JOURNAL OF MATHEMATICAL PHYSICS VOLUME 43, NUMBER 9 SEPTEMBER 2002

# The entanglement of purification

Barbara M. Terhal[a)]
*Institute for Quantum Information, California Institute of Technology, Pasadena, California 91125 and IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*

Michał Horodecki
*Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80–952 Gdańsk, Poland*

Debbie W. Leung
*IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*

David P. DiVincenzo
*Institute for Quantum Information, California Institute of Technology, Pasadena, California 91125 and IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*

(Received 6 February 2002; accepted for publication 16 May 2002)

We introduce a measure of both quantum as well as classical correlations in a quantum state, the entanglement of purification. We show that the (regularized) entanglement of purification is equal to the entanglement cost of creating a state $\rho$ asymptotically from maximally entangled states, with negligible communication. We prove that the classical mutual information and the quantum mutual information divided by two are lower bounds for the regularized entanglement of purification. We present numerical results of the entanglement of purification for Werner states in $\mathcal{H}_2 \otimes \mathcal{H}_2$. © *2002 American Institute of Physics.*
[DOI: 10.1063/1.1498001]

## I. INTRODUCTION

The theory of quantum entanglement aims at quantifying and characterizing uniquely quantum correlations. It does so by analyzing how entangled quantum states can be processed and transformed by quantum operations. A crucial role in the theory is played by the class of local operations and classical communication (LOCC), since quantum entanglement is nonincreasing under these operations. Indeed, by considering this class of operations we are able to neatly distinguish between the quantum entanglement and the classical correlations that are present in the quantum state.

Given the success of this theory, we may be daring enough to ask whether we can similarly construct a theory of purely classical correlations in quantum states and their behavior under local or nonlocal processing. At first sight, such an effort seems doomed to fail since merely local actions can convert quantum entanglement into classical correlations. Namely, Alice and Bob who possess an entangled state $|\psi\rangle = \Sigma_i \sqrt{\lambda_i} |a_i\rangle \otimes |b_i\rangle$ with Schmidt coefficients $\lambda_i$ can, by local measurements, obtain a joint probability distribution with mutual information equal to $H(\lambda)$. Thus it does not seem possible to separate the classical correlations from the entanglement if we try to do this in an operational way. Note that it may be possible to separate quantum and classical correlations in a nonoperational way (see, for example, Ref. 1 or 2. The drawback of such an approach is that no connection is made to the dynamical processing of quantum information, which is precisely what has made the theory of quantum entanglement so elegant and innovative. An

a)Electronic mail: terhal@watson.ibm.com

 

operational approach to the quantification of quantum and classical correlations was recently formulated in Ref. 3.

In this article we propose to treat quantum entanglement and classical correlation in a unified framework, namely we express both correlations in units of entanglement. Such a theory of "all" correlations may have potential applications outside quantum information theory as well. Researchers have started to look at entanglement properties of many-particle systems for example at (quantum) phase transitions (see, for example, Ref. 4 and references therein). Instead of considering the entanglement of formation in these studies, one may choose to look at the behavior of a complete correlation measure. In this article we introduce such a measure, called the entanglement of purification. We would like to emphasize that our correlation measure is *not* an entanglement measure, but a measure of correlations expressed in terms of the entanglement of a pure state.

It has been the experience in (quantum) information theory that questions in the asymptotic approximate regime are easier to answer than exact nonasymptotic queries. Thus we ask how to create a bipartite quantum state $\rho$ in the asymptotic regime, allowing approximation, from an initial supply of EPR-pairs by means of *local operations and asymptotically vanishing communication*. This latter class of operations will be denoted as LOq [local operations with $o(n)$ communication in the asymptotic regime] versus the class LO for strictly local operations. We can properly define this formation cost $E_{LOq}$ as follows:

$$E_{LOq}(\rho)=\lim_{\epsilon\to 0}\inf\left\{\frac{m}{n}\;\middle|\;\exists\,\mathcal{L}_{LOq}, D(\mathcal{L}_{LOq}(|\Psi_-\rangle\langle\Psi_-|^{\otimes m}),\rho^{\otimes n})\leq\epsilon\right\}. \tag{1}$$

Here $|\Psi_-\rangle$ is the singlet state in $\mathcal{H}_2\otimes\mathcal{H}_2$ and $\mathcal{L}_{LOq}$ is a local superoperator using $o(n)$ quantum communication. $D$ is the Bures distance $D(\rho,\rho')=2\sqrt{1-F(\rho,\rho')}$ and the square-root-fidelity is defined as $F(\rho,\rho')=\mathrm{Tr}(\sqrt{\rho^{1/2}\rho'\rho^{1/2}})$.[5] We could have allowed classical instead of quantum communication in our definition (our results will not depend on this choice), so we may as well call all communication quantum communication.

Before we consider this entanglement cost for mixed states, we observe that by allowing asymptotically vanishing communication, we have preserved the interconvertibility result for pure states.[6] This is due to the fact that both the process of entanglement dilution as well as entanglement concentration can be accomplished with no more than asymptotically vanishing amount of communication, see Ref. 7.

We see that the cost $E_{LOq}(\rho)$ of creating the state $\rho$ is defined analogously to the entanglement cost $E_c(\rho)$,[8,9] with the restriction that Alice and Bob can only do a negligible amount of communication. It is immediate that $E_{LOq}(\rho)$ will in general be larger than $E_c(\rho)$. In particular, for a separable density matrix, $E_c(\rho)=0$, whereas we will show that for any correlated (i.e., not of the form $\rho_{AB}=\rho_A\otimes\rho_B$) density matrix $E_{LOq}(\rho)>0$. The entanglement cost $E_c$ was found[9] to be equal to

$$E_c(\rho)=\lim_{n\to\infty}\frac{E_f(\rho^{\otimes n})}{n}, \tag{2}$$

where $E_f(\rho)$ is the entanglement of formation.[8] We will similarly find an expression for $E_{LOq}$,

$$E_{LOq}=\lim_{n\to\infty}\frac{E_p(\rho^{\otimes n})}{n}\equiv E_p^{\infty}(\rho), \tag{3}$$

where $E_p(\rho)$ is a new quantity, the entanglement of purification of $\rho$.

Our article is organized in the following manner. We start by defining the entanglement of purification and deriving some basic properties of this new function, such as continuity and monotonicity under local operations. We will relate the entanglement of purification to the problem of minimizing the entropy of a state under a local TCP (trace-preserving completely positive) map. With these tools in hand, we can prove our main result, Theorem 2. Then we spend some

time proving the mutual information lower bounds for $E_{LOq}(\rho)$. We also compare our correlation measure with the induced Holevo correlation measures $C_{A/B}$ that were introduced in Ref. 1. We prove that for Bell-diagonal states the correlation measure $C_A$ is equal to the classical capacity of the related one-qubit Pauli channel. At the end of the article we present our numerical results for $E_p(\rho)$ where $\rho$ is a Werner state on $\mathcal{H}_2 \otimes \mathcal{H}_2$. The proofs of the lemmas and theorems in this article are all fairly straightforward and use many basic properties of entropy and mutual information (concavity, subadditivity of entropy, nonincrease of mutual information under local actions, etc.).

