Correspondence

A Low-Rate Bound on the Reliability of a Quantum Discrete Memoryless Channel

Alexander Barg, *Senior Member, IEEE*

***Abstract*—We extend a low-rate improvement of the random coding bound on the reliability of a classical discrete memoryless channel (DMC) to its quantum counterpart. The key observation that we make is that the problem of bounding below the error exponent for a quantum channel relying on the class of stabilizer codes is equivalent to the problem of deriving error exponents for a certain symmetric classical channel.**

***Index Terms*—Additive channels, depolarizing channel, Gilbert–Varshamov bound, method of types, stabilizer codes.**

I. 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 [10]. 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 [9]. Here we show that this bound can be improved for low noise and low values of the transmission rate. In Section II, we give precise definitions of the quantum discrete memoryless channel (QDMC), codes, decoding, and error probability. Section III contains a brief review of stabilizer codes and their decoding. It turns out that if we restrict ourselves to the class of stabilizer codes, then the bounds on their distortion exponent also follow from the corresponding classical results. In particular, in Section IV, we give a short proof of the result of [9]. The link to the classical results motivates us to derive a low-rate error exponent for a QDMC (Section V). A condition when it improves the random coding bound of [9] 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.

II. 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. In the following, by $\mathcal{X}$ 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 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

$$X|i\rangle = |(i-1) \bmod d\rangle, \quad Z|j\rangle = \omega^j |j\rangle,$$

and ω is a primitive dth root of unity.

Manuscript received March 25, 2002; revised July 31, 2002. This work was supported in part by the Binational Science Foundation (USA–Israel) under Grant 1999099.

The author is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: abarg@research.bell-labs.com).

Communicated by P. W. Shor, Associate Editor for Quantum Information Theory.

Digital Object Identifier 10.1109/TIT.2002.805080

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

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

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

$$W(v) = a_v a_v^* \quad (v \in \mathcal{X}), \qquad \sum_v W(v) = 1.$$

We note that this definition is derived from the general definition of the quantum channel Φ which is a trace-preserving completely positive map on the set of density operators on H_n. It is known that any such map Φ can be written as

$$\Phi(S) = \sum_k A_k S A_k^*$$

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*

$$\mathcal{W} = \left\{\sqrt{1-p}\, I,\ \sqrt{p/3}\, \sigma_x,\ \sqrt{p/3}\, \sigma_z,\ \sqrt{p/3}\, \sigma_y\right\}$$

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:

$$\mathcal{W} = \left\{\sqrt{1-p}\, I;\ \sqrt{\frac{p}{q-1}}\, E_{i,j},\ i, j \in \mathbb{F}_d\right\}.$$

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 Φ and a given recovery operator $\mathcal{R}$ equals

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

where the minimum is taken over all orthonormal bases B of the code. In particular, for the QDMC defined above, $\Phi = \mathcal{W}^{\otimes n}$. In what follows, 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

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

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

$$E(R, \mathcal{W}) = \liminf_{n \to \infty} E(n, R, \mathcal{W}).$$

Let

$$H_m(Q) = -\sum_{x \in \mathcal{X}} Q(x) \log_m Q(x)$$

be the entropy of a probability distribution Q on $\mathcal{X}$. For two probability distributions P and Q, their information divergence is given by

$$D_m(Q\|P) = \sum_{x\in\mathcal{X}} Q(x)\log_m \frac{Q(x)}{P(x)}$$

(if the base of the logarithms and exponents below is omitted, it is equal to d).

The following theorem was proved in [9].

Theorem 1 [9]: For any rate $R \geq 0$ and any QDMC $\mathcal{W}$

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

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 $P(u) = |\{i\colon x_i = u\}|/n$, $u \in \mathcal{X}$. 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 $\mathsf{T}_P(\mathcal{X}^n)$. It is clear that

$$|\mathsf{T}_P(\mathcal{X}^n)| = \exp_q(n(H_q(P) + o(1))).$$

Let $\mathcal{P}(\mathcal{X}^n)$ be the set of all types on $\mathcal{X}^n$. Obviously

$$|\mathcal{P}(\mathcal{X}^n)| = \binom{n+q-1}{q-1} \leq n^q \qquad (n, q \geq 2).$$

For any $x \in \mathcal{X}^n$ and any stochastic matrix $V\colon \mathcal{X} \to \mathcal{Y}$, the *$V$-shell* of x is defined as the set $\mathsf{T}_V(x) \subset \mathcal{Y}^n$ formed by those y whose conditional type is V. This means that for any such y its type is $T(y) = PV$, where PV is the probability distribution on $\mathcal{Y}$ given by

$$PV(y) = \sum_{x\in\mathcal{X}} P(x)V(y|x).$$

III. 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 $(,)$-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 $N_\mathcal{E} = \{N_x,\ x \in \mathcal{E}\}$ 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 $(x_{i,1}, x_{i,2}) \in (F_d)^2$. 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 [9] based on a result in [12].

