Distillation of local purity from quantum states

I. Devetak*
Electrical Engineering Department, University of Southern California, Los Angeles, California 90089, USA
(Received 22 February 2005; published 7 June 2005)

Recently Horodecki *et al.* [Phys. Rev. Lett. **90**, 100402 (2003)] introduced an important quantum information processing paradigm, in which two parties sharing many copies of the same bipartite quantum state distill local pure states by means of local unitary operations assisted by a one-way (two-way) completely dephasing channel. Local pure states are a valuable resource from a thermodynamical point of view, since they allow thermal energy to be converted into work by local quantum heat engines. We give a simple information-theoretical characterization of the one-way distillable local purity, which turns out to be closely related to a previously known operational measure of classical correlations, the one-way distillable common randomness.

DOI: 10.1103/PhysRevA.71.062303 PACS number(s): 03.67.Hk, 05.70.−a

I. INTRODUCTION

One of the primary tasks of quantum information theory is to explore the operational reductions between information processing resources such as shared entanglement or quantum channels, including both noisy and noiseless varieties. For instance, entanglement distillation [1] involves transforming a large number of noisy bipartite quantum states ρ^{AB}, shared between two distant parties Alice and Bob, into pure ebits $|\Phi^+\rangle = 1/\sqrt{2}(|0\rangle|0\rangle + |1\rangle|1\rangle)$ at the best possible conversion rate. This conversion task is naturally defined within the LOCC (local operations and classical communication) paradigm: Alice and Bob are allowed *at no cost* to (i) locally add pure state ancillas to their quantum systems, (ii) perform local unitary operations, and (iii) communicate classically. In a slight refinement of this paradigm, one could assign a cost for one-way classical communication, leading to trade-offs between the amount of entanglement distilled and the classical communication invested [2]. The communication theorist still feels at home with this modification: after all, classical communication is a valuable bipartite resource and should not be taken for granted. It is only recently that attention has been given to *local* resources, in particular local pure states [3].

Local pure states can be seen as valuable from a *thermodynamical* perspective. Although we use the language of quantum states, the phenomenon is essentially classical. Landauer [4] was the first to observe that work was required to erase a bit of information—i.e., to reset a system from an unknown state to a known (pure) state. Conversely, a supply of pure states can be used as "fuel" to increase the amount of useful work extractable from a system at non-zero temperature [5,6]. This is achieved by reversibly transferring entropy from the system to the pure states, thereby "cooling" the system [7].

Having an appreciation for the value of pure states, it is natural to ask about the different ways in which they can be produced. In [3,8] the idea of manipulating and concentrating "purity" already existing in a diluted form, rather than performing work to create it, was introduced. This is very much analogous to entanglement distillation: given a noisy resource one wishes to remove impurities from it. There is a local and distributed version of this problem. In the local scenario, which we call *purity concentration,* Alice is given a large supply of states ρ^A and her task is to extract pure qubit states using only unitary operations. The maximal asymptotic rate at which this can be done is given by the difference between the size of the system A (in qubits) and its von Neumann entropy [9]. In the distributed scenario—*local purity distillation*—Alice and Bob share a supply of bipartite states ρ^{AB} and they wish to distill local pure states using CLOCC (*closed* local operations and classical communication) [8], a modification of the LOCC paradigm that disallows unrestricted consumption of local pure states. Horodecki *et al.* [3] had previously obtained some bounds on this problem, both for the one-way and two-way CLOCC cases.

In this paper we investigate the two scenarios in detail. Our main result pertains to the distributed case; we give an information-theoretical expression for the optimal one-way distillable local purity. This quantity turns out to be related to a previously known operational measure of classical correlations, the one-way distillable common randomness [10]. Section II is devoted to establishing the notation. Section III treats the local scenario, reproducing the results of [9] in a somewhat more rigorous coding-theoretical language. The two-party distributed scenario is considered in Sec. IV and our main result is proved. Section V discusses how to embed purity distillation and the CLOCC paradigm in the existing formalism for quantum Shannon theory, and concludes with open questions. Appendix A collects a number of auxiliary inequalities used throughout the paper.

II. NOTATION AND DEFINITIONS

Recall the notion of an *ensemble* of quantum states $\mathcal{E} = (p(x), \rho_x^B)_{x\in\mathcal{X}}$: the quantum system B is in the state ρ_x^B with probability $p(x)$. The ensemble $\mathcal{E}$ is equivalently represented by a *classical-quantum* system [10] XB in the state

$$\rho^{XB} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|^X \otimes \rho_x^B, \tag{1}$$

where $\mathcal{H}_X$ has a preferred orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$. X plays the dual role of an auxiliary quantum system in the state

*Electronic address: devetak@usc.edu

$\Sigma_x p(x)|x\rangle\langle x|$ and of a random variable with distribution p and cardinality $|X| := |\mathcal{X}|$. For a multiparty state such as ρ^{XB}, the reduced density operator ρ^B is defined by $\mathrm{Tr}_X \rho^{XB}$. Conversely, we call ρ^{XB} an *extension* of ρ^B. A pure extension is conventionally called a *purification*.