## II. ENTANGLEMENT OF PURIFICATION

We define the entanglement of purification:

*Definition 1: Let $\rho$ be a bipartite density matrix on $\mathcal{H}_A \otimes \mathcal{H}_B$. Let $|\psi\rangle \in \mathcal{H}_{AA'} \otimes \mathcal{H}_{BB'}$. The entanglement of purification $E_p(\rho)$ is defined as*

$$E_p(\rho) = \min_{\psi: \mathrm{Tr}_{A'B'}|\psi\rangle\langle\psi| = \rho} E_f(|\psi\rangle\langle\psi|), \tag{4}$$

where $E_f(|\psi\rangle\langle\psi|)$ is the entanglement of $|\psi\rangle$ which is equal to the von Neumann entropy $S(\sigma_{BB'}) = -\mathrm{Tr}\sigma_{BB'} \log \sigma_{BB'}$ where $\sigma_{BB'} = \mathrm{Tr}_{AA'}|\psi\rangle\langle\psi|$. Let $\{\lambda_i, |\psi_i\rangle\}$ be the eigenvalues and eigenvectors of $\rho_{AB}$. The "standard purification" of $\rho$ is defined as

$$|\psi_s\rangle = \sum_i \sqrt{\lambda_i} |\psi_i\rangle_{AB} \otimes |0\rangle_{A'} |i\rangle_{B'}. \tag{5}$$

Every purification of $\rho$ can be written as $|\psi\rangle = (I_{AB} \otimes U_{A'B'})|\psi_s\rangle$ for some unitary operator $U_{A'B'}$ on $A'$ and $B'$. Therefore, Eq. (4) can be rephrased as

$$E_p(\rho) = \min_{U_{A'B'}} E((I_{AB} \otimes U_{A'B'})|\psi_s\rangle\langle\psi_s|(I_{AB} \otimes U_{A'B'})^\dagger) \tag{6}$$

$$= \min_{U_{A'B'}} S(\mathrm{Tr}_{AA'}(I_{AB} \otimes U_{A'B'})|\psi_s\rangle\langle\psi_s|(I_{AB} \otimes U_{A'B'})^\dagger))$$

$$= \min_{\Lambda_{B'}} S((I_B \otimes \Lambda_{B'})(\mu_{BB'}(\rho))), \tag{7}$$

where we have taken the trace over $A$ and $A'$ to obtain Eq. (7),

$$\mu_{BB'}(\rho) = \mathrm{Tr}_{AA'}|\psi_s\rangle\langle\psi_s|, \tag{8}$$

and $\Lambda_{B'}(\nu) \equiv \mathrm{Tr}_{A'} U_{A'B'}(\nu_{B'} \otimes |0\rangle\langle 0|_{A'}) U^\dagger_{A'B'}$. The minimization in Eq. (7) is over all possible TCP maps $\Lambda_{B'}$ since every TCP map can be implemented by performing a unitary transformation on the system and some ancilla and tracing over the ancilla. Note that the minimizations over $U_{A'B'}$ and $\Lambda_{B'}$ are equivalent. Equations (6) and (7) provide two different formulations of the same minimization. Conceptually the first formulation is based on purifications of $\rho$ and variation over $U_{A'B'}$. The second formulation is based on extensions of $\rho$, $\sigma_{ABB'}$, such that $\mathrm{Tr}_{B'}\sigma_{ABB'} = \rho_{AB}$, and variation over $\Lambda_{B'}(\nu)$. Both formulations will be used throughout the article.

The idea of bipartite purifications was considered in Ref. 10 where the authors proved that every correlated state has, in our language, a nonzero entanglement of purification. If we would have included mixed states in the minimization in Eq. (4) and used the entanglement of formation as the entanglement measure, then the defined quantity would be equal to the entanglement of formation of $\rho$, since the optimal extension of $\rho$ is $\rho$ itself.

We put some simple bounds on $E_p(\rho)$. Intuitively, "the amount of quantum correlation in a state is smaller than or equal to the total amount of correlation," or $E_f(\rho) \leqslant E_p(\rho)$. To prove this lower bound, let $|\psi_\rho\rangle = \Sigma_{i,j}|i\rangle_{A'}|j\rangle_{B'} \otimes |\psi_{ij}\rangle$ be the purification that achieves the minimum in Eq.

(4). Alice and Bob locally measure the labels $i_{A'}$ and $j_{B'}$ of the state $|\psi_\rho\rangle$ such that they obtain $|\psi_{ij}\rangle$ with probability $p_{ij}=\langle\psi_{ij}|\psi_{ij}\rangle$. Since entanglement is nonincreasing under local operations, we have

$$E_f(\rho)\leqslant\sum_{ij} p_{ij}E\left(\frac{|\psi_{ij}\rangle\langle\psi_{ij}|}{p_{ij}}\right)\leqslant E_p(\rho). \tag{9}$$

It is immediate that we have equality between the entanglement of formation and the entanglement of purification for pure states, where the optimal purification of a pure state is the pure state itself.

An easy upper bound is $E_p(\rho)\leqslant E(|\psi_s\rangle\langle\psi_s|)=S(\rho_A)$, where $\rho_A=\mathrm{Tr}_B(\rho)$ is the reduced density matrix in $A$. This corresponds to $U_{A'B'}=I_{A'B'}$ or equivalently $\Lambda_{B'}=I_{B'}$ on the rhs of Eq. (6) or (7). Applying the same argument with $AA'$ and $BB'$ interchanged, we obtain

$$E_p(\rho)\leqslant\min(S(\rho_A),S(\rho_B)), \tag{10}$$

where the purifications correspond to either completely purifying the state on $A'$ or on $B'$. In general this is not the optimal purification, as we will see in Sec. V.

The entanglement of purification is neither convex nor concave, unlike the entanglement of formation. For instance, a mixture of product states, each with zero entanglement of purification, need not have zero entanglement of purification (for example, consider an equal mixture of $|00\rangle$ and $|11\rangle$). On the other hand, the completely mixed state has zero entanglement of purification equal to zero yet it is a mixture of four Bell states, each with one ebit of entanglement of purification.

