# Entanglement of assistance and multipartite state distillation

John A. Smolin,[1,*] Frank Verstraete,[2,3,†] and Andreas Winter[4,‡]
[1]*IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA*
[2]*Institute for Quantum Information, Caltech 107-81, Pasadena, California 91125, USA*
[3]*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str.1, 85748 Garching, Germany*
[4]*Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom*
(Received 10 June 2005; published 17 November 2005)

We find that the asymptotic entanglement of assistance of a general bipartite mixed state is equal to the smaller of its two local entropies. Our protocol gives rise to the asymptotically optimal Einstein-Podolsky-Rosen (EPR) pair distillation procedure for a given tripartite pure state, and we show that it actually yields EPR and Greenberger-Horne-Zeilinger (GHZ) states; in fact, under a restricted class of protocols, which we call "one-way broadcasting," the GHZ rate is shown to be optimal. This result implies a capacity theorem for quantum channels where the environment helps transmission by broadcasting the outcome of an optimally chosen measurement. We discuss generalizations to $m$ parties and show (for $m=4$) that the maximal amount of entanglement that can be localized between two parties is given by the smallest entropy of a group of parties of which the one party is a member, but not the other. This gives an explicit expression for the asymptotic localizable entanglement and shows that any nontrivial ground state of a spin system can be used as a perfect quantum repeater if many copies are available in parallel. Finally, we provide evidence that any unital channel is asymptotically equivalent to a mixture of unitaries and any general channel to a mixture of partial isometries.

DOI: 10.1103/PhysRevA.72.052317 PACS number(s): 03.67.Mn

## I. MULTIPARTITE QUANTUM STATES

One of the big ongoing program of quantum information theory is the classification of multipartite (pure) quantum states $\psi^{AB\cdots Z}$ and the understanding of the possible transformations between them allowing only local operations and classical communication (LOCC). As entanglement presents a resource that can be used for, e.g., quantum communication, it is especially interesting to study the asymptotic Shannon-theoretic limit. In this scenario, a few parties hold asymptotically many copies of identical states distributed among them, only joint operations between the particles at the same site and classical communication between the parties are allowed, and the conversion of states occurs with vanishing errors in the asymptotic limit.

In the bipartite case, this question is well understood: every pure state $\psi^{AB}=|\psi\rangle\langle\psi|^{AB}$ is asymptotically reversibly equivalent to maximally entangled [Einstein-Podolsky-Rosen (EPR)] states,

$$|\Phi_2\rangle^{AB}=\frac{1}{\sqrt{2}}(|0\rangle^A|0\rangle^B+|1\rangle^A|1\rangle^B),$$

at rate $E(\psi)=S(\psi^A)=S(A)$, the entropy of entanglement [1] (Note our notation convention $\psi^A=\mathrm{Tr}_B\,\psi^{AB}$ for the restriction of the state $\psi^{AB}$ to $A$.) So not only can we quantify the exact yield of the useful EPR states, but the latter serve as a normal form in general.

For multipartite states, the situation becomes more complex: there does not exist any longer a single state suited as a "gold standard;" e.g., it is quite evident that an EPR state $|\Phi_2\rangle^{AB}$ can never be equivalent to any quantity of Greenberger-Horne-Zeilinger (GHZ) states of three parties,

$$|\Gamma\rangle^{ABC}=\frac{1}{\sqrt{2}}(|0\rangle^A|0\rangle^B|0\rangle^C+|1\rangle^A|1\rangle^B|1\rangle^C).$$

So one has to aim at a "(minimal) reversible entanglement generating set" (MREGS) [2], about which little is known, except that apart from the easy candidates of $|\Phi_2\rangle^{AB}$, $|\Phi_2\rangle^{BC}$, $|\Phi_2\rangle^{AC}$, and $|\Gamma\rangle^{ABC}$ [3], an MREGS has to contain at least another state and possibly infinitely many. See [4] for an instructive case study.

Usually the two parts of the multiparty entanglement program—classification and possible transformations—are viewed as one, but as we have seen, the first is really much harder: this is because it involves studying the transformations between pairs of states which are *asymptotically reversible*.

In this paper, we have a more modest goal: we want to go from (many copies of) a given state to particular, interesting states, like the EPR and GHZ states. To be precise, one would like to "distill" as many as possible of these target states, with high fidelity in the limit of $n\to\infty$, and will care primarily about optimality of these processes and not so much for reversibility.

