

Barbara M. Terhal<sup>1,2</sup>, Michał Horodecki<sup>3</sup>, Debbie W. Leung<sup>2</sup> and David P. DiVincenzo<sup>1,2</sup>

<sup>1</sup> *Institute for Quantum Information, California Institute of Technology, Pasadena, CA 91125, USA*

<sup>2</sup> *IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA*

<sup>3</sup> *Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland*

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 classical 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  $\mathcal{H}_2 \otimes \mathcal{H}_2$ .

## 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 non-increasing 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  $|\psi\rangle = \sum_i \sqrt{\lambda_i} |a_i\rangle \otimes |b_i\rangle$  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. [?] or [?]. 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.

If we want to deal with classical correlations in an operational way at all, we will have to treat quantum entanglement and classical correlation in a unified framework, namely we should express both correlations in units of entanglement. It is the goal of this paper to develop this framework.

It has been the experience in (quantum) information theory that questions in the asymptotic approximate regime are easier to answer than exact non-asymptotic 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:

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

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  $F(\rho, \rho') = \text{Tr}(\rho^{1/2} \rho' \rho^{1/2})$ . 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 [?]. 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. [?].

We see that the cost  $E_{LOq}(\rho)$  of creating the state  $\rho$  is defined analogously to the entanglement cost  $E_c(\rho)$  [?], 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 [?] to be equal to

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

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

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

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

Our paper 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. [?]. We prove that for Bell-diagonal states the correlation measure  $C_A$  is equal to the classical capacity of the related 1-qubit Pauli channel. At the end of the paper 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 paper 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

$$E_p(\rho) = \min_{\psi: \text{Tr}_{A'B'}|\psi\rangle\langle\psi|=\rho} E(|\psi\rangle\langle\psi|). \quad (4)$$

Let  $\{\lambda_i, |\psi_i\rangle\}$  be the eigenvalues and eigenvectors of  $\rho_{AB}$ . The “standard purification” of  $\rho$  is defined as

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

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:

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

$$\begin{aligned} &= \min_{U_{A'B'}} S(\text{Tr}_{AA'}(I_{AB} \otimes U_{A'B'})|\psi_s\rangle\langle\psi_s|(I_{AB} \otimes U_{A'B'})^\dagger)) \\ &= \min_{\Lambda_{B'}} S((I_B \otimes \Lambda_{B'})\mu_{BB'}(\rho)), \end{aligned} \quad (7)$$

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

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

and  $\Lambda_{B'}(\nu) \equiv \text{Tr}_{A'}U_{A'B'}(\nu_{B'} \otimes |0\rangle\langle 0|_{A'})U_{A'B'}^\dagger$ . 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  $\text{Tr}_{B'}\sigma_{ABB'} = \rho_{AB}$ , and variation over  $\Lambda_{B'}(\nu)$ . Both formulations will be used throughout the paper.

The idea of bipartite purifications was considered in Ref. [?] where the authors proved that every correlated state has, in our language, a nonzero entanglement of purification.

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  $E_f(\rho) \leq E_p(\rho)$ . To prove this lower bound, let  $|\psi_\rho\rangle = \sum_{i,j} |i\rangle_{A'} |j\rangle_{B'} \otimes |\psi_{ij}\rangle$  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

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

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

$$E_p(\rho) \leq \min(S(\rho_A), S(\rho_B)), \quad (10)$$

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 Section 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 4 Bell states, each with 1 ebit of entanglement of purification. We emphasize that the entanglement of purification is neither an entanglement measure nor entanglement monotone. Instead, it is a measure of combined quantum and classical correlations in a bipartite quantum state expressed in units of entanglement.

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. This observation allows 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  $d_{AB} = d_{AB}$ . 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 Choi [?] shows that an extremal TCP map  $\Lambda : B(\mathcal{H}_{d_1}) \rightarrow B(\mathcal{H}_{d_2})^1$  can be written with at most  $d_1$  operations elements, that is, has the form

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

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

---

<sup>1</sup>We have a special case when  $d_2 = \infty$ . The Stinespring theorem [?] 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.

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  $d_{B'} = d_{AB}^2$ .  $\square$

It is interesting that a similar minimization problem was encountered in Ref. [?]. 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  $D(\rho, \sigma) \leq \epsilon$ . Then*

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

for small enough  $\epsilon$ .

