rho^*)}\right.\\ &\qquad\left. + (q-1)(\epsilon/(q-1))^{1/(1+\rho^*)}\Big]\right)^{1+\rho^*}\\ &= \delta^*\log_q\frac{\delta^*}{q-1} + (1-\delta^*)\log_q(1-\delta^*)\\ &\quad -\delta^*\log_q\frac{\epsilon}{q-1} - (1-\delta^*)\log_q(1-\epsilon) \qquad (62)\\ &= \mathcal{D}_q(\delta^*\|\epsilon)\\ &= \mathcal{D}_q(\mathcal{H}_q^{-1}(1-r)\|\epsilon) \qquad (63)\end{aligned}$$

where (62) follows by (60), and (61) was used in the last equality. Since

$$\delta|_{\rho=0} = \epsilon, \delta|_{\rho=1} = \frac{\sqrt{\epsilon(q-1)}}{\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)}} \text{ and } \delta|_{\rho\to\infty} = 1-q^{-1}$$

(63) gives the sphere-packing exponent for $r \in (0, C_q)$, and the random coding exponent for $r \in [r_2, C_q)$, where

$$r_2 = 1 - \mathcal{H}_q\left(\frac{\sqrt{\epsilon(q-1)}}{\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)}}\right). \tag{64}$$

As shown in [15], $\rho = 1$ maximizes $\hat{E}_0(\rho)$ for rates less than r_2. Hence, for $r \in (0, r_2]$

$$\begin{aligned}E_{\mathrm{rc}}(r) &= \hat{E}_0(1)\\ &= -r + 1 - 2\log_q[\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)}]\\ &= C_q^{(1/2)} - r,\end{aligned}$$

which completes the proof of (23) and (24). □

Expurgated Exponent

Let us define (see [15, p. 153])

$$F_x(\rho, \boldsymbol{Q}) \triangleq \sum_{i=0}^{q-1}\sum_{j=0}^{q-1} Q(i)Q(j)\left[\sum_{l=0}^{q-1}\sqrt{P(l|i)P(l|j)}\right]^{1/\rho} \tag{65}$$

$$E_x(\rho, \boldsymbol{Q}) \triangleq -\rho\log_q F_x(\rho, \boldsymbol{Q}) \tag{66}$$

$$\hat{E}_x(\rho) \triangleq -\rho r + \max_{\boldsymbol{Q}} E_x(\rho, \boldsymbol{Q}) \tag{67}$$

$$E_{\mathrm{ex}}(r) \triangleq \sup_{\rho\geq 1}\hat{E}_x(\rho). \tag{68}$$

Again

$$\mathrm{argmax}_{\boldsymbol{Q}}\, E_0(\rho, \boldsymbol{Q}) = \mathrm{argmin}_{\boldsymbol{Q}}\, F_0(\rho, \boldsymbol{Q}).$$

For a q-ary symmetric channel (see (2))

$$\sum_{l=0}^{q-1}\sqrt{P(l|i)P(l|j)} = \begin{cases}1, & i = j\\ 2\sqrt{(1-\epsilon)\frac{\epsilon}{q-1}} + (q-2)\frac{\epsilon}{q-1}, & i \neq j.\end{cases} \tag{69}$$

Let us define

$$\omega \triangleq 2\sqrt{(1-\epsilon)\frac{\epsilon}{q-1}} + (q-2)\frac{\epsilon}{q-1}. \tag{70}$$

Alternatively, ω can be expressed in terms of $\mathcal{H}_q^{(1/2)}(\epsilon)$. Using

$$\begin{aligned}1+(q-1)\omega &= 1 + 2\sqrt{(1-\epsilon)\epsilon(q-1)} + (q-2)\epsilon\\ &= (\sqrt{1-\epsilon}+\sqrt{(q-1)\epsilon})^2\\ &= q^{\mathcal{H}_q^{(1/2)}(\epsilon)}\end{aligned} \tag{71}$$

we obtain

$$\omega = \frac{q^{\mathcal{H}_q^{(1/2)}(\epsilon)} - 1}{q-1}. \tag{72}$$

Using ω, (65) can be rewritten as

$$F_x(\rho, \boldsymbol{Q}) = \sum_{i=0}^{q-1} Q(i)^2 + \omega^{1/\rho}\left(1 - \sum_{i=0}^{q-1} Q(i)^2\right).$$

