Quantum Rate-Distortion Theory for Memoryless Sources

Igor Devetak, *Student Member, IEEE*, and Toby Berger, *Fellow, IEEE*

Invited Paper

This paper is dedicated to Aaron's loving wife, Nusha Wyner, who actively and selflessly nourished his lifelong devotion both to the intellectual discipline of information theory and to the activities of the IEEE Information Theory Society.

***Abstract*—We formulate quantum rate-distortion theory in the most general setting where classical side information is included in the tradeoff. Using a natural distortion measure based on entanglement fidelity and specializing to the case of an unrestricted classical side channel, we find the exact quantum rate-distortion function for a source of isotropic qubits. An upper bound we believe to be exact is found in the case of biased sources. We establish that in this scenario optimal rate-distortion codes produce no entropy exchange with the environment of any individual qubit.**

***Index Terms*—Entanglement, entanglement fidelity, quantum information theory, quantum rate-distortion theory, qubit, rate-distortion theory, source coding.**

I. Introduction

THE quantum lossless source coding theorem specifies the minimum rate, called the *entropy* and measured in code qubits per source qubit, to which a quantum source can be compressed subject to the requirement that the source qubits can be recovered *perfectly* from the code qubits. In realistic applications, we may be able to tolerate imperfect recovery of the source qubits at the receiver, in which case we would seek to minimize the rate required to achieve a specified level of distortion. Equivalently, we may be required to use a rate R less than the entropy of the source, in which case we would seek to minimize the distortion subject to this rate constraint. Here, the distortion measure is a user-defined function of the input and the reconstruction, the precise form of which depends on the nature of the application.

Analysis of the tradeoff between rate and distortion is the subject matter of *rate-distortion theory*. Classical rate-distortion theory [1] is an important and fertile area in information theory. Considering that coding theorems for both noiseless [2], [11], and noisy [3] quantum *channels* have been established some years ago, it is surprising that little effort has been put into developing quantum rate-distortion theory. The purpose of this paper is to fill that gap.

To be completely general one must allow for a classical side channel containing information gathered while manipulating the source qubits, and include the corresponding classical rate r, measured in bits per source qubit, in the tradeoff. It has been shown in [4] that at zero distortion, no classical side information can help reduce the quantum rate below the von Neumann entropy of the source. This turns out not to be the case for positive distortion d. Therefore, one must speak of a two-dimensional tradeoff manifold $R(d, r)$. Here we introduce this general formulation for the first time. However, we focus mainly on the scenario of unrestricted classical side information, i.e., $r = \infty$, and refer to $R(d) \equiv R(d, \infty)$ as the rate-distortion function. This clearly provides a lower bound on achievable R for the same distortion d but restricted classical rate r.

In classical information theory the rate-distortion function has the simple form

$$R(D) = \min_{Y:\, E_{X,Y} d(X,Y) \leq D} I(X; Y) \tag{1}$$

where X is a random variable distributed like a typical source letter, Y is a random variable jointly distributed with X that is used to construct approximations to the source output and ranges over an alphabet possibly different from the source alphabet, $E_{X,Y}$ denotes expectation with respect to the joint distribution of X and Y, $d(\cdot, \cdot)$ is a suitably defined distortion function, and $I(X; Y)$ is the average mutual information between X and Y. The relevant information-like quantity playing the role in the quantum channel capacity formula is the coherent information $I_c(\rho, \mathcal{E})$ [5] to be defined in the next section. The natural first impulse is to assume that the same quantity should appear in quantum rate-distortion theory. Indeed, Barnum [6] has derived a lower bound on $R(d, 0)$ based on coherent information. This bound is far from tight, however, since the coherent information often is negative for distortions considerably smaller that which can be achieved with the receiver is sent no qubits at

Manuscript received December 31, 2000; revised October 17, 2001. This work was supported in part by the ARO-administered MURI Grant DAAG-19-99-1-0125 and by NSF under Grant CCR-9980616. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Washington, DC, June 24–28, 2001.

The authors are with the Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: igor@ece.cornell.edu; berger@ece.cornell.edu).

Communicated by S. Shamai, Guest Editor.
Publisher Item Identifier S 0018-9448(02)04018-X.

all. (A comparable problem does not occur in channel capacity calculations because the maximization procedure invoked there ensures positivity.) In view of this, we pursue quantum rate-distortion from first principles using a natural distortion measure based on entanglement fidelity that was introduced in [6].

We define the problem in Section II, where we also provide relevant background on quantum operations, entropies, and fidelity measures. In Section III, we find the rate-distortion function for a restricted class of coding procedures; in Section IV, we argue that the optimum coding scheme belongs to this class. Section V describes a simple physical realization of the optimal coding procedure. Speculations are left for the final section.

II. Definitions

Let us recall some basic definitions of quantum information theory [7], [8]. A general quantum information source is described by a density matrix ρ^Q of a quantum system Q. This density matrix may result from the system being prepared in certain pure states with respective probabilities. Alternatively, we may view our quantum system Q as a part of a larger system RQ which includes a *reference system* R which always may be constructed such that the overall state is pure $|\Psi^{RQ}\rangle$ and ρ^Q results from restricting to Q, i.e.,

$$\rho^Q = \mathrm{tr}_R(|\Psi^{RQ}\rangle\langle\Psi^{RQ}|). \tag{2}$$

Next consider a quantum process acting on the source ρ^Q

$$\rho^Q \rightarrow \hat{\mathcal{E}}(\rho^Q) \equiv \frac{\mathcal{E}(\rho^Q)}{\mathrm{tr}\,(\mathcal{E}(\rho^Q))} \tag{3}$$

with a general quantum operation $\mathcal{E}$ of the form

$$\mathcal{E}(\rho^Q) = \sum_{i=1}^{k} A_i \rho^Q A_i^\dagger. \tag{4}$$

Note that the action of $\mathcal{E}$ is completely determined by the set of operation elements $\{A_i\}$. A useful way to think about the quantum process is to embed RQ into an even larger space RQE by adding an *environment* E, initially in a pure state $|s\rangle$ and hence decoupled from RQ. Then a well-known representation theorem [7], [8] states that a general quantum process $\mathcal{E}$ may be realized by performing a unitary transformation U^{QE} entangling Q and E, followed by projecting via P^E onto the environment alone, and then tracing out R and E; i.e.,

$$\mathcal{E}(\rho^Q) = c\,\mathrm{tr}_{RE}(P^E U^{QE} \cdot(|\Psi^{RQ}\rangle\langle\Psi^{RQ}| \otimes |s\rangle\langle s|)U^{QE\dagger}P^E) \tag{5}$$

where c is a positive constant. Although the theorem refers to a mathematical construction, it provides physical insight. For instance, it enables one to define the entropy exchange [7], [3]

$$S_e(\rho^Q, \mathcal{E}) \equiv S(\rho^{E'}) = S(\rho^{RQ'}). \tag{6}$$

Here, $S(\sigma) \equiv -\mathrm{tr}\,(\sigma \log_2 \sigma)$ is the von Neumann entropy and $\rho^{E'}$ and $\rho^{RQ'}$ denote the states of E and RQ, respectively, after the operation. The equality in (6) holds because the whole system RQE remains in a pure state after the process, as a consequence of which $S_e(\rho^Q, \mathcal{E})$ measures the amount of "disorder," or "noise," introduced into the system RQ by virtue of its having become entangled with E, and *vice versa*.

