A LOW-RATE BOUND ON THE RELIABILITY OF A QUANTUM DISCRETE MEMORYLESS CHANNEL

ALEXANDER BARG

ABSTRACT. We extend a low-rate improvement of the random coding bound on the reliability of a classical discrete memoryless channel to its quantum counterpart.

1. Introduction

Derivation of error bounds in quantum information theory is usually performed by translation of the standard methods from its classical counterpart. Error exponents for the classical-quantum channel (transmission of orthogonal states) were derived in [8]. Here we are concerned with the so-called quantum-quantum channel which is the standard universe for quantum error-correcting codes. An exponential upper bound on the distortion (error) probability was derived in a recent paper [7]. Here we show that this bound can be improved for low noise and low values of the transmission rate. In Sect. 2 we give precise definitions of the quantum discrete memoryless channel (henceforth QDMC), codes, decoding, and error probability. Sect. 3 contains a brief review of stabilizer codes and their decoding. It turns out that if we restrict ourselves to the the class of stabilizer codes, then the bounds on their distortion exponent also follow from the corresponding classical results. In particular, in Sect. 4 we give a short proof and a somewhat more convenient form of the result of [7]. The link to the classical results motivates us to derive a low-rate error exponent for a QDMC (Sect. 5). A condition when it improves the random coding bound of [7] is given. We conclude by specializing the results to the case of a depolarizing channel and showing a concrete improvement for low code rates in the case of low noise.

2. Preliminaries

A quantum d-ary digit, a qudit, is a d-dimensional complex space $H = \mathbb{C}^d$, where d will be assumed a prime number. Below by $X$ we denote the finite field $\mathbb{F}_q$, where $q = d^2$. We consider transmission of unit-length state vectors $|\psi\rangle$ from the $d^n$-dimensional space $H_n = H \otimes^n$. Let us fix some orthonormal basis of $H_n$ and write it as $(|0\rangle, |1\rangle, \ldots, |d-1\rangle)$. A unitary basis of error operators (an error basis, for short) is defined as 

$$E_{i,j} = X^i Z^j, i, j \in \mathbb{F}_d,$$

where $X$ is defined as $X|i\rangle = |(i-1)\text{mod} d\rangle$, $Z|j\rangle = \omega^j|j\rangle$, and $\omega$ is a primitive $d$th root of unity.

A quantum discrete memoryless channel $W$ is defined as an arbitrary collection of operators of the form $(A_u, u \in X)$, where

$$A_u = \sum_{v \in X} a_{uv} E_v$$

and where the complex row vectors $a_v = (a_{uv}, u \in X)$ define a probability distribution on $X$ given by

$$W(v) = a_v a^*_v \quad (v \in X), \quad \sum_v W(v) = 1.$$
We note that this definition is derived from the general definition of the quantum channel \( \Phi \) which is a trace-preserving completely positive map on the set of density operators on \( H_n \). It is known that any such map \( \Phi \) can be written as

\[
\Phi = \sum_k A_k S A_k^*
\]

for some set of operators \( A_k \), where \( S \) is a density operator on \( H_n \) (the so-called Kraus representation of the channel). The absence of memory in the channel is reflected by the fact that the operators \( A_k \) can be written as tensor products of operators on \( H \).

As an example, let \( d = 2 \) and consider the so-called depolarizing channel \( \mathcal{W} = \{ \sqrt{1-p}I, \sqrt{p/3}\sigma_x, \sqrt{p/3}\sigma_z, \sqrt{p/3}\sigma_y \} \), where \( (\sigma_x, \sigma_z, \sigma_y) \) is the set of Pauli matrices. This channel acts on qubits by phase flips, amplitude flips, or combinations of both applied with probability \( p/3 \) each. More generally, for any \( d \) we can define a depolarizing channel as follows: \( \mathcal{W} = \{ \sqrt{1-p}I; \sqrt{p/q}E_{i,j}, i, j \in [p_d] \} \).