The ensemble $\mathcal{E}$ may come about by performing a POVM $\Lambda=(\Lambda_x)_x$, $\Sigma\Lambda_x=I$, on the A part of a bipartite state ρ^{AB}, in which case $p(x)=\mathrm{Tr}(\Lambda_x\rho^A)$ and $\rho_x^B=p(x)^{-1}\mathrm{Tr}_A[(\Lambda_x^A\otimes I^B)\rho^{AB}]$. Equivalently, Λ may be thought of as a quantum map $\Lambda:\mathcal{H}_A\to\mathcal{H}_X$, sending ρ^{AB} to ρ^{XB}. A classical map $f:\mathcal{X}\to\mathcal{Y}$ may similarly be viewed as a quantum one $f:\mathcal{H}_X\to\mathcal{H}_Y$,

$$f(\rho)=\sum_{x\in\mathcal{X}}\langle x|\rho|x\rangle|f(x)\rangle\langle f(x)|^Y,$$

where $\mathcal{H}_Y$ has a preferred orthonormal basis $\{|y\rangle\}_{y\in\mathcal{Y}}$.

Define the von Neumann entropy of a quantum state ρ by $H(\rho)=-\mathrm{Tr}(\rho\log\rho)$, where log is the base 2 logarithm. We write $H(A)_\sigma=H(\sigma^A)$, omitting the subscript when the reference state is clear from the context. The Shannon entropy $-\Sigma_x p(x)\log p(x)$ of the random variable X is thus equal to the von Neumann entropy $H(X)$ of the system X. Define the conditional entropy

$$H(A|B)=H(B)-H(AB),$$

(quantum) mutual information

$$I(A;B)=H(A)+H(B)-H(AB),$$

and conditional mutual information

$$I(A;B|X)=I(A;BX)-I(A;X).$$

For a sequence $x_1,\ldots,x_n$ of classical indices x_i we use the shorthand notation x^n, and $\rho_{x^n}:=\otimes_i\rho_{x_i}$. For an integer μ define $[\mu]=\{1,\ldots,\mu\}$.

The trace norm of an operator is defined as

$$\|\omega\|_1=\mathrm{Tr}\sqrt{\omega^\dagger\omega},$$

which for ω Hermitian amounts to the sum of the absolute values of the eigenvalues of ω. We say that two states ρ and ω are ϵ close if

$$\|\rho-\omega\|_1\leq\epsilon.$$

We loosely refer to an isometry $U:\mathcal{H}_A\to\mathcal{H}_B\otimes\mathcal{H}_C$ as a unitary operation under the assumption that A may be written as a composite system BC. For a POVM $\Lambda=(\Lambda_x)_x$ acting on a composite system AB we say that it is *rank 1 on A* if, for all x, Λ_x is of the form

$$\Lambda_x^{AB}=|\phi_x\rangle\langle\phi_x|^A\otimes I^B.$$

Throughout the paper, $|0\rangle^A$ will denote a standard pure state on the system A.

III. LOCAL SCENARIO: PURITY CONCENTRATION

We begin by formally defining a purity concentration code. Alice has n copies of a state ρ^A defined on a system A of dimension d_A. In other words, Alice has a n-partite quantum system $A^n=A_1,\ldots,A_n$ with Hilbert space $\mathcal{H}_{A^n}=\mathcal{H}_{A_1}\otimes\cdots\otimes\mathcal{H}_{A_n}$ in a tensor power state $\rho^{\otimes n}$. An (n,ϵ) *purity concentration code* consists of a unitary operation $U:\mathcal{H}_{A^n}\to\mathcal{H}_{A_p}\otimes\mathcal{H}_{A_g}$ such that, for $\sigma^{A_pA_g}=U(\rho^{\otimes n})$,

$$\|\sigma^{A_p}-|0\rangle\langle 0|^{A_p}\|_1\leq\epsilon. \tag{2}$$

The *rate* of the code is defined by $R=(1/n)\log d_{A_p}$, where d_{A_p} is shorthand for $\dim\mathcal{H}_{A_p}$. A rate R is said to be *achievable* if for all ϵ, $\delta>0$ and sufficiently large n there exists an (n,ϵ) code with rate $R-\delta$. The *purity* $\kappa(\rho)$ (also referred to as "information" in [3]) is defined as the supremum over all achievable rates R.

The following theorem, previously proven in [9], gives an information-theoretical expression for κ.

Theorem 1. The purity of the state ρ^A of the d_A dimensional quantum system A is

$$\kappa(\rho^A)=\log d_A-H(A)_\rho.$$

Proof. We start by proving the "converse"—i.e. the $\leqslant$ direction of the theorem. Consider a general (n,ϵ) purity concentration protocol. Obviously,

$$\log d_{A_p}=n\log d_A-\log d_{A_g}.$$

The second term is bounded as

$$\log d_{A_g}\geqslant H(A_g)\geqslant H(A_pA_g)-H(A_p)=nH(A)-H(A_p)$$
$$\geqslant nH(A)-\frac{1}{e}-n\epsilon\log d_A. \tag{3}$$

The second inequality follows from the subadditivity of von Neumann entropy (A4), and the third inequality is Fannes' inequality (A3) applied to (2). Hence,

$$R=\frac{1}{n}\log d_{A_p}\leqslant\log d_A-H(A)+\delta,$$

where without loss of generality $\delta\geqslant 1/en+\epsilon\log d_A$.