A convenient expression in terms of the original operation elements $\{A_i\}$ is

$$S_e(\rho^Q, \mathcal{E}) = S(W) = -\mathrm{tr}\,(W \log_2 W) \tag{7}$$

with

$$W_{ij} = \frac{\mathrm{tr}\,(A_i \rho^Q A_j^\dagger)}{\mathrm{tr}\,(\mathcal{E}(\rho^Q))}. \tag{8}$$

Observe that if there is only one operation element (or, equivalently, if they are all the same), then the entropy exchange is zero. The noise interpretation of S_e is also evident from the formula for coherent information

$$I_c(\rho^Q, \mathcal{E}) = S(\hat{\mathcal{E}}(\rho^Q)) - S_e(\rho^Q, \mathcal{E}) \tag{9}$$

that appears in the channel capacity formula. Comparing $I_c(\rho^Q, \mathcal{E})$ to its classical counterpart

$$I(X;Y) = H(Y) - H(Y|X)$$

we see that $S_e(\rho^Q, \mathcal{E})$ plays a role analogous to the noise term $H(Y|X)$.

We end this brief review with the definition of *entanglement fidelity*, denoted by $F_e(\rho^Q, \mathcal{E})$ and defined by

$$F_e(\rho^Q, \mathcal{E}) = \langle\Psi^{RQ}|(I_R \otimes \mathcal{E})(|\Psi^{RQ}\rangle\langle\Psi^{RQ}|)|\Psi^{RQ}\rangle. \tag{10}$$

The entanglement fidelity tells us how well the system's state and the system's entanglement with its surroundings R, which do not participate directly in the quantum process, are preserved under the operation in question. Like any meaningful quantity, it has an expression which is manifestly independent of which purification R is employed, namely,

$$F_e(\rho^Q, \mathcal{E}) = \frac{\sum_i |\mathrm{tr}\,(A_i \rho^Q)|^2}{\mathrm{tr}\,(\mathcal{E}(\rho^Q))}. \tag{11}$$

We now augment Barnum's formulation of the $r = 0$ case [6] to allow for classical side information. First we restrict attention to independent and identically distributed (i.i.d.) sources with density matrix ρ, so that $\rho^{(n)} \equiv \rho^{\otimes n}$. An (n, R, r) *rate-distortion code* consists of an encoding operation $\mathcal{C}^{(n)}$ from the source space $\rho^{(n)}$ to a block of $\lfloor nR \rfloor$ qubits and $\lfloor nr \rfloor$ bits (henceforth abbreviated to nR and nr, respectively), and a decoding operation $\mathcal{D}^{(n)}$ acting in the reverse direction. Here $R \leq 1$, so in effect we are compressing the n source qubits to nR qubits and then decompressing them back to n qubits with the help of nr bits of information gathered during the compression phase, in an attempt to recover the original with the maximum possible fidelity consistent with the values of R and r.

For the rate-distortion code $(\mathcal{C}^{(n)}, \mathcal{D}^{(n)})$, Barnum defines a natural distortion based on entanglement fidelity, namely,

$$d_e\left(\rho^{(n)}, \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}\right) \equiv \sum_{\alpha=1}^{n} \frac{1}{n}(1 - F_e(\rho, \mathcal{T}^\alpha)) \tag{12}$$

with $\mathcal{T}^\alpha$ being the marginal operation on the αth copy of ρ induced by the encoding–decoding operation

$$\mathcal{T}^\alpha(\sigma) \equiv \mathrm{tr}_{1,\ldots,\alpha-1,\alpha+1,\ldots,n}\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(\rho\otimes\rho\cdots\otimes\rho\otimes\sigma\otimes\rho\cdots\otimes\rho). \quad (13)$$

We say that a rate-distortion triplet (R, r, d) is *achievable* for a given ρ iff there exists a sequence of (n, R, r) rate-distortion codes $(\mathcal{C}^{(n)}, \mathcal{D}^{(n)})$ such that

$$\lim_{n\to\infty} d_e\left(\rho^{(n)}, \mathcal{D}^{(n)}\circ\mathcal{C}^{(n)}\right) \leq d. \quad (14)$$

Then, the *rate-distortion manifold* $R(d, r)$ is defined as the infimum of all R for which (R, r, d) is achievable.

In the following, we approach the problem of finding $R(d, r)$ from first principles. Without loss of generality, the encoding procedure may be divided into two steps. In the first step, the encoder manipulates blocks of qubits of size n via some quantum operation

$$\mathcal{E}(\rho^{(n)}) = \sum_{i=1}^{k} A_i\rho^{(n)}A_i^\dagger.$$

For $\mathcal{E}$ to be physical its operation elements $\{A_i\}$ must satisfy the trace preserving condition

$$\sum_{i=1}^{k} A_i^\dagger A_i = I.$$

Define quantum operations

$$\mathcal{E}_{A_i}(\rho^{(n)}) = A_i\rho^{(n)}A_i^\dagger.$$

A given decomposition $\{A_i\}$ of unity implies that $\lambda_i = \mathrm{tr}\,(\mathcal{E}_{A_i}(\rho^{(n)}))$ is the probability that the non-trace-preserving operation $\mathcal{E}_{A_i}$ is the one that will be performed. Quantum mechanics forbids the encoder to have *control* over which of the k operations will get performed, but afterwards the encoder can obtain *information* about which one actually took place. This information is embodied in the index random variable I taking integer values i, $1\leq i\leq k$ with respective probabilities λ_i. In general, some or all of this information may be available to the decoder, embodied in the random variable $J = f(I)$, a deterministic function of I. Further define

$$\overline{S} = E_J S\left(E_{I|J}\hat{\mathcal{E}}_{A_I}\left(\rho^{(n)}\right)\right) \quad (15)$$

the average output von Neumann entropy conditional on the knowledge of J (i.e., from the point of view of somebody who knows the value of J but not the value of I). Given R and r, the goal is to choose $\mathcal{E}$ and f so that the distortion is minimized while keeping $\overline{S}\leq nR$ and $H(J)\leq nr$.

In the second step, we take a large number N of such blocks, group them according to the value of J, and process each group in the standard lossless coding way [2], [11], [9] in order to get a string of at most NnR qubits in the limit of large N. The decoding procedure is just reversing the second step, which the lossless coding theorem assures us can be done with effectively perfect fidelity in the limit as $N\to\infty$ (for fixed n), and using the Nnr bits of classical information about the values of J for each block so the decoder may unscramble them properly. Finally, the rate-distortion function will be achieved in the combined limit of large n and large N.

Since the distortion depends only on the operation elements A_i, the choice of f affects only the tradeoff between R and r. Using the concavity of von Neumann entropy [10] and the fact that $E_J E_{I|J} = E_I$, we have the following inequalities:

$$E_I S\left(\hat{\mathcal{E}}_{A_I}\left(\rho^{(n)}\right)\right) \leq \overline{S} \leq S\left(E_I\hat{\mathcal{E}}_{A_I}\left(\rho^{(n)}\right)\right) = S\left(\mathcal{E}\left(\rho^{(n)}\right)\right). \quad (16)$$

The upper bound is attained when $f = \mathrm{const}$, i.e., when no classical side information is allowed. The lower bound on $\overline{S}$ is attained when f is the identity map, in which case

$$H(J) = H(I) = -\sum_{i=1}^{k}\lambda_i\log_2\lambda_i$$

is maximum. An intuitive argument for the latter is that, from the point of view of the decoder, only single-element operations $\mathcal{E}_{A_i}$ have been performed; these in turn have zero entropy exchange, which we interpreted as noise. Whenever the decoder lacks information about the value of I, the entropy exchange of the block is strictly positive.

We shall henceforth concentrate on the case of maximal classical rate r, thus reducing the problem to finding the tradeoff function $R_n(d)$ between

$$\overline{S} = \sum_{i=1}^{k}\lambda_i S(\hat{\mathcal{E}}_{A_i}(\rho^{(n)}))$$

and the distortion $d_e(\rho^{(n)}, \mathcal{E})$. The rate-distortion function is given by the limit $R(d) = \lim_{n\to\infty}R_n(d)$. In the next section, we analyze the $n = 1$ case. Subsequently, we demonstrate the perhaps surprising fact that $n = 1$ already attains the $R(d)$ curve.