A quantum code \( \mathcal{Q} \) is a linear subspace of \( H_n \). The rate of \( \mathcal{Q} \) is defined as \( R = R(\mathcal{Q}) := (\log_d K)/n \), where \( K \) is the dimension of \( \mathcal{Q} \). Let \( \mathcal{R} \) be a recovery operator, i.e., another completely positive trace-preserving map on \( H_n \), restricted to \( \mathcal{Q} \). The fidelity of the code \( \mathcal{Q} \) for a given channel \( \Phi \) and a given recovery operator \( \mathcal{R} \) equals

\[
F(\mathcal{Q}, \{ \Phi, \mathcal{R} \}) = \frac{1}{K} \min_{B \subseteq \mathcal{Q}} \sum_{\psi \in B} \langle \psi | \mathcal{R} \Phi [|\psi\rangle \langle \psi|] | \psi \rangle,
\]

where the minimum is taken over all orthonormal bases \( B \) of the code. In particular, for the QDMC defined above, \( \Phi = \mathcal{W}^\otimes n \). Below we will omit the recovery operator from the notation.

For a given rate \( R \) we wish to define the reliability (exponent) of a QDMC \( \mathcal{W} \). Let

\[
E(n, R, \mathcal{W}) = \sup_{\mathcal{Q} \subseteq H_n: R(\mathcal{Q}) \geq R} -\frac{1}{n} \log_d (1 - F(\mathcal{Q}, \mathcal{W}))
\]

be the error exponent for the rate \( R \) and code length \( n \). Let

\[
E(R, \mathcal{W}) = \liminf_{i \to \infty} E(n_i, R, \mathcal{W}),
\]

where the limit is taken over all sequences of codes \( \mathcal{Q}_{n_i}, i = 1, 2, \ldots \) such that \( R(\mathcal{Q}_{n_i}) \geq R \) for all \( i \) greater than some \( i_0 \).

Let \( H_m(\mathcal{Q}) = -\sum_{x \in \mathcal{X}} Q(x) \log_m Q(x) \) be the entropy of a probability distribution \( Q \) on \( \mathcal{X} \). For two probability distributions \( P \) and \( Q \), their divergence is given by \( D_m(\mathcal{Q} || \mathcal{P}) = \sum_{x \in \mathcal{X}} Q(x) \log_m \frac{Q(x)}{P(x)} \) (if the base of the logarithms and exponents below is omitted, it is equal to \( d \)).

The following theorem was proved in [7].

**Theorem 1.** [7] For any rate \( R \geq 0 \) and any QDMC \( \mathcal{W} \)

\[
(1) \quad E(R, \mathcal{W}) \geq E_r(R, \mathcal{W}) = \min_V |D(V || W) + |1 - H(V) - R|^+|,
\]

where the minimum is taken with respect to all probability distributions on \( \mathcal{X} \) and \( |a|^+ := \max(a, 0) \).

Since \( E_r(R, \mathcal{W}) > 0 \) for \( 0 \leq R < 1 - H(W) \), this result also implies a lower bound of \( 1 - H(W) \) on the capacity of the channel \( \mathcal{W} \).

Given a vector \( x \in \mathcal{X}^n \), we can define an empirical probability distribution \( P \) on \( \mathcal{X} \) given by \( P(u) = |\{i : x_i = u\}|/n, u \in \mathcal{X} \). Below we call it the type of the vector \( x \) and write \( T(x) = P \). The type of the all-zero vector will be denoted by \( P_0 \); we have \( P_0(u) = \delta_{u,0} \). The set of all sequences of a given type \( P \) will be denoted as \( T_P \). It is clear that

\[
|T_P| = \exp_q(n(H_q(P) + o(1)))).
\]
Let \( \mathcal{P}(\mathcal{X}^n) \) be the set of all types on \( \mathcal{X}^n \). Obviously,
\[
|\mathcal{P}(\mathcal{X}^n)| = \binom{n + q - 1}{q - 1} \leq n^q \quad (n, q \geq 2).
\]

3. Stabilizer codes and their decoding

The construction of stabilizer quantum codes in [2], [3] is as follows. Consider the vector space \( V_n = (\mathbb{F}_d \times \mathbb{F}_d)^n \). Write a typical vector \( x \in V_n \) as \( (x_1, x_1', x_2, x_2', \ldots, x_n, x_n') \) and consider a standard symplectic form on \( V_n \) defined by
\[
(x, y) = \sum_{i=1}^{n} x_i y_i' - x_i' y_i.
\]

Now let \( \mathcal{C} \subset \mathcal{X}^n \) be an additive code, i.e., an additive subgroup of \( \mathcal{X}^n \) and define \( \mathcal{C}^\perp \) as the set of vectors in \( (\mathbb{F}_q^n) \cong V_n \) that are \((\cdot, \cdot)\)-orthogonal to every vector in \( \mathcal{C} \). Suppose that the number of vectors in \( \mathcal{C} \) is \( q^k \) so that the rate of \( \mathcal{C} \) equals \( R(\mathcal{C}) = k/n \). We then have \( |\mathcal{C}^\perp| = q^{n-k} \).