## II. THE TASK(S)

Given many copies of a (pure) tripartite state $\psi^{ABC}$, which "standard" entangled states like EPR states $\Phi_2$ between any pair of parties or GHZ states $\Gamma$, can the three parties distill by

*Electronic address: smolin@watson.ibm.com
†Electronic address: fverstraete@ist.caltech.edu
‡Electronic address: a.j.winter@bris.ac.uk

 

local operations and classical communication (LOCC)?

We shall focus on three scenarios in succession: first, we study optimal distillation of EPR states between a given pair of players from a tripartite state; second, we recast the protocol as one of distilling EPR and GHZ states at the same time and show that for this target set and a restricted class of protocols it gives optimal yield; and third, we look at $m$-partite states and how many EPR states between a prescribed pair of players can be distilled by local measurements on the other $m-2$ parties.

Scenario 1 was studied in [5–7] under the name of "entanglement of assistance" in a nonasymptotic setting: given a mixed state $\psi^{AB}=\mathrm{Tr}_C(|\psi^{ABC}\rangle\langle\psi^{ABC}|)$ shared between $A$ and $B$, there exists a unique purification $\psi^{ABC}$ up to local unitary operations on $C$, and the question was asked how much EPR-type entanglement can be created between $A$ and $B$ when $C$ is doing local measurements and communicates the results to $A$ and $B$. In this paper we completely solve that question in the asymptotic setting.

About scenario 2 very little has appeared in the literature, except for upper-capacity bounds—e.g., [3]—and a few (qubit) protocols which, however, remain largely in the single-copy setting [8,9].

The third scenario has been studied in the context of spin chains under the name of *localizable entanglement* [10]. The present work will reveal some intriguing connections between the concept of entropy of a block of spins and the entanglement length in spin systems.

The main results are as follows.

*Theorem 1*. Given a pure tripartite state $\psi^{ABC}$, then the optimal EPR rate distillable between $A$ and $B$ with the help of $C$ under LOCC is

$$E_A^\infty(\psi^{ABC}) = \min\{S(A),S(B)\}.$$

(Our notation is such that the first two parties obtain EPR states and the remaining is the helper.) This is the asymptotic entanglement of assistance [6].

Writing the tripartite state as $|\psi\rangle^{ABC}=\sum_j\sqrt{q_j}|\psi_j\rangle^{AB}|j\rangle^C$, with orthogonal $|j\rangle$—corresponding to to a pure state decomposition $\psi^{AB}=\sum_j q_j|\psi_j\rangle\langle\psi_j|^{AB}$—let $\bar{E}=\sum_j q_j E(\psi_j)$ be the average entanglement of the pure state decomposition. Define finally $\chi=\min\{S(A),S(B)\}-\bar{E}$.

*Theorem 2*. Let $\psi^{ABC}$ be a pure tripartite state. Then, for $\epsilon,\delta>0$ and sufficiently large $n$, there exists a protocol involving only an instrument on $C^n$ and broadcast of the measured result, followed by local operations on $A^n$ and $B^n$, which effects the transformation

$$(\psi^{ABC})^{\otimes n} \rightarrow (\Gamma^{ABC})^{\otimes n(\chi-\delta)} \otimes (\Phi_2{}^{AB})^{\otimes n(\bar{E}-\delta)},$$

with fidelity $1-\epsilon$.

Observe that the minimal value of $\bar{E}$ above is the *entanglement of formation* $E_F(\psi^{AB})$ of the mixed state $\psi^{AB}$ between Alice and Bob [11]. In the limit of many copies we have to substitute the *entanglement cost* $E_C(\psi^{AB})$ [12]. This outcome is better than theorem 1, as we can always (irreversibly) turn GHZ states into EPR states and achieve the previous EPR rate. Theorem 9 in Sec. V shows that the corresponding GHZ rate $\min\{S(A),S(B)\}-E_C(\psi^{AB})$ is indeed optimal under an important class of protocols.