III. The Rate-Distortion Function for $n = 1$

Let us temporarily restrict attention to $k = 1$, so that (4) becomes $\mathcal{E}(\sigma) = A\sigma A^\dagger$, and also temporarily ignore the trace-preserving constraint. First a technical lemma.

Lemma 1: Let Δ and Λ be positive diagonal matrices whose diagonal elements are given in a nonascending order. Then for any unitary U and V, the inequality $|\mathrm{tr}\,(U\Delta V\Lambda)| \leq \mathrm{tr}\,(\Delta\Lambda)$ holds.

Proof: Consider the Cauchy–Schwartz inequality for the Hilbert–Schmidt inner product [8] $\langle A, B\rangle \equiv \mathrm{tr}\,(AB^\dagger)$, namely,

$$|\mathrm{tr}\,(AB^\dagger)|^2 \leq \mathrm{tr}\,(AA^\dagger)\mathrm{tr}\,(BB^\dagger). \quad (17)$$

Since Δ and Λ are positive we have $\Delta = \sqrt{\Delta\Delta^\dagger}$ and $\Lambda = \sqrt{\Lambda\Lambda^\dagger}$. Setting $A = \sqrt{\Delta}V\sqrt{\Lambda}$ and $B = \sqrt{\Delta}U^\dagger\sqrt{\Lambda}$, we find that

$$|\mathrm{tr}\,(U\Delta V\Lambda)|^2 \leq \mathrm{tr}\,(U\Delta U^\dagger\Lambda)\mathrm{tr}\,(V\Delta V^\dagger\Lambda) \quad (18)$$

so the lemma will hold for general unitary U and V provided it holds when $V = U^\dagger$. Next, denote the elements of U and diagonal elements of Δ and Λ by $\{u_{ij}\}$, $\{\delta_i\}$, and $\{\lambda_i\}$, respectively. Defining the matrix P with elements $p_{ij} = |u_{ij}|^2$, we have

$$tr(U\Delta U^\dagger\Lambda) = \sum_{i,j} u_{ij}\delta_j u_{ij}^*\lambda_i = \sum_{i,j} p_{ij}\delta_j\lambda_i. \quad (19)$$

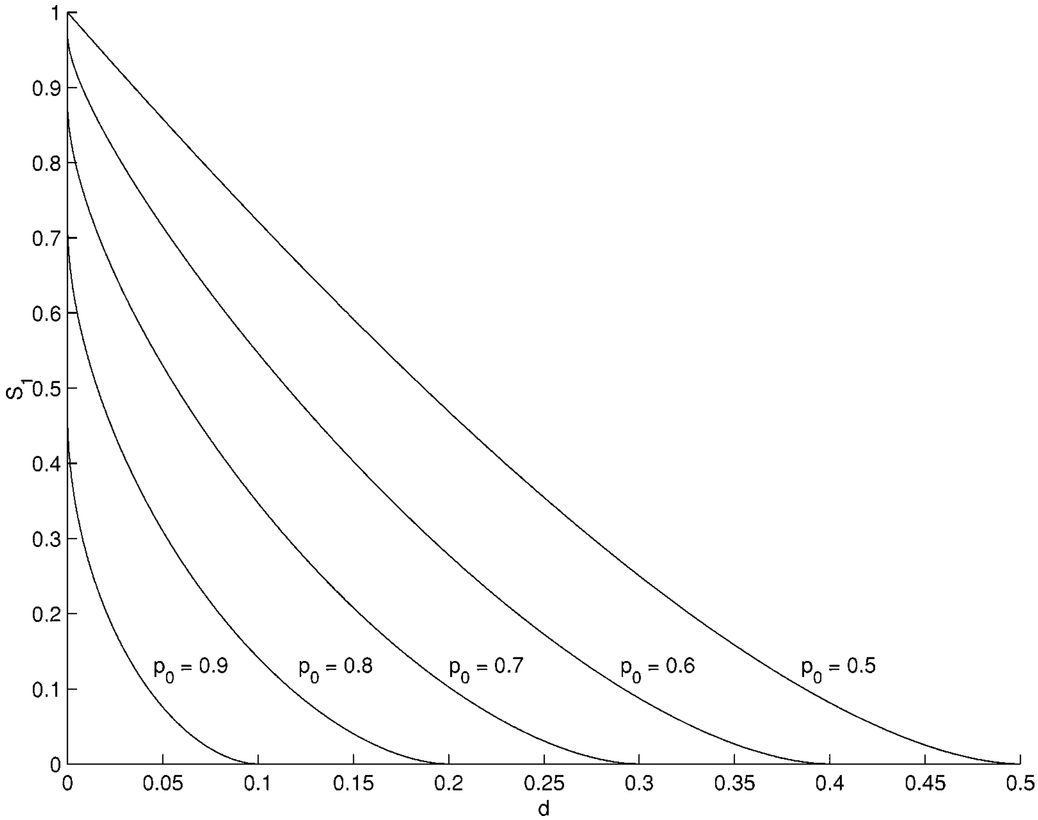

Fig. 1. Lower bound $S_1(d)$ on the single-qubit rate-distortion function for $p_0 = 0.5,\ 0.6,\ 0.7,\ 0.8$, and 0.9.

Since elements of each row and column of P add up to 1, P is a stochastic matrix, and hence a convex combination of permutation matrices [10]. So the maximum value of $tr(U\Delta U^\dagger \Lambda)$ is equal to $\sum_i \delta_i' \lambda_i$ with δ_i' a permutation of the δ_i. By Chebyshev's inequality, $P = I$ corresponds to the optimum permutation; this is especially easy to see for 2×2 matrices for which the ordering condition implies $(\lambda_1 - \lambda_2)(\delta_1 - \delta_2) \geq 0$, or $\lambda_1\delta_1 + \lambda_2\delta_2 \geq \lambda_1\delta_2 + \lambda_2\delta_1$. Therefore, $U = V = I$ maximizes $|\mathrm{tr}\,(U\Delta V\Lambda)|$; hence the lemma is proved.

Theorem 1: For all single qubit quantum operations $\mathcal{E}_A(\rho) = A\rho A^\dagger$, there exists a quantum operation $\mathcal{E}_D(\rho) = D\rho D^\dagger$ with $[D, \rho] = 0$ and D positive, of the same output entropy and smaller or equal distortion.

Proof: We work in the basis $\{|0\rangle, |1\rangle\}$ in which ρ is diagonal, so $\rho = p_0|0\rangle\langle 0| + p_1|1\rangle\langle 1|$ with $p_0 + p_1 = 1$ and $p_0 \geq p_1$. It is easy to see that any complex matrix A can be expressed as a product $A = UD\rho^{1/2}V\rho^{-1/2}$, where U and V are unitary and D is diagonal-positive and hence commutes with ρ. This follows from applying the polar decomposition of any complex matrix B, namely, $B = U\Delta V$. Here U and V are unitary, Δ is diagonal-positive with nonascending elements, and we choose $B = A\rho^{1/2}$ and $D = \Delta\rho^{-1/2}$. Such a decomposition ensures that $A\rho A^\dagger = U(D\rho D^\dagger)U^\dagger$, so that $tr(A\rho A^\dagger) = tr(D\rho D^\dagger)$ and $S(\hat{\mathcal{E}}_A) = S(\hat{\mathcal{E}}_D)$. In addition, since both $\Delta = D\rho^{1/2}$ and $\rho^{1/2}$ are diagonal-positive with nonascending elements, Lemma 1 asserts that $|\mathrm{tr}\,(A\rho)| \leq |\mathrm{tr}\,(D\rho)|$. Combining the above with the single-qubit distortion formula

$$d_e(\rho, \mathcal{E}_A) = 1 - \frac{|\mathrm{tr}\,(A\rho)|^2}{\mathrm{tr}\,(A\rho A^\dagger)} \tag{20}$$

we see that the operation $\mathcal{E}_D$ has the same output entropy but a distortion that is less than or equal to that of $\mathcal{E}_A$, thus proving the statement of the theorem. □

Since A is defined only up to a multiplicative constant, Theorem 1 implies a complete parametrization for the unphysical $n = k = 1$ curve, which we denote here by $S_1(d)$. It is easy to see that in the $\{|0\rangle, |1\rangle\}$ basis the matrix

$$A = \begin{pmatrix} \cos\theta & 0 \\ 0 & \sin\theta \end{pmatrix}, \qquad \theta \in \left[0, \frac{\pi}{4}\right] \tag{21}$$

interpolates between the zero distortion limit $A = I$, where $S = S(\rho)$, and the zero entropy limit $A = |0\rangle\langle 0|$, where we replace the source with the pure "best guess" state $|0\rangle\langle 0|$.

