Quantum capacity is properly defined without encodings

Howard Barnum,[1,*] John A. Smolin,[2,†] and Barbara M. Terhal[3,‡]
[1]*School of Natural Sciences, Hampshire College and Institute for Science and Interdisciplinary Studies, Amherst, Massachusetts 01022*
[2]*IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York 10598*
[3]*Faculteit Wiskunde Informatica, Natuur- en Sterrenkunde, Universiteit van Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands*
and Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
(Received 18 November 1997)

We show that no source encoding is needed in the definition of the capacity of a quantum channel for carrying quantum information. This allows us to use the coherent information maximized over all sources and block sizes, but not encodings, to bound the quantum capacity. We perform an explicit calculation of this maximum coherent information for the quantum erasure channel and apply the bound in order find the erasure channel's capacity without relying on an unproven assumption as in an earlier paper.
[S1050-2947(98)00911-1]

PACS number(s): 03.67.Hk, 03.65.Bz, 89.70.+c

I. INTRODUCTION

In recent years the field of quantum information theory has emerged. One of the central issues in this field is the concept of quantum channel capacity. Several papers have discussed the capacity of noisy quantum channels to carry quantum information [1–6]. Unfortunately, defining and calculating the quantum capacity has turned out to be difficult because of the specific (and sometimes odd) features of quantum information. Various types of capacities of quantum channels have also been defined, such as the capacity of a quantum channel to carry classical information [7,8], the capacities of quantum channels to carry quantum information with the assistance of classical side channels [4], and a capacity based on a quantum analog of the Shannon mutual information [9]. Here we will concentrate on just one type of quantum capacity.

Barnum, Nielsen, and Schumacher [5] have given a definition of quantum capacity $Q_E(\chi)$ of a channel χ in terms of the entanglement fidelity and the von Neumann entropy $S(\rho)\equiv -\mathrm{Tr}\ \rho \log_2 \rho$ of the source's density matrix ρ. The entanglement fidelity of a density matrix ρ relative to a linear trace-preserving completely positive map $\mathcal{E}$ [10] is defined as

$$F_e(\rho,\mathcal{E})=\langle \eta|(\mathcal{I}\otimes\mathcal{E})(|\eta\rangle\langle\eta|)|\eta\rangle, \tag{1}$$

where $|\eta\rangle$ is any purification of ρ. A *purification* [11] of any density matrix ρ in a Hilbert space $\mathcal{H}$ is any pure state $|\eta\rangle$ in a tensor product space $\mathcal{H}_A\otimes\mathcal{H}_B$ such that $\mathrm{Tr}_A|\eta\rangle\langle\eta|=\rho$. In Eq. (1) the identity operates on the purification space $\mathcal{H}_A$ and $\mathcal{E}$ operates on $\mathcal{H}_B$. Note that $F_e(\rho,\mathcal{E})$ is independent of the choice of purification [1].

Definition 1. The entanglement capacity Q_E of a channel χ is

$$Q_E(\chi)\equiv\sup\left\{ q:\forall\ \ \epsilon>0,\exists\ \ \mathcal{E},\mathcal{D},\rho,N:\frac{S(\rho)}{N}=q,\right.$$
$$\left. F_e(\rho,\mathcal{D}\circ\chi^{\otimes N}\circ\mathcal{E})>1-\epsilon\right\}. \tag{2}$$

That is, roughly, Q_E is the highest entropy per use of the channel that can be sent reliably using block coding. Here the density operator ρ is on a block of N copies of the input Hilbert space and the encoding and decoding operations $\mathcal{E}$ and $\mathcal{D}$ (which are linear trace-preserving completely positive maps) act on such block density operators. The definition requires that arbitrarily high entanglement fidelities may be achieved, possibly by going to larger and larger block size N. It does not require, however, that arbitrarily high fidelity be achievable for some fixed block size N. It is immediately apparent from the definition that one may bound this capacity below by some constant r (for rate) by exhibiting a sequence (in N) of source density operators and coding schemes such that the entropy of the source operators goes to r and the entanglement fidelity of the operators under the total operation goes to 1 with large N. We will say such a sequence of triplets $(\rho,\mathcal{E},\mathcal{D})$ *achieves the rate r*.

The definition of Q_E uses the entropy of the source ρ as a measure of the information that is sent through the channel rather than the entropy of the output signal $(\mathcal{D}\circ\chi^{\otimes N}\circ\mathcal{E})(\rho)$. One might argue that since capacity is about sending entropy to the channel output one should consider a definition Q^{out} in which the entropy of the output signal appears in place of the entropy of the input ρ as in Q_E. However, in general, as the decoding process $\mathcal{D}$ need not be unitary (and indeed it cannot be if it is to extract the noise from the output signal) it can map the signal onto an arbitrarily large Hilbert space and the output entropy can become unboundedly large. This implies that Q^{out} is not a good measure of the total amount of information that is sent through the channel. The problem is that for any pure state there exist density matrices of high fidelity relative to that pure state that have arbitrarily high entropy. Consider the density matrix $\rho=(1-\epsilon)|\psi\rangle\langle\psi| +(\epsilon/n)\Sigma_{i=1}^{n}|i\rangle\langle i|$ with the $|i\rangle$'s an orthonormal set of vec-

*Electronic address: hbarnum@hampshire.edu
†Electronic address: smolin@watson.ibm.com
‡Electronic address: terhal@phys.uva.nl

tors orthogonal to $|\psi\rangle$. This density matrix has entropy $H_2(\epsilon)+\epsilon \log_2 n$ and fidelity $1-\epsilon$ relative to $|\psi\rangle$ for any ϵ and any n. $[H_2(\epsilon)=-\epsilon \log_2(\epsilon)-(1-\epsilon)\log_2(1-\epsilon')$ is the binary entropy function.]