Before we present continuity bounds for the entanglement of purification, we analyze the optimization problem of Eq. (4) in more detail. We can omit doubly stochastic maps $\Lambda_{B'}$ in the optimization in Eq. (7) since they never decrease the entropy. Furthermore, the von Neumann entropy is concave, so that the optimum in Eq. (7) can always be achieved when $\Lambda_{B'}$ is an *extremal* TCP map. An extremal TCP map is a TCP map that cannot be expressed as a convex combination of other TCP maps. Choi[11] has proved that an extremal TCP map with input dimension $d$ has at most $d$ operation elements in its operator-sum representation. This result will allow us to upper bound the dimensions of the optimal purifying Hilbert spaces, as stated in the following lemma.

*Lemma 1: Let $\rho$ act on a Hilbert space of dimension $d_{AB}=d_Ad_B$. The minimum of Eq. (4) can always be achieved by a state $\psi$ for which the dimension of $A'$ is $d_{A'}=d_{AB}$ and the dimension of $B'$ is $d_{B'}=d_{AB}^2$ (or vice versa).*

*Proof:* We use the formulation of the entanglement of purification as an optimization of a TCP map in Eq. (7). Since the density matrix $\mu_{BB'}(\rho)$ is on $\mathcal{H}_{d_B}\otimes\mathcal{H}_{d_{AB}}$, the optimal map $\Lambda_{B'}$ maps $\mathcal{H}_{d_{AB}}$ into a space of some unspecified dimension. The optimal map $\Lambda_{B'}$ can be assumed to be extremal. Theorem 5 of Ref. 11 shows that an extremal TCP map $\Lambda:B(\mathcal{H}_{d_1})\to B(\mathcal{H}_{d_2})$ (Refs. 12 and 13) can be written with at most $d_1$ operations elements, that is, has the form

$$\Lambda(\rho)=\sum_{i=1}^{d_1} V_i\rho V_i^\dagger. \tag{11}$$

In our case $d_1=d_{AB}$. Consider implementing the TCP map by applying a unitary operation $U$ to the input state with an ancilla appended. In our case, this ancilla can be taken as Alice's purifying system $A'$, and $U$ acts on $A'B'$. The dimension of the ancilla $A'$ can always be taken to be the number of operation elements. Thus we have $d_{A'}=d_{AB}$. The $B'$ dimension is equal to the output dimension $d_2$ of the optimal map $\Lambda$, which is unconstrained by the extremality condition. However, we note that the operator $\Lambda(\rho)$ of Eq. (11) has a rank of at most $d_{AB}^2$. This is obtained by observing that the range of this operator is exactly that of the vectors given by all the columns of the matrices $V_i$ for all $i$ (the $V_i$ matrices have $d_1$ columns and $d_2$ rows). Thus, there exists a unitary operator $U$ that permits the construction of a new map $\Lambda'=U\Lambda$ whose output is confined

to the first $d_1^2$ dimensions of the output space. The operator $U$ may be obtained explicitly via a Gram–Schmidt procedure applied to the column vectors of the $V_i$ matrices. $\Lambda'$ is also optimal, since the entropy of Eq. (7) is not changed by a unitary operation. Since the output space of $\Lambda'$ has dimension $d_1^2$, we conclude that $d_{B'}$ can be taken to be $d_{B'}=d_{AB}^2$. □

It is interesting to note that a similar minimization problem was encountered in Ref. 14. There the goal was to use a set of noisy states for classical information transmission and we wanted to minimize the coherent information divided by the entropy of a quantum state under the action of a local map.

**Theorem 1 (continuity of the entanglement of purification):** *Let $\rho$ and $\sigma$ be two density matrices on $\mathcal{H}_{d_A}\otimes\mathcal{H}_{d_B}$ with Bures distance $D(\rho,\sigma)\leqslant\epsilon$. Then*

$$|E_p(\rho)-E_p(\sigma)|\leqslant 20D(\rho,\sigma)\log d_{AB}-D(\rho,\sigma)\log D(\rho,\sigma), \tag{12}$$

*for small enough $\epsilon$.*

*Proof:* Let $|\psi'_\sigma\rangle$ and $|\psi'_\rho\rangle$ be the purifications of $\rho$ and $\sigma$ which achieve the maximum[5] in

$$F(\rho,\sigma)=\max_{\psi_\sigma,\psi_\rho}|\langle\psi_\sigma|\psi_\rho\rangle|. \tag{13}$$

Let $|\phi_\rho\rangle$ and $|\phi_\sigma\rangle$ correspond to the optimal purifications of $\rho$ and $\sigma$ with respect to $E_p$. There exists a unitary transformation $U$ relating $|\psi'_\rho\rangle$ to $|\phi_\rho\rangle$, i.e., $(U\otimes\mathbf{1})|\psi'_\rho\rangle=|\phi_\rho\rangle$. We define the (nonoptimal) purification $|\psi_\sigma\rangle$ as $(U\otimes\mathbf{1})|\psi'_\sigma\rangle=|\psi_\sigma\rangle$. Now we have

$$E_p(\sigma)-E_p(\rho)=E(|\phi_\sigma\rangle\langle\phi_\sigma|)-E(|\phi_\rho\rangle\langle\phi_\rho|)\leqslant E(|\psi_\sigma\rangle\langle\psi_\sigma|)-E(|\phi_\rho\rangle\langle\phi_\rho|). \tag{14}$$

We use continuity of entanglement,[15,16] Lemma 1 (which indicates that the pure state has support on a space of dimension at most $d_{AB}^4$), and the fact that $|\langle\psi_\sigma|\phi_\rho\rangle|=|\langle\psi'_\sigma|\psi'_\rho\rangle|=F(\rho,\sigma)$ to bound

$$E_p(\sigma)-E_p(\rho)\leqslant 5D(\rho,\sigma)\log d_{AB}^4-2D(\rho,\sigma)\log D(\rho,\sigma) \tag{15}$$

for small enough $D(\rho,\sigma)$. We can obtain the full bound in Eq. (12) by alternatively relating $|\psi'_\sigma\rangle$ to the optimal purification $|\phi_\sigma\rangle$ by a unitary transformation $U$. □

It is fairly straightforward to prove monotonicity of the entanglement of purification from monotonicity of entanglement:

*Lemma 2 (monotonicity of the entanglement of purification): The entanglement of purification of a density matrix $\rho$ is nonincreasing under strictly local operations. Let Alice carry out a local TCP map $\mathcal{S}_A$ on the state $\rho$. We have*

$$E_p((\mathcal{S}_A\otimes\mathbf{1})(\rho))\leqslant E_p(\rho). \tag{16}$$

*Let Alice carry out a local measurement on $\rho$ through which she obtains the state $\rho_i$ with probability $p_i$. We have*

$$\sum_i p_i E_p(\rho_i)\leqslant E_p(\rho). \tag{17}$$

*Let $\mathcal{L}_{LOq}$ be a local operation assisted by $m$ qubits of communication. The entanglement of purification obeys the equation*

$$E_p(\mathcal{L}_{LOq}(\rho))\leqslant E_p(\rho)+m. \tag{18}$$

*Proof:* Let $|\psi_\rho\rangle$ be the optimal purification of $\rho$. This optimal purification is related to *some* purification of $(\mathcal{S}_A\otimes\mathbf{1})(\rho)$ by a unitary transformation on Alice's system only. Then Eq. (16) follows from the fact that entanglement is nonincreasing under local partial traces. The state