To prove the "direct coding theorem" (the $\geqslant$ direction), consider the typical projector [11] $\Pi_{\rho,\delta}^n$ commuting with $\rho^{\otimes n}$ with the property that, for all ϵ, $\delta>0$ and sufficiently large n,

$$\mathrm{Tr}\rho^{\otimes n}\Pi_{\rho,\delta}^n\geqslant 1-\epsilon,$$

while $\mathrm{Tr}\,\Pi_{\rho,\delta}^n\leqslant n[H(\rho)+\delta]$. The coding theorem now follows from lemma 1 below. □

Lemma 1. Let Π be a projector with $\mathrm{Tr}\,\Pi=d_1$ and ρ a state that commutes with Π, both defined on a d_1d_2-dimensional Hilbert space $\mathcal{H}_A$. If $\mathrm{Tr}\,\rho\Pi\geqslant 1-\epsilon$, then there exists a unitary $U:\mathcal{H}_A\to\mathcal{H}_B\otimes\mathcal{H}_C$, with $\dim\mathcal{H}_B=d_1$ and $\dim\mathcal{H}_C=d_2$, such that

$$\|U\rho U^\dagger-(\Pi\rho\Pi)^B\otimes|0\rangle\langle 0|^C\|_1\leq\epsilon.$$

Proof. Let $\{|i\rangle\}_{i\in[d_1d_2]}$ be a basis for A such that

$$\Pi^A=\sum_{i=1}^{d_1}|i\rangle\langle i|^A.$$

Viewing A as a composite system BC, with basis $\{|i\rangle\otimes|j\rangle\}_{i\in[d_1],j+1\in[d_2]}$, define U to satisfy $U|i\rangle^A=|i\rangle^B|0\rangle^C$ for all $i\in[d_1]$. The lemma follows from

$$\|\Pi\rho\Pi - \rho\| \leqslant \epsilon.$$

□

IV. BIPARTITE SCENARIO: LOCAL PURITY DISTILLATION

We now consider the bipartite scenario where Alice and Bob share many copies of a some state ρ^{AB}. Their task is to distill local pure qubit states by means of protocols involving only closed local operations and classical communication. More precisely, Alice and Bob may perform local unitary operations and are allowed unlimited use of a completely dephasing channel in both directions. A dephasing channel is given by the map $\mathcal{P}:\mathcal{H}_X \to \mathcal{H}_X$,

$$\mathcal{P}(\rho) = \sum_x |x\rangle\langle x|\rho|x\rangle\langle x|,$$

where $\{|x\rangle\}$ is an orthonormal basis for $\mathcal{H}_X$. The term "closed" refers to Alice and Bob not being given free access to local pure state ancillas; this is the main difference between CLOCC and the more familiar LOCC relevant for entanglement distillation [1]. A catalytic variation of CLOCC, which we denote by CLOCC′, allows Alice and Bob to borrow local pure state ancillas, but they have to return them at the end of the protocol. Similarly define the 1-CLOCC and 1-CLOCC′ paradigms with the bidirectional communication replaced by a one-way dephasing channel from Alice to Bob. In [3] yet another paradigm, NLOCC (*noisy* local operations and classical communication) was used, which allows both parties unlimited access to maximally mixed local states. This additional resource will prove to be useless for our purposes.

Our main focus will be on the 1-CLOCC′ paradigm as it turns out to be amenable to information theoretical characterization. We proceed to formally define a local purity distillation code. Alice and Bob share n copies of the state ρ^{AB}, embodied in the shared quantum system A^nB^n, and Alice also has access to some quantum system C of dimension d_C, initially in a pure state $|0\rangle^C$. An (n,ϵ) *(catalytic) one-way local purity distillation code* consists of

(i) a unitary operation $U_A:\mathcal{H}_{A^n}\otimes\mathcal{H}_C \to \mathcal{H}_{A_p}\otimes\mathcal{H}_X$ on Alice's side,

(ii) a dephasing channel $\mathcal{P}:\mathcal{H}_X \to \mathcal{H}_X$ from Alice to Bob,

(iii) a unitary operation $U_B:\mathcal{H}_{B^n}\otimes\mathcal{H}_X \to \mathcal{H}_{B_p}\otimes\mathcal{H}_{B_g}$ on Bob's side,

such that, for

$$\sigma^{A_pB_pB_g} = (U_B \circ \mathcal{P} \circ U_A)((\rho^{AB})^{\otimes n} \otimes |0\rangle\langle 0|^C),$$

$$\|\sigma^{A_pB_p} - |0\rangle\langle 0|^{A_p} \otimes |0\rangle\langle 0|^{B_p}\|_1 \leqslant \epsilon. \tag{4}$$

The *rate* of the code is defined by $R=(1/n)(\log d_{A_pB_p} - \log d_C)$. The *catalyst rate* is $(1/n)\log d_C$. A rate R is said to be *achievable* if for all ϵ, $\delta>0$ and sufficiently large n there exists an (n,ϵ) code with rate $R-\delta$. The *one-way local purity* $\kappa_{\to}(\rho^{AB})$ is defined as the supremum over all achievable rates R.