Another quantity that has been of interest is the coherent information [1,3].

Definition 2. The coherent information of a density matrix ρ and a linear trace-preserving completely positive map $\mathcal{E}$ is

$$I_c(\rho,\mathcal{E})=S(\mathcal{E}(\rho))-S_{\text{env}}(\rho,\mathcal{E}), \tag{3}$$

where $S_{\text{env}}(\rho,\mathcal{E})$ is the final entropy of an initially pure environment implementing χ [10].

Barnum, Nielsen, and Schumacher [5] have shown that

$$Q_E \leqslant I_{\max} \equiv \sup_N \max_{\rho,\mathcal{E}} \frac{I_c(\rho,\chi^{\otimes N}\circ\mathcal{E})}{N}. \tag{4}$$

It has been conjectured [1,3,5] that this bound is an equality.

Notice that the definition of Q_E includes a supremum over encodings. This is required to give a most general definition of a channel capacity, but it is surprising from a physical point of view. Any unitary encoding of a source is equivalent to using a different source and since the supremum also includes the source, the unitary encoding could be left out. The coherent information, due to the failure of the pipelining inequality, *can* increase by using nonunitary encoding (see [5]), which suggests the necessity of the supremum over nonunitary encodings in the capacity definition. However, a nonunitary encoding intuitively corresponds to adding noise to the signal, which seems unlikely to improve the quality of the output signal. This illustrates the complexity of the issue. In this paper we resolve this matter by showing that the supremum over encodings can be omitted from the definition of capacity, though we do not know if the maximization over encodings can be omitted from $I_{\max}$.

Another issue is the continuity of the quantum channel capacity in the parameters of channel χ. It is not known whether Q_E or Q_P is continuous. It was stated in [6] that the capacity of the erasure channel is $Q=\max\{0,1-2p\}$. This result was derived by bounding the capacity both from below and from above with $\max\{0,1-2p\}$. The derivation of the upper bound, however, assumed the capacity to be continuous as a function of p, which has not been proved. We will use the results in this paper to prove the capacity in an alternative way, thus resolving the continuity question for the erasure channel. A similar proof of the capacity of the erasure channel was carried out independently by Cerf [12] using a different definition of the quantum channel capacity.

In this paper we prove the following.

(i) The maximization over encodings $\mathcal{E}$ in the definition of Q_E is not necessary. In other words, we find that

$$Q_E = Q_E^{\text{no encoding}}, \tag{5}$$

where $Q_E^{\text{no encoding}}$ is defined exactly as is Q_E, except without the encoding map $\mathcal{E}$ over encodings. See Sec. II.

(ii) The quantum capacity Q_E is bounded from above by the maximum coherent information without *source encoding*

$$Q_E \leqslant \lim_{N\to\infty} \max_{\rho} \frac{I_c(\rho,\chi^{\otimes N})}{N}. \tag{6}$$

See Sec. III.

(iii) The quantum capacity of the erasure channel [6] is given by $Q_E=Q_P=\max\{1-2p,0\}$ as in [6]. See Sec. III.

II. Q_E IS WELL DEFINED WITHOUT SOURCE ENCODING

Consider a situation where the sequence of triplets $(\rho,\mathcal{D},\mathcal{E})$ achieves Q_E and the $\mathcal{E}$'s may be nonunitary. We will show that there exists another sequence of triplet $(\rho',\mathcal{T}\circ\mathcal{D},\mathcal{I})$ that achieves the capacity Q_E, where $\mathcal{T}$ is an additional decoding step. We thus replace the nonunitary *encoding* by a not-necessarily-unitary *decoding*. We will do this by showing that for any triplet $(\rho,\mathcal{D},\mathcal{E})$ with a given entropy and with a given entanglement fidelity when used with the channel χ, there exists another triplet $(\rho',\mathcal{T}\circ\mathcal{D},\mathcal{I})$ whose entropy and entanglement fidelity are both close to those of the original triplet.

A. Preliminaries

We will need the following two lemmas.