$|\psi_i\rangle = A_i \otimes I_B |\psi\rangle / \sqrt{\langle\psi|A_i^\dagger A_i \otimes I_B|\psi\rangle}$, where $A_i$ corresponds to a measurement outcome of Alice, is some purification of $\rho_i$. The entanglement is nonincreasing under local operations and thus

$$E_p(\rho) = E(|\psi_\rho\rangle\langle\psi_\rho|) \geqslant \sum_i p_i E(|\psi_i\rangle\langle\psi_i|) \geqslant \sum_i p_i E_p(\rho_i). \tag{19}$$

For the last inequality, let Alice and Bob start with the entangled state $|\psi_\rho\rangle$ and carry out their LOq protocol. By subadditivity of entropy, the entanglement of this state can increase by at most $m$ bits when $m$ qubits of communication are sent (back and forth). Thus the entanglement of the final state which is some purification of $\mathcal{L}_{LOq}(\rho)$ is smaller than or equal to $E_p(\rho)+m$. □

Now we are ready to prove our main theorem:

**Theorem 2:** *The entanglement cost of $\rho$ on $\mathcal{H}_d \otimes \mathcal{H}_d$ without classical communication equals* $E_{LOq}(\rho) = E_p^\infty(\rho)$.

*Proof:* The inequality $E_{LOq}(\rho) \leqslant E_p^\infty(\rho)$ uses entanglement dilution. Let $k$ be the number of copies of $\rho$ for which the regularized entanglement of purification $E_p^\infty$ is achieved. One way of making many $(p)$ copies of $\rho^{\otimes k}$ out of EPR pairs and $o(p) \leqslant o(pk)$ classical communication is to first perform entanglement dilution on the EPR pairs so as to create (an approximation to) the purification $|\psi\rangle^{\otimes p}$ and then trace over the additional registers to get $\rho^{\otimes kp}$. The other inequality $E_p^\infty(\rho) \leqslant E_{LOq}(\rho)$ can be proved from monotonicity and continuity of the entanglement of purification. We start with $n$ EPR pairs which have $E_p$ equal to $n$. The LOq process for creating an approximation $\tilde{\rho}_k$ to $\rho^{\otimes k}$ using $o(k)$ qubits of communication increases the entanglement of purification by at most $o(k)$ bits, see Lemma 2, or $E_p(\tilde{\rho}_k) \leqslant n + o(k)$. Using the continuity of Theorem 1 and dividing the last inequality by $k$ and taking the limit $k \to \infty$ gives $E_p^\infty(\rho) \leqslant E_{LOq}(\rho)$. □

## III. MUTUAL INFORMATION LOWER BOUNDS

The entanglement cost $E_{LOq}$ is a measure of the quantum and classical correlations in a quantum state. The quantum and classical mutual information of a quantum state are similar measures that capture correlations in a quantum state. How do these measures relate to the new correlation measure? The quantum mutual information $I_q(\rho_{AB})$ is defined as

$$I_q(\rho_{AB}) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}). \tag{20}$$

We define the classical mutual information of a quantum state $I_c(\rho_{AB})$ as

$$I_c(\rho_{AB}) = \max_{M_A : p_A, M_B : p_B} H(p_A) + H(p_B) - H(p_{AB}). \tag{21}$$

Here local measurements $M_A$ and $M_B$ give rise to local probability distributions $p_A$ and $p_B$. The classical mutual information of a quantum state is the maximum classical mutual information that can be obtained by local measurements by Alice and Bob. Both quantum as well as classical mutual information share the important property that they are nonincreasing under local operations (LO) by Alice and Bob. For the classical mutual information, this basically follows from the definition Eq. (21). The definition itself as a maximum over local measurements makes sense since the classical mutual information of a probability distribution is nonincreasing under local manipulations of the distribution. The proof of this well known fact is analogous to the proof for the quantum mutual information which we will give here for completeness.

We can write the quantum mutual information as

$$I_q(\rho_{AB}) = S(\rho_{AB} || \rho_A \otimes \rho_B), \tag{22}$$

where $S(.||.)$ is the relative entropy. The relative entropy is nonincreasing under any map $\Lambda$ (cf. Ref. 17), i.e.,

$$S(\Lambda(\rho_{AB})||\Lambda(\rho_A\otimes\rho_B))\leqslant S(\rho_{AB}||\rho_A\otimes\rho_B). \tag{23}$$

When $\Lambda$ is of a local form, i.e., $\Lambda_A\otimes\Lambda_B$, the lhs of this equation equals the quantum mutual information of the state $(\Lambda_A\otimes\Lambda_B)(\rho_{AB})$ and thus the inequality $I_q((\Lambda_A\otimes\Lambda_B)(\rho_{AB}))\leqslant I_q(\rho_{AB})$ is proved.

### Proof of lower bounds

We show that the quantities $I_q(\rho)/2$ and the regularized classical information $I_c^\infty(\rho)=\lim_{n\to\infty}(I_c(\rho^{\otimes n})/n)$ are both lower bounds for the entanglement cost $E_{LOq}$. The argument is similar to the proof of the $E_p^\infty$ lower bound on $E_{LOq}$ in Theorem 2. [The reasoning is in fact a special case of Theorem 4 in Ref. 18 (cf. Ref. 19) applied to the class LOq instead of the original LOCC.]