A quantity of particular interest is the *classical deficit*

$$\Delta^c_{\to}(\rho^{AB}) = \kappa_{\to}(\rho^{AB}) - \kappa(\rho^A) - \kappa(\rho^B).$$

This quantity (or, rather, its bidirectional version) was introduced in [12] and advertised as a measure of classical correlations in the state ρ^{AB}.

Example 1. Assume that Alice and Bob are given a bit of common randomness, which is represented by the state

$$\overline{\Phi}^{AB} = \frac{1}{2}(|0\rangle\langle 0|^A \otimes |0\rangle\langle 0|^B + |1\rangle\langle 1|^A \otimes |1\rangle\langle 1|^B).$$

Alice sends her system to Bob through the dephasing channel, which leaves it intact. Bob performs the controlled unitary

$$U^{AB} = |0\rangle\langle 0|^A \otimes I + |1\rangle\langle 1|^A \otimes V^B,$$

where $V|1\rangle = |0\rangle$, leaving the B system in the state $|0\rangle^B$. This gives $\kappa_{\to} = \Delta^c_{\to} = 1$.

Our main result is contained in the following theorem.

Theorem 2. The local one-way purity of a state ρ^{AB} defined on a system of dimension $d_A \times d_B$ is given by

$$\kappa_{\to}(\rho^{AB}) = \log d_A + \log d_B - H(A)_\rho - H(B)_\rho + D_{\to}(\rho^{AB}),$$

with

$$D_{\to}(\rho^{AB}) = \lim_{n\to\infty} \frac{1}{n} D^{(1)}_{\to}((\rho^{AB})^{\otimes n})$$

and

$$D^{(1)}_{\to}(\rho^{AB}) = \max_{\Lambda} I(X;B)_{(\Lambda\otimes I)(\rho)}. \tag{5}$$

The maximization is over all rank-1 POVM's $\Lambda:\mathcal{H}_A \to \mathcal{H}_X$.

Corollary 1:

$$\Delta^c_{\to}(\rho^{AB}) = D_{\to}(\rho^{AB}).$$

The quantity $D^{(1)}_{\to}(\rho^{AB})$ first appeared in [13], where it was proposed, on heuristic grounds, as a measure of classical correlations in the state ρ^{AB}. Its "regularized" version $D_{\to}(\rho^{AB})$ [12] was given operational meaning in [10] where it was shown to be equal to the *one-way distillable common randomness* (1-DCR) of ρ^{AB}. The 1-DCR is the maximum conversion rate from ρ^{AB} into bits of common randomness, achievable with 1-LOCC, *in excess* of the classical communication invested.

In [10], the additivity of $D^{(1)}_{\to}$ was shown for a separable state σ^{AB} and arbitrary ρ^{AB},

$$D^{(1)}_{\to}(\rho^{AB}\otimes\sigma^{AB}) = D^{(1)}_{\to}(\rho^{AB}) + D^{(1)}_{\to}(\sigma^{AB}). \tag{6}$$

Therefore, adding local maximally mixed states $\sigma^{AB} = (d_Ad_B)^{-1}I^A\otimes I^B$, for which $D^{(1)}_{\to}(\sigma^{AB})=0$ does not affect the 1-DCR or the classical deficit. Moreover, for separable states ρ^{AB} the classical deficit is efficiently computable, as

$$D_{\to}(\rho^{AB}) = D^{(1)}_{\to}(\rho^{AB}).$$

From [10] we know additivity to hold for the case of pure states $|\phi\rangle^{AB}$, and it is easily seen that [3]

$$\Delta^{c}_{\rightarrow}(\phi^{AB}) = E(\phi^{AB}) \coloneqq H(A)_{\phi},$$

where E is the unique measure of entanglement for pure states. Additivity also holds for Bell-diagonal states [14,15]. The general question of the additivity of $D^{(1)}_{\rightarrow}$ is known to be equivalent to several other open additivity problems in quantum information theory [16–18], including that of the Holevo capacity of quantum channels.

In proving theorem 2, we shall need two lemmas. The first is from [19].

Lemma 2. Consider a classical-quantum system X^nB^n in the state $(\rho^{XB})^{\otimes n}$, where ρ^{XB} is given by Eq. (1). For any ϵ, $\delta>0$ and sufficiently large n, there exist

(i) a set $\mathcal{S}$ in $\mathcal{X}^n$ with

$$\Pr\{X^n \notin \mathcal{S}\} \leqslant \epsilon, \tag{7}$$

(ii) a bijection $f:[\mu]\times[\lambda]\rightarrow\mathcal{S}$, where $\lambda \leqslant 2^{n[H(X)-I(X;B)+\delta]}$ and $\mu\lambda \leqslant 2^{n[H(X)+\delta]}$,

(iii) a collection of POVM's $(\Upsilon^{(l)})_{l\in[\lambda]}$ [each $\Upsilon^{(l)} = (\Upsilon^{(l)}_m)_m$ is a POVM], such that

$$\mathrm{Tr}\rho^{B}_{f(m,l)}\Upsilon^{(l)}_m \geqslant 1-\epsilon, \quad \forall\, m,l. \tag{8}$$

The above lemma says that a highly probable set $\mathcal{S}$ of sequences x^n can be covered by λ disjoint sets $\mathcal{S}_l$, $l\in[\lambda]$, of size μ in such a way that, given the index l, the identity of a particular sequence in $\mathcal{S}_l$ may be reliably inferred from a measurement on B^n.