*Theorem 3*. For an $m$-party state $\psi^{ABC_1\cdots C_{m-2}}$, the optimal rate $R$ of EPR states distillable between $A$ and $B$ with the help of the $C_i$ via LOCC satisfies

$$R \leq \min_{\mathcal{S}\subset\{C_1,\ldots,C_{m-2}\}} S(A\mathcal{S}). \qquad (1)$$

This bound is achievable for all $m$, which is therefore the expression for the asymptotic version of the localizable entanglement. Furthermore, the bound is achieved by a protocol where each helper $C_i$ takes a single turn in which he measures his state and communicates the result to the remaining parties.

We prove this here for $m=4$; the general proof is given in [13]. Observe that the right-hand side in Eq. (1) is the minimum pure state entanglement over all bipartite cuts of the systems which separate Alice and Bob.

The remainder of the paper is structured as follows: In Sec. III we present a protocol and prove theorem 1. Section IV presents an application of this first result to quantum transmission with a classical helper in the channel environment. In Sec. V we show how to make the basic protocol coherent, such that it also gives GHZ states, and prove theorem 2. Its GHZ rate we prove to be optimal under a subclass of protocols which we call *one-way broadcast*. Then, in Sec. VI we generalize the basic protocol to more than one helper (proof of theorem 3 for $m=4$) and discuss the connection between the concept of localizable entanglement and entropy of blocks of spins. Finally, in Sec. VII we discuss possible applications and/or extensions of our main result to asymptotic normal forms of quantum channels and conclude in Sec. VIII.

## III. ASYMPTOTIC ENTANGLEMENT OF ASSISTANCE

In [5,6], the following quantity was introduced under the name of *entanglement of assistance* of a bipartite mixed state $\rho^{AB}$ (with purification $\psi^{ABC}$):

$$E_A(\rho^{AB}) := E_A(\psi^{ABC}) := \max\left\{\sum_i p_i E(\psi_i^{AB}) : \rho^{AB} = \sum_i p_i \psi_i^{AB}\right\}.$$

The idea is that by varying a measurement, i.e. a positive operator valued measure (POVM) on $C$, the helper Charlie can effect any pure state ensemble decomposition $\rho^{AB}=\sum_i p_i\psi_i^{AB}$ for Alice and Bob's state [14]. In this sense, $E_A$ gives the maximum amount of entanglement obtainable between Alice and Bob with the (remote) help from Charlie. Of course, we are primarily interested in the operational asymptotic rate of EPR states, $E_A^\infty$, which will turn out to be given by the regularization of $E_A$:

$$E_A^\infty(\rho) = \lim_{n\rightarrow\infty}\frac{1}{n}E_A(\rho^{\otimes n}).$$

Now we argue the upper bound $E_A^\infty(\rho^{AB})\leq S(A)$, which was noted in [6], operationally: whatever can be done under three-party LOCC is contained in protocols which allow general transformations on $BC$ and LOCC with respect to the cut

$A$ vs $BC$. But in this latter formulation, we are in a bipartite pure state situation, for which the maximum yield of EPR states is well known to be $S(A)$ [1]. By an identical argument, we have the same bound with $S(B)$, and hence we obtain

$$E_A^\infty(\rho^{AB}) \leq \min\{S(A),S(B)\}. \quad (2)$$

Note that this upper bound is not additive under general tensor products (compare [6]): consider strictly mixed states $\rho^{AB}=|0\rangle\langle 0|^A \otimes \rho^B$ and $\sigma^{A'B'}=\sigma^{A'}\otimes|0\rangle\langle 0|^{B'}$; they have both $E_A=0$, because the entropy upper bound is 0. However, $E_A(\rho\otimes\sigma)>0$. A less trivial example of superadditivity of $E_A$ is given in [6] for two copies of the same state. Here is a very easy one.

*Example 4 (superadditivity of $E_A$)*. Consider the three-qutrit determinant (or Aharonov) state

$$|\alpha\rangle^{ABC} = \frac{1}{\sqrt{6}}(|012\rangle + |120\rangle + |201\rangle - |210\rangle - |102\rangle - |021\rangle).$$