We start with a number, say $k$, of EPR pairs which have $I_q=2k$ and $I_c$ equal to $k$.[20] In the limit of large $n$, the ratio $k/n$ is the entanglement cost $E_{LOq}(\rho)$. We apply the LOq map $\mathcal{L}$ which uses $o(n)$ communication to obtain an approximation $\tilde{\rho}_n$ to $\rho^{\otimes n}$. Since the quantum mutual information and the classical mutual information can only increase by $o(n)$ by the process $\mathcal{L}$ applied to the initial EPR pairs, see Lemma 3, it follows that

$$I_q(\tilde{\rho}_n)\leqslant o(n)+2k, \tag{24}$$

and similarly

$$I_c(\tilde{\rho}_n)\leqslant o(n)+k. \tag{25}$$

The last step is to relate the mutual informations of $\tilde{\rho}_n$ to the mutual information of $\rho^{\otimes n}$. For this, we need a continuity result of the form

$$|I_{q/c}(\sigma)-I_{q/c}(\rho)|\leqslant C\log d\|\rho-\sigma\|_1+O(1) \tag{26}$$

for $\rho$, $\sigma$ on $\mathcal{H}_d$, $\|\rho-\sigma\|_1$ sufficiently small and $C$ is some constant.[21] Below we will prove these desired continuity results. We can divide Eqs. (24) and (25) by $n$ and take the limit of large $n$. We use the continuity relation of Eq. (26) and the fact that in the large $n$ limit $\tilde{\rho}_n$ tends to $\rho^{\otimes n}$. Thus we have

$$\lim_{n\to\infty}\frac{I_q(\rho^{\otimes n})}{n}=I_q(\rho)\leqslant 2E_{LOq}(\rho), \tag{27}$$

where we used that the quantum mutual information is additive, and similarly

$$I_c^\infty(\rho)\leqslant E_{LOq}(\rho). \tag{28}$$

What remains is to prove the continuity relations and the nonincrease modulo $o(n)$ under LOq operations.

### Continuity of mutual information

The continuity of the quantum mutual information $I_q(\rho)$ can be proved by invoking Fannes' inequality[22] and Ruskai's proof of nonincrease of the trace-distance under TCP maps.[23] Let $\rho$ and $\sigma$ be two density matrices which are close, i.e., $\|\rho-\sigma\|_1=\mathrm{Tr}|\rho-\sigma|\leqslant\epsilon$ for sufficiently small $\epsilon$. We have

$$|I_q(\rho_{AB})-I_q(\sigma_{AB})|\leqslant|S(\rho_A)-S(\sigma_A)|+|S(\rho_B)-S(\sigma_B)|+|S(\sigma_{AB})-S(\rho_{AB})|, \tag{29}$$

which can be bounded as

$$|I_q(\rho_{AB})-I_q(\sigma_{AB})|\leqslant 3\log d_{AB}\|\rho_{AB}-\sigma_{AB}\|_1+3\,\eta(\|\rho_{AB}-\sigma_{AB}\|_1), \tag{30}$$

where $\eta(x)=-x\log x$ and $\|\rho-\sigma\|_1\leqslant\frac{1}{3}$.

It is not hard to prove the continuity of the classical information of a quantum state, again using the nonincrease of $\|.\|_1$ under TCP maps. Let $M_A^\rho$ and $M_B^\rho$ be the optimal measurement achieving the classical mutual information $I_c(\rho)$. Under this measurement the states $\rho$ and $\sigma$, which is, say, close to $\rho$, go to probability distributions $p^\rho(i,j)$ and $p^\sigma(i,j)$ which are close again, i.e., $\|p^\rho-p^\sigma\|_1\leqslant\|\rho-\sigma\|_1$. We have that

$$I_c(\sigma)-I_c(\rho)\leqslant I(p^\sigma)-I(p^\rho)\leqslant\log k\|p^\rho-p^\sigma\|_1+O(1), \tag{31}$$

where $k$ is the number of joint outcomes in the optimal measurement $(M_A^\rho,M_B^\rho)$ and $I$ is the classical mutual information of a joint probability distribution. The last inequality in Eq. (31) could in principle be derived from Fannes' inequality, using diagonal matrices, but it is a standard continuity result in information theory[24] as well. To finish the argument, we should argue that $k$, the number of joint measurement outcomes, is bounded. The classical mutual information $I$ is a concave function of the joint probability $p(i,j)$.[24] Therefore only extremal measurements $M_A$ and $M_B$ need to be considered in the optimization over measurements. An extremal measurement has at most $d^2$ outcomes when acting on a space of dimension $d$ (Ref. 25) and thus $k\leqslant d_{AB}^2$. The same argument, interchanging $\sigma$ and $\rho$, can be used to upperbound $I_c(\rho)-I_c(\sigma)$.

*Lemma 3 (monotonicity properties of mutual information): Let $\mathcal{L}$ consist of a series of local operations assisted by m qubits of two-way communication. The quantum mutual information obeys the inequality*

$$I_q(\mathcal{L}(\sigma))\leqslant I_q(\sigma)+2m, \tag{32}$$

*for all states $\sigma$. For the classical mutual information we have*

$$I_c(\mathcal{L}(|\psi\rangle\langle\psi|)\leqslant I_c(|\psi\rangle\langle\psi|)+m, \tag{33}$$

*for all pure states $|\psi\rangle$.*

*Proof:* Let us first consider the quantum mutual information. We can decompose the two-way scheme $\mathcal{L}$ into a sequence of one-way schemes. It is sufficient to prove for such a one-way scheme using $m$ qubits of communication, say from Alice to Bob, that

$$I_q(\mathcal{L}(\sigma))\leqslant I_q(\sigma)+2m. \tag{34}$$

Alice's local action can consist of adding an ancilla $A'$ in some state and apply a TCP map to the systems $AA'$ thus obtaining the state $\sigma_{AA':B}$. Such an action does not increase the quantum nor classical mutual information as we showed before. Now Alice sends system $A'$ to Bob. We have

$$\begin{aligned} I_q(\sigma_{AB})\geqslant I_q(\sigma_{AA':B}) &= S(AA')+S(B)-S(AA'B)\\ &\geqslant S(AA')-S(A')+S(BA')-S(AA'B)\\ &\geqslant S(A)-2S(A')+S(BA')-S(AA'B)\\ &= I_q(\sigma_{A:BA'})-2S(A'), \end{aligned} \tag{35}$$

where we used $|S(A)-S(B)|\leqslant S(AB)\leqslant S(A)+S(B)$. The quantum mutual information of the final state is $I_q(\sigma_{A:BA'})$. Since $S(A')\leqslant m$, we obtain the needed inequality. Alice could send only a part of ancilla $A'$, but this does not change the bound.

Let us now consider the classical mutual information. We may convert the entire process $\mathcal{L}$ into a coherent process $\mathcal{L}$ where all the measurements are deferred to the end; this does not change the amount of communication that Alice and Bob carry out. Thus, prior to the measurements Alice and Bob have converted the pure state $|\psi\rangle$ into some pure state $|\phi\rangle$ whose local entropy is at most $E+m$ where $E$ is the entanglement of the state $|\psi\rangle$, which is equal to $I_c(|\psi\rangle\langle\psi|)$ (see Ref. 20). Now Alice and Bob locally measure and/or trace out some registers which are operations that do

not increase the classical mutual information. Therefore the final state $\mathcal{L}(|\psi\rangle\langle\psi|)$ has a classical mutual information that is bounded by the initial classical mutual information plus $m$. □

*Remark:* Note that Eq. (32) for the quantum mutual information applies to both pure and mixed states while we have found mixed states that violate Eq. (33) for the classical mutual information.