This curve, easily verified to be convex, is shown in Fig. 1 for several values of p_0. It is parametrized as

$$\begin{aligned} S_1(\Delta) &= h_2\left(\frac{p_0(1+\cos\Delta)}{(p_0+p_1)+(p_0-p_1)\cos\Delta}\right) \\ d(\Delta) &= \frac{p_0p_1(1-\sin\Delta)}{(p_0+p_1)+(p_0-p_1)\cos\Delta} \end{aligned} \tag{22}$$

where $\Delta \in [0, \frac{\pi}{2}]$. Here

$$h_2(\lambda) \equiv -\lambda\log_2(\lambda) - (1-\lambda)\log_2(1-\lambda)$$

is the Shannon binary entropy function. When $p_0 = \frac{1}{2}$ the above simplifies to

$$S_1(d) = h_2\left(\frac{1}{2} + \sqrt{d(1-d)}\right).$$

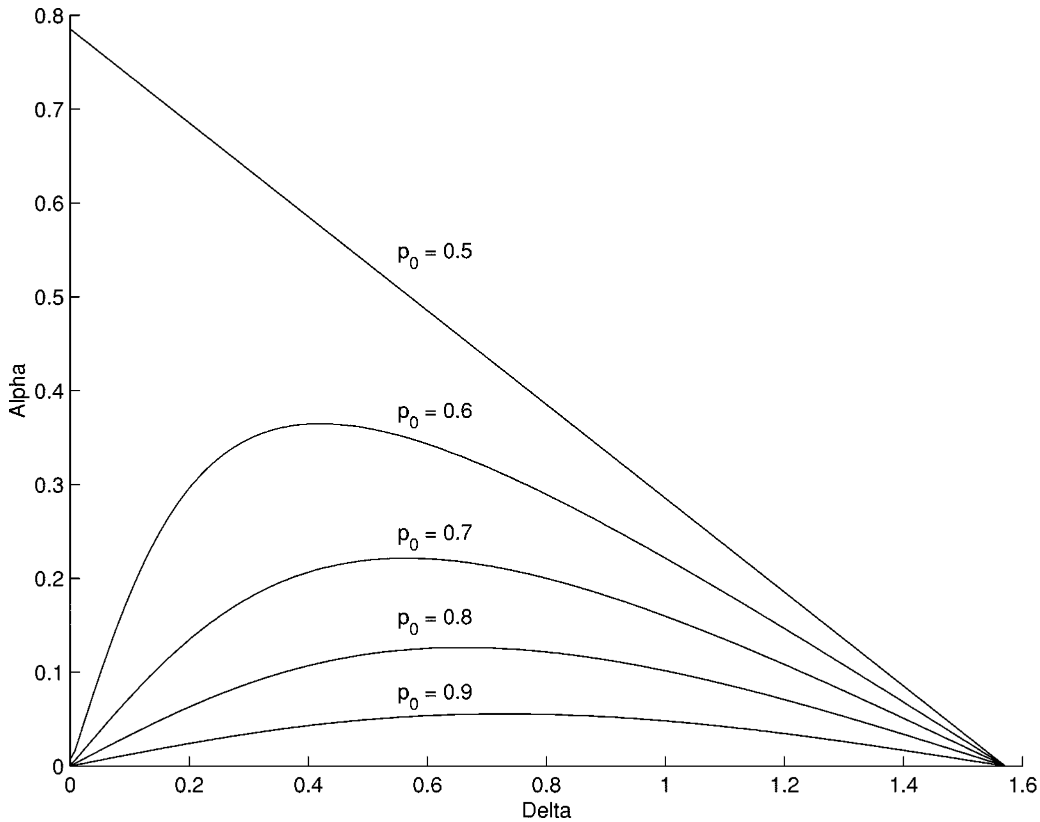

Fig. 2. The function $\alpha(\Delta)$ plotted for $p_0 = 0.5,\ 0.6,\ 0.7,\ 0.8$, and 0.9.

$S_1(d)$ serves as a lower bound for $R_1(d)$ since, for any decomposition of unity $\sum_i A_i^\dagger A_i = I$ and $\lambda_i = \mathrm{tr}\,(\mathcal{E}_{A_i}(\rho))$, we have

$$\overline{S} = \sum_{i=1}^{k} \lambda_i S\left(\hat{\mathcal{E}}_{A_i}(\rho)\right) \geq \sum_{i=1}^{k} \lambda_i S_1(d_e(\rho, \mathcal{E}_{A_i})) \geq S_1\left(\sum_{i=1}^{k} \lambda_i d_e(\rho, \mathcal{E}_{A_i})\right) \tag{23}$$

by the convexity of $S_1(d)$. In the case of $p_0 = \frac{1}{2}$, due to isotropy, this lower bound is attainable with $k = 2$

$$A_1 = \begin{pmatrix} \cos\theta & 0 \\ 0 & \sin\theta \end{pmatrix}, \quad A_2 = \begin{pmatrix} \sin\theta & 0 \\ 0 & \cos\theta \end{pmatrix}, \quad \theta \in \left[0,\ \frac{\pi}{4}\right]. \tag{24}$$

The case $p_0 > \frac{1}{2}$ is not as obvious. First, we would like to show that $k = 2$ suffices. We fix A_1 and vary A_i, $2 \leq i \leq k$. We use Lagrange multipliers μ and $\Lambda = (\lambda_{m,n})$ and seek the minimum of

$$\sum_{i=2}^{k} \mathrm{tr}\,(A_i \rho A_i^\dagger) S\left(\frac{A_i \rho A_i^\dagger}{tr(A_i \rho A_i^\dagger)}\right) - \mu \sum_{i=2}^{k} |\mathrm{tr}\,(A_i\rho)|^2 - \sum_{i=2}^{k} \mathrm{tr}\,(\Lambda A_i^\dagger A_i). \tag{25}$$

Differentiating $\overline{S}$ with respect to A_i and $A_i^\dagger$ and setting this to zero, we obtain an equation involving only A_i, $A_i^\dagger$, μ, and Λ, so evidently a solution is obtained for $A_2 = \cdots = A_k$. This has the same effect as retaining only A_2, so $k = 2$ includes natural solutions to the extremum problem. Motivated by the $p_0 = \frac{1}{2}$ case, we conjecture that the global *minimum* is among them.