We begin with a pair of codes \( \mathcal{C}^\perp \subset \mathcal{C} \subset \mathcal{X}^n \) and a set \( \mathcal{E} \subset \mathcal{X}^n \) such that
\[
\forall x, y \in \mathcal{E} (y - x \in \mathcal{C}) \Rightarrow (x = y).
\]

According to this definition, we can take at most one error vector per coset of \( \mathcal{X}_n/\mathcal{C} \) and therefore, the maximum size of the set \( \mathcal{E} \) equals \( q^{n-k} \). It is possible to construct a quantum code \( \mathcal{Q} \subset H_n \) of (complex) dimension \( d^{2k-n} \) which is an invariant subspace of the set of error operators \( N_\mathcal{E} = \{N_x, x \in \mathcal{C}^\perp\} \) given by
\[
N_x = \bigotimes_{i=1}^{n} N_{x_i},
\]
where for every \( i \) the operator \( N_{x_i} = E_{x_i,1,x_i,2} \) is an element of the error basis determined by the representation of the coordinate \( x_i \in \mathbb{F}_q \) of \( x \) as a pair of elements \( (x_{i,1}, x_{i,2}) \in (\mathbb{F}_d)^2 \). Moreover, there are \( d^{2(n-k)} \) such invariant subspaces whose orthogonal direct sum equals \( H_n \). Thus the rate \( R \) of the stabilizer code \( \mathcal{Q} \) is related to the rate of \( \mathcal{C} \) as \( R = 2R(\mathcal{C}) - 1 \).

A stabilizer code \( \mathcal{Q} \) is \( \mathcal{E} \)-error-correcting in the sense that the action of any error operator from the set \( N_\mathcal{E} \) can be removed from the transmitted state. The received state \( w \) is measured with respect to the set of pairwise orthogonal operators \( P_i \), each being an orthogonal projector on the subspace of \( H_n \) that corresponds to a coset of \( \mathcal{X}^n/\mathcal{C} \). Then within this coset we find one of the most probable error vectors and recover the transmitted state by applying the inverse error operator.

The following bound on the fidelity of a given stabilizer code \( \mathcal{Q} \) was proved in [7] based on a result in [10].

**Theorem 2.** [7] Let \( \mathcal{Q} \) be an \( \mathcal{E} \)-error-correcting stabilizer code used over a QDMC \( \mathcal{W} \). Then
\[
1 - F(\mathcal{Q}, \mathcal{W}) \leq \sum_{x \notin \mathcal{E}} W^n(x).
\]

This theorem provides a link between the quantum and the classical setting.

Note that there is substantial freedom in the choice of the error set \( \mathcal{E} \). To derive our result, we will take \( \mathcal{E} \) as follows. As pointed out above, the channel \( \mathcal{W} \) defines a probability distribution \( W \) on \( \mathcal{X} \). For an additive code \( \mathcal{C} \) consider the quotient space \( \mathcal{X}^n/\mathcal{C} \). From each coset \( S \) we take one of the vectors \( y = y(S) \) whose probability \( W(y) = \prod W(y_i) \) is the smallest in \( S \). Finally, we take \( \mathcal{E} = \bigcup_y y(S) \).

We conclude this section by deriving a general analog of the weight distribution and of the Gilbert-Varshamov bound for additive codes over \( \mathcal{X} \). For \( q = 4 \) and the Hamming weight distribution this result was proved in [1].
Theorem 3. For any rate $R(C) > 0$ and any $\delta > 0$, there exists an additive code $C \subset X^n$ of size $\exp(nR)$ such that $C^\perp \subset C$ and for any type $P \neq P_0$,

$$\{|x \in (C \cap T_P)|\} \leq \exp_q[nR(C) + H_q(P) - 1 - \delta].$$

In particular, for any $x \in C \setminus \{0\}$ with $T(x) = P$ we have

$$R(C) \geq 1 - H_q(P) - \delta.$$

Proof. Let

$$S_{n,k} = \{C \in X^n : n^{-1} \log_q |C| = k, C^\perp \subset C\}.$$  

It was observed several times in the literature (e.g., [2], [7]) that every vector $x \in X^n \setminus \{0\}$ is contained in the same number of codes in $S_{n,k}$. Denote this number by $B$ and let $S = |S_{n,k}|$. Counting in two ways the sum of sizes of all the codes in $S_{n,k}$ we obtain $(q^n - 1)B = (q^k - 1)S$. Let us fix a type $P$. Clearly,