Let us state the final result once more:

*Corollary 1:* $E_{LOq}(\rho) \geq I_q(\rho)/2$ *and* $E_{LOq} \geq I_c^{\infty}(\rho)$.

With this corollary we can show that the LOq-entanglement cost of any correlated density matrix $\rho$ is nonzero.[26] Indeed, the quantum mutual information $I_q(\rho)$ of a correlated density matrix is strictly larger than zero, since $S(\rho_{AB})$ is strictly less than $S(\rho_A)+S(\rho_B)$ (equality is only obtained when $\rho_{AB}=\rho_A \otimes \rho_B$) and therefore $E_{LOq}(\rho)>0$.

We present a simple example for which $E_{LOq}(\rho)=E_p^{\infty}(\rho)>I_q(\rho)/2$.

*Example 1 (All correlation is classical correlation): Consider the separable state* $\rho = \Sigma_i p_i |a_i\rangle\langle a_i| \otimes |b_i\rangle\langle b_i|$ *where* $\langle a_i|a_j\rangle = \delta_{ij}$ *and* $\langle b_i|b_j\rangle = \delta_{ij}$. *In this case* $I_q(\rho)/2 = H(p)/2$. *However, we can show that* $E_p(\rho) \geq H(p)$. *We have* [*cf. Eq. (8)*] $\mu(\rho) = \Sigma_i p_i |b_i\rangle\langle b_i| \otimes |i\rangle\langle i|$. *Under some local TCP map* $\Lambda$ *we obtain a state* $\mu' = \Sigma_i p_i |b_i\rangle\langle b_i| \otimes \rho_i$ *where* $\rho_i$ *are density matrices. The entropy of* $\mu'$ *equals* $S(\mu') = \Sigma_i p_i S(\rho_i) + H(p) \geq H(p)$. *The entanglement of purification* $E_p(\rho)$ *may be nonadditive, so we have to consider* $E_p(\rho^{\otimes n})$. *We have* $\mu(\rho^{\otimes n}) = \mu^{\otimes n}$ *and now* $\mu' = \Sigma_{i_1,\ldots,i_n} p_{i_1} \cdots p_{i_n} |i_1,\ldots,i_n\rangle\langle i_1,\ldots,i_n| \otimes \rho_{i_1,\ldots,i_n}$. *Again the von Neumann entropy of* $\mu'$ *is larger than or equal to* $nH(p)$. *Note that in this example we do achieve the classical mutual information lower bound.*

Here is an example where the upper and lower bounds fix the (regularized) entanglement of purification:

*Example 2: Let* $\rho$ *be an equal mixture of the state* $|\Psi_0\rangle = (1/\sqrt{2})(|00\rangle + |11\rangle)$ *and* $|\Psi_1\rangle = (1/\sqrt{2})(|00\rangle - |11\rangle)$. *Alice and Bob can get one bit of classical mutual information by both measuring in the* $\{0,1\}$ *basis. Thus* $E_{LOq}(\rho) \geq I_c(\rho) = 1$, *but* $E_{LOq}(\rho) \leq S(\rho_A) \leq 1$, *Eq. (10). Therefore* $E_{LOq} = 1$.

## IV. OTHER CORRELATION MEASURES: THE LOCALLY INDUCED HOLEVO INFORMATION

In Ref. 1 the authors considered the locally induced Holevo information as a measure of classical correlations in the state. It is defined either with respect to Alice's measurement ($C_A$) or Bob's measurement ($C_B$)

$$C_{A/B}(\rho) = \max_{M_A/M_B} S\left(\sum_i p_i^{B/A} \rho_i^{B/A}\right) - \sum_i p_i^{B/A} S(\rho_i^{B/A}), \tag{36}$$

where $M_A(M_B)$ on $\rho$ gives reduced density matrices $\rho_i^B(\rho_i^A)$ with probability $p_i^B(p_i^A)$. The classical mutual information $I_c^{\infty}(\rho)$ will in general be less than these quantities, since to achieve the Holevo information one may have to do coding. In Ref. 1 it was shown that $C_{A/B}$ are nonincreasing under local operations. We leave it as an exercise for the reader to prove continuity and nonincrease modulo $o(n)$ under LOq operations (applied to some pure state), thus showing that the regularized versions of these two quantities are also lower bounds for $E_{LOq}$.

### Bell-diagonal states

We show that for Bell-diagonal states $\rho_{\text{Bell}}$ the quantity $C_A$ (equal to $C_B$ by symmetry of the Bell-diagonal states) is equal to the classical capacity of the corresponding qubit channels. By the previous arguments this give us some lower bounds on the regularized entanglement of purification of these states. The Bell-diagonal states are of the following form,

$$\rho_{\text{Bell}} = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|, \tag{37}$$

where $\Psi_{0\ldots3}$ are the four Bell states where $|\Psi_0\rangle$ is $(1/\sqrt{2})(|00\rangle+|11\rangle)$. The corresponding channel, the so called generalized depolarizing channel, or Pauli channel, is of the form

$$\Lambda_\rho(\cdot)=\sum_i p_i\sigma_i(\cdot)\sigma_i, \tag{38}$$

where $\sigma_0=\mathbf{1}$, and $\sigma_{1,2,3}$ are the three Pauli matrices. It is known[27] that all two qubit states with maximally mixed subsystems are Bell-diagonal, up to a unitary transformation $U_A\otimes U_B$. From the isomorphism between states and channels,[8,11,28] it follows that all unital channels are of the form (38) (cf. Ref. 29), up to unitary transformations applied before and after the action of the channel. The classical one-shot capacity of the quantum channel $\Lambda$ is given by[30,31]

$$C_1(\Lambda)=\sup_{\{q_i,\rho_i\}}\chi(\{q_i,\Lambda(\rho_i)\}), \tag{39}$$

where $\chi$ is the Holevo function of the ensemble

$$\chi(\{q_i,\rho_i\})=S\left(\sum_i q_i\rho_i\right)-\sum_i q_iS(\rho_i). \tag{40}$$

The optimal states $\rho_i$ that achieve the capacity $C_1$ are always pure states, moreover it can be shown[29] that the ensemble $\{q_i,|\psi_i\rangle\}$ that achieves $C_1$ for unital one-qubit channels satisfies

$$\sum_i q_i|\psi_i\rangle\langle\psi_i|=\frac{1}{2}\mathbf{1}. \tag{41}$$

Let us argue that $C_A(\rho)=C_1(\Lambda)$ for a Bell-diagonal state $\rho_{\mathrm{Bell}}=(\mathbf{1}_A\otimes\Lambda_\rho)(|\Psi_0\rangle\langle\Psi_0|)$. Alice's POVM measurement on this state commutes with the channel $\Lambda_\rho$. By doing a measurement on $|\Psi_0\rangle$ she can create any pure-state-ensemble on system $B$, obeying the relation Eq. (41). This ensemble is then sent through the channel $\Lambda_\rho$. If the ensemble is optimal for $C_1$, then its Holevo information $\chi$ equals $C_1$ and thus $C_A=C_1$.