The following technical lemma is a corollary of the measurement compression theorem [20], and is proved in Appendix B.

Lemma 3. Given the system A^nB^n in the state $(\rho^{AB})^{\otimes n}$ and a rank-1 POVM $\Lambda:\mathcal{H}_A\rightarrow\mathcal{H}_X$, for any ϵ, $\delta>0$ and sufficiently large n, there exists

(i) a decomposition $A^n=A_1A_2$ such that

$$H(A_1) \leqslant n\epsilon,$$

(ii) a POVM $\tilde{\Lambda}:\mathcal{H}_{A^n}\rightarrow\mathcal{H}_K$ which is rank 1 on A_2 and

$$\log|K| \leqslant n[H(A)+\delta], \tag{9}$$

$$I(K;B^n)_{\omega} \geqslant n[I(X;B)_{\rho}-\epsilon], \tag{10}$$

where

$$\rho^{XB} = (\Lambda\otimes I)(\rho^{AB}), \tag{11}$$

$$\omega^{KB^n} = (\tilde{\Lambda}\otimes I)(\rho^{AB})^{\otimes n}. \tag{12}$$

Proof of theorem 2. First, let us prove the converse. Consider a general (n,ϵ) purity distillation protocol. We know that

$$\log d_{A_pB_p} - \log d_C = n(\log d_A + \log d_B) - \log d_{B_g}.$$

Assume, w.l.o.g., $\delta \geqslant 1/en + \epsilon\log(d_Ad_B)$. The entropic quantities below refer to the overall quantum state at a stage of the protocol which is implicit from the subsystems involved. For instance, the system B^n exists only before U_B is applied.

$$\begin{aligned}\log d_{B_g} &\geqslant H(B_g) \geqslant H(B_pB_g) - H(B_p) = H(XB^n) - H(B_p)\\ &\geqslant H(X) + H(B^n|X) - \frac{1}{e} - n\epsilon\log d_A \geqslant H(A_g)\\ &\quad + H(B^n|X) - \frac{1}{e} - n\epsilon\log d_A \geqslant nH(A) + H(B^n|X)\\ &\quad - n\delta.\end{aligned}$$

The second inequality is subadditivity (A4), the third is Fannes' inequality (A3) and (4), the fourth follows from the fact that dephasing cannot decrease entropy [21], and the fifth follows along the lines of Eq. (3). Hence,

$$\begin{aligned}R &= \frac{1}{n}(\log d_{A_pB_p} - \log d_C) \leqslant \log d_A + \log d_B - H(A) - H(B)\\ &\quad + \frac{1}{n}I(X;B^n) + \delta.\end{aligned}$$

The idea behind the direct coding theorem is that there are two potential sources of purity. The first comprises the locally concentrable purity for the two parties, from Sec. III and is responsible for the $\kappa(\rho^A)+\kappa(\rho^B)$ term. The second comes from the classical correlations present in the system and gives rise to the $D_{\rightarrow}(\rho^{AB})$ term. Roughly speaking, Alice sends her part of the classical correlations through the dephasing channel; Bob then takes advantage of the redundancy, as in example 1, to distill purity.

We start by considering a special case. Assume that the system A can be divided into subsystems $A=A_1A_2$ such that $H(A_1)\leqslant\tau$ and that Λ is rank 1 on A_2. We show that we can achieve a rate arbitrarily close to

$$\log d_A + \log d_B - \tau - H(X)_{\rho} - H(B)_{\rho} + I(X;B)_{\rho},$$

with ρ given by Eq. (11). Consider a sufficiently large n and the induced decomposition $A^n=A_1^nA_2^n$. The purity distillation protocol comprises of the following steps.

(i) First, Alice applies the protocol from theorem 1 to A_1^n, yielding a subsystem A_{1p} of size $n[\log d_{A_1}-\tau-\delta]$ qubits, in a state ϵ close to $|0\rangle^{A_{1p}}$.

(ii) The measurement $\Lambda^{\otimes n}$ may be implemented by borrowing $n\log d_X$ qubit ancillas (in some fixed state $|0\rangle^{X^n}$), performing some unitary operation U on the system $A_2^nX^n$, and completely dephasing the system X^n in a fixed basis $\{|x^n\rangle\}$. Here we let Alice perform this measurement *coherently*—i.e., by omitting the dephasing step (the channel $\mathcal{P}$ will later do this for us). Since $\Lambda^{\otimes n}$ is rank 1 on A_2^n, this results in a state of the form

$$\sum_{x^n}\sqrt{p(x^n)}|x^n\rangle^{X^n}|\psi_{x^n}\rangle^{A_2^n}|\phi_{x^n}\rangle^{R^n},$$

where R^n is the "reference system" that purifies the initial state of A_2^n. She then performs the controlled unitary

$$\sum_{x^n}|x^n\rangle\langle x^n|^{X^n}\otimes V^{A_2^n}_{x^n},$$

where $V_{x^n}|\psi_{x^n}\rangle=|0\rangle$, leaving A_2^n in the state $|0\rangle^{A_2^n}$.