The restriction $\alpha^{AB}$ is proportional to the projector onto the $(3\times 3)$-antisymmetric subspace, and it is well known that this subspace consists entirely of "singlets"—i.e., states $|v\rangle|v'\rangle - |v'\rangle|v\rangle$, with $\langle v|v'\rangle=0$. Hence $E_A(\alpha)=1$ [and by the way also $E_F(\alpha^{AB})=E_C(\alpha^{AB})=1$ [15]]. However, $E_A(\alpha\otimes\alpha)\geq 2.5$, since $\alpha\otimes\alpha$ can be presented as a uniform mixture of states $(U_1^{A_1}\otimes U_2^{A_2}\otimes U_1^{B_1}\otimes U_2^{B_2})|\varphi\rangle$, with

$$|\varphi\rangle^{A_1A_2B_1B_2} = \frac{1}{\sqrt{8}}[(|01\rangle - |10\rangle)^{A_1B_1}\otimes(|01\rangle - |10\rangle)^{A_2B_2} + (|12\rangle - |21\rangle)^{A_1B_1}\otimes(|12\rangle - |21\rangle)^{A_2B_2}].$$

It is easily established that the Schmidt spectrum of this state is $[\frac{1}{4},\frac{1}{4},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8}]$, so its entropy of entanglement is $E(\varphi)=2.5$.

This example contains a valuable insight: for a given single-copy decomposition of $\rho$, one can form superpositions of tensor products of component states and increase the entanglement. A little consideration reveals that this is so because the tensor products have some local distinguishability. Hence, in the general case we should try to enforce local distinguishability of the states we put in superposition.

*Proof of theorem 1.* Write $|\psi\rangle^{ABC}=\sum_j\sqrt{q_j}|\psi_j\rangle^{AB}|j\rangle^C$, with an orthonormal basis $\{|j\rangle\}$ of $C$. Let

$$\chi_A = \chi\{(q_j,\psi_j^A)\} = S(A) - \sum_j q_j S(\psi_j^A), \quad (3)$$

$$\chi_B = \chi\{(q_j,\psi_j^B)\} = S(B) - \sum_j q_j S(\psi_j^B) \quad (4)$$

denote the Holevo information of the given ensembles; observe the common term

$$\sum_j q_j S(\psi_j^A) = \bar{E} = \sum_j q_j S(\psi_j^B).$$

We may assume without loss of generality that $S(A)\leq S(B)$, and hence $\chi_A\leq\chi_B$.

For $n$ copies of $\psi$, the sequences $J=j_1,\ldots,j_n$ and, consequently, the states $|J\rangle=|j_1\rangle\cdots|j_n\rangle$ fall into (polynomially many) *type classes:* we say that $J$ is of type $P$ (which is a probability distribution on the letters $j$) if $j$ occurs exactly $nP(j)$ times in $J$. This is relevant because the probability $q_J=q_{j_1}\cdots q_{j_n}$, the product of the letter probabilities, of a sequence is constant across a type class. We can write the state as

$$(|\psi\rangle^{ABC})^{\otimes n} = \sum_J \sqrt{q_J}|\psi_J\rangle^{A^nB^n}|J\rangle^{C^n},$$

where $|\psi_J\rangle=|\psi_{j_1}\rangle\cdots|\psi_{j_n}\rangle$ and $A^n=A_1A_2\cdots A_n$ are Alice's $n$ copies of system $A$, etc. The goal of Charlie's strategy will be to project this state down to a superposition of terms $|\psi_J\rangle^{A^nB^n}$ which are as orthogonal as possible on both Alice's and Bob's systems: because then Alice's (say) reduced state is roughly an orthogonal mixture of the states $\psi_J^{A^n}$ and we can easily calculate its entropy.

More precisely, Charlie's measurement consists of two steps: first, a projection into the subspaces of constant type—say, $P$:

$$\Pi(P) \coloneqq \mathrm{span}\{J{:}J \text{ is of type } P\}.$$

Note that, for any $\eta>0$, with probability $1-\epsilon$, $\|P-q\|_1\leq\eta$, if only $n$ is sufficiently large (otherwise, abort). Here, $\|\cdot\|_1$ is the total variational distance (or one-norm distance) of probability distributions. By Fannes' inequality (stated below as lemma 5), then (with $\delta=-\eta\log\eta+2\eta\log d$),

$$S\left(\sum_j P(j)\psi_j^A\right) - \sum_j P(j)S(\psi_j^A) \geq \chi_A - \delta.$$

