Distillation Protocols: Output Entanglement and Local Mutual Information

Michał Horodecki,[1] Jonathan Oppenheim,[1,2] Aditi Sen(De),[3] and Ujjwal Sen[3]
[1]*Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland*
[2]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom*
[3]*Institut für Theoretische Physik, Universität Hannover, D-30167 Hannover, Germany*
(Received 22 June 2004; published 20 October 2004)

A complementary behavior between local mutual information and average output entanglement is derived for arbitrary bipartite ensembles. This leads to bounds on the yield of entanglement in distillation protocols that involve disinguishing. This bound is saturated in the hashing protocol for distillation, for Bell-diagonal states.

DOI: 10.1103/PhysRevLett.93.170503 PACS numbers: 03.67.Hk, 03.67.Mn

Introduction.—Distillation of entanglement [1,2] is a key issue in attaining nonclassical tasks in quantum communication protocols [3]. In a typical communication protocol, entanglement must be shared between distant partners (Alice and Bob). Since channels are invariably noisy, the partners usually end up with mixed state entanglement, which must then be distilled into pure form via local operations and classical communication (LOCC), to make them amenable to the envisaged quantum communication protocol.

The aim of this paper is twofold. We obtain an upper bound on local mutual information, I^{LOCC}, of arbitrary bipartite ensembles. We then use this bound to provide bounds on the yield of entanglement in any distillation protocol that uses local distinguishing of ensembles of states. The obtained bounds are then compared with the yield in the existing distillation protocols (e.g. [1,2,4]) and similar generalizations thereof, and also in some other cases, in which the distillation is based on a distinguishability protocol [5,6]. As a spin-off, we obtain a complementarity relation between local mutual information and average output entanglement.

Generalized universal Holevo-like upper bound on local mutual information.—To begin, we obtain a generalized Holevo-like bound on local mutual information for arbitrary bipartite ensembles. Suppose then that a source prepares the ensemble $\mathcal{R} = \{p_x, \varrho_x^{AB}\}$ and sends the A part to Alice and the B part to Bob. The task of Alice and Bob is to estimate the identity x of the sent state. If Alice and Bob are together, so that they are allowed to perform global operations, the mutual information is bounded by the Holevo quantity [7], $\chi_{\mathcal{R}} = S(\varrho) - \sum_x p_x S(\varrho_x)$, where ϱ is the average ensemble state $\sum_x p_x \varrho_x$. $S(\cdot)$ is the von Neumann entropy and is defined for a state ϱ as $S(\varrho) = -\mathrm{tr}\varrho\log_2\varrho$. We will however need the following result [8,9], which is a generalization of the Holevo bound on mutual information.

Lemma 1: *If a measurement on ensemble $Q = \{p_x, \varrho_x\}$ produces result y with probability p_y, and leaves a postmeasurement ensemble $Q^y = \{p_{x|y}, \varrho_{x|y}\}$, then the mutual information I (between the identity of state in the ensemble and measurement outcome) extracted from the measurement has the following bound:*

$$I \leq \chi_Q - \overline{\chi}_{Q^y}. \tag{1}$$

Here $\overline{\chi}_{Q^y}$ is the average Holevo bound for the possible postmeasurement ensembles, i.e., $\sum_y p_y \chi_{Q^y}$.

Suppose now that Alice and Bob are far apart, so that they are able to perform only local operations and communicate classically between the operations. In this scenario, universal Holevo-like upper bound on local mutual information for an arbitrary bipartite ensemble $\{p_x, \varrho_x^{AB}\}$ was obtained in [9]

$$I^{\mathrm{LOCC}} \leq S(\varrho^A) + S(\varrho^B) - \max_{Z=A,B} \sum_x p_x S(\varrho_x^Z). \tag{2}$$