Restricting attention to $k = 2$, we concentrate on the case where A_1 and A_2 are diagonal and use the parametrization

$$A_1 = \begin{pmatrix} \cos\alpha & 0 \\ 0 & \cos(\alpha+\Delta) \end{pmatrix}, \quad A_2 = \begin{pmatrix} \sin\alpha & 0 \\ 0 & \sin(\alpha+\Delta) \end{pmatrix}, \quad \Delta \in \left[0,\ \frac{\pi}{2}\right] \tag{26}$$

and $d = 2p_0p_1(1 - \cos\Delta)$. Here α is function of Δ such that

$$\overline{S} = \sum_{i=1}^{2} \mathrm{tr}\,(A_i \rho A_i^\dagger) S\left(\frac{A_i \rho A_i^\dagger}{tr(A_i \rho A_i^\dagger)}\right) \tag{27}$$

is minimized. Differentiating with respect to α, we arrive at

$$\begin{aligned} &2p_0p_1 \sin\Delta \left(\log_2\left(\frac{p_1 \cos^2(\alpha+\Delta)}{p_0 \cos^2\alpha}\right)\right. \\ &\quad \cdot \frac{\cos\alpha\cos(\alpha+\Delta)}{p_0\cos\alpha + p_1\cos(\alpha+\Delta)} \\ &\quad \left. + \log_2\left(\frac{p_1 \sin^2(\alpha+\Delta)}{p_0 \sin^2\alpha}\right) \frac{\sin\alpha\sin(\alpha+\Delta)}{p_0\sin\alpha + p_1\sin(\alpha+\Delta)}\right) \\ &\quad + (p_0 \sin 2\alpha + p_1 \sin 2(\alpha+\Delta)) \\ &\quad \cdot \left(h_2\left(\frac{p_0 \sin^2\alpha}{p_0\sin^2\alpha + p_1\sin^2(\alpha+\Delta)}\right)\right. \\ &\quad \left. - h_2\left(\frac{p_0\cos^2\alpha}{p_0\cos^2\alpha + p_1\cos^2(\alpha+\Delta)}\right)\right) = 0 \end{aligned}$$

which we solve numerically. The function $\alpha(\Delta)$ is plotted in Fig. 2 for several values of p_0. We also plot the corresponding

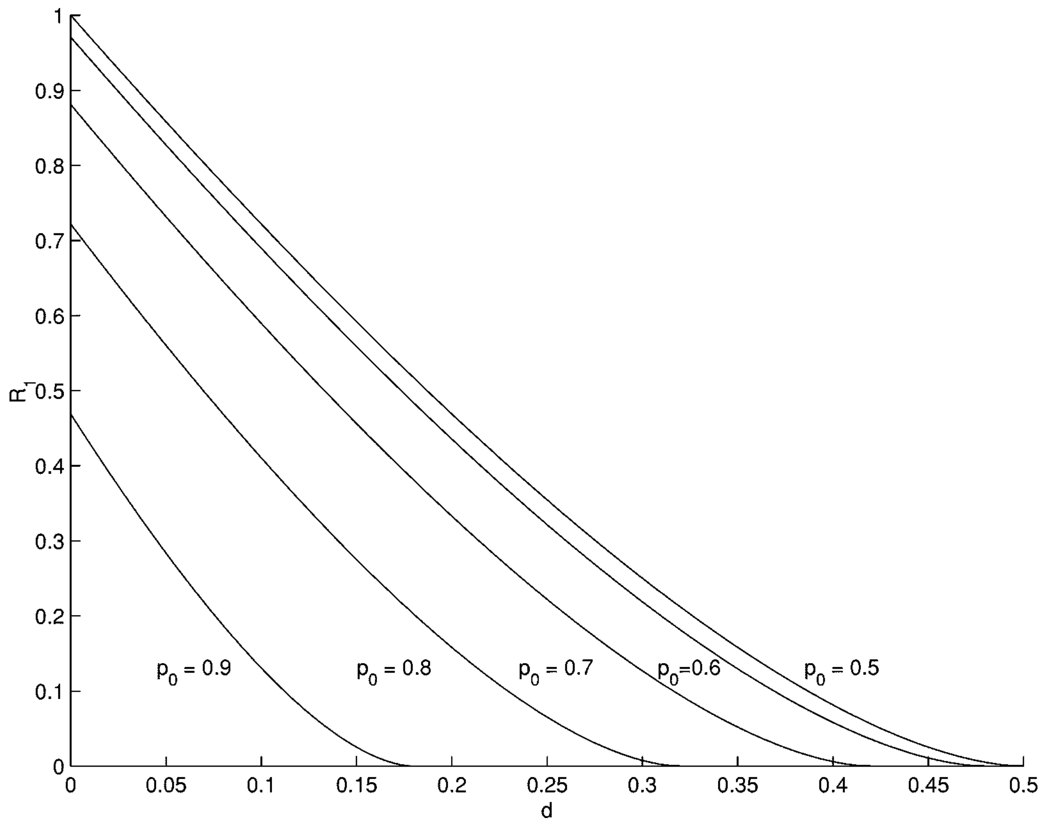

Fig. 3. The single-qubit rate-distortion function $R_1(d)$ plotted for $p_0 = 0.5,\ 0.6,\ 0.7,\ 0.8$, and 0.9.

rate-distortion curves in Fig. 3. The curves are convex and approach $d_{\max} = 2p_0p_1$ with zero slope. Note that the $p_0 = \frac{1}{2}$ solution is precisely the one obtained previously, namely, $S_1(d)$.

Now we show that this diagonal solution is optimal with respect to local perturbations of the $\{A_i\}$. Recall that we wish to find the optimal tradeoff between $\overline{S}$ and $d = 1 - \sum_i |\mathrm{tr}\,(A_i\rho)|^2$ under the constraint $\sum_i A_i^\dagger A_i = I$. Notice that both $\overline{S}$ and the trace-preserving condition are invariant under the transformation $A_i \to U_iA_i$ where U_i are unitary matrices. Furthermore, $|\mathrm{tr}\,(U_iA_i\rho)| \leq |\mathrm{tr}\,(A_i\rho)|$ when $A_i\rho$ is positive (see Lemma 2 below), and we may always pick U_i to achieve this upper bound. This can be seen from the polar decomposition $A_i\rho = V_iD_iW_i$ and choosing $U_i = (V_iW_i)^{-1}$. Therefore, we restrict attention to positive $A_i\rho$ and use a new parametrization

$$A_1 = f\begin{pmatrix} \frac{\lambda\cos\theta}{p_0} & \frac{x\sin\theta}{p_1} \\ \frac{x^*\sin\theta}{p_0} & \frac{(1-\lambda)\cos\theta}{p_1} \end{pmatrix}$$
$$A_2 = f\begin{pmatrix} \frac{\mu\sin\theta}{p_0} & -\frac{x\cos\theta}{p_1} \\ -\frac{x^*\cos\theta}{p_0} & \frac{(1-\mu)\sin\theta}{p_1} \end{pmatrix} \tag{28}$$

in terms of θ and complex x. Here λ and μ are functions of $|x|$ determined by the conditions

$$\lambda^2\cos^2\theta + \mu^2\sin^2\theta = \frac{p_0^2}{f^2} - |x|^2 \tag{29}$$

$$(1-\lambda)^2\cos^2\theta + (1-\mu)^2\sin^2\theta = \frac{p_1^2}{f^2} - |x|^2 \tag{30}$$

and $d = 1 - f^2$. We see from the expansion about $x = 0$ that λ and μ are both quadratic in $|x|$. It is also easy to see that the traces and determinants of the $A_i\rho A_i^\dagger$ (and hence the eigenvalues) also have no terms linear in x. Expanding to second order about the optimal diagonal solution, we verify that $\overline{S}$ is indeed at a local maximum with respect to varying x. We thus conclude our argument that the $n = 1$ rate-distortion curves $R_1(d)$ are those depicted in Fig. 3.

IV. The Rate-Distortion Function for General n

Now we move to the general n case and argue that we cannot do any better than $R_1(d)$. We have n qubits with joint density operator $\rho^{\otimes n}$, and we consider appropriate combinations of quantum operations $\mathcal{E}_A(\rho^{\otimes n}) = A(\rho^{\otimes n})A^\dagger$. We work in the basis $\mathcal{B}^n = \{|0\rangle, |1\rangle\}^n$ with $|0\rangle$ and $|1\rangle$ defined as before. In this basis, the system operator A is given by

$$A = \begin{pmatrix} B & K \\ L & C \end{pmatrix} \tag{31}$$

where the B, K, L, and C are $2^{n-1} \times 2^{n-1}$ matrices acting on the last $n-1$ qubits. It is easy to verify that the restriction $\mathcal{E}^>$ of $\mathcal{E}$ to the last $n-1$ qubits is given by the set $\{\sqrt{p_0}B,\ \sqrt{p_1}K,\ \sqrt{p_0}L,\ \sqrt{p_1}C\}$ of operation elements.