$$\sum_{P' \in P^n(X) : H_q(P') \leq H_q(P)} |T_{P'}| \leq n^q q^{nH_q(P)}. $$

Thus as long as $n^q q^{nR_q(P)}B < S$ or

$$n^q q^{nH_q(P)} < q^n - 1 < q^n(1-R),$$

there exists a code $C \in S_{n,k}$ such that for every $x \in C \setminus \{0\}$ we have $H_q(T(x)) \geq H_q(P)$. This proves the last part of the claim.

For any $P \neq P_0$ the average number of code vectors of type $P$ in a code $C \in S_{n,k}$ equals

$$E|\{x \in (C \cap T_P)\}| = \frac{B|T_P|}{S} = \exp_q[nR(C) + H_q(P) - 1 + o(1)].$$

Since there are no more than $n^q$ different types, this proves the first part of the claim. 

4. THE RANDOM CODING BOUND

Let $X$ be an input and $Y$ an output alphabet of a classical DMC given by a stochastic matrix $W(y|x)$. Suppose that $X \subset Y$ and that $Y$ is an abelian group, written additively. A channel is called additive if $W(y|x)$ depends only on the difference $y-x$, i.e., $W(y|x) = W(y-x)$ (the last term is actually $W(y-x|0)$, but below we abuse the notation slightly and use unconditional distributions). Note that an additive channel $W$ is symmetric in the sense that every row is a permutation of a fixed probability vector, and the same is true with respect to every column. By Theorem 2 the problem of bounding from below the reliability exponent of the QDMC is now reduced to the corresponding classical problem for the additive DMC with $Y = X$.

Gallager’s general bound for the reliability exponent of a DMC $W$ has the form [4]

$$(2) \quad E_r(R', W) = \max_P \min_W [D(V\|W|P) + |I(P,V) - R'|^+] ,$$

where $V$ is any stochastic $|X| \times |Y|$ matrix,

$$D(V\|W|P) = \sum_{x,y} P(x)V(y|x)\log \frac{V(y|x)}{W(y|x)}$$

is the conditional divergence,

$$I(P,V) = \sum_{x,y} P(x)V(y|x)\log \frac{V(y|x)}{\sum_x P(x)V(y|x)}$$

is the mutual information, and $R' = n^{-1} \log_q |C|$. Note that in this form the bound relies on the maximum mutual information decoder, which was employed in [7] and which is different from the
decoder defined in Sect. 3. Now, if the channel $W$ is symmetric, the maximizing input distribution $P$ in (2) is known to be uniform [4]: $P(z) = |X|^{-1}$ for any $z \in X$. Let us substitute this in (2). Clearly, $D(V\|W|P) = D(V\|W)$, where on the right-hand side $V$ and $W$ are probability distributions on $X$ given by $W(z) = W(y|x), V(z) = V(y|x)$ for any $y, x$ such that $z = x - y$.

Further, 

$$I(P, V) - R' = -|X|^{-1} \sum_{z \in X} H(V) + \log |X| - 1 - R = 1 - R - H(V),$$

where we have used the relation $R' = 2R(C) = 1 + R$. Together with Theorem 3 this implies Theorem 1.

We also note that if we return from (2) to Gallager’s original form of the random coding bound (by a method outlined in [4, pp.192-193]), the exponent (1) can be written in a somewhat more convenient form. Namely:

**Theorem 4.** Under the conditions of Theorem 1

$$E_r(R, W) = \max_{0 \leq \rho \leq 1} [-\rho R + E_0(\rho, W)],$$

where

$$E_0(\rho, W) = 2\rho - (1 + \rho) \log \sum_{x \in X} W(x)^{1/(1+\rho)}.$$

In the classical setting, the line of thought realized in Theorem 1 would correspond to an attempt to prove error bounds for a general DMC relying on the class of additive codes. It is well known (e.g., [4]) that this approach produces good results only when the optimizing probability distribution on the input alphabet is uniform. However, the classical channel derived from a general QDMC for stabilizer codes turns out to be additive and hence symmetric. Hence the lower bounds on the reliability exponent thus obtained are arguably rather strong. Note however that the lower capacity bound $C \geq 1 - H(W)$ implied by Theorem 1 when specialized to the depolarizing channel (see Example in the next section) falls below the best currently known estimate of [5].