Here $\varrho_x^{A(B)} = \mathrm{tr}_{B(A)}(\varrho_x^{AB})$, and $\varrho^{A(B)} = \mathrm{tr}_{B(A)}\sum_x p_x \varrho_x^{AB}$. In this paper, we will prove a generalization of this bound. Precisely, we show that

$$I^{\mathrm{LOCC}} \leq S(\varrho^A) + S(\varrho^B) - \sum_x p_x S(\varrho_x^B) - \sum_{a,b,\ldots,(n)} p_{a,b,\ldots,(n)} S\left(\sum_x p_{x|ab,\ldots,(n)} \varrho^A_{x|ab,\ldots,(n)}\right). \tag{3}$$

Here $\{p_{x|ab,\ldots,(n)}, \varrho^{AB}_{x|ab,\ldots,(n)}\}$ is the postmeasurement ensemble obtained after the measurement in the nth step, and $p_{a,b,\ldots,(n)}$ is the probability of the sequence of measurement outcomes in steps $1, 2, \ldots, n$. Our generalization in (3) is related to the previous bound in (2), in a similar way as Lemma 1 is related to the original Holevo bound.

We will now prove the inequality in (3). To start the protocol for obtaining the identity x of the given ensemble $\mathcal{R} = \{p_x, \varrho_x^{AB}\}$, Alice makes a measurement [10], and suppose that she obtains an outcome a, with probability p_a. Suppose that the postmeasurement ensemble (for outcome a at Alice) is $\mathcal{R}_a = \{p_{x|a}, \varrho^{AB}_{x|a}\}$.

The results presented in this paper are in terms of mutual information, which when maximized over all measurement strategies gives the "accessible information". All the results are of course true for the extreme case of the best measurement strategy (for attaining

 © 2004 The American Physical Society

maximal mutual information), but are true also for any other nonextreme measurement strategy. The mutual information gathered from the measurement of Alice has the following bound due to Lemma 1: $I_1^A \leq \chi_{\mathcal{R}^A} - \overline{\chi}_{\mathcal{R}_a^A}$. Here $\chi_{\mathcal{R}^A}$ is the Holevo quantity of the A part of the ensemble $\mathcal{R}$, i.e., of the ensemble $\mathcal{R}^A = \{p_x, \varrho_x^A\}$. And $\chi_{\mathcal{R}_a^A}$ is the Holevo quantity of the A part of the ensemble $\mathcal{R}_a$. The subscript 1 in I_1^A indicates that the information is extracted from the first measurement.

After Alice communicates her result to Bob, his ensemble is $\mathcal{R}_a^B = \{p_{x|a}, \varrho_{x|a}^B\}$, with $\varrho_x^B = \mathrm{tr}_A(\varrho_x^{AB})$. Suppose now that Bob performs a measurement and obtains outcome b with probability p_b, so that the postmeasurement ensemble (at his part) is $\mathcal{R}_{ab}^B = \{p_{x|ab}, \varrho_{x|ab}^B\}$, where $\varrho_{x|ab}^B = \mathrm{tr}_A(\varrho_{x|ab}^{AB})$. So (again due to Lemma 1), the information extracted in Bob's measurement has the following bound: $I_2^B \leq \overline{\chi}_{\mathcal{R}_a^B} - \overline{\chi}_{\mathcal{R}_{ab}^B}$.

This procedure of measuring and communicating the result goes on for an arbitrary number of steps, and by the chain rule for mutual information (see, e.g., [11]), the mutual information obtained in all steps is $I^{\mathrm{LOCC}} = I_1^A + I_2^B + I_3^A + \cdots$. Note that this quantity depends on the measurement strategy followed by Alice and Bob.

Now we (repeatedly) use the following facts: (i) The von Neumann entropy is concave (i.e., $S(p_1\varrho_1 + p_2\varrho_2) \geq p_1 S(\varrho_1) + p_2 S(\varrho_2)$, for arbitrary density matrices ϱ_1 and ϱ_2, and probabilities p_1 and p_2) and positive. (ii) A measurement on one subsystem cannot change the state at a distant subsystem. (iii) The average change (initial minus final) of von Neumann entropy due to a measurement on one subsystem cannot be less than the average change in a distant subsystem. So, for example, after the first measurement by Alice, we have $\sum_x p_x S(\varrho_x^A) - \sum_a p_a \sum_x p_{x|a} S(\varrho_{x|a}^A) \geq \sum_x p_x S(\varrho_x^B) - \sum_a p_a \sum_x p_{x|a} S(\varrho_{x|a}^B)$. (iv) The Holevo quantity is positive.