Theorem 2 [9]: 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 which will be pivotal in our argument.

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 $W^n(y) = \prod W(y_i)$ is the largest in S. Finally, we take $\mathcal{E} = \cup_S y(S)$.

We conclude this section by deriving a general analog of the weight distribution and of the Gilbert–Varshamov bound for additive self-orthogonal codes over $\mathcal{X}$. For $q = 4$ and the Hamming weight distribution this result was proved in [1].

Theorem 3: For any rate $R(\mathcal{C}) > 0$ and any $\delta > 0$, there exists an additive code $\mathcal{C} \subset \mathcal{X}^n$ of size $\exp(nR(\mathcal{C}))$ such that $\mathcal{C}^\perp \subset \mathcal{C}$ and for any type $P \neq P_0$

$$|\mathcal{C} \cap \mathsf{T}_P| \leq \exp_q[n(R(\mathcal{C}) + H_q(P) - 1 + \delta)]. \quad (2)$$

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

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

Proof: Let

$$S_{n,k} = \{\mathcal{C} \in \mathcal{X}^n\colon \log_q|\mathcal{C}| = k, \mathcal{C}^\perp \subset \mathcal{C}\}.$$

It was observed several times in the literature (e.g., [2], [9]) that every vector $x \in \mathcal{X}^n \setminus \{0\}$ 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

$$\sum_{P'\in\mathcal{P}^n(\mathcal{X})\colon\ H_q(P')\leq H_q(P)} |\mathsf{T}_{P'}| \leq n^q q^{nH_q(P)}.$$

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

$$n^q q^{nH_q(P)} < \frac{q^n - 1}{q^k - 1} = q^{n(1-R(\mathcal{C}))}(1 + o(1))$$

there exists a code $\mathcal{C} \in S_{n,k}$ such that for every $x \in \mathcal{C} \setminus \{0\}$ 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 $\mathcal{C} \in S_{n,k}$ equals

$$\frac{1}{S}\sum_{\mathcal{C}\in S_{n,k}} |\{x \in (\mathcal{C} \cap \mathsf{T}_P)\}| = \frac{B|\mathsf{T}_P|}{S} = \exp_q[n(R(\mathcal{C}) + H_q(P) - 1 + o(1))].$$

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

IV. The Random Coding Bound

Let $\mathcal{X}$ be an input and $\mathcal{Y}$ an output alphabet of a classical discrete memoryless channel (DMC) given by a stochastic matrix $W(y|x)$. Suppose that $\mathcal{X} = \mathcal{Y}$ and that $\mathcal{Y}$ 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 a QDMC is now *reduced to the corresponding classical problem for a symmetric, additive DMC*. With this observation, Theorem 1 follows by a combination of standard arguments; so having in mind the reader quite familiar with error exponents of classical channels we could as well stop here. In the interest of staying self-contained we will supply some more details.

A. General Form of the Random Coding Exponent

For any type $P \in \mathcal{P}(\mathcal{X}^n)$ and any stochastic $|\mathcal{X}| \times |\mathcal{Y}|$ matrix V, let

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

be the conditional divergence and

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

be the mutual information between $x \in \mathsf{T}_P(\mathcal{X}^n)$ and $y \in \mathsf{T}_V(x)$. The following theorem (reformulated slightly from [4]) gives one of the general forms of the error exponent of a classical DMC.

Theorem 4: For a given type $P \in \mathcal{P}(\mathcal{X}^n)$ let $A \subset \mathsf{T}_P(\mathcal{X}^n)$, $|A| = d^{(R'-\epsilon)n}$ be a code such that for every stochastic matrix $\tilde{V}\colon \mathcal{X} \to \mathcal{X}$

$$|\{(x_i, x_j) \in A \times A\colon x_j \in \mathsf{T}_{\tilde{V}}(x_i)\}| \leq \exp[n(R' - I(P, \tilde{V}))]. \quad (3)$$

Suppose that A is used over a DMC $W\colon \mathcal{X} \to \mathcal{Y}$ together with a maximum mutual information decoder. Then the exponent $E(A, W)$ of the maximum error probability $\max_{x\in A} p_e$ satisfies $E(A, W) \geq E_r(P, R', W)$, where

$$E_r(P, R', W) = \min_V [D(V\|W|P) + |I(P, V) - R'|^+] \quad (4)$$

and where V runs over the set of all channels $\mathcal{X} \to \mathcal{Y}$.