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

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

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 (non-optimal) purification  $|\psi_\sigma\rangle$  as  $(U \otimes \mathbf{1})|\psi'_\sigma\rangle = |\psi_\sigma\rangle$ . Now we have

$$E_p(\sigma) - E_p(\rho) = E(|\phi_\sigma\rangle\langle\phi_\sigma|) - E(|\phi_\rho\rangle\langle\phi_\rho|) \leq E(|\psi_\sigma\rangle\langle\psi_\sigma|) - E(|\phi_\rho\rangle\langle\phi_\rho|). \quad (14)$$

We use continuity of entanglement [?,?], 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

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

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  $U$ .  $\square$

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*

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

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

$$\sum_i p_i E_p(\rho_i) \leq E_p(\rho). \quad (17)$$

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

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

*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  $|\psi_i\rangle = \frac{A_i \otimes I_B |\psi\rangle}{\sqrt{\langle \psi | A_i^\dagger A_i \otimes I_B | \psi \rangle}}$  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

$$E_p(\rho) = E(|\psi_\rho\rangle\langle\psi_\rho|) \geq \sum_i p_i E(|\psi_i\rangle\langle\psi_i|) \geq \sum_i p_i E_p(\rho_i). \quad (19)$$

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  $E_p(\rho) + m$ .  $\square$

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  $E_{LOq}(\rho) \leq E_p^\infty(\rho)$  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  $o(p) \leq o(pk)$  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  $E_p^\infty(\rho) \leq E_{LOq}(\rho)$  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  $E_p(\tilde{\rho}_k) \leq n + o(k)$ . Using the continuity of Theorem 1 and dividing the last inequality by  $k$  and taking the limit  $k \rightarrow \infty$  gives  $E_p^\infty(\rho) \leq E_{LOq}(\rho)$ .  $\square$

### 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

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

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

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

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 non-increasing 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 non-increasing 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

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

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

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

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

#### A. Proof of Lower bounds

We show that the quantities  $I_q(\rho)/2$  and the regularized classical information  $I_c^\infty(\rho) = \lim_{n \rightarrow \infty} \frac{I_c(\rho^{\otimes n})}{n}$  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. [?] (cf. Ref. [?]) 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  $k$ <sup>2</sup>. 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

---

<sup>2</sup>One can prove that  $I_c \leq k$  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.

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

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

and similarly

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

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

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

for  $\rho, \sigma$  on  $\mathcal{H}_d$ ,  $\|\rho - \sigma\|_1$  sufficiently small and  $C$  is some constant<sup>3</sup>. 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

$$\lim_{n \rightarrow \infty} \frac{I_q(\rho^{\otimes n})}{n} = I_q(\rho) \leq 2E_{LOq}(\rho), \quad (27)$$

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

$$I_c^\infty(\rho) \leq E_{LOq}(\rho). \quad (28)$$

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

### 1. Continuity of Mutual Informations

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

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

which can be bounded as

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

where  $\eta(x) = -x \log x$  and  $\|\rho - \sigma\|_1 \leq 1/3$ .

It is not hard to prove the continuity of the classical information of a quantum state, again using the nonincrease of  $\|\cdot\|_1$  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.  $\|p^\rho - p^\sigma\|_1 \leq \|\rho - \sigma\|_1$ . We have that

$$I_c(\sigma) - I_c(\rho) \leq I(p^\sigma) - I(p^\rho) \leq \log k \|p^\rho - p^\sigma\|_1 + O(1), \quad (31)$$

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 [?] 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  $p(i, j)$  [?]. 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  $d$  [?] and thus  $k \leq d_{AB}^2$ . The same argument, interchanging  $\sigma$  and  $\rho$ , can be used to upperbound  $I_c(\rho) - I_c(\sigma)$ .

---

<sup>3</sup>We can alternatively write down a continuity relation using the Bures distance. Since the trace-distance  $\|\cdot\|_1$  and the Bures distance are equivalent distances, one continuity relation follows from the other and vice versa.

**Lemma 3 (Monotonicity Properties of Mutual Information)** *Let  $\mathcal{L}$  consist of a series of local operations assisted by  $m$  qubits of 2-way communication. The quantum mutual information obeys the inequality*

$$I_q(\mathcal{L}(\sigma)) \leq I_q(\sigma) + 2m, \quad (32)$$

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

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

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

*Proof:* Let us first consider the quantum mutual information. We can decompose the 2-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

$$I_q(\mathcal{L}(\sigma)) \leq I_q(\sigma) + 2m. \quad (34)$$

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

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

where we used  $|S(A) - S(B)| \leq S(AB) \leq S(A) + S(B)$ . The quantum mutual information of the final state is  $I_q(\sigma_{A:BA'})$ . Since  $S(A') \leq m$ , 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 footnote 2). 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  $m$ .  $\square$ .

*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<sup>4</sup>. 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  $\rho = \sum_i p_i |a_i\rangle\langle a_i| \otimes |b_i\rangle\langle b_i|$  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))  $\mu(\rho) = \sum_i p_i |b_i\rangle\langle b_i| \otimes |i\rangle\langle i|$ . Under some local TCP map  $\Lambda$  we obtain a state  $\mu' = \sum_i p_i |b_i\rangle\langle b_i| \otimes \rho_i$  where  $\rho_i$  are density matrices. The entropy of  $\mu'$  equals  $S(\mu') = \sum_i p_i S(\rho_i) + H(p) \geq H(p)$ . 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  $\mu' = \sum_{i_1, \dots, i_n} p_{i_1} \dots p_{i_n} |i_1 \dots i_n\rangle\langle i_1 \dots i_n| \otimes \rho_{i_1, \dots, i_n}$ . 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  $|\Psi_0\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$  and  $|\Psi_1\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$ . Alice and Bob can get 1 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  $E_{LOq} = 1$ .