Then after n steps of measurements, we obtain the inequality (3).

We have assumed that the last measurement is performed by Alice. The last term of the bound (3) is a contribution from this last measurement by Alice. We will see below that the final result is free from this asymmetry. Moreover, for the same measurements, but using the above items (i)-(iv) in a different way, one can reach the inequality (3), but with A and B interchanged, i.e., we also have

$$I^{\mathrm{LOCC}} \leq S(\varrho^A) + S(\varrho^B) - \sum_x p_x S(\varrho_x^A) - \sum_{a,b,\ldots,(n-1)} p_{a,b,\ldots,(n-1)} S\Bigg(\sum_x p_{x|ab,\ldots,(n-1)} \times \varrho_{x|ab,\ldots,(n-1)}^B\Bigg). \quad (4)$$

Note that now the last term is a contribution from the next-to-last measurement, which (due to the assumption that Alice performed the last measurement) is performed by Bob. Inequalities (3) and (4) give us upper bounds on local mutual information, for *arbitrary* bipartite ensembles. These inequalities are true for any measurement strategy of Alice and Bob. In particular, they are true for the one which maximizes I^{LOCC}. This is then the so-called locally accessible information ($I_{\mathrm{acc}}^{\mathrm{LOCC}}$).

The last terms in the bounds on local mutual information in inequalities (3) and (4) respectively are negative quantities, due to the positivity of von Neumann entropy. Leaving it out, we have the inequality (2).

Input and output entanglements.—We now try to write the bounds on local mutual information in (3) and (4) in a more revealing form. To that end, note that the von Neumann entropy of either of the local density matrices of a bipartite state is no smaller than the entanglement of formation [2], and the entanglement of formation is a lower bound for any asymptotically consistent measure of bipartite entanglement [12].

Then, the last term in the upper bound of Eq. (4) is $\leq -\sum_{a,b,\ldots,(n-1)} p_{a,b,\ldots,(n-1)} E(\sum_x p_{x|ab,\ldots,(n-1)} \varrho_{x|ab,\ldots,(n-1)}^{AB})$, which in turn [by the fact that entanglement cannot increase (on average) under LOCC] is no greater than

$$-\sum_{a,b,\ldots,(n)} p_{a,b,\ldots,(n)} E\Bigg(\sum_x p_{x|ab,\ldots,(n)} \varrho_{x|ab,\ldots,(n)}^{AB}\Bigg), \quad (5)$$

where E denotes any asymptotically consistent measure of bipartite entanglement. The last term of (3) is directly $\leq$ the right-hand side of (5), by the fact that the von Neumann entropy of local density matrix is $\geq$ any asymptotic entanglement measure. The right-hand side of (5) (without the minus sign) is just the average entanglement that we obtain at the output in the n step local measurement protocol between Alice and Bob. We denote it by $\overline{E}_{\mathrm{out}}$. Note that from here on, the results are independent of whether it was Alice or Bob who ended the protocol.

Referring back to the inequalities (3) and (4), we have

$$I^{\mathrm{LOCC}} \leq S(\varrho^A) + S(\varrho^B) - \max_{Z=A,B} \sum_x p_x S(\varrho_x^Z) - \overline{E}_{\mathrm{out}}. \quad (6)$$