Lemma 1. Given two bipartite pure states $|\psi\rangle$ and $|\phi\rangle$ in a Hilbert space $\mathcal{H}=\mathcal{H}_A\otimes\mathcal{H}_B$ with $|\langle\psi|\phi\rangle|^2\geqslant 1-\epsilon$, then

$$|S(\mathrm{Tr}_A|\psi\rangle\langle\psi|)-S(\mathrm{Tr}_A|\phi\rangle\langle\phi|)| \leqslant 2\sqrt{\epsilon}\,\log_2 d+1 \tag{7}$$

for all $\epsilon<\frac{1}{36}$ where d is the dimension of $\mathcal{H}_B$.

Proof. We will use an inequality from Fannes [13] involving the L_1 norm. The L_1 norm of an operator A, indicated by $\|A\|$, is defined by

$$\|A\| \equiv \mathrm{Tr}|A| \equiv \mathrm{Tr}\sqrt{A^\dagger A}. \tag{8}$$

We also define the function $\eta(x)=-x\log_2 x$ and let ρ_1,ρ_2 be density matrices in $\mathcal{H}_B$. We than have from [13] (when $\|\rho_1-\rho_2\|<\frac{1}{3}$)

$$|S(\rho_1)-S(\rho_2)| \leqslant \|\rho_1-\rho_2\|\log_2 d+\eta(\|\rho_1-\rho_2\|). \tag{9}$$

For our purposes, we may note that for $x<\frac{1}{3}$, $\eta(x)<\log_2 3/3<1$ and we use the weaker inequality

$$|S(\rho_1)-S(\rho_2)| \leqslant \log_2 d\|\rho_1-\rho_2\|+1. \tag{10}$$

For two commuting density matrices ρ_1 and ρ_2 we have $\|\rho_1-\rho_2\|=\Sigma_i|\lambda_i^{(1)}-\lambda_i^{(2)}|$, with $\lambda_i^{(1,2)}$ the eigenvalues of density matrices ρ_1,ρ_2 respectively. Since the entropy difference is invariant under independent unitary rotations of each density matrix,

$$|S(\rho_1)-S(\rho_2)| \leqslant \log_2 d\sum_i |\lambda_i^{(1)}-\lambda_i^{(2)}|+1, \tag{11}$$

where we have rearranged the eigenvalues in order of size. It is known [14] that

$$\sum_i |\lambda_i^{(1)}-\lambda_i^{(2)}| \leqslant 2\sqrt{1-B(\lambda^{(1)},\lambda^{(2)})}, \tag{12}$$

where B is the Bhattacharyya-Wootters overlap [15], defined by

$$B(\lambda^{(1)},\lambda^{(2)}) \equiv \left(\sum_i \sqrt{\lambda_i^{(1)}\lambda_i^{(2)}} \right)^2. \tag{13}$$

The fidelity between two density matrices ρ_1, ρ_2 can be defined as the maximum inner product between all purifications $|\zeta_1\rangle, |\zeta_2\rangle$ of ρ_1 and ρ_2:

$$F(\rho_1,\rho_2) = \max_{|\zeta_1\rangle,|\zeta_2\rangle} |\langle \zeta_1|\zeta_2\rangle|^2. \tag{14}$$

Since, given the eigenvalues of two density operators, the fidelity is maximized by choosing their eigenvectors to be the same (assigned to eigenvalues in order of size)

$$B(\lambda^{(1)},\lambda^{(2)}) \geq F(\rho_1,\rho_2). \tag{15}$$

Hence

$$|S(\rho_1)-S(\rho_2)| \leq 2\sqrt{1-F(\rho_1,\rho_2)}\ \log_2 d+1 \tag{16}$$

when

$$2\sqrt{1-F(\rho_1,\rho_2)} < \tfrac{1}{3} \tag{17}$$

and by the definition of $F(\rho_1,\rho_2)$ (which includes a maximization) we have that

$$|S(\rho_1)-S(\rho_2)| \leq 2\sqrt{1-|\langle \psi|\phi\rangle|^2}\ \log_2 d+1, \tag{18}$$

where $|\psi\rangle$ and $|\phi\rangle$ are purifications of ρ_1 and ρ_2, i.e., $\mathrm{Tr}_A|\psi\rangle\langle\psi| = \rho_1$ and $\mathrm{Tr}_A|\phi\rangle\langle\phi| = \rho_2$. This holds whenever $F(\rho_1,\rho_2) > 1-\frac{1}{36}$, which is certainly true whenever $|\langle\psi|\phi\rangle|^2 > 1-\frac{1}{36}$. ■

Lemma 2. Given a bipartite pure state $|\phi\rangle$ and density matrix ρ in Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ with $\langle\phi|\rho|\phi\rangle \geq 1-\epsilon$ and $\epsilon < \frac{1}{72}$, then

$$|S(\mathrm{Tr}_A|\phi\rangle\langle\phi|) - S(\mathrm{Tr}_A\rho)| \leq 2\sqrt{2\epsilon}\ \log_2 \dim \mathcal{H}_B + 2, \tag{19}$$

and similarly for system B, and thus

$$|S(\mathrm{Tr}_A\rho)-S(\mathrm{Tr}_B\rho)| \leq 4\sqrt{2\epsilon}\ \log_2 \max\{\dim\mathcal{H}_A, \dim\mathcal{H}_B\} + 4. \tag{20}$$

Proof. We can write

$$\rho = (1-\epsilon')|\phi_{\max}\rangle\langle\phi_{\max}| + \epsilon'\rho', \tag{21}$$

with $\epsilon' \leq \epsilon$. This is obtained by diagonalizing ρ and noting that the largest eigenvalue of a density matrix is always no smaller than the largest diagonal element of the matrix [16]. $|\phi_{\max}\rangle$ is the eigenvector of ρ corresponding to its largest eigenvalue.