We first restrict attention to processes with A diagonal in the $\mathcal{B}^n$ basis.

Theorem 2: General n-qubit trace-preserving processes with operation elements $\{A_i\}$ diagonal in the $\mathcal{B}^n$ basis cannot perform below the single-qubit rate-distortion curve $R_1(d)$.

Proof: We prove the theorem using induction on n. It is true for $n = 1$ by the results of the previous section. Let us now assume it holds for n, and then show its validity for $n+1$. We work in the $\mathcal{B}^{n+1}$ basis where A_i is represented by a $2^{n+1} \times 2^{n+1}$ dimensional matrix

$$A_i = \begin{pmatrix} \frac{1}{\sqrt{p_0}} B_i & \\ & \frac{1}{\sqrt{p_1}} C_i \end{pmatrix} \tag{32}$$

with B_i and C_i both diagonal $2^n \times 2^n$ matrices acting on the last n qubits. Then the projection of $\mathcal{E}_{A_i}$ onto the last n qubits is $\mathcal{E}^{>}_{A_i}(\rho^{\otimes n}) = B_i \rho^{\otimes n} B_i^\dagger + C_i \rho^{\otimes n} C_i^\dagger$. We also have from (32) that

$$\mathcal{E}_{A_i}(\rho^{\otimes n+1}) = \begin{pmatrix} B_i \rho^{\otimes n} B_i^\dagger & \\ & C_i \rho^{\otimes n} C_i^\dagger \end{pmatrix}. \tag{33}$$

Then the normalized projection of $\mathcal{E}_{A_i}$ onto the first qubit is

$$\hat{\mathcal{E}}^1_{A_i}(\rho) = \begin{pmatrix} \lambda_i & \\ & 1-\lambda_i \end{pmatrix} \tag{34}$$

where $\lambda_i = \mathrm{tr}\,(\mathcal{E}_{B_i}(\rho^{\otimes n}))/tr(\mathcal{E}_{A_i}(\rho^{\otimes n+1}))$.

The average distortion associated with the coding procedure defined by the $\{A_i\}$ is

$$d = \frac{n}{n+1}\, d^{>} + \frac{1}{n+1}\, d^1 \tag{35}$$

where

$$d^{>} = \sum_i tr(\mathcal{E}_{B_i}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{B_i}) + \mathrm{tr}\,(\mathcal{E}_{C_i}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{C_i}) \tag{36}$$

and

$$d^1 = \sum_i d_e(\rho, \mathcal{E}^1_{A_i}). \tag{37}$$

Using the simple identity

$$S(\lambda\rho_1 \oplus (1-\lambda)\rho_2) = \lambda S(\rho_1) + (1-\lambda) S(\rho_2) + h_2(\lambda) \tag{38}$$

we find that

$$S(\hat{\mathcal{E}}_{A_i}(\rho^{\otimes n+1}) = \lambda_i S(\hat{\mathcal{E}}_{B_i}(\rho^{\otimes n})) + (1-\lambda_i) S(\hat{\mathcal{E}}_{C_i}(\rho^{\otimes n})) + h_2(\lambda_i). \tag{39}$$

Hence

$$\begin{aligned} &\frac{1}{n+1} \sum_i \mathrm{tr}\,(\mathcal{E}_{A_i}(\rho^{\otimes n+1})) S(\hat{\mathcal{E}}_{A_i}(\rho^{\otimes n+1})) \\ &= \frac{n}{n+1} \Bigg(\frac{1}{n} \sum_i \mathrm{tr}\,(\mathcal{E}_{B_i}(\rho^{\otimes n})) S(\hat{\mathcal{E}}_{B_i}(\rho^{\otimes n})) \\ &\qquad + \mathrm{tr}\,(\mathcal{E}_{C_i}(\rho^{\otimes n})) S(\hat{\mathcal{E}}_{C_i}(\rho^{\otimes n})) \Bigg) \\ &\quad + \frac{1}{n+1} \sum_i \mathrm{tr}\,(\mathcal{E}^1_{A_i}(\rho)) S(\hat{\mathcal{E}}^1_{A_i}(\rho)) \\ &\geq \frac{n}{n+1}\, R_1(d^{>}) + \frac{1}{n+1}\, R_1(d^1) \geq R_1(d). \end{aligned} \tag{40}$$

The equality comes from (34), (39), and the fact that

$$tr(\mathcal{E}_{A_i}(\rho^{\otimes n+1})) = \mathrm{tr}\,(\hat{\mathcal{E}}^1_{A_i}(\rho));$$

the first inequality comes from the inductive hypothesis, and the second inequality is a consequence of convexity of $R_1(d)$ and (35). Hence, the rate for $\{A_i\}$ is greater than or equal to $R_1(d)$ at the same distortion, as claimed. □

Finally, it remains to show that diagonal processes are optimal for general n. This may be shown exactly in the case $p_0 = \frac{1}{2}$ due to its many simplifying features. We begin with two lemmas.

Lemma 2: Given matrices $\{Y_i\}$ with $\sum_i Y_i^\dagger Y_i = I$ and positive D, the inequality $\sum_i |\mathrm{tr}\,(Y_i D)|^2 \leq |\mathrm{tr}\,(D)|^2$ holds.

Proof: We use the fact that $D = \sqrt{DD^\dagger}$ for D positive and employ the Cauchy–Schwartz inequality (17) to write

$$\begin{aligned} \sum_i |\mathrm{tr}\,(Y_i D)|^2 &= \sum_i \left| \mathrm{tr}\, \left(\left(Y_i \sqrt{D} \right) \sqrt{D^\dagger} \right) \right|^2 \\ &\leq \sum_i \mathrm{tr}\,(Y_i D Y_i^\dagger) \mathrm{tr}\,(D) = |\mathrm{tr}\,(D)|^2. \end{aligned} \tag{41}$$

The last equality comes from the cyclicity and linearity of trace. □

Lemma 3: Given operators $\{Y_i\}$ acting on n qubits with $\sum_i Y_i^\dagger Y_i = I$ and positive D, diagonal in the $\mathcal{B}^n$ basis, we have the inequality

$$\sum_i \mathrm{tr}\,(\mathcal{E}_{Y_i D}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{Y_i D}) \geq \mathrm{tr}\,(\mathcal{E}_D(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_D).$$

Proof: We again use induction. The $n = 1$ case follows from Lemma 2. Assuming the lemma holds for n we prove it for $n+1$. Consider $2^{n+1} \times 2^{n+1}$-dimensional matrices $\{Y_i\}$, and let

$$Y_i = \begin{pmatrix} E_i & F_i \\ G_i & H_i \end{pmatrix} \qquad D = \begin{pmatrix} \frac{1}{\sqrt{p_0}} D_0 & \\ & \frac{1}{\sqrt{p_1}} D_1 \end{pmatrix} \tag{42}$$

with E_i, etc., of dimension $2^n \times 2^n$. $\sum_i Y_i^\dagger Y_i = I$ implies that

$$\sum_i (E_i^\dagger E_i + G_i^\dagger G_i) = I \tag{43}$$

and similarly for F_i and H_i. The restriction $\mathcal{E}^{>}_{Y_i D}$ of $\mathcal{E}_{Y_i D}$ onto the last n qubits is described by the set $\{E_i D_0, F_i D_1, G_i D_0, H_i D_1\}$. Then

$$\begin{aligned} &\sum_i \mathrm{tr}\,(\mathcal{E}^{>}_{Y_i D}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}^{>}_{Y_i D}) \\ &= \sum_i \mathrm{tr}\,(\mathcal{E}_{E_i D_0}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{E_i D_0}) \\ &\quad + \mathrm{tr}\,(\mathcal{E}_{F_i D_1}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{F_i D_1}) \\ &\quad + \mathrm{tr}\,(\mathcal{E}_{G_i D_0}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{G_i D_0}) \\ &\quad + \mathrm{tr}\,(\mathcal{E}_{H_i D_1}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{H_i D_1}) \\ &\geq \sum_i \mathrm{tr}\,(\mathcal{E}_{D_0}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{D_0}) \\ &\quad + \mathrm{tr}\,(\mathcal{E}_{D_1}(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}_{D_1}) \\ &= \mathrm{tr}\,(\mathcal{E}^{>}_D(\rho^{\otimes n})) d_e(\rho^{\otimes n}, \mathcal{E}^{>}_D). \end{aligned} \tag{44}$$