---

<sup>4</sup>Note that this does not directly follow from the result in Ref. [?], since the entanglement of purification may be nonadditive.

In Ref. [?] 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 ( $C_B$ )

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

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. [?] 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  $E_{LOq}$ .

### A. Bell-diagonal states

We show that for Bell-diagonal states  $\rho_{Bell}$  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

$$\rho_{Bell} = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|, \quad (37)$$

where  $\Psi_{0...3}$  are the four Bell states where  $|\Psi_0\rangle$  is  $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ . The corresponding channel, –the so called generalized depolarizing channel–, or Pauli channel, is of the form

$$\Lambda_\rho(\cdot) = \sum_i p_i \sigma_i(\cdot) \sigma_i, \quad (38)$$

where  $\sigma_0 = \mathbf{1}$ , and  $\sigma_{1,2,3}$  are the three Pauli matrices. It is known [?] 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 [?, ?, ?], it follows that all unital channels are of the form (38) [cf. [?]], up to unitary transformations applied before and after the action of the channel. The classical 1-shot capacity of the quantum channel  $\Lambda$  is given by [?, ?]

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

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

$$\chi(\{q_i, \rho_i\}) = S(\sum_i q_i \rho_i) - \sum_i q_i S(\rho_i). \quad (40)$$

The optimal states  $\rho_i$  that achieve the capacity  $C_1$  are always pure states, moreover it can be shown [?] that the ensemble  $\{q_i, |\psi_i\rangle\}$  that achieves  $C_1$  for unital 1-qubit channels satisfies

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

Let us argue that  $C_A(\rho) = C_1(\Lambda)$  for a Bell-diagonal state  $\rho_{Bell} = (\mathbf{1}_A \otimes \Lambda_\rho)(|\Psi_0\rangle\langle\Psi_0|)$ . 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 1-qubit channels  $C_1$  is given by [?, ?]

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

We can perform the minimization 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



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

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

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

we obtain

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

It was shown by King [?] that the classical capacity of unital 1-qubit channels is equal to the one shot capacity, or  $C_1 = C_1^\infty \equiv \lim_{n \rightarrow \infty} \frac{1}{n} C_1(\rho^{\otimes n})$ . 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  $4 \times 4$  unitary  $U_{A'B'}$  picked according to the Haar measure. In the second case, the initial point was determined by a random  $16 \times 4$  isometry picked according to a parameterization derived from Ref. [?]. We did not explore the largest dimensions permitted by Lemma 1, which would have required an optimization over a  $64 \times 4$  isometry.

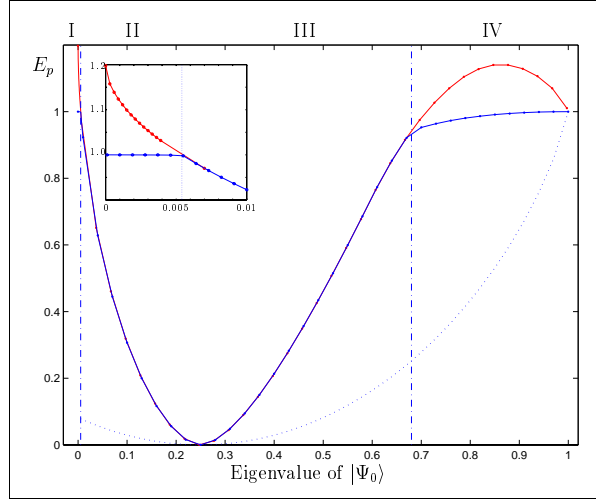


FIG. 1. Numerical bounds on  $E_p$  for Werner states. In the upper curve we restrict to  $\dim(A') = \dim(B') = 2$ ; for the next curve, we permit  $\dim(A') = \dim(B') = 4$ . The inset shows the curious behavior of  $E_p$  around the point where the eigenvalue of  $|\Psi_0\rangle$  approaches zero.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

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  $e = 1/4$  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  $0 \leq e \leq 0.005$ . II) In the range  $0.005 \leq e \leq 0.25$  we find an optimal purification of the form

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

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  $e = 1/4$ . III) In the range  $0.25 \leq e \leq 0.69$  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  $0.69 \leq e \leq 1$  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:

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

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

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

achieved by a TCP map  $\Lambda_{CD} = \mathcal{S} \otimes \mathcal{S}$ ? This problem is similar again to the additivity question encountered in Ref. [?] 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. Winter and Wilms [?] 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 (assuming that the purification of  $\rho$  is not in the hands of the eavesdropper Eve). This quantity is at least  $I_c^\infty(\rho)$  and it can be proved that  $K \geq \max(C_A, C_B)$  as well, but in general  $K$  can be even higher [?]. On the other hand, the secret key rate  $K$  is smaller than or equal to the regularized entanglement of purification.

## VII. 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 numbers 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 KBN grant No. 2 P03B 103 16 and 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. M.H. would 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).