5. **Expurgation exponent for a QDMC**

Let $Q \subset H_n$ be a stabilizer code of rate $R = R(Q)$ used over a QDMC $W$ together with the decoder defined in Sect. 3. Define the $W$-weight of a letter $x \in X$ as

$$|x|_W = -\log \sum_{e \in X} \sqrt{W(e)W(e-x)},$$

where $\log 0 = -\infty$ by definition.

**Theorem 5.**

$$E(R, W) \geq E_x(R, W) = \min_{P : H(P) \geq 1-R} \left[ \sum_{x \in X} P(x) |x|_W - (R + H(P) - 1) \right].$$

**Proof.** We start with the code $C$ whose existence is proved in Theorem 3. Let $Q$ be the stabilizer quantum code associated with it. By Theorem 2

$$1 - F(Q, W) = \sum_{e \notin E} W(e) = \sum_{x \in C \setminus \{0\}} \sum_{y \in X^n_{W(y-x) \geq W(y)}} W(y) \leq \sum_{x \in C \setminus \{0\}} \sum_{y \in X^n} \sqrt{W(y)W(y-x)}.$$
where last step follows because the channel is memoryless. Conclude by computing the logarithm and substituting the relation $2R(C) = 1 + R$. 

Note that it is possible that $E_x(R, W)$ becomes infinite for $R < R_\infty(W) > 0$, which means that for rates $R < R_\infty(W)$ errors outside the set $E$ occur with probability zero. The quantity $R_\infty(W)$ is called the zero-error capacity of the channel $W$. Shannon’s classical example of a channel with $R_\infty(W) > 0$ [6, p. 532] is given by the additive channel with $\mathcal{X} = \mathbb{Z}_5$ and $W(x) = W(x + 1) = 1/2$. Clearly, $R_\infty(W) > 0$ if and only if $|x|_W = 0$ for some $x \in \mathcal{X}$. A channel is called indivisible if this condition does not hold, and hence $R_\infty(W) = 0$.

The function $E_x$ can be transformed to a different form, also due to Gallager [6]:

$$E_x(R, W) = \sup_{\rho \geq 1} \left[ - \rho R + E_{ex}(\rho, W) \right],$$

where

$$E_{ex}(\rho, W) = -\rho \log_d \frac{1}{d^2} \sum_{x \in \mathcal{X}} \left( \sum_{e \in \mathcal{X}} \sqrt{W(e)W(e + x)} \right)^{1/\rho}.$$

Let us state a condition for the bound $E_{ex}(R, W)$ to improve the result of Theorem 1. As remarked above, the optimizing probability distribution on $\mathcal{X}$ for the random coding bound (2) in our case is uniform. Moreover, the exponent $E_x$ is also derived under the same assumption (recall that the code $C$ is linear). It is known [6] that for one and the same input distribution and for code rates $R < \partial E_{ex}(\rho, W)/\partial \rho|_{\rho=1}$ the function $E_x(R, W)$ is greater than $E_\rho(R, W)$, so in this region of rates Theorem 5 improves the result of Theorem 1. Hence if the point $R_x = \partial E_{ex}(\rho, W)/\partial \rho|_{\rho=1} > 0$ then there is a nonempty interval of code rates where $E_x(R, W) > E_\rho(R, W)$. Note that typically such an interval exists only for low noise level in the channel. To make an analogy with the classical case, the improvement takes place if the value of the code rate $R(C)$ that corresponds to $R_x$ is greater than $1/2$. In the range where it improves the bound (1), the exponent $E_x(R, W)$ can be written as

$$E_x(R, W) = \min_{P : H(P) = 1 - R} E |X|_W,$$

where $X$ is a random variable on $\mathcal{X}$ distributed according to $P$.