(iii) Were Alice to perform the von Neumann measurement on X^n, the resulting state of the system X^nB^n would be

$$(\rho^{XB})^{\otimes n} = \sum_{x^n} p(x^n)|x^n\rangle\langle x^n|^{X^n} \otimes \rho_{x^n}^{B^n}.$$

Choose the set $\mathcal{S}$, bijection f and collection of POVM's $(\Upsilon^{(l)})_l$ as in lemma 2. Define $\Pi' = \Sigma_{x^n \in \mathcal{S}}|x^n\rangle\langle x^n|^{X^n} \otimes I^{B^n}$. By (7) and the proof of lemma 1, there is a unitary operation (acting on Alice's system only) that takes $(\rho^{XB})^{\otimes n}$ to a state 2ϵ close to $|0\rangle\langle 0|^{X'} \otimes \theta'^{MLB^n}$ with $d_{X'} = (\mu\lambda)^{-1} d_{X^n}$ and

$$\theta'^{MLB^n} = \sum_{m,l} p(m,l)|m\rangle\langle m|^M \otimes |l\rangle\langle l|^L \otimes \rho_{f(m,l)}^{B^n}.$$

The $p(m,l)$ is some probability distribution associated with a composite random variable ML. Alice performs said unitary.

(iv) Alice sends the ML system through the dephasing channel, leaving MLB^n in a state θ^{MLB^n} which is 2ϵ close to θ'^{MLB^n}.

(v) For each l one can define a unitary $W_l^{B^nM}$, a coherent version of the measurement $\Upsilon^{(l)}$, which upon measurement "outcome" m performs the transformation $|m\rangle^M \mapsto |0\rangle^M$. Explicitly, $W_l^{B^nM}$ is w.l.o.g. of the form $\Sigma_{m,m'}|m'\rangle\langle m|^M \otimes Y_{m'm}^{B^n}$. Choosing $Y_{0m} = (\Upsilon_m^{(l)})^{1/2}$ and the remaining $Y_{m'm}$ to satisfy unitarity leaves W_l with the desired property. Defining

$$\sigma_{ml}^{B^nM} = W_l^{B^nM}(\rho_{f(m,l)}^{B^n} \otimes |m\rangle\langle m|^M),$$

the measurement success criterion (8) of lemma 2 becomes

$$\langle 0|\sigma_{ml}^M|0\rangle \geqslant 1 - \epsilon.$$

By (A2),

$$\left\| \sum_{m,l} p(m,l)\sigma_{ml}^M - |0\rangle\langle 0|^M \right\|_1 \leqslant 2\sqrt{\epsilon}. \tag{13}$$

Bob applies the controlled unitary

$$W^{LB^nM} = \sum_l |l\rangle\langle l|^L \otimes W_l^{B^nM},$$

which, by (13), maps θ'^{MLB^n} to a state whose M part is $2\sqrt{\epsilon}$ close to $|0\rangle^M$. Since θ^{MLB^n} is 2ϵ close to θ'^{MLB^n}, upon application of W its M part becomes $(2\epsilon + 2\sqrt{\epsilon})$ close to $|0\rangle^M$, by the triangle inequality (A1).

(vi) By the gentle operator lemma (see Appendix A), performing W perturbs the B system very little, leaving it in a state $(\epsilon + \sqrt{8\epsilon})$ close to $(\rho^B)^{\otimes n}$. Bob applies the protocol from theorem 1 to B^n, yielding a subsystem B_p of size $n(d_B - H(B) - \delta)$ qubits, in a state $(2\epsilon + \sqrt{8\epsilon})$-close to $|0\rangle^{B_P}$.

In summary, the protocol consumes a catalyst of $n \log d_X$ qubits, while returning a system of size

$$n[\log d_{A_1} - \tau - \delta] + n \log d_{A_2} + n \log d_X - \log(\mu\lambda) + \log \mu$$
$$+ n[d_B - H(B) - \delta]$$

qubits, in a state which is $[7\epsilon + (2 + \sqrt{8})\sqrt{\epsilon}]$ close to pure. This corresponds to a purity distillation rate of at least

$$\log d_A + \log d_B - \tau - H(X) - H(B) + I(X;B) - 3\delta,$$

while the classical communication rate required was $n^{-1}\log(\mu\lambda) \leqslant H(X) + \delta$ bits per copy.

To prove the general statement of the theorem we shall rely on lemma 3 and "double blocking." Let n' be sufficiently large for lemma 3 to apply with respect to the *optimal* Λ achieving $D_{\rightarrow}^{(1)}(\rho^{AB})$ in (5). We shall apply the special-case protocol described above to the block system $A^{n'} = A_1A_2$ and block measurement $\tilde{\Lambda}$, obtaining a rate of

$$\log d_A + \log d_B - \frac{1}{n'}H(K) - H(B) + \frac{1}{n'}I(K;B^n) - 3\delta - \epsilon.$$

By lemma 3 and (5), this is bounded from below by

$$\log d_A + \log d_B - H(A) - H(B) + D_{\rightarrow}^{(1)}(\rho^{AB}) - 4\delta - 2\epsilon.$$

The classical communication rate required for this protocol is $H(A) + 2\delta$.