The inequality comes from the inductive hypothesis and (43). Finally, this result is invariant under permutations of the qubits; averaging over all permutations yields

$$\sum_i \mathrm{tr}\,(\mathcal{E}_{Y_i D}(\rho^{\otimes n+1})) d_e(\rho^{\otimes n+1}, \mathcal{E}_{Y_i D}) \geq \mathrm{tr}\,(\mathcal{E}_D(\rho^{\otimes n+1})) d_e(\rho^{\otimes n+1}, \mathcal{E}_D). \tag{45}$$

This proves the lemma. □

Theorem 3: General n-qubit processes cannot perform below the single-qubit entropy-distortion curve $S_1(d)$ in the case of isotropic sources ($p_0 = \frac{1}{2}$).

Proof: This is an immediate consequence of Lemma 3. We ignore the trace-preserving condition for the time being and consider $\mathcal{E}_A(\rho^{\otimes n}) = A(\rho^{\otimes n})A^\dagger$. Then we use the polar decomposition $A = UDV$ with U and V unitary and D diagonal-positive. Using the fact that $\rho = \frac{1}{2}I$, it is easy to see that

$$tr(\mathcal{E}_A(\rho^{\otimes n})) = \mathrm{tr}\,(\mathcal{E}_D(\rho^{\otimes n}))$$
$$S(\hat{\mathcal{E}}_A(\rho^{\otimes n})) = S(\hat{\mathcal{E}}_D(\rho^{\otimes n}))$$

and

$$d_e(\rho^{\otimes n}, \mathcal{E}_A) = d_e(\rho^{\otimes n}, \mathcal{E}_{VUD}).$$

Then from Lemma 3 with $m = 1$ and $Y_1 = VU$, we get

$$d_e(\rho^{\otimes n}, \mathcal{E}_A) \geq d_e(\rho^{\otimes n}, \mathcal{E}_D).$$

Therefore, there is a diagonal map that does at least as well as $\mathcal{E}_A$. From a trivial variation on Theorem 2 (note that the trace-preserving condition plays no role in the proof), this diagonal map cannot do better than the $n = k = 1$ curve $S_1(d)$ which is attainable for $p_0 = \frac{1}{2}$. Having established that the optimal $\mathcal{E}_A$ yields the convex $S_1(d)$, using the same argument as in (23) we see that reintroducing the trace-preserving condition does not affect our result. Hence the theorem is proved. □

We conjecture that the theorem also holds for the case $p_0 > \frac{1}{2}$, and we now present some evidence to support this conjecture. It again suffices to show that diagonal processes are optimal for general n.

- Consider perturbing a process defined by $2^n \times 2^n$-dimensional diagonal matrices $\{A_i\}$ with $\sum_i A_i^\dagger A_i = I$ by a general matrix $\{Q_i\}$ with diagonal elements all equal to zero. It is easy to see that to *linear* order, the trace-preserving condition still holds, and both average entropy and distortion remain unchanged. Hence, all diagonal processes are local extrema with respect to off-diagonal perturbations.
- In Theorem 2 we never used the fact that B_i and C_i were diagonal, so a more general class of operators given by (32), in $\mathcal{B}^n$ or any other basis obtained by permutations of the qubits, lies above the $R_1(d)$ curve.
- A straightforward modification of Theorem 3 shows that diagonal processes D_i do better than $U_i D_i$, where U_i is any unitary operator (note that the trace-preserving condition still holds).
- By iterating the argument preceding Theorem 2, the restriction of a general n-qubit operation onto a single qubit involves 2^{n-1} operation elements which greatly increases the entropy exchange with the environment of that qubit. Essentially, individual qubits act as the environment for each other, and entangling them creates noise. On the other hand, as in classical information theory, the benefit of entangling (correlating) the qubits is a reduction in entropy since $S(\mathcal{E}(\rho^{\otimes n})) \leq \sum_\alpha S(\mathcal{E}^\alpha(\rho))$, where $\mathcal{E}^\alpha$ is the restriction of $\mathcal{E}$ to the αth qubit. There is a competition between these two effects, and the former wins, as we have proven rigorously for $p_0 = \frac{1}{2}$. In this sense, however, there is nothing special about $p_0 = \frac{1}{2}$. If anything, we would expect the entropy to be even harder to reduce via quantum operations for $p_0 > \frac{1}{2}$ than for $p_0 = \frac{1}{2}$ because it is lower to start with.

V. Physical Realization of the $R(d)$ Curve

We now elaborate on how our coding procedure may be realized physically. For the lossy part of the coding we need to provide an ancilla qubit in a definite state. We first apply a unitary transformation entangling the ancilla with the source qubit, and then measure the ancilla. In the basis $\{|0\rangle_A|0\rangle_Q, |0\rangle_A|1\rangle_Q, |1\rangle_A|0\rangle_Q, |1\rangle_A|1\rangle_Q\}$, the unitary transformation is given by the matrix

$$U = \begin{pmatrix} \cos\alpha & & -\sin\alpha & \\ & \cos(\alpha+\Delta) & & -\sin(\alpha+\Delta) \\ \sin\alpha & & \cos\alpha & \\ & \sin(\alpha+\Delta) & & \cos(\alpha+\Delta) \end{pmatrix} \tag{46}$$

with $\Delta \in [0, \frac{\pi}{2}]$ and $\alpha = \alpha(\Delta)$ as defined before. The ancilla is prepared in the $|0\rangle_A$ state so that the initial density operator for the ancilla-source system is

$$\Xi = \begin{pmatrix} \rho & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{47}$$

Then

$$U\Xi U^\dagger = \begin{pmatrix} A_1\rho A_1^\dagger & A_1\rho A_2^\dagger \\ A_2\rho A_1^\dagger & A_2\rho A_2^\dagger \end{pmatrix} \tag{48}$$

where A_1 and A_2 are as defined in (26). We then measure the ancilla qubit. If the outcome is $|0\rangle_A$, we know the map $\rho \rightarrow \hat{\mathcal{E}}_{A_1}(\rho)$ has been performed and we label the qubit as belonging to type 1. Similarly, if the outcome is $|1\rangle_A$, we know the map $\rho \rightarrow \hat{\mathcal{E}}_{A_2}(\rho)$ has transpired and label the qubit to be of type 2. In the end we perform two Schumacher encodings, one on all the qubits of the first type and a separate one on all the bits of the second type. When decoding, we need information about the sequence of operations performed. The rate of classical information required for this is $r = h_2(tr(A_1\rho A_1^\dagger))$. These classical rates are plotted for several values of p_0 in Fig. 4.

VI. Discussion

We have shown that for the distortion measure in question and when allowed an unrestricted classical side channel, optimum quantum rate-distortion codes are separable into a lossy part involving single qubit operations followed by standard Schumacher lossless coding of large blocks of qubits.