Here is the plan for the proof. We will first bound $|S(\mathrm{Tr}_A\rho) - S(\mathrm{Tr}_A|\phi_{\max}\rangle\langle\phi_{\max}|)|$. Then we will argue that $|\phi_{\max}\rangle$ has high fidelity with respect to $|\phi\rangle$ and use Lemma 1 to bound $|S(\mathrm{Tr}_A|\phi\rangle\langle\phi|) - S(\mathrm{Tr}_A|\phi_{\max}\rangle\langle\phi_{\max}|)|$, which will finally give us a bound on $|S(\mathrm{Tr}_A|\phi\rangle\langle\phi|) - S(\mathrm{Tr}_A\rho)|$.

Recall the property of the entropy [17]

$$\sum_i \lambda_i S(\rho_i) \leq S\left(\sum_i \lambda_i\rho_i\right) \leq \sum_i \lambda_i S(\rho_i) - \sum_i \lambda_i \log_2 \lambda_i, \tag{22}$$

with $\Sigma_i\lambda_i = 1$ and ρ_i density matrices. Taking the partial trace of Eq. (21) and using Eq. (22) one can derive that

$$\begin{aligned} &\epsilon' S(\mathrm{Tr}_A\rho') - \epsilon' S(\mathrm{Tr}_A|\phi_{\max}\rangle\langle\phi_{\max}|) \\ &\quad \leq S(\mathrm{Tr}_A\rho) - S(\mathrm{Tr}_A|\phi_{\max}\rangle\langle\phi_{\max}|) \\ &\quad \leq \epsilon' S(\mathrm{Tr}_A\rho') - \epsilon' S(\mathrm{Tr}_A|\phi_{\max}\rangle\langle\phi_{\max}|) + H_2(\epsilon') \end{aligned} \tag{23}$$

and thus

$$|S(\mathrm{Tr}_A\rho) - S(\mathrm{Tr}_A|\phi_{\max}\rangle\langle\phi_{\max}|)| \leq \epsilon\ \log_2 \dim \mathcal{H}_B + 1. \tag{24}$$

To prove that $|\phi\rangle$ and $|\phi_{\max}\rangle$ have high fidelity we use Eq. (21) and $\langle\phi|\rho|\phi\rangle \geq 1-\epsilon$ to write

$$\langle\phi|\rho|\phi\rangle = (1-\epsilon')|\langle\phi|\phi_{\max}\rangle|^2 + \epsilon'\langle\phi|\rho'|\phi\rangle \geq 1-\epsilon. \tag{25}$$

The inner product $\langle\phi|\rho'|\phi\rangle$ is no bigger than one and $\epsilon' \leq \epsilon$, so we can rearrange things to get

$$|\langle\phi|\phi_{\max}\rangle|^2 \geq 1-2\epsilon. \tag{26}$$

Thus, by Lemma 1 we can bound

$$|S(\mathrm{Tr}_A|\phi\rangle\langle\phi|) - S(\mathrm{Tr}_A|\phi_{\max}\rangle\langle\phi_{\max}|)| \leq \sqrt{2\epsilon}\ \log_2 \dim\mathcal{H}_B + 1. \tag{27}$$

Therefore we find, with Eq. (24) and (27),

$$|S(\mathrm{Tr}_A|\phi\rangle\langle\phi|) - S(\mathrm{Tr}_A\rho)| \leq 2\sqrt{2\epsilon}\ \log_2 \dim\mathcal{H}_B + 2. \tag{28}$$

Finally, using $\mathrm{Tr}_A|\phi\rangle\langle\phi| = \mathrm{Tr}_B|\phi\rangle\langle\phi|$ for all pure states and Eq. (19), we immediately have Eq. (20). ■

B. The main theorem

Theorem 1. Suppose ρ is a density operator on a Hilbert space $\mathcal{H}_A$ and $\mathcal{E},\mathcal{D}$ linear trace-preserving completely positive operations such that

$$F_e(\rho, \mathcal{D}\circ\chi^{\otimes N}\circ\mathcal{E}) \geq 1-\epsilon. \tag{29}$$

Then there exist a density operator ρ' and a linear trace-preserving completely positive operation $\mathcal{T}$ such that

$$F_e(\rho', \mathcal{T}\circ\mathcal{D}\circ\chi^{\otimes N}) \geq 1-2\epsilon \tag{30}$$

and

$$|S(\rho)-S(\rho')| \leq 2\sqrt{2\epsilon}\ \log_2 \dim\mathcal{H}_A + 2. \tag{31}$$

The proof consists of two parts. First we show if there exists a source ρ that has high entanglement fidelity using some encoding $\mathcal{E}$ and decoding $\mathcal{D}$, we can always find another source ρ' that has a high entanglement fidelity as well, but has additional decoding instead of encoding. Second we show that this new source ρ' has very nearly the same von Neumann entropy as ρ.