Remarks:

1) This theorem is a generalization of a classical fact of coding theory, that "binary linear codes of rate R and weight distribution $A_w \leq 2^{n(R-1)}\binom{n}{w}$, $w = 1, 2, \ldots, n$ achieve the random coding exponent of the binary symmetric channel."

2) The best bound on the reliability exponent of the channel W is obtained by computing the maximum on P in (4). The quantity $E(R', W) = \max_P E_r(P, R', W)$ is usually called the random coding exponent of W.

3) The maximum mutual information decoder, which is used to prove this result and which was employed in [9], is different from the decoder defined in Section III.

B. Additive Channels and Codes

Recall that in our problem $\mathcal{X}$ is an additive group and that $\mathcal{Y} = \mathcal{X}$. Further, since the channel W is symmetric, the maximizing input distribution P in (4) is known to be uniform [6]: $P_u(x) = |\mathcal{X}|^{-1}$ for any $x \in \mathcal{X}$.

Let us substitute P_u into the condition (3) on the "distance distribution" of the code A. Let x be a vector such that $T(x) = P_u$ and let $\tilde{V}$ be a stochastic matrix such that $\mathsf{T}_{\tilde{V}}(x) \cap \mathsf{T}_{P_u}(\mathcal{X}^n) \neq \emptyset$. Then for any letter $x \in \mathcal{X}$, the sum $\sum_{y\in\mathcal{X}} \tilde{V}(y|x) = 1$. We compute

$$I(P_u, \tilde{V}) = \log|\mathcal{X}| - H(\tilde{V}|P_u)$$

so the upper bound in (3) takes the form

$$|\{(x_i, x_j) \in A \times A\colon x_j \in \mathsf{T}_{\tilde{V}}(x_i)\}| \leq \exp[n(R' + H(\tilde{V}|P_u) - \log|\mathcal{X}|)]. \quad (5)$$

Now consider the code $\mathcal{C}$ from Theorem 3. Almost all of its codewords are of type P_u and nearby types (types close to it in some suitable metric, say, the ℓ_1-distance). We claim that the "distance distribution" of the code $\mathcal{C}$ satisfies (5). Since the code is additive, it suffices to consider matrices $\tilde{V}$ such that $\tilde{V}(y|x)$ depends only on the difference $y - x$. Any such matrix defines a distribution $\tilde{V}(z) = \tilde{V}(z|0)$ on $\mathcal{X}$. Using this in (5), we observe that this condition reduces to the condition (2) satisfied by the "weight" distribution of $\mathcal{C}$. Now recall from [5] that the function $E_r(P, R', W)$ is uniformly continuous on P and that, on account of the channel and code being additive, the error probability of decoding does not depend on the transmitted codeword. Therefore, for growing n the error exponent of the code $\mathcal{C}$ attains the bound $E(R', W)$. This proves Theorem 1.

Transforming the exponent (4) to the form (1) is a matter of calculation. Indeed, let us substitute P_u in (4). Clearly, $D(V\|W|P) = D(V\|W)$, where on the right-hand side V and W are probability distributions on $\mathcal{X}$ given by $W(z) = W(y|x)$, $V(z) = V(y|x)$ for any y, x such that $z = x - y$. Further

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

where we have used the relation $R' = 2R(\mathcal{C}) = 1 + R$.

C. Further Observations

1) By the same token, the capacity of the quantum channel $\mathcal{W}$ is bounded below by the capacity of the classical symmetric channel W. Again, the mutual information is maximized for the uniform input distribution, which implies the bound $\mathcal{C} \geq 1 - H(W)$ independently of the results on error exponents. Note, however, that when this result is specialized to the depolarizing channel (see the Example in the next section), it falls below the best currently known estimate of [7].

2) If we return from (4) to Gallager's original form of the random coding bound (by a method outlined in [5, pp. 192–193]), the exponent (1) can be written in a somewhat more convenient form as follows.