It is possible to write Eq. (3) in an even more revealing way. Note that $S(\varrho^A) + S(\varrho^B) \leq N$, where N is the number of qubits (two-dimensional quantum systems) in the Alice-Bob system. That is, $N = \log_2 d_A d_B$, where d_A and d_B are, respectively, the dimensions of the Hilbert spaces of Alice's and Bob's particles. Moreover, we have $S(\varrho_x^B) \geq \mathcal{E}(\varrho_x^{AB})$, where again $\mathcal{E}$ denotes any asymptotically consistent measure of bipartite entanglement [2,12]. The quantity $\sum_x p_x \mathcal{E}(\varrho_x^{AB})$ is the average input (initial) entanglement in the Alice-Bob system. We denote it by $\overline{\mathcal{E}}_{in}$. We use a separate notation for the asymptotic entanglement measure for the input states than that in the output states, to underline the fact that they can be different measures. It is known that there exist several asymptotically consistent measures of bipartite entanglement

(see [13]). We will come back to this point later. So finally we have

$$I^{\mathrm{LOCC}} \leq N - \overline{\mathcal{E}}_{\mathrm{in}} - \overline{E}_{\mathrm{out}}. \tag{7}$$

Equation (7) can also be obtained from Eq. (4), with the additional assumption of monotonicity under LOCC of E. Before connecting above bounds on local mutual information with entanglement distilled in distillation protocols, let us note some interesting features of these inequalities.

Complementarity between extracted and unused information.—One way of interpreting the result in Eq. (7) is to note that the terms I^{LOCC} and $\overline{E}_{\mathrm{out}}$ depend on the measurement protocol followed by Alice and Bob. The other two terms (N and $\overline{\mathcal{E}}_{\mathrm{in}}$) are fixed for a given ensemble. So, writing the inequality as $I^{\mathrm{LOCC}} + \overline{E}_{\mathrm{out}} \leq N - \overline{\mathcal{E}}_{\mathrm{in}}$, we see that the left-hand side can be interpreted as a sum of "extracted information" (I^{LOCC}) and "unused information" ($\overline{E}_{\mathrm{out}}$). Independently (i.e., considered separately), the extracted and unused informations depend on the measurement strategy followed by Alice and Bob. However for all strategies, the sum of the extracted and unused informations is bounded by $N - \overline{\mathcal{E}}_{\mathrm{in}}$.

On bound entanglement with nonpositive partial transpose.—Another interesting feature of the inequality (7) is that the entanglement measures E and $\mathcal{E}$ need not be the same measures. They must only be no greater than the von Neumann entropy of either of the local density matrices. In particular, any asymptotically consistent measure of bipartite entanglement satisfies such conditions (see [13]). This may have nontrivial consequences. For example, we may require that $\mathcal{E}$ must be a convex function, and keep E to be such that it need not necessarily be convex [14]. The only entanglement measure for which there is some evidence for nonconvexity is for distillable entanglement [2], and this is related to the phenomenon of bound entanglement [15]. Precisely, it was shown in Ref. [16] that distillable entanglement can be proven to be nonconvex, if there exists a certain bound entangled state [17], having nonpositive partial transpose (NPT) [18]. Bound entanglement, and more particularly NPT bound entanglement, is not a well understood phenomenon of quantum mechanics. We believe that the inequality (7), may have important consequences for NPT bound entangled states. The point that we make here is also to be seen with respect to the fact that below we actually relate the output entanglement E_{out} to entanglement distilled in different distillation protocols, and bound entanglement is precisely that entanglement which cannot be distilled.

Bound on entanglement distillable via protocols correcting all errors.—We will now consider distillation protocols based on full distinction between the possible pure states in a decomposition of m copies a bipartite state ρ. Suppose therefore that Alice and Bob share m copies of the state ρ given by

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{8}$$

where $|\psi_i\rangle$ are eigenvectors of ρ. Alice and Bob can imagine that they actually share some string of the form $\psi_{i_1} \otimes \ldots \otimes \psi_{i_m}$. Now we propose the following strategy for distillation. Alice and Bob try to fully distinguish between all strings. That is, they apply some LOCC operation that tells them what is the string that they share. Usually during such distinguishing, they destroy the string to some degree. For example, the protocol of distinguishing two pure orthogonal states, given in [5], destroys the states completely. Yet in the hashing protocol for distilling entanglement, Alice and Bob are able to distinguish strings without destroying all entanglement they share [2].