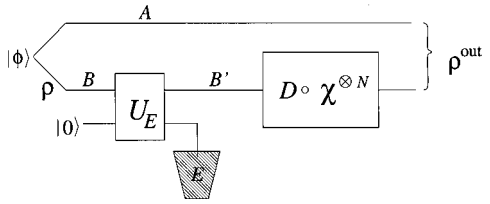

FIG. 1. General encoding-channel-decoding system. U_E is the unitary operation of the encoder (the associated environment E makes the whole action of the encoder nonunitary in general).

Let $|\phi\rangle$ be a purification of ρ in Hilbert space $\mathcal{H}_A\otimes\mathcal{H}_B$. See Fig. 1. Any linear trace-preserving completely positive map, including nonunitary operations, can be written as a unitary operator that operates on the original system along with an ancillary system (often referred to as an environment), as in Fig. 1. Thus, for the case of the nonunitary encoder, some quantum system E that is in general entangled with the AB' system will remain in the encoder. Since this system is not to be sent through the channel it may be measured in an orthogonal basis giving result i with probability p_i and leaving the AB' system in a pure state $|\psi_i\rangle$. After the channel operates on the B' system and the decoding process is performed, one is left with $\rho_i^{\text{out}}=[\mathcal{I}_A\otimes(\mathcal{D}\circ\chi^{\otimes N})_B](|\psi_i\rangle\langle\psi_i|)$. [To simplify the notation we will hereafter write $\mathcal{I}_A\otimes(\mathcal{D}\circ\chi^{\otimes N})_B$ as $\mathcal{D}\circ\chi^{\otimes N}$.] The whole encoding-channel-decoding process results in a high entanglement fidelity so that

$$F_e(\rho,\mathcal{D}\circ\chi^{\otimes N}\circ\mathcal{E})=\sum_i p_i\langle\phi|(\mathcal{D}\circ\chi^{\otimes N})(|\psi_i\rangle\langle\psi_i|)|\phi\rangle\geqslant 1-\epsilon. \tag{32}$$

For at least one value of i it must be that

$$\langle\phi|(\mathcal{D}\circ\chi^{\otimes N})(|\psi_i\rangle\langle\psi_i|)|\phi\rangle\geqslant 1-\epsilon. \tag{33}$$

Thus the unitary encoder that simply takes $|\phi\rangle$ and rotates it to $|\psi_i\rangle$ is sufficient to achieve a high entanglement fidelity. Hereafter the i subscript will be dropped from $|\psi_i\rangle$ and ρ_i^{out}.

We are now, however, left in the odd situation in which the unitary encoder operates on both the B and A systems. We have thus so far only traded nonunitarity for this odd form of unitarity. This situation is shown in Fig. 2. We will show that instead of using $|\phi\rangle$ as input, we can use the unencoded $|\psi\rangle$ as input if we do an additional decoding step. The following lemma will be of use.

Lemma 3. Given a density matrix ρ in Hilbert space $\mathcal{H}_A\otimes\mathcal{H}_B$, then there exists a purification $|\Psi\rangle$ of $\text{Tr}_B\,\rho$ into Hilbert space $\mathcal{H}_A\otimes\mathcal{H}_B\otimes\mathcal{H}_C$ with $\dim\mathcal{H}_C=\dim\mathcal{H}_A+1$ and

$$\langle\Psi|(\rho\otimes|0^C\rangle\langle 0^C|)|\Psi\rangle=\lambda_{\max}^2, \tag{34}$$

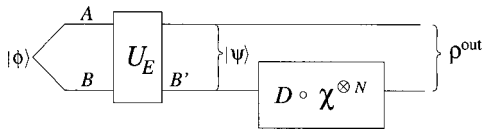

FIG. 2. Channel with unitary encoder acting on both the A and B systems.

where $\lambda_{\max}$ is the largest eigenvalue of ρ.

Proof. We can write $\rho\otimes|0^C\rangle\langle 0^C|$ as

$$\rho\otimes|0^C\rangle\langle 0^C|=\lambda_{\max}|\phi_{\max}\rangle\langle\phi_{\max}|\otimes|0^C\rangle\langle 0^C|+(1-\lambda_{\max})\rho'\otimes|0^C\rangle\langle 0^C|, \tag{35}$$

where $|\phi_{\max}\rangle$ is the eigenvector of ρ corresponding to $\lambda_{\max}$. Take

$$|\Psi\rangle=\sqrt{\lambda_{\max}}|\phi_{\max}\rangle\otimes|0^C\rangle+\sqrt{1-\lambda_{\max}}\sum_{i=1}^{\dim\mathcal{H}_A}\sqrt{\mu_i}|i^A\rangle\otimes|0^B\rangle\otimes|i^C\rangle, \tag{36}$$

where $|i^A\rangle$ and μ_i are the eigenvectors and eigenvalues of $\text{Tr}_B\,\rho$ and $\langle 0^C|i^C\rangle=0$. Thus $\langle\Psi|(\rho\otimes|0^C\rangle\langle 0^C|)|\Psi\rangle=\lambda_{\max}^2$.■