Second, for each such type $P$, letting $N=\lfloor 2^{n(\chi_A-2\delta)}\rfloor$, define states depending on a set $\mathcal{J}=\{J^{(0)},\ldots,J^{(N-1)}\}$ of sequences of type $P$ and a number $\alpha=0,\ldots,N-1$:

$$|t_{\mathcal{J}}(\alpha)\rangle = \frac{1}{\sqrt{N}}\sum_{\beta=0}^{N-1} e^{2\pi i\alpha\beta/N}|J^{(\beta)}\rangle.$$

Clearly, with an appropriate constant $c>0$, the collection $((c/N)|t_{\mathcal{J}}(\alpha)\rangle\langle t_{\mathcal{J}}(\alpha)|)_{\mathcal{J},\alpha}$ forms a POVM on the type $P$ subspace; i.e., these operators sum up to $\Pi(P)$. This is the second (rank-1) POVM of Charlie.

By the Holevo-Schumacher-Westmoreland (HSW) theorem, stated as lemma 6 below, the vast majority of the sets $\mathcal{J}$ are good codes for the classical-quantum channel $j\mapsto\psi_j^A$ and simultaneously for $j\mapsto\psi_j^B$. For a good code $\mathcal{J}$ and any $\alpha$, consider the projected state $|\vartheta\rangle$ of $AB$ (up to normalization), dropping the superscript $n$ from the registers:

$$|\vartheta\rangle = \frac{1}{\sqrt{N}}\sum_{\beta=0}^{N-1} e^{-2\pi i\alpha\beta/N}|\psi_{J(\beta)}\rangle^{AB}.$$

Because Alice and Bob have good decoders for $\beta$—i.e., POVM's $(D_\beta^A)_\beta$ and $(D_\beta^B)_\beta$—they can locally extract $\beta$ with high reliability. We can always think of these measurements as (local) isometries—for example, for Alice,

$$V_A = \sum_\beta \sqrt{D^A_\beta} \otimes |\beta\rangle^{B'},$$

and a similar expression $V_B$ for Bob. This pair of unitaries takes the state $|\vartheta\rangle$ to

$$|\tilde{\vartheta}\rangle = (V_A \otimes V_B)|\vartheta\rangle = \sum_{\beta,\gamma} \left(\sqrt{D^A_\beta} \otimes \sqrt{D^B_\gamma}|\vartheta\rangle\right)^{AB} \otimes |\beta\rangle^{B'}|\gamma\rangle^{C'}.$$

Since we have (with high probability) a good code both for Alice's and Bob's channels, we expect that

$$|\tilde{\vartheta}\rangle \approx \frac{1}{\sqrt{N}}\sum_{\beta=0}^{N-1} e^{-2\pi i\alpha\beta/N}|\psi_{J(\beta)}\rangle^{A^nB^n}|\beta\rangle^{A'}|\beta\rangle^{B'};$$

i.e., the fidelity between both states is close to 1, which indeed can be shown (see the proof of theorem 2 below). Here we need something only slightly weaker:

When tracing over $BB'$, we can assume that Bob's POVM is actually performed (i.e., the register $B'$ observed); using the fact that both Alice's and Bob's POVM's have average error probability $\leq \epsilon$, we get

$$\left\| \tilde{\vartheta}^A - \frac{1}{N}\sum_\beta \sqrt{D^A_\beta}\psi^A_{J(\beta)}\sqrt{D^A_\beta} \otimes |\beta\rangle\langle\beta|^{A'} \right\|_1 \leq 2(2\epsilon),$$

with the trace norm $\|\cdot\|_1$ on (density) operators. Furthermore, by the gentle measurement lemma 7, stated below for convenience, this yields (for $\epsilon \leq 1$)

$$\left\| \tilde{\vartheta}^A - \frac{1}{N}\sum_\beta \psi^A_{J(\beta)} \otimes |\beta\rangle\langle\beta|^{A'} \right\|_1 \leq 4\epsilon + \sqrt{8\epsilon} \leq 7\sqrt{\epsilon}.$$

Hence, for the entropy (choosing $\epsilon$ and $\eta$ small enough),

$$\begin{aligned} S(\vartheta^A) = S(\tilde{\vartheta}^A) &\geq S\left(\frac{1}{N}\sum_\beta \psi^A_{J(\beta)} \otimes |\beta\rangle\langle\beta|^{A'}\right) - n\delta \\ &= \log N + \frac{1}{N}\sum_\beta E(\psi^{AB}_{J(\beta)}) - n\delta \\ &= \log N + n\sum_j P(j)E(\psi^{AB}_j) - n\delta \geq n(S(A) - 4\delta), \end{aligned}$$

where we have used the Fannes inequality, the fact that all $J^{(\beta)}$ have the same type $P$, and Fannes inequality once more.