In the case of full distinguishing (in some distillation protocol P), the accessible information is $mS(\rho)$. The initial entanglement per input pair is equal to $\overline{S}_A \equiv \sum_i p_i S(\rho_i^A)$, where ρ_i^A is the local density matrix of $|\psi_i\rangle$. Since we have full distinguishing, the final entanglement is pure entanglement, so that it can be converted reversibly by LOCC, into singlets $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ [19]. Thus the output entanglement is the entanglement D_P that has been distilled in such protocol P. Using the inequality (6) we have then

$$S \leq S_A + S_B - \overline{S}_A - D_P, \tag{9}$$

where for ease of notation, we have used the notations $S \equiv S(\rho)$, $S_A \equiv S(\mathrm{Tr}_B\rho)$, and $S_B \equiv S(\mathrm{Tr}_A\rho)$. This gives

$$D_P \leq S_A + S_B - S - \overline{S}_A. \tag{10}$$

Note that since $|\psi_i\rangle$ are pure, $\overline{S}_A = \sum_i p_i S(\mathrm{Tr}_B|\psi_i\rangle\langle\psi_i|) = \sum_i p_i S(\mathrm{Tr}_A|\psi_i\rangle\langle\psi_i|) = \overline{S}_B$. So the last term in the above inequality (10) can be replaced by $\overline{S}_B$. For the case of Bell-diagonal states (i.e. states that are diagonal in the canonical maximally entangled basis [20]), we have $S_A = S_B = \overline{S}_A = \log_2 d$ so that in that case, inequality (10) gives us

$$D_P(\rho) \leq \log_2 d - S(\rho). \tag{11}$$

This result is compatible with the fact that the quantity $\log_2 d - S(\rho)$ can be attained by hashing methods that reveal all errors [2,4].

It is also instructive to consider a hypothetical protocol, in which Alice and Bob would divide their m systems into two groups G_1 and G_2 of length m_1 and $m - m_1$ respectively. Now by applying some LOCC actions, Alice and Bob would aim to get to know the identities of the states of systems from G_1, while G_2 would serve as a resource to do this and would be destroyed during protocol. The protocol differs from the previous one, as in the present case Alice and Bob do not aim to distinguish between states of systems from this latter group.

Suppose now that such a protocol (P') exists. Then the output entanglement is $m_1\overline{S}_A$, the input one is $m\overline{S}_A$, while

the mutual information is equal to $m_1 S(\rho)$. The entanglement $D_{P'}$ distillable in this protocol is therefore equal to the output entanglement divided by m:

$$D_{P'} = \frac{m_1 \overline{S}_A}{m}.$$

We obtain the following constraint for $r \equiv \frac{m_1}{m}$:

$$r \leq \frac{S_A + S_B - \overline{S}_A}{S + \overline{S}_A} \tag{12}$$

which finally leads to

$$D_{P'} \leq \frac{S_A + S_B - \overline{S}_A}{S + \overline{S}_A} \overline{S}_A. \tag{13}$$

(We remember that $\overline{S}_A = \overline{S}_B$.) For Bell-diagonal states it gives the following bound:

$$D_{P'}(\rho) \leq \frac{(\log_2 d)^2}{\log_2 d + S(\rho)}. \tag{14}$$

(For Bell-diagonal states in $2 \otimes 2$, this reduces to $D_{P'}(\rho) \leq \frac{1}{1+S(\rho)}$.) The bound is always nonzero, even for separable states. This means that the inequality (6) is not the only restriction on local mutual information in this complicated situation. This is however not surprising, as in the considered protocol, we assumed that using a part of the string, we can get the whole information about the rest of the string, but nothing about the used part. What one expects is that at the some point, one perhaps would also gain some information about the used part. Note here that the bound in (14) is for those distillation protocols in which one bases on a distinguishing protocol.