Thus the Jacobian is a diagonal matrix (in fact, it is the identity matrix scaled)

$$\left[\frac{\partial^2}{\partial Q(i)\partial Q(j)}F_x(\rho, \boldsymbol{Q})\right] = 2(1-\omega^{1/\rho})\boldsymbol{I}$$

which is positive-definite for $\epsilon \in (0, 1-q^{-1})$, since (72) implies that $\omega < 1$, unless $\epsilon = 1 - q^{-1}$. Hence, setting

$$\frac{\partial}{\partial Q(j)}(F_x(\rho, \boldsymbol{Q}) - \lambda\sum_{i=0}^{q-1} Q(i)) = 0$$

yields a minimum. By symmetry, the minimizing distribution is uniform. Thus

$$\begin{aligned}\hat{E}_x(\rho) &= -\rho r - \rho\log_q[q^{-1} + \omega^{1/\rho}(1-q^{-1})]\\ &= \rho(1 - r - \log_q[1+(q-1)\omega^{1/\rho}])\end{aligned} \tag{73}$$

which implies

$$\frac{d}{d\rho}\hat{E}_x(\rho) = 1 - r - \log_q[1+(q-1)\omega^{1/\rho}] + \rho\frac{1}{1+(q-1)\omega^{1/\rho}}(q-1)\omega^{1/\rho}\frac{1}{\rho^2}\log_q\omega] \quad (74)$$

$$\begin{aligned}\frac{d}{d\rho}\hat{E}_x(\rho) &= 1 - r + \frac{1}{1+(q-1)\omega^{1/\rho}}\log_q\left(\frac{1}{1+(q-1)\omega^{1/\rho}}\right) \\ &\quad + \frac{(q-1)\omega^{1/\rho}}{1+(q-1)\omega^{1/\rho}}\log_q\left(\frac{\omega^{1/\rho}}{1+(q-1)\omega^{1/\rho}}\right) \\ &= 1 - r - \mathcal{H}_q(\delta) \end{aligned} \quad (75)$$

where

$$\delta = \frac{(q-1)\omega^{1/\rho}}{1+(q-1)\omega^{1/\rho}}. \quad (76)$$

Now, since

$$\begin{aligned}\frac{\partial\delta}{\partial\rho} &= -\left(\frac{1}{\rho^2}\right)\Big[\Big((q-1)\omega^{1/\rho}\ln\omega\Big)\Big[1+(q-1)\omega^{1/\rho}\Big] \\ &\quad - (q-1)\omega^{1/\rho}\Big[(q-1)\omega^{1/\rho}\ln\omega\Big]\Big] \\ &\quad \cdot \Big[1+(q-1)\omega^{1/\rho}\Big]^{-2} \\ &= \frac{\delta(1-\delta)}{\rho}\mathcal{H}_q'(\delta)\ln q\end{aligned}$$

we obtain

$$\begin{aligned}\frac{d^2}{d\rho^2}\hat{E}_x(\rho) &= -\mathcal{H}_q'(\delta)\frac{\partial\delta}{\partial\rho} \\ &= -\frac{\delta(1-\delta)}{\rho}(\mathcal{H}_q'(\delta))^2 \ \ln \ q \\ &< 0\end{aligned}$$

for all values of δ corresponding to $\rho \geq 1$. Thus the stationary point ρ^*, found by setting $d\hat{E}_x/d\rho = 0$, is a maximum. Instead of solving explicitly for ρ^*, we obtain a parametric expression for the error exponent in terms of $\delta^* = \delta|_{\rho=\rho^*}$. From (75), we have

$$r = 1 - \mathcal{H}_q(\delta^*) \quad (77)$$

which when substituted into (73) gives

$$\begin{aligned}\hat{E}_x(\rho^*) &= \rho^*(\mathcal{H}_q(\delta^*) - \log_q[1+(q-1)\omega^{1/\rho^*}]) \\ &= \rho^*(\delta^*\log_q(q-1) - \delta^*\log_q\delta^* \\ &\quad - (1-\delta^*)\log_q(1-\delta^*) - \log_q[1+(q-1)\omega^{1/\rho^*}]) \\ &= -\delta^*\log_q\left(\frac{\delta^*}{q-1}[1+(q-1)\omega^{1/\rho^*}]\right)^{\rho^*} \\ &\quad - \rho^*(1-\delta^*)\log_q((1-\delta^*)[1+(q-1)\omega^{1/\rho^*}]) \quad (78) \\ &= -\delta^*\log_q\omega \\ &= -\mathcal{H}_q^{-1}(1-r)\log_q\omega \quad (79)\end{aligned}$$

where (78) follows by (76), and (79) follows by (77) and (71). Since

$$\delta|_{\rho=1} = 1 - \frac{1}{1+(q-1)\omega} = 1 - q^{-\mathcal{H}_q^{(1/2)}(\epsilon)}$$

and $\delta|_{\rho\to\infty} = 1 - q^{-1}$, (79) gives the expurgated exponent for $r \in (0, r_1]$, where

$$r_1 = 1 - \mathcal{H}_q(1 - q^{-\mathcal{H}_q^{(1/2)}(\epsilon)}). \quad (80)$$

For $r > r_1$, $\hat{E}_x(\rho)$ is a decreasing function of ρ, since $d\hat{E}_x/d\rho < 0$ in this case (see (75)). Hence, for $r > r_1$

$$\begin{aligned}E_{\rm ex}(r) &= \hat{E}_x(1) = 1 - r - \log_q[1+(q-1)\omega] \\ &= 1 - r - \mathcal{H}_q^{(1/2)}(\epsilon) = C_q^{(1/2)} - r.\end{aligned} \quad (81)$$