*Lemma 5 (Fannes inequality [16]).* For any states $\rho$ and $\sigma$ on a $d$-dimensional Hilbert space: if $\|\rho - \sigma\|_1 \leq \epsilon \leq 1/e$, then $|S(\rho) - S(\sigma)| \leq \eta(\epsilon) + \epsilon \log d$, with $\eta(x) = -x \log x$.

*Lemma 6 (HSW theorem [17]).* For a classical-quantum channel $W: x \mapsto W_x$ on the Hilbert space $\mathcal{H}$ and a probability distribution $P$, let $U^{(i)}$ be independent and indentically distributed (i.i.d.) uniformly random from the sequences of length $n$ of type $P$. Then for every $\epsilon, \delta > 0$ and sufficiently large $n$, if $\ln N \leq n(\chi\{(P(x), W_x)\} - \delta)$,

$$\Pr\left\{\mathcal{C} = (U^{(i)})_{i=1}^N \text{ has error at most } \epsilon\right\} \geq 1 - \epsilon.$$

Here we call a collection of codewords "$\epsilon$-good" if there exists a POVM $(D_i)_{i=1}^N$ on $\mathcal{H}^{\otimes n}$ such that

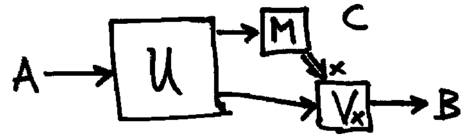


FIG. 1. Alice prepares an input to (many copies of) the isometry $U$, which gives part of the state to Bob and part to Charlie. The latter measures a POVM $M$ on his system and classically communicates his result $x$ to Bob, who executes a unitary $V_x$ depending on Charlie's message to recover Alice's sent state.

$$\frac{1}{N}\sum_{i=1}^N \mathrm{Tr}(W^n_{U^{(i)}}D_i) \geq 1 - \epsilon.$$

(In this form, the theorem is proved in [18].) □

*Lemma 7 (gentle measurements [19]).* Let $\rho$ be a state (actually, $\rho \geq 0$ and $\mathrm{Tr}\,\rho \leq 1$ are enough) and $0 \leq X \leq \mathbb{1}$, such that $\mathrm{Tr}(\rho X) \geq 1 - \epsilon$. Then, $\|\rho - \sqrt{X}\rho\sqrt{X}\|_1 \leq \sqrt{8\epsilon}$. □

## IV. CHANNEL CAPACITY WITH CLASSICAL HELPER IN THE ENVIRONMENT

Gregoratti and Werner [20] have considered the following channel model with helper in the environment:

$$U: \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_C,$$

described by an isometry from Alice's input system $A$ to the combination of Bob's output system $B$ and the environment $C$. Assume that the environment system may be measured and the classical results of the observation forwarded to Bob—attempting to help him in error correcting quantum information sent from Alice.

We are interested in the quantum capacity of this scenario from Alice to Bob, in the asymptotic limit of block-coded information (and collectively measured environment). The setup is illustrated in Fig. 1.

We want to mention a related model, discussed by Hayden and King [21], where the objective is to transmit classical information rather than quantum. Of course, the corresponding capacity will usually be higher, since the helper in the environment can learn part of the message and forward this information to Bob.

*Theorem 8.* The *environment-assisted quantum capacity* of a noisy quantum channel $T: A \to B$ is

$$Q_A(T) = \max_\rho \min\{S(\rho), S(T(\rho))\}.$$

The same capacity is obtained allowing unlimited LOCC between Alice, Bob, and Charlie.