Conclusions.—We have shown that it is possible to obtain bounds on the yield in distillation protocols, basing on distinguishability, of bipartite states, from a complementarity connecting local mutual information with average output entanglement, for the case of bipartite ensembles. For Bell-diagonal states, saturation of this bound is obtained in the hashing protocol for distillation. It is consistent with results of [21], where to beat hashing bound, degenerate codes were applied. Whether any distillation protocol is a distinguishing process remains an open question.

M. H. is supported by the Polish Ministry of Scientific Research and Information Technology under Grant No. PBZ-MIN-008/P03/2003 and by EC grants RESQ and QUPRODIS. J. O. is supported by EC grant PROSECCO. A. S. and U. S. acknowledge support from the Alexander von Humboldt Foundation.

Note added.—After completion of our work, we came to know of the recent related work in Ref. [22].

[1] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).

[2] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[3] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[4] K. G. H. Vollbrecht and M. M. Wolf, Phys. Rev. A **67**, 012303 (2003).

[5] J. Walgate, A. J. Short, L. Hardy, and V. Vedral, Phys. Rev. Lett. **85**, 4972 (2000); S. Virmani, M. F. Sacchi, M. B. Plenio, and D. Markham, Phys. Lett. A **288**, 62 (2001); Y.-X. Chen and D. Yang, Phys. Rev. A **64**, 064303 (2001); **65**, 022320 (2002).

[6] Henceforth, by "distinguishing", we will mean "local distinguishing".

[7] J. P. Gordon, in *Proceedings of the International School of Physics "Enrico Fermi, Course XXXI"*, edited by P. A. Miles (Academic Press, NY, 1964), p. 156; L. B. Levitin, in *Proceedings VI National Conference Information Theory, Tashkent* (1969), p. 111; A. S. Holevo, Prob. Peredachi Inf. **9**, 3 1973 [Problems of Information Transmission (Engl Trans) **9**, 110 (1973)].

[8] B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. Lett. **76**, 3452 (1996).

[9] P. Badziąg, M. Horodecki, A. Sen(De), and U. Sen, Phys. Rev. Lett. **91**, 117901 (2003).

[10] The results obtained are independent of whether it was Alice or Bob who started the protocol.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, (Wiley, New York, 1991).

[12] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **84**, 2014 (2000); See also M. Donald, M. Horodecki, and O. Rudolph, J. Math. Phys. (N.Y.) **43**, 4252 (2002).

[13] M. Horodecki, Quantum Inf. Comput. **1**, 3 (2001).

[14] An entanglement measure E is said to be convex, if for all bipartite density matrices ϱ_1 and ϱ_2, and probabilities p_1 and p_2, we have $E(p_1\varrho_1 + p_2\varrho_2) \leq p_1 E(\varrho_1) + p_2 E(\varrho_2)$. This inequality essentially says that forgetting of information is not useful to increase entanglement. If this inequality does not hold for some choice of the density matrices (and probabilities), then the corresponding entanglement measure is said to be nonconvex.

[15] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **80**, 5239 (1998); P. Horodecki, Phys. Lett. A **232**, 333 (1997).

[16] P. W. Shor, J. A. Smolin, and B. M. Terhal, Phys. Rev. Lett. **86**, 2681 (2001).

[17] D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, and A. V. Thapliyal, Phys. Rev. A **61**, 062312 (2000); W. Dür, J. I. Cirac, M. Lewenstein, and D. Bruss, Phys. Rev. A **61**, 062313 (2000).

[18] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996); M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).

[19] C. H. Bennett, H. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).

[20] The canonical maximally entangled states in $d \otimes d$ are $\frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} e^{2\pi i j n/d} |j\rangle |(j+m) \mathrm{mod}\ d\rangle$, $(n, m = 0, \ldots, d-1)$. For $d = 2$, it is the familiar Bell basis.

[21] P. W. Shor and J. A. Smolin, quant-ph/9604006.

[22] S. Ghosh, P. Joag, G. Kar, S. Kunkri, and A. Roy, quant-ph/0403134.