Theorem 5: Let

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

Then

$$E_r(R, \mathcal{W}) = 1 - R - \log\left(\sum_{x\in\mathcal{X}} \sqrt{W(x)}\right)^2 \qquad \left(0 \le R < \left.\frac{\partial E_0}{\partial \rho}\right|_{\rho=1}\right)$$

and

$$E_r(R, \mathcal{W}) = \max_{0\le\rho\le 1} [-\rho R + E_0(\rho, W)] \qquad \left(\left.\frac{\partial E_0}{\partial \rho}\right|_{\rho=1} \le R \le 1 - H(W)\right).$$

3) 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 [6], [5] that this approach produces good results only when the optimizing probability distribution on the input alphabet is uniform. 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.

V. Expurgation Exponent for a QDMC

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

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

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

Theorem 6:

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

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

$$\begin{aligned} 1 - F(\mathcal{Q}, \mathcal{W}) &= \sum_{e\notin\mathcal{E}} W^n(e) = \sum_{x\in\mathcal{C}\setminus\{0\}} \sum_{\substack{y\in\mathcal{X}^n \\ W^n(y-x)\ge W^n(y)}} W^n(y) \\ &\le \sum_{x\in\mathcal{C}\setminus\{0\}} \sum_{y\in\mathcal{X}^n} \sqrt{W^n(y)W^n(y-x)} \\ &= \sum_{P\in\mathcal{P}(X^n)} \sum_{x\in\mathcal{C}\cap\mathsf{T}_P(\mathcal{X}^n)} \sum_y \sqrt{W^n(y)W^n(y-x)} \\ &\le \sum_{P\in\mathcal{P}(X^n)} \exp_d[2n(R(\mathcal{C}) + H_q(P) - 1 + o(1)) \\ &\quad - n\sum_{x\in\mathcal{X}} P(x)|x|_W] \end{aligned}$$

where the last step follows because the channel is memoryless. Conclude by computing the logarithm and substituting the relation $2R(\mathcal{C}) = 1 + R$. □

Note that it is possible that $E_x(R, \mathcal{W})$ becomes infinite for $R \downarrow R_\infty(\mathcal{W}) > 0$, which means that for rates $R < R_\infty(\mathcal{W})$ errors outside the set $\mathcal{E}$ occur with probability zero. The quantity $R_\infty(\mathcal{W})$ gives a lower bound on the zero-error capacity of the channel $\mathcal{W}$. Shannon's classical example of a channel with $R_\infty(\mathcal{W}) > 0$ [8, p. 532] is given by the additive channel with $\mathcal{X} = \mathbb{Z}_5$ and $W(x) = W(x+1) = 1/2$. Clearly, $R_\infty(\mathcal{W}) > 0$ 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 $R_\infty(\mathcal{W}) = 0$.

The function E_x can be transformed to a different form, also due to Gallager [8]

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

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 $E_{ex}(R, \mathcal{W})$ to improve the result of Theorem 1. As remarked earlier, the optimizing probability distribution on $\mathcal{X}$ for the random coding bound (4) in our case is uniform. Moreover, the exponent E_x is also derived under the same assumption. It is known [8] that for one and the same input distribution and for code rates $R < \partial E_{ex}(\rho, W)/\partial\rho|_{\rho=1}$ the function $E_x(R, \mathcal{W})$ is greater than $E_r(R, \mathcal{W})$, so in this region of rates Theorem 6 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 $E_x(R, \mathcal{W}) > E_r(R, \mathcal{W})$. 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 $R(\mathcal{C})$ that corresponds to R_x is greater than $1/2$. In the range where it improves the bound (1), the exponent $E_x(R, \mathcal{W})$ can be written as

$$E_x(R, \mathcal{W}) = \min_{P:\, H(P)=1-R} \mathsf{E}|X|_W \tag{6}$$

where X is a random variable on $\mathcal{X}$ distributed according to P. This follows by the Gilbert–Varshamov bound of Theorem 3.

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

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

where

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

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

$$\left[\left(\sum_{e\in\mathcal{X}} \sqrt{W(x-e)W(x'-e)}\right)^{1/\rho}\right]$$

is nonnegative definite for every $\rho \ge 1$ [11], and turns into a difficult problem otherwise. For the channel to be nonnegative definite it is sufficient that for every pair of distinct vectors (x, x') the sum on e in the expression for $E_x(\rho, P, \mathcal{W})$ takes one and the same value

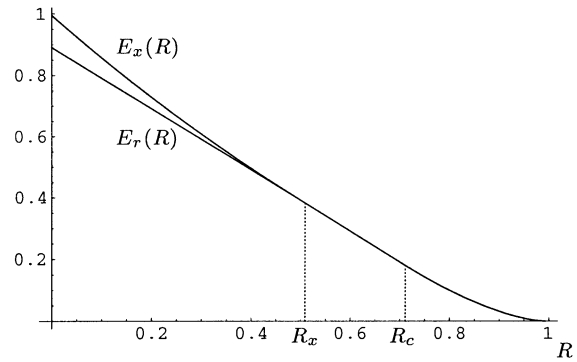

Fig. 1. Error exponents for the depolarizing channel with $d = 2$ and $p = 0.0005$. For $0 \le R \le R_x$ the function E_x gives a stronger bound than E_r.