Combining (79) and (81), we obtain (25). □

Maximum of the Random Coding and Expurgated Exponents

It is easy to argue that both

$$E_{\rm rc}(r) \geq C_q^{(1/2)} - r \quad \text{and} \quad E_{\rm ex}(r) \geq C_q^{(1/2)} - r \quad (82)$$

for all rates $r \in (0, 1)$, since for every r, the value $\rho = 1$ corresponding to $C_q^{(1/2)} - r$ is in the range of maximization in both (56) and (68). Hence, provided that $0 \leq r_1 \leq r_2 \leq C_q^{(1/2)} \leq C_q$ holds, (26) is obtained. Consulting (80), the properties of the entropy function imply that $0 \leq r_1$. To see that $r_1 \leq r_2$, we rewrite (80) as

$$r_1 = 1 - \mathcal{H}_q\left(1 - \frac{1}{(\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)})^2}\right)$$

and (64) as

$$r_2 = 1 - \mathcal{H}_q\left(1 - \frac{\sqrt{1-\epsilon}}{\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)}}\right).$$

Then $r_1 \leq r_2$ is equivalent to

$$\begin{aligned}1 &\leq \sqrt{1-\epsilon}(\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)}) \\ &= 1 - \epsilon + \sqrt{\epsilon(1-\epsilon)(q-1)} \\ \epsilon &\leq (1-\epsilon)(q-1) \\ \epsilon &\leq 1 - q^{-1}.\end{aligned}$$

Next, we show $r_2 \leq C_q^{(1/2)}$. By the concavity of log, for any $\delta \in [0, 1]$

$$\begin{aligned}\mathcal{H}_q(\delta) &= -\delta\log_q\frac{\delta}{q-1} - (1-\delta)\log_q(1-\delta) \\ &\geq -\log_q\left(\frac{1}{q-1}\delta^2 + (1-\delta)^2\right).\end{aligned}$$

Substituting this with $\delta = \frac{\sqrt{\epsilon(q-1)}}{\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)}}$ in (64), we can upper-bound r_2 as

$$\begin{aligned}r_2 &\leq 1 + \log_q\Bigg(\frac{1}{q-1}\ \frac{\epsilon(q-1)}{(\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)})^2} \\ &\qquad + \frac{1-\epsilon}{(\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)})^2}\Bigg) \\ &= 1 - 2\log_q(\sqrt{1-\epsilon}+\sqrt{\epsilon(q-1)}) \\ &= C_q^{(1/2)}\end{aligned}$$

which is what we wanted to prove. The remaining inequality $C_q^{(1/2)} \leq C_q$ can be equivalently stated as $\mathcal{H}_q^{(1/2)}(\epsilon) \geq \mathcal{H}_q(\epsilon)$, which follows by Jensen's inequality. □

REFERENCES

[1] K. Zeger and V. Manzella, "Asymptotic bounds on optimal noisy channel quantization via random coding," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1926–1938, Nov. 1994.

[2] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1412–1424, Sept. 1997.

[3] B. Hochwald, "Tradeoff between source and channel coding on a Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 3044–3055, Nov. 1998.

[4] J. N. Pierce, "Limit distribution of the minimum distance of random linear codes," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 595–599, Oct. 1967.

[5] J. T. Coffey and R. M. Goodman, "Any code of which we cannot think is good," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1453–1461, Nov. 1990.

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

[7] M. A. Tsfasman, S. G. Vlădut, and T. Zink, "Modular curves, Shimura curves, Goppa codes, better than Varshamov-Gilbert bound," *Math. Nachr.*, vol. 109, pp. 21–28, 1982.

[8] G. L. Katsman, M. A. Tsfasman, and S. G. Vlăduţ, "Modular curves and codes with a polynomial construction," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 353–355, Mar. 1984.

[9] C. Voss and T. Høholdt, "An explicit construction of a sequence of codes attaining the Tsfasman–Vlăduţ–Zink bound the first steps," *IEEE Trans. Inform. Theory*, vol. 43, pp. 128–135, Jan. 1997.

[10] M. A. Tsfasman, "Algebraic-geometric codes and asymptotic problems," *Discr. Appl. Math.*, vol. 33, pp. 241–256, 1991.

[11] M. A. Tsfasman and S. G. Vlăduţ, *Algebraic-Geometric Codes*. Dordrecht, The Netherlands: Kluwer, 1991.

[12] T. Uyematsu and E. Okamoto, "A construction of codes with exponential error bounds on arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 43, pp. 992–996, May 1997.

[13] A. Méhes and K. Zeger, "Performance of quantizers on noisy channels using structured families of codes," *IEEE Trans. Inform. Theory*, vol. 46, Nov. 2000, to be published.

[14] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 139–149, Mar. 1982.

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

[16] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inform. Theory*, vol. 41, pp. 26–34, Jan. 1995.

[17] S. N. Litsyn and M. A. Tsfasman, "A note on lower bounds," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 705–706, Sept. 1986.

[18] S. N. Litsyn, "New upper bounds on error exponents," *IEEE Trans. Inform. Theory*, vol. 45, pp. 385–398, Mar. 1999.