Since $\langle\phi|\rho^{\text{out}}|\phi\rangle\geqslant 1-\epsilon$ we have [as in Eq. (21)] $\lambda_{\max}\geqslant 1-\epsilon$. Take $|\Psi\rangle$ also purifying $\text{Tr}_B(\rho^{\text{out}})$ as in the lemma. Then

$$\langle\Psi|(\rho^{\text{out}}\otimes|0^C\rangle\langle 0^C|)|\Psi\rangle\geqslant(1-\epsilon)^2\geqslant 1-2\epsilon. \tag{37}$$

Since $|\psi\rangle$ purifies $\text{Tr}_B(\rho^{\text{out}})$, so does $|\psi_0\rangle\equiv|\psi\rangle\otimes|0^C\rangle$. As $|\Psi\rangle$ and $|\psi_0\rangle$ both purify $\text{Tr}_B(\rho)\otimes|0^C\rangle\langle 0^C|$, they are related by a unitary transformation $U=\mathcal{I}_A\otimes U_{BC}$ acting only on $\mathcal{H}_B$ and $\mathcal{H}_C$ [18],

$$U|\Psi\rangle=|\psi_0\rangle. \tag{38}$$

Substituting this into Eq. (37) and writing $\rho_0^{\text{out}}\equiv\rho^{\text{out}}\otimes|0^C\rangle\langle 0^C|$, we obtain

$$\langle\psi_0|U\rho_0U^\dagger|\psi_0\rangle\geqslant 1-2\epsilon. \tag{39}$$

We will now rid ourselves of the C system. As

$$\langle\psi|\text{Tr}_C U\rho_0^{\text{out}}U^\dagger|\psi\rangle=\langle\psi\otimes 0^C|U\rho_0U^\dagger|\psi\otimes 0^C\rangle+\sum_{i\neq 0}\langle\psi\otimes i^C|U\rho_0U^\dagger|\psi\otimes i^C\rangle, \tag{40}$$

with $\langle\psi\otimes i^C|U\rho_0^{\text{out}}U^\dagger|\psi\otimes i^C\rangle\geqslant 0$ since $U\rho_0^{\text{out}}U^\dagger$ is a density matrix, we can rewrite Eq. (39) as

$$\langle\psi|\text{Tr}_C U\rho_0^{\text{out}}U^\dagger|\psi\rangle\geqslant 1-2\epsilon. \tag{41}$$

Let us define $\mathcal{T}(\rho^{\text{out}})$ be the linear trace-preserving completely positive map implemented by appending a $|0^C\rangle$ state to ρ^{out}, rotating using U, and then tracing out the C system. What we have done is replaced $|\phi\rangle$ with $|\psi\rangle$ and added the decoding stage $\mathcal{T}$ and still achieved high entanglement fidelity. In other words, writing $\rho'\equiv\text{Tr}_A|\psi\rangle\langle\psi|$, we have

$$F_e(\rho',\mathcal{T}\circ\mathcal{D}\circ\chi^{\otimes N})\geqslant 1-2\epsilon. \tag{42}$$

Achieving a high entanglement fidelity alone is not sufficient. It is also necessary to show that $\rho'\equiv\text{Tr}_A|\psi\rangle\langle\psi|)$ has entropy close enough to that of $\rho\equiv\text{Tr}_A|\phi\rangle\langle\phi|)$ to achieve the same capacity. Using Eqs. (33) and (19) we know that

$$|S(\text{Tr}_B|\phi\rangle\langle\phi|)-S(\text{Tr}_B\rho^{\text{out}})|\leqslant 2\sqrt{2\epsilon}\,\log_2\,\dim\mathcal{H}_A+2 \tag{43}$$

for $\epsilon < \frac{1}{72}$. Since $\mathrm{Tr}_B \rho^{\mathrm{out}} = \mathrm{Tr}_B |\psi\rangle\langle\psi|$, $S(\mathrm{Tr}_B |\psi\rangle\langle\psi|) = S(\mathrm{Tr}_A |\psi\rangle\langle\psi|) = S(\rho')$, and $S(\mathrm{Tr}_B |\phi\rangle\langle\phi|) = S(\mathrm{Tr}_A |\phi\rangle\langle\phi|) = S(\rho)$ we have

$$|S(\rho) - S(\rho')| \leq 2\sqrt{2\epsilon}\, \log_2 \dim \mathcal{H}_A + 2. \tag{44}$$

This proves the theorem. The application to channel capacity is straightforward. As we can always purify a density matrix in a Hilbert space of dimension d into a Hilbert space of dimension d^2, the dimension $\dim \mathcal{H}_A$ can be set to $(\dim \chi)^N$, where $\dim \chi$ is the dimension on which χ acts. Since the definition of quantum capacity Q_E [Eq. (2)] has an N in the denominator, it is clear that Eq. (44) is strong enough to make $Q_E = Q_E^{\text{no encoding}}$.