(the so-called equidistant channels [11]). 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. 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:

u	00	01	02	10	11	12	20	21	22
$W(u)$	0	0.49	0	0.01	0.01	0	0.49	0	0,

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 $\rho \ge 1.37$. □

Example: Let us specialize the results of Theorems 1 and 6 for the case of the d-ary depolarizing channel $\mathcal{W}$. Let us denote the reliability exponent of $\mathcal{W}$ by $E(R, p)$. The result can be expressed in a closed form. Let

$$h(x) = -x \log_q \frac{x}{q-1} - (1-x) \log_q x$$

$$D(x\|y) = x \log_q \frac{x}{y} + (1-x) \log_q \frac{1-x}{1-y}$$

$$\delta_0(x) = h^{-1}(1-x).$$

We have

$$E(R(\mathcal{Q}), \mathcal{W}) \ge 2E_\ell((1+R)/2, p)$$

where

$$E_\ell(r,p) = -\delta_0(r) \log_q \gamma_q(p) \qquad (0 \le r \le r_x)$$

$$E_\ell(r,p) = D(\rho_0\|p) + r_c - r \qquad (r_x \le r \le r_c)$$

$$E_\ell(r,p) = D(\delta_0(r)\|p) \qquad (r_c \le r \le 1 - h(p))$$

$$r_x = 1 - h\left(\rho_0\left(2 - \frac{q\rho_0}{q-1}\right)\right), \qquad r_c = 1 - h(\rho_0)$$

$$\rho_0 = \frac{\sqrt{p(q-1)}}{\sqrt{p(q-1)} + \sqrt{1-p}}$$

$$\gamma_q(p) = p\frac{q-2}{q-1} + 2\sqrt{\frac{p(1-p)}{q-1}}. \tag{7}$$

This reliability exponent can be obtained from Theorems 5 and 6 or computed directly starting with codes whose existence is proved in Theorem 3. The expurgation exponent (7) is straightforward from (6). If $R_x := 2r_x - 1 > 0$, then from (7) we obtain an improvement over the result of Theorem 1 in the interval of values of R between zero and R_x. It turns out that this condition is satisfied for low noise levels (see an example in Fig. 1). For $d = 2$, the expurgation bound improves the random coding exponent for $0 < p \le 0.004$. □

Acknowledgment

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

References

[1] A. Ashikhmin, A. Barg, E. Knill, and S. Litsyn, "Quantum error detection, II," *IEEE Trans. Inform. Theory*, vol. 46, pp. 789–800, May 2000.

[2] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, "Quantum error correction and orthogonal geometry," *Phys. Rev. Lett.*, vol. 78, no. 3, pp. 405–408, 1997.

[3] ——, "Quantum error correction via codes over GF(4)," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1369–1387, July 1998.

[4] I. Csiszár, "The method of types," *IEEE Trans. Inform. Theory (Information Theory: 1948–1998)*, vol. 44, pp. 2505–2523, Oct. 1998.

[5] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Channels*. Budapest, Hungary: Akadémiai Kiadó, 1981.

[6] R. L. Dobrushin, "Asymptotic optimality for grouped and systematic codes for certain channels," *Teor. Verojatnost. i Primenen*, vol. 8, pp. 52–66, 1963.

[7] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin, "Quantum capacity of very noisy channels," *Phys. Rev. A*, vol. 57, pp. 830–839, 1998.

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

[9] M. Hamada, "Lower bounds on the quantum capacity and highest error exponent of general memoryless channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2547–2557, Sept. 2002.

[10] A. S. Holevo, "Reliability function of general classical-quantum channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2256–2261, Sept. 2000.

[11] F. Jelinek, "Evaluation of expurgated bound exponents," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 501–505, Sept. 1968.

[12] J. Preskill, *Lecture Notes for Physics 229: Quantum Information and Computation*. Pasadena, CA: Calif. Inst. Technol., 1999.