Finally, a third layer of blocking allows us to replace $D_{\rightarrow}^{(1)}$ by $D_{\rightarrow}$, and we are done. □

It is not hard to see that the above protocol may be bootstrapped to make the catalyst rate arbitrarily small. Moreover, if $\kappa(\rho^A) > 0$, a catalyst is not needed at all (see also [15]).

V. DISCUSSION

The question of counting local resources in standard quantum-information-theoretical tasks, such as entanglement distillation, was recently raised by Bennett [22]. In particular, it is desirable to extend the theory of *resource inequalities* [2] to include the manipulation of local resources. Recall the notation from [10] in which $[c \rightarrow c]$, $[q \rightarrow q]$ and $[qq]$ stand for a bit of classical communication, a qubit of quantum communication, and ebit of entanglement, respectively. There it was implicit that local pure ancillas could be added for free, which makes a classical channel and a dephasing quantum channel operationally equivalent. To define a "closed" version of this formalism, one must identify $[c \rightarrow c]$ with a dephasing qubit channel and introduce a new resource, a local pure qubit state $|0\rangle$ w.l.o.g. in Bob's possession. This resource may be written as either $[q]$ or $[c]$, as there is little distinction between classical and quantum for strictly local resources. The main result of our paper may be written succinctly as

$$\{qq\} + H(A)_\rho[c \rightarrow c] \geqslant \kappa_{\rightarrow}(\rho^{AB})[q],$$

where $\{qq\}$ represents the noisy static resource ρ^{AB}, and $\kappa_{\rightarrow}(\rho^{AB})$ is given by theorem 2. Regarding entanglement distillation, closer inspection of the optimal one-way protocol from [23] reveals that (i) only a negligible rate of pure state ancillas need be consumed and, (ii) moreover, the locally concentratable purity $\kappa(\rho^A) + \kappa(\rho^B)$ is available without affecting the entanglement distillation rate.

Whether the above holds for general quantum-Shannon-theoretic problems remains to be investigated.

We conclude with a list of open problems.

(i) It would be interesting to find the optimal trade-off between the local purity distilled and the one-way classical communication (dephasing) invested. In particular, does the problem reduce to the 1-DCR trade-off curve from [10]? Also, one could consider purity distillation assited by quantum communication [24].

(ii) We have seen that purity distillation and common randomness distillation are intimately related. Is there a nontrivial trade-off between the two, or is it always optimal to (linearly) interpolate between the known purity distillation and common randomness distillation protocols? One could also consider the simultaneous distillation of purity and other resources, such as entanglement (see [12]).

(iii) Clearly, one would like a formula for the two-way distillable local purity. Solving this problem in the sense of the present paper appears to be difficult; Ref. [15] gives a formula involving maximizations over a class of states which is, alas, rather hard to characterize. A more tractable question is whether the relationship established between distillable purity and distillable common randomness carries over to the two-way scenario.

ACKNOWLEDGMENTS

We are grateful to Charles Bennett, Guido Burkard, David DiVincenzo, Aram Harrow, Barbara Terhal, and John Smolin for useful discussions. We also thank Michał and Paweł Horodecki, Jonathan Oppenheim, and Barbara Synak for comments on the manuscript and sharing their unpublished results on purity distillation [15,25]. This work was supported in part by the NSA under the U.S. Army Research Office (ARO), Grant Nos. DAAG55-98-C-0041 and DAAD19-01-1-06.

APPENDIX A: MISCELLANEOUS INEQUALITIES

For states ρ, ω, and σ, the triangle inequality holds:

$$\|\rho-\omega\|_1+\|\omega-\sigma\|_1\geqslant\|\rho-\sigma\|_1. \tag{A1}$$

The following bound [26] relates trace distance and fidelity:

$$\|\rho-|\phi\rangle\langle\phi|\|_1\leqslant 2\sqrt{1-\langle\phi|\rho|\phi\rangle}. \tag{A2}$$

The gentle operator lemma [27] says that a POVM element that succeeds on a state with high probability does not disturb it much.

Lemma 4. For a state ρ and operator $0\leqslant\Lambda\leqslant I$, if $\mathrm{Tr}(\rho\Lambda)\geqslant 1-\lambda$, then

$$\|\rho-\sqrt{\Lambda}\rho\sqrt{\Lambda}\|_1\leqslant\sqrt{8\lambda}.$$

The same holds if ρ is only a subnormalized density operator. □

For two states ρ and ω defined on a d-dimensional Hilbert space, Fannes' inequality [28] reads

$$|H(\rho)-H(\sigma)|\leqslant\frac{1}{e}+\log d\|\rho-\omega\|_1. \tag{A3}$$

An important property of von Neumann entropy is *subadditivity*

$$H(B)\geqslant H(AB)-H(A). \tag{A4}$$

APPENDIX B: PROOF OF LEMMA 3

By the proof of the measurement compression theorem [20], for any ϵ, $\delta>0$ and sufficiently large n there is an ensemble of rank-1 sub-POVM's $(p_s,\tilde{\Lambda}^{(s)}:\mathcal{H}_{A^n}\to\mathcal{H}_K)_s$ and a classical map $g:\mathcal{H}_S\otimes\mathcal{H}_K\to\mathcal{H}_{X^n}$ such that