III. A CORRECT PROOF OF THE CAPACITY OF THE ERASURE CHANNEL

In this section we will provide a correct upper bound of the capacity of the erasure channel that is ''proved'' incorrectly in [6] by making use of the unproven assumption that the quantum channel capacity is continuous. By providing a correct upper bound the entire capacity is restored, as the upper bound coincides with the correct lower bound given in [6]. We work here with Q_E rather than the definition of capacity in terms of a protected subspace employed in [6], but these two definitions of capacity have been shown to be equivalent [19,20]. Cerf independently provided a similar correct upper bound [12] using a slightly different definition of capacity, which we expect is also equivalent.

Barnum, Nielsen, and Schumacher [5] have shown that

$$Q_E^{\text{no encoding}} \leq I_{\max}^{\text{no encoding}} \equiv \lim_{N\to\infty} \max_{\rho} \frac{I_c(\rho, \chi^{\otimes N})}{N}. \tag{45}$$

Together with the results of Sec. II that $Q_E = Q_E^{\text{no encoding}}$, we now have

$$Q_E \leq \lim_{N\to\infty} \max_{\rho} \frac{I_c(\rho, \chi^{\otimes N})}{N}. \tag{46}$$

A quantum erasure channel with erasure probability p maps an input quantum bit (qubit) ρ to $(1-p)\rho + p|3\rangle\langle 3|$, where $|3\rangle$ is an orthogonal direction to the $|1\rangle, |2\rangle$ space in which ρ resides. In [6] it was shown correctly that $Q_P = 0$ for $p \geq \frac{1}{2}$. Thus we will consider here only channels with $p < \frac{1}{2}$.

Recall the definition of the coherent information

$$I_c(\rho, \chi^{\otimes N}) = S(\chi^{\otimes N}(\rho)) - S_{\text{env}}(\rho, \chi^{\otimes N}). \tag{47}$$

For the erasure channel we can write

$$I_c(\rho, \chi^{\otimes N}) = \sum_{k=0}^{N} p^k (1-p)^{N-k} \sum_{i=1}^{\binom{N}{k}} [S(\rho_i) - S(\rho_{\bar{i}})], \tag{48}$$

where i designates a particular set of $N-k$ qubits and $\bar{i}$ the complement of the set i. ρ_i is defined as $\rho_i = \mathrm{Tr}_{\bar{i}} \rho$. This expression is obtained by noticing that the density matrix for the receiver is block diagonal, where the block labeled with (i,k) is of the form

$$p^k (1-p)^{N-k} \rho_i. \tag{49}$$

Thus the entropy of the block (i,k) is $p^k(1-p)^{N-k} S(\rho_i)$. The total entropy of such a block diagonal density matrix $S(\chi^{\otimes N}(\rho))$ is equal to the sum of the entropy of the blocks plus the entropy of choosing among the blocks. The expression $\underline{S}_{\text{env}}(\rho, \chi^{\otimes N})$ will be the same as $S(\chi^{\otimes N}(\rho))$, but with i and $\bar{i}$ interchanged (what is not erased the environment gets and vice versa). Subtracting the two entropies will result in Eq. (48).

We split the sum over k into two terms I_+ and I_-, which we will bound separately,

$$I_+ = \sum_{k=0}^{\lfloor N/2 \rfloor} p^k (1-p)^{N-k} \sum_{i=1}^{\binom{N}{k}} [S(\rho_i) - S(\rho_{\bar{i}})] \tag{50}$$

and

$$I_- = \sum_{k=\lfloor N/2 \rfloor + 1}^{n} p^k (1-p)^{N-k} \sum_{i=1}^{\binom{N}{k}} [S(\rho_i) - S(\rho_{\bar{i}})]. \tag{51}$$

Each term in I_- can be at most

$$S(\rho_i) - S(\rho_{\bar{i}}) \leq N - k. \tag{52}$$

To bound I_+ we will rewrite the sum over the sets i in such a way that we can use the subadditivity property of the von Neumann entropy. The idea is to pairwise match terms in Eq. (50). We match $S(\rho_i)$ with a term $S(\rho_{\bar{j}})$ and $S(\rho_{\bar{i}})$ with $\underline{S}(\rho_j)$ where we take the set of (qubits) j such that $j \subset i$ and $i \subset j$. For these matching sets, we can use subadditivity,

$$\begin{aligned} S(\rho_i) - S(\rho_{\bar{j}}) &\leq N - 2k, \\ S(\rho_j) - S(\rho_{\bar{i}}) &\leq N - 2k. \end{aligned} \tag{53}$$