For unital one-qubit channels $C_1$ is given by[29,32]

$$C_1(\Lambda)=1-\min_\psi S(\Lambda(|\psi\rangle\langle\psi|)). \tag{42}$$

We can perform the minization in the last inequality and we obtain the following formula for the capacity of a Pauli channel or the induced Holevo information of the Bell-diagonal states,

$$C_A(\rho_{Bell})=C_1(\Lambda_\rho)=1-H(1-\lambda), \tag{43}$$

where $\lambda$ is the sum of the two largest probabilities $p_i$ and $H(.)$ is the binary entropy function $H(x)=-x\log x-(1-x)\log(1-x)$. For two-qubit Werner states of the form

$$\rho_W=e|\Psi_0\rangle\langle\Psi_0|+(1-e)/3\sum_{i=1}^{3}|\Psi_i\rangle\langle\Psi_i|, \tag{44}$$

we obtain

$$\begin{aligned} C_A&=1-H\left(\frac{1+2e}{3}\right)\quad \text{for } e\in\left[\frac{1}{4},1\right],\\ C_A&=1-H\left(\frac{2-2e}{3}\right)\quad \text{for } e\in\left[0,\frac{1}{4}\right]. \end{aligned} \tag{45}$$

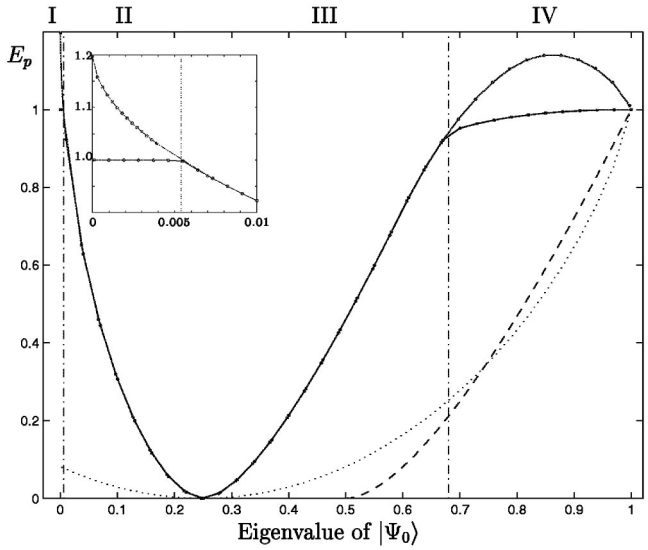


FIG. 1. Numerical bounds on $E_p$ for Werner states. In the upper curve we restrict to $\dim(A')=\dim(B')=2$; for the next curve, we permit $\dim(A')=\dim(B')=4$. The inset shows the curious behavior of $E_p$ around the point where the eigenvalue of $|\Psi_0\rangle$ approaches zero. The dotted curve is the $C_A$ lower bound of Sec. IV A. The dashed curve is the entanglement of formation lower bound which vanishes when the eigenvalue is smaller than or equal to $\frac{1}{2}$.

It was shown by King[32] that the classical capacity of unital one-qubit channels is equal to the one shot capacity, or $C_1=C_1^{\infty}\equiv\lim_{n\to\infty}(1/n)\,C_1(\rho^{\otimes n})$. Therefore $C_A=C_A^{\infty}=C_1$, which is a lower bound on $E_{LOq}$.

## V. WERNER STATES

A numerical minimization based on Eq. (6) was performed for the Werner states Eq. (44) for $E_p$. We plot the results as a function of the $|\Psi_0\rangle$ eigenvalue $e$ in Fig. 1. We permitted various output dimensions; The two curves shown have $\dim(A')=\dim(B')=2$ and $\dim(A')=\dim(B')=4$. In the first case, the initial variable of the minimization was determined by a random $4\times4$ unitary $U_{A'B'}$ picked according to the Haar measure. In the second case, the initial point was determined by a random $16\times4$ isometry picked according to a parameterization derived from Ref. 33. We did not explore the largest dimensions permitted by Lemma 1, which would have required an optimization over a $64\times4$ isometry.

It is evident from the numerics presented in the figure that the $C_A$ bound of Eq. (45) is not achieved for the Werner states: the $C_A$ lower bound is only tight at the trivial points $e=\frac{1}{4}$ and $e=1$. Our results indicate that $E_p$ is a very complex function, neither concave nor convex, with several distinct regimes. In fact, we find four different regimes in our numerics: (I) In this regime the standard purification of Eq. (5) appears to be optimal, so the $U$ of Eq. (6) is the identity, and the purifying dimensions are $\dim(A')=1$ and $\dim(B')=4$. This regime only extends over a tiny range, approximately $0\leqslant e\leqslant0.005$. (II) In the range $0.005\leqslant e\leqslant0.25$ we find an optimal purification of the form

$$\sqrt{e}|\Psi_0\rangle_{AB}|\Psi_0\rangle_{A'B'}+\sqrt{\frac{1-e}{3}}(|\Psi_1\rangle_{AB}|\Psi_1\rangle_{A'B'}+|\Psi_2\rangle_{AB}|\Psi_2\rangle_{A'B'}+|\Psi_3\rangle_{AB}|\Psi_3\rangle_{A'B'}). \tag{46}$$

In this region the $E_p$ curve is given by $E_p=-x\log x-(1-x)\log((1-x)/3)$, with $x=(1+2e-2\sqrt{3}\sqrt{e(1-e)})/12$. Here the purifying dimensions are $\dim(A')=2$ and $\dim(B')=2$. Of course $E_p$ drops to zero for the completely mixed state at $e=\frac{1}{4}$. (III) In the range $0.25\leqslant e\leqslant0.69$ we also find purifying dimensions $\dim(A')=2$ and $\dim(B')=2$, but we were unable to determine the

analytical form of the purifying state or of $E_p$. (IV) In the range $0.69 \leqslant e \leqslant 1$ the purifying dimensions were $\dim(A') = 2$ and $\dim(B') = 3$. Again, we were unable to come to any analytical understanding of the result. Of course, $E_p = 1$ for $e = 1$, corresponding to the pure maximally entangled state.