Our result has the following interpretation: the rate-distortion curve is achieved by quantum operations that produce no entropy exchange with the environment of any individual qubit. We do not expect zero entropy exchange to be optimal for more general distortion measures. Since our distortion measure, which is based on the concept of entanglement fidelity, emphasizes preserving the state of RQ, it forbids any increase of the entropy of RQ which means it forbids any entropy exchange. We also do not believe $n = 1$ to be optimal when restrictions on r are imposed since, as remarked in Section II, the entropy exchange is positive as long as there is uncertainty in the value of the index random variable I.

Let us examine the action of our quantum map on normalized pure states. If we picture $|0\rangle$ and $|1\rangle$ as orthogonal vectors then, depending on which of the two operations has been performed,

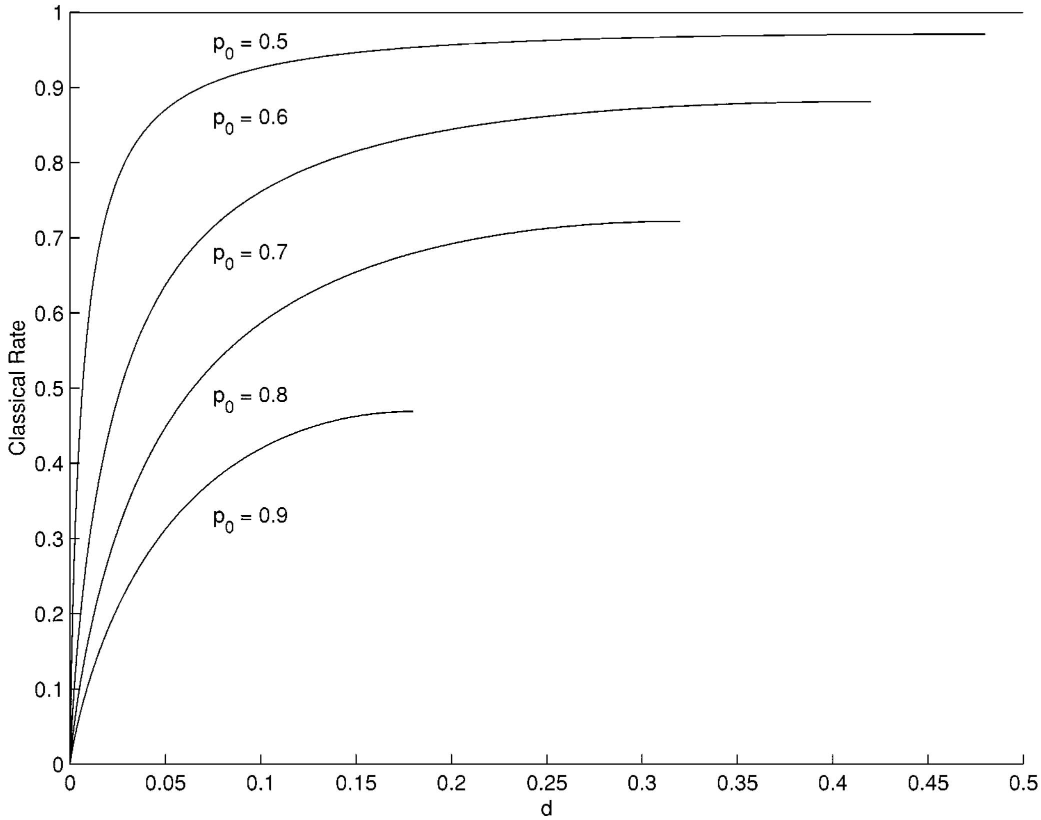

Fig. 4. The classical information rate needed to decode versus d for $p_0 = 0.5, 0.6, 0.7, 0.8,$ and 0.9.

the map rotates our pure state vector toward $|0\rangle$ or toward $|1\rangle$. The source is biased toward $|0\rangle$, which it produces with a higher probability than $|1\rangle$. The first type of operation produces qubits biased even more toward $|0\rangle$, hence causing a decrease in entropy. The second type does the opposite and perhaps even increases the entropy for $p_0 > \frac{1}{2}$; however, it has to occur a certain fraction of the time in order to obey the trace-preserving condition, which says that the total probability of performing *some* operation must be equal to 1 regardless of the input state. On average, the entropy does decrease, while the discrepancy between the initial and final state increases. The $R(d)$ curve is thus swept out.

Notice that our quantum $R(d)$ curve first reaches $R = 0$ at $d_{\max} = 2p_0p_1$, as opposed to the classical value $d_{\max} = p_1$ associated with reconstructing the source bit with the best guess at its value. This, too, is due to our choice of fidelity measure: replacing the original qubit with a fresh one prepared in the state $|0\rangle$ destroys the entanglement with the original reference system. The best we can do is project onto $|0\rangle$ with probability p_0 and otherwise project onto $|1\rangle$.

We do not expect a general expression resembling the classical prescription (1) for the rate-distortion function that is valid for all distortion measures to exist for quantum rate distortion. Our reason for this lies in the richness of distortion measures which vary in their degree of "quantumness." The one we have used based on entanglement fidelity evidently has a highly quantum nature. On the other hand, we could view ρ as being realized by a specific ensemble like $\mathcal{Q} = \{(|0\rangle, p_0), (|1\rangle, p_1)\}$, and use as our distortion measure the corresponding average pure state distortion measure $\overline{d}(\mathcal{Q}^n, \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)})$ based on the average pure state fidelity $\overline{F}(\mathcal{Q}, \mathcal{E})$, namely,

$$\overline{F}(\mathcal{Q}, \mathcal{E}) = p_0\langle 0|\mathcal{E}(|0\rangle\langle 0|)|0\rangle + p_1\langle 1|\mathcal{E}(|1\rangle\langle 1|)|1\rangle. \quad (49)$$

Here we are able to attain zero distortion merely by sending classical information—the measurement results in the $\{|0\rangle, |1\rangle\}$ basis. If we do not allow storing classical information, then the appropriate cross section of the rate-distortion manifold becomes the classical rate-distortion function for the Hamming measure, namely, $R(d, 0) = S(\rho) - h_2(d)$.

One could also investigate more general ensembles, as well as distortion measures tied to specific quantum cryptography protocols. Finally, the work presented here naturally generalizes to systems with more than two degrees of freedom.

ACKNOWLEDGMENT

The authors would like to thank K. Banaszek, H. Barnum, D. Mermin, I. Walmsley, and anonymous referees for valuable comments and particularly for pointing out problem formulation inadequacies in earlier versions of the manuscript.

REFERENCES

[1] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[2] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, p. 2738, 1995.
[3] S. Lloyd, "Capacity of the noisy quantum channel," *Phys. Rev. A*, vol. 55, p. 1613, 1996.
[4] H. Barnum, P. Hayden, R. Jozsa, and A. Winter, "On the reversible extraction of classical information from a quantum source," LANL preprint quant-ph/0011072.

[5] B. Schumacher and M. A. Nielsen, "Quantum data processing and error correction," *Phys. Rev. A*, vol. 54, p. 2629, 1996.
[6] H. Barnum, "Quantum rate-distortion coding," *Phys. Rev. A*, vol. 62, p. 42309, 2000.
[7] B. Schumacher, "Sending entanglement through noisy quantum channels," *Phys. Rev. A*, vol. 54, p. 2614, 1995.
[8] H. Barnum, M. A. Nielsen, and B. Schumacher, "Information transmission through a noisy quantum channel," *Phys. Rev. A*, vol. 57, p. 4153, 1998.
[9] I. L. Chuang and D. S. Modha, "Reversible arithmetic coding for quantum data compression," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1104–1116, May 2000.
[10] A. Wehrl, "General properties of entropy," *Rev. Mod. Phys.*, vol. 50, p. 221, 1978.
[11] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Mod. Opt.*, vol. 41, p. 2343, 1994.