(i) $\Sigma_k\tilde{\Lambda}_k^{(s)}\leqslant\Pi$, where the index k ranges over $[2^{n[H(A)+\delta]}]$, and Π is a projector commuting with $(\rho^A)^{\otimes n}$ such that $\mathrm{Tr}\,\Pi\leqslant 2^{n[H(A)+\delta]}$ and $\mathrm{Tr}(\rho^A)^{\otimes n}\Pi\geqslant 1-\epsilon$,

(ii)

$$\|(\rho^{XB})^{\otimes n}-\sigma^{X^nB^n}\|_1\leqslant\epsilon, \tag{B1}$$

where

$$\sigma^{X^nB^n}=(g\otimes I^{B^n})\Omega^{SKB^n},$$

$$\Omega^{SKB^n}=\sum_s p(s)|s\rangle\langle s|^S\otimes[(\Lambda^{(s)}\otimes I^{B^n})(\rho^{AB})^{\otimes n}],$$

for some probability distribution $p(s)$.

Each sub-POVM $\tilde{\Lambda}^{(s)}$ may be augmented by no more than $2^{n[H(A)+\delta]}$ rank-1 elements to satisfy equality

$$\sum_k\tilde{\Lambda}_k^{(s)}=\Pi.$$

The proof of lemma 1 and Fannes' inequality (A3) implies the existence of a decomposition $A^n=A_1A_2$ such that

$$H(A_1)\leqslant\frac{1}{e}+n\epsilon\log d_A,$$

while $\tilde{\Lambda}^{(s)}$ is now viewed as a rank-1 POVM on A_2 such that (B1) still holds for $\tilde{\Lambda}^{(s)}$.

Define $\epsilon'=3/ne+2\epsilon\log(d_Xd_B)$. Then

$$nI(X;B)_\rho\leqslant I(X^n;B^n)_\sigma-n\epsilon'\leqslant I(KS;B^n)_\Omega-n\epsilon'=I(S;B^n)_\Omega+I(K;B^n|S)_\Omega-n\epsilon'=I(K;B^n|S)_\Omega-n\epsilon'.$$

The first inequality is a triple application of Fannes' inequality, and the second is by the data processing inequality (see, e.g., [21]). The last line is by locality: the state of B^n is independent of which measurement $\tilde{\Lambda}^{(s)}$ gets applied to A^n. Thus there exists a particular s such that (10) is satisfied for $\tilde{\Lambda}=\tilde{\Lambda}^{(s)}$. □

[1] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[2] I. Devetak, A. W. Harrow, and A. Winter, e-print quant-ph/0308044.
[3] M. Horodecki, K. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen (De), and U. Sen, Phys. Rev. Lett. **90**, 100402 (2003).
[4] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[5] L. Szilard, Z. Phys. **53**, 840 (1929).
[6] C. H. Bennett, Int. J. Phys. **21**, 905 (1982).
[7] S. Lloyd, Phys. Rev. A **56**, 3374 (1997).
[8] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **89**, 180402 (2002).
[9] M. Horodecki, P. Horodecki, and J. Oppenheim, Phys. Rev. A **67**, 062104 (2003).
[10] I. Devetak and A. Winter, IEEE Trans. Inf. Theory **50**, 3183 (2004).
[11] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[12] J. Oppenheim, K. Horodecki, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **68**, 022307 (2003).
[13] L. Henderson and V. Vedral, J. Phys. A **34**, 6899 (2001).
[14] B. M. Terhal, M. Horodecki, D. W. Leung, and D. P. DiVincenzo, J. Math. Phys. **43**, 4286 (2002).
[15] M. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen (De), U. Sen, and B. Synak e-print quant-ph/0410090.
[16] I. Devetak and A. Winter, Proc. R. Soc. London, Ser. A **461**, 207 (2005); Phys. Rev. Lett. **93**, 080501 (2004).
[17] K. Matsumoto, T. Shimono, and A. Winter, Commun. Math. Phys. **246** (3), 427 (2004).
[18] M. Koashi and A. Winter, Phys. Rev. A **69**, 022309 (2004).
[19] I. Devetak and A. Winter, Phys. Rev. A **68**, 042301 (2003).
[20] A. Winter, e-print quant-ph/0109050.
[21] M. A. Nielsen and I. L. Chuang, *Quantum Information and Quantum Computation* (Cambridge University Press, Cambridge, England, 2001).
[22] C. H. Bennett (unpublished).
[23] I. Devetak and A. Winter, Proc. R. Soc. London, Ser. A (to be published), e-print quant-ph/0306078; Phys. Rev. Lett. (to be published), e-print quant-ph/0307053.
[24] J. Oppenheim, M. Horodecki, and R. Horodecki, Phys. Rev. Lett. **90**, 010404 (2003).
[25] B. Synak and M. Horodecki, e-print quant-ph/0403167.
[26] C. A. Fuchs and J. van de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).
[27] A. Winter, Commun. Math. Phys. **244** (1), 157 (2004).
[28] M. Fannes, Commun. Math. Phys. **31**, 291 (1973).