## VI. CONCLUSION

We have shown that the entanglement cost $E_{LOq}(\rho)$ is equal to the regularized entanglement of purification. It is an open question whether the entanglement of purification is additive:

$$E_p(\rho \otimes \rho) \stackrel{?}{=} E_p(\rho) + E_p(\rho). \tag{47}$$

In the alternative formulation using the state $\mu(\rho)$ the additivity question is the following. Is the minimum in

$$\min_{\Lambda_{CD}} S((I_{AB} \otimes \Lambda_{CD})(\mu_{AC} \otimes \mu_{BD})), \tag{48}$$

achieved by a TCP map $\Lambda_{CD} = \mathcal{S} \otimes \mathcal{S}$? This problem is similar again to the additivity question encountered in Ref. 14 where a local map could possibly lower the ratio of the coherent information and the entropy of many copies of a state together.

It is interesting not only to ask the formation question with respect to this class LOq, but also consider ''the distillation'' question. One can consider different versions. For example, how much entanglement can we distill from $\rho$ using $o(n)$ communication? One would expect that this quantity $D_{LOq}(\rho)$ is always zero for states for which the entanglement cost $E_c$ (using *LOCC*) is lower than the distillable entanglement $D$. We do not have a proof of this statement, relating irreversibility to a need for classical communication.

Instead of trying to convert the correlations in $\rho$ back to entanglement, we may ask what classical correlations Alice and Bob can establish using $\rho$. We could allow Alice and Bob to perform an asymptotically vanishing amount of communication in this extraction process. A little bit of communication could potentially increase the classical mutual information in a quantum state by a large amount (when the classical correlation is initially ''hidden''), thus this may not be the best problem to pose. Researchers[34,35] have investigated the possibly more interesting problem of the *secret* key $K$ that Alice and Bob can establish given $\rho$ where one allows arbitrary public classical communication between the parties. There is again more than one version of this problem, one in which Eve possesses the purification of the density matrix[34] and a situation in which Eve is initially uncorrelated with the density matrix. In Ref. 36 a general framework is developed to address these issues also in the multipartite setting.

Quite recently, entanglement properties of bipartite density matrices were studied by looking at mixed extensions of the density matrix.[37] It would be interesting to explore the connection between our results here on the entanglement of purification and this other approach.

## ACKNOWLEDGMENTS

B.M.T., D.W.L., and D.P.D. are grateful for the support of the National Security Agency and the Advanced Research and Development Activity through Army Research Office Contract Nos DAAG55-98-C-0041 and DAAD19-01-C-0056 and partial support from the National Reconnaissance Office. This work was also supported in part by the National Science Foundation under Grant. No. EIA-0086038. M.H. acknowledges the support of EU grant EQUIP, Contract No. IST-1999-11053. We thank Charles Bennett, Paweł Horodecki, Ryszard Horodecki and John Smolin for a pleasant IBM lunch discussion on this topic. The concept of entanglement of purification was raised in a discussion of M.H. with Ryszard Horodecki. M.H. would also like to thank Robert Alicki and Ryszard Horodecki for stimulating discussions. B.M.T. would like to thank Andreas

Winter for interesting discussions about the secret key rate $K$ and its relation to other correlation measures. D.W.L. would like to thank Charles Bennett and John Smolin for discussions on mixed state inputs that violate Eq. (33).

[1] L. Henderson and V. Vedral, J. Phys. A **34**, 6899 (2001); quant-ph/0105028.
[2] N. J. Cerf and C. Adami, Phys. Rev. A **56**, 3470 (1997); quant-ph/9609024.
[3] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, quant-ph/0112074.
[4] T. Osborne and M. Nielsen, quant-ph/0202162.
[5] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).
[6] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).
[7] H.-K. Lo and S. Popescu, Phys. Rev. Lett. **83**, 1459 (1999); quant-ph/9902045.
[8] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996); quant-ph/9604024.
[9] P. Hayden, M. Horodecki, and B. M. Terhal, J. Phys. A **34**, 6891 (2001); quant-ph/0008134.
[10] J. Bouda and V. Buzek, to appear in Phys. Rev. A; quant-ph/0112015.
[11] M.-D. Choi, Linear Algebr. Appl. **10**, 285 (1975).
[12] We have a special case when $d_2=\infty$. The Stinespring theorem (Ref. 13) implies that we have an operator-sum representation of such a map. Then Choi's results on extremality apply, bounding the number of operation elements, from which the final result can be proved.
[13] W. F. Stinespring, Proc. Am. Math. Soc. **6**, 211 (1955).
[14] M. Horodecki, P. Horodecki, R. Horodecki, D. W. Leung, and B. M. Terhal, Quantum Inf. Comput. **1**, 70 (2001); quant-ph/0106080.
[15] H. Barnum, J. A. Smolin, and B. M. Terhal, Phys. Rev. A **58**, 3496 (1998); quant-ph/9711032.
[16] M. A. Nielsen, Phys. Rev. A **61**, 064301 (2000); quant-ph/9908086.
[17] V. Vedral and M. Plenio, Phys. Rev. A **57**, 1619 (1998); quant-ph/9707035.
[18] M. Horodecki, Quantum Inf. Comput. **1**, 3 (2001).
[19] M. J. Donald, M. Horodecki, and O. Rudolph, quant-ph/0105017.
[20] One can prove that $I_c \leqslant k$ by observing that any local measurement that is not projecting in the Schmidt basis is a noisy version of the measurement that does project in the Schmidt basis. In other words, the probability distribution of any set of local measurements can be obtained from the probability distribution of the Schmidt basis measurement by local processing, which does not increase the classical mutual information.
[21] We can alternatively write down a continuity relation using the Bures distance. Since the trace distance $\|.\|_1$ and the Bures distance are equivalent distances, one continuity relation follows from the other and vice versa.
[22] M. Fannes, Commun. Math. Phys. **31**, 291 (1973).
[23] M.-B. Ruskai, Rep. Math. Phys. **6**, 1147 (1994).
[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[25] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Dordrecht, 1993).
[26] Note that this does not directly follow from the result in Ref. 10, since the entanglement of purification may be nonadditive.
[27] M. Horodecki and R. Horodecki, Phys. Rev. A **54**, 1838 (1996); quant-ph/9607007.
[28] A. Jamiołkowski, Rev. Mod. Phys. **3**, 275 (1972).
[29] C. King and M.-B. Ruskai, IEEE Trans. Inf. Theory **47**, 192 (2001); quant-ph/9911079.
[30] B. Schumacher and M. Westmoreland, Phys. Rev. A **56**, 131 (1997).
[31] A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998); quant-ph/9611023.
[32] C. King, quant-ph/0103156.
[33] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Phys. Rev. Lett. **73**, 58 (1994).
[34] N. Gisin and S. Wolf, quant-ph/0005042.
[35] A. Winter and R. Wilmink, unpublished manuscript, private communication.
[36] N. J. Cerf, S. Massar, and S. Schneider, quant-ph/0202103.
[37] R. Tucci, quant-ph/0101123.