The way to do the pairwise matching is the following. Pick $N-2k$ qubits out of the total set of N qubits. These are the qubits that two matching sets will have in common. Then pick a subset of k qubits out of the remaining $2k$. Together with the $N-2k$ qubits, these will form set i. The set j is made from the remaining k qubits and the $N-2k$ overlap qubits. In this way each set is matched to another one. However, we have counted the sets multiple times. Each set is counted $2\binom{N-k}{k}$ times. Dividing by this number will thus give us the original sum. Thus we have derived that

$$I_+ \leq \sum_{k=0}^{\lfloor N/2 \rfloor} p^k (1-p)^{N-k} \binom{N}{k} (N-2k). \tag{54}$$

We will take I_+ and I_- together and use

$$\sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} k = Np \tag{55}$$

to get

$$I_c(\rho,\chi^{\otimes N}) \leqslant N(1-p) - \sum_{k=0}^{\lfloor N/2 \rfloor} \binom{N}{k} p^k (1-p)^{N-k} k. \quad (56)$$

We will use a property of binomial distributions

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{\lfloor N/2 \rfloor} \binom{N}{k} p^k (1-p)^{N-k} k = p \quad \text{for } p < \tfrac{1}{2}. \quad (57)$$

This implies

$$\lim_{N\to\infty} \max_{p} \frac{I_c(\rho,\chi^{\otimes N})}{N} \leqslant 1-2p \quad (58)$$

(note that this bound is achieved by taking $\rho = \mathcal{I}/2^N$) and therefore [with Eq. (46)]

$$Q_P \leqslant Q_E \leqslant 1-2p. \quad (59)$$

In [6] a constructive lower bound on Q_P has been established,

$$Q_P \geqslant 1-2p. \quad (60)$$

Together with our upper bound we prove the capacity of the erasure channel

$$Q_E = Q_P = \max\{1-2p, 0\}. \quad (61)$$

IV. DISCUSSION AND OPEN PROBLEMS

An important open question is the conjecture of the equality of $I_{\max}$ and the channel capacity. The conjecture would be flawed if $I_{\max} \neq I_{\max}^{\text{no encoding}}$ since we have shown that the latter upper bounds the capacity.

Equation (61) for the capacity of the erasure channel is a continuous function of p, but a resolution of the problem of the continuity of capacity for general channels is to be desired. If the channel capacity turns out not to be continuous, this would once again show a curious characteristic of quantum information. On the other hand, if the capacity were proven continuous, the quite general method for bounding the quantum capacity introduced in [4] and applied incorrectly in [6] would be restored. For example, the quantum cloning results in [21] could be used to improve the bound on the capacity of the quantum depolarizing channel.

In [4] it was shown that the quantum capacities with and without a classical forward side channel are equal in the case of perfect error correction ($\epsilon = 0$). A proof similar to the one in Sec. II can be used to show that this is true for Q_E even in the case of an asymptotically perfect correction as in the definition of quantum capacity.

ACKNOWLEDGMENTS

The authors would like to thank Charles Bennett, David DiVincenzo, and Michael Nielsen for helpful discussions. J.A.S. would like to thank the Army Research Office for financial support. B.M.T. would like to thank the NWO for financial support from the SIR program. H.B. thanks the NSF for financial support under Grant No. PHY-9722614, Office for Naval Research Grant No. N00014-93-1-0116, the Institute for Scientific Interchange, Turin, Italy, and Elsag-Bailey for financial support.

[1] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).
[2] B. Schumacher and M. A. Nielsen, Phys. Rev. A **54**, 2629 (1996).
[3] S. Lloyd, Phys. Rev. A **55**, 1613 (1997).
[4] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[5] H. Barnum, M. A. Nielsen, and B. Schumacher, Phys. Rev. A **57**, 4153 (1998); e-print quant-ph/9702049.
[6] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, Phys. Rev. Lett. **78**, 3217 (1997).
[7] A. S. Holevo, e-print quant-ph/9611023.
[8] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997).
[9] C. Adami and N. J. Cerf, Phys. Rev. A **56**, 3470 (1997).
[10] The most general transformation on a density matrix allowed by quantum mechanics is a linear trace-preserving completely positive operation. This is equivalent to the unitary operations including an environment initially in a known state $|0\rangle$ as shown in Fig. 1.
[11] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).
[12] N. J. Cerf, Phys. Rev. A **57**, 3330 (1998); e-print quant-ph/9707023.
[13] M. Ohya and D. Petz, *Quantum Entropy and its Use* (Springer-Verlag, Berlin, 1983).
[14] The inequality is due to C. H. Kraft; cf. T. Kailath, IEEE Trans. Commun. Technol. **COM-15**, 52 (1967).
[15] A. Bhattacharyya, Bull. Calcutta Math. Soc. **35**, 99 (1943); W. K. Wootters, Phys. Rev. D **23**, 357 (1981).
[16] This is a special case of Ky Fan's inequality; cf. [17].
[17] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).
[18] L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A **183**, 14 (1993).
[19] H. Barnum, Ph.D. thesis, University of New Mexico, 1997 (unpublished).
[20] H. Barnum, E. Knill, and M. A. Nielsen, e-print quant-ph/9809010.
[21] D. Bruss, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello, and J. A. Smolin, Phys. Rev. A **57**, 2368 (1998); e-print quant-ph/9705038.