**Remark.** The general form of the function $E_x(R, W)$ for a given additive, indivisible channel $W$ is as follows:

$$E_x(R, W) = \max_P \sup_{\rho \geq 1} \left[ - \rho R + E_{ex}(\rho, P, W) \right],$$

where

$$E_{ex}(\rho, P, W) = -\rho \log_d \sum_{x,x' \in \mathcal{X}} P(x)P(x') \left( \sum_{e \in \mathcal{X}} \sqrt{W(x - e)W(x' - e)} \right)^{1/\rho}.$$

Optimization on the input distribution $P$ in this expression is easy if the $q \times q$ matrix

$$\left[ \left( \sum_{e \in \mathcal{X}} \sqrt{W(x - e)W(x' - e)} \right)^{1/\rho} \right]$$
Figure 1. Error exponents for the depolarizing channel with $d = 2$ and $p = 0.0005$.

For $0 \leq R \leq R_x$ the function $E_x$ gives a stronger bound than $E_r$.

is nonnegative definite for every $\rho \geq 1$ [9], and turns into a difficult problem otherwise. For the channel to be nonnegative definite it is sufficient that for every pair $(x, x')$ the sum on $e$ in the expression for $E_x(\rho, P, W)$ takes one and the same value (the so-called equidistant channels [9]). For equidistant channels the maximum on $P$ is achieved for the uniform distribution $P(x) = 1/q, x \in \mathcal{X}$. For instance, the $d$-ary depolarizing channel is equidistant. More generally, any binary QDMC $d = 2$ is trivially equidistant, and hence the optimum is again attained for $P$ uniform. However, there are many examples of not nonnegative definite additive, indivisible channels. For instance, let $d = 3$.

Consider the channel given by the following probability distribution:

\[
\begin{array}{cccccccccc}
u & 00 & 01 & 02 & 10 & 11 & 12 & 20 & 21 & 22 \\
W(u) & 0 & 0.49 & 0.01 & 0.01 & 0 & 0.49 & 0 & 0 & 1
\end{array}
\]

where $u \in (\mathbb{F}_9)^+ \cong \mathbb{Z}_3 \times \mathbb{Z}_3$. It is easily verified that this channel is not nonnegative definite for for $\rho \geq 1.37$.

**Example.** Let us specialize the results of Theorems 1 and 5 for the case of the $d$-ary depolarizing channel $W$. Let us denote the reliability exponent of $W$ by $E(R, p)$. The result can be expressed in a closed form. Let

\[
\begin{align*}
h(x) &= -x \log_q \frac{x}{q - 1} - (1 - x) \log_q x \\
D(x \parallel y) &= x \log_q \frac{x}{y} + (1 - x) \log_q \frac{1 - x}{1 - y} \\
\delta_0(x) &= h^{-1}(1 - x).
\end{align*}
\]

We have

\[
E(R(Q), W) \geq 2E_\ell((1 + R)/2, p),
\]

where

\[
\begin{align*}
E_\ell(r, p) &= -\delta_0(r) \log_q \gamma_q(p) & (0 \leq r \leq r_x) \\
E_\ell(r, p) &= D(\rho_0 \parallel p) + r_c - r & (r_x \leq r \leq r_c) \\
E_\ell(r, p) &= D(\delta_0(\rho_0) \parallel p) & (r_c \leq r \leq 1 - h(\rho_0)),
\end{align*}
\]

\[
\begin{align*}
r_x &= 1 - h\left(\rho_0\left(2 - \frac{q\rho_0}{q - 1}\right)\right), \quad r_c = 1 - h(\rho_0),
\end{align*}
\]
\[ \rho_0 = \frac{\sqrt{p(q-1)}}{\sqrt{p(q-1)} + \sqrt{1-p}}, \quad \gamma_q(p) = p \frac{q-2}{q-1} + 2 \sqrt{\frac{p(1-p)}{q-1}}. \]

This reliability exponent can be obtained from Theorems 4, 5 or computed directly starting with codes whose existence is proved in Theorem 3. The expurgation exponent is straightforward from (3). If \( R_\chi := 2r_\chi - 1 > 0 \), we obtain an improvement over the result of Theorem 1 in the interval of values of \( R \) between zero and \( R_\chi \). It turns out that this condition is satisfied for low noise level (see an example on Fig. 1). For \( d = 2 \) the expurgation bound improves the random coding exponent for \( 0 < p \leq 0.004 \).

\[ \square \]

**Acknowledgment.** The author is grateful to A. Ashikhmin and G. Kramer for helpful discussions.

**References**


**Bell Labs, Lucent Technologies, 700 Mountain Avenue, Rm. 2C-375, Murray Hill, NJ 07974 USA**

**E-mail address:** abarg@research.bell-labs.com