*Proof.* Let us first deal with the converse: whatever the detailed strategy, Alice will eventually input the $A^n$ part of some state $|\Phi\rangle^{A'A^n}$ into the channel (there is no loss of generality in assuming that the players keep all ancillas around and hence the state pure). After the channel, the three players share the state

$$|\psi\rangle^{A'B^nC^n} = (\mathbb{1} \otimes U^{\otimes n})|\Phi\rangle^{A'A^n}.$$

By the same argument as for the upper bound in theorem 1, the pure state entanglement between Alice and Bob cannot exceed either

$$S(\psi^{A^n}) \leq \sum_k S(\psi^{A_k}) = \sum_k S(\Phi^{A_k}) \leq nS(\rho^A)$$

or

$$S(\psi^{B^n}) \leq \sum_k S(\psi^{B_k}) = \sum_k S(T(\Phi^{A_k})) \leq nS(T(\rho^A)),$$

with $\rho^A=(1/n)\Sigma_k\Phi^{A_k}$.

For the direct part, let $\rho$ be the optimal input state for the maximum in the theorem and denote a purification of it $|\phi\rangle^{A'A}$, which is used in the following as "test state."

Let Charlie pick, for some $n$, an optimal measurement $(M_x)_x$ for the entanglement of assistance of ($n$ copies of) $|\psi\rangle=(\mathbb{1}\otimes U)|\phi\rangle^{A'A}$, according to theorem 1. Then we can define a new quantum channel

$$T':A^n \to B^nB'$$

$$\varphi \mapsto \sum_x \mathrm{Tr}_{C^n}[(M_x\otimes\mathbb{1})(U\varphi U^*)]\otimes|x\rangle\langle x|^{B'},$$

which, by theorem 1, has on the test state $|\phi\rangle^{\otimes n}$ the coherent information [22]

$$\begin{aligned} I(A'\rangle B^nB') &= S(B^nB') - S(A'B^nB') \\ &\geq n(\min\{S(\rho),S(T(\rho))\} - \delta). \end{aligned}$$

Invoking the quantum channel coding theorem [23–25], there are block codes for $T'$ achieving this rate asymptotically. □

## V. GHZ DISTILLATION

Now we will show how to modify the protocol of theorem 1 by "making it coherent" (after the model of [23,26]) such that part of its yield is in the form of GHZ states. We shall freely use the notation introduced in the proof of theorem 1.

*Proof of theorem 2*. By possibly embedding $C$ into a larger space, we can write $|\psi\rangle^{ABC}=\Sigma_j\sqrt{q_j}|\psi_j\rangle^{AB}|j\rangle^C$, for any pure state decomposition of $\psi^{AB}$ into an ensemble $\{q_j,\psi_j^{AB}\}$.

Consider sets $\mathcal{J}=\{J^{(0)},\ldots,J^{(N-1)}\}$ of $N=\lfloor 2^{n(\chi_A-2\delta)}\rfloor$ (integer part) type-$P$ sequences of length $n$, with, as before, $\|P-q\|_1\leq\eta$. Now construct the projectors

$$\Theta_{\mathcal{J}} = \sum_{\alpha=0}^{N-1}|t_{\mathcal{J}}(\alpha)\rangle\langle t_{\mathcal{J}}(\alpha)|,$$

so that we have a POVM $(c\Theta_{\mathcal{J}})_{\mathcal{J}}$, a coarse graining of the measurement used in the proof of theorem 1.

Charlie's measurement is again in two parts: first he measures the type subspace $\Pi(P)$, and $P$ is close to $q$ as above with high probability (otherwise abort). Then he measures $(c\Theta_{\mathcal{J}})_{\mathcal{J}}$; if the operator $\Theta_{\mathcal{J}}$ acts, the projected state is (dropping the superscript $n$ from the register names)

$$|\zeta\rangle^{ABC} = \frac{1}{\sqrt{N}}\sum_{\beta=0}^{N-1}|\psi_{J^{(\beta)}}\rangle^{AB}|J^{(\beta)}\rangle^C.$$

Most of the sets $\mathcal{J}$ are good codes for both the channels $j\mapsto\psi_j^A,\psi_j^B$. Hence, with large probability, we can use the same local isometries $V_A$ and $V_B$ as before to extract $\beta$ with little state disturbance and a local unitary $V_C$ mapping $|J^{(\beta)}\rangle\mapsto|\beta\rangle$. These isometries map $|\zeta\rangle$ to

$$|\widetilde{\zeta}\rangle^{AA'BB'C'} \approx \frac{1}{\sqrt{N}}\sum_{\beta=0}^{N-1}|\psi_{J^{(\beta)}}\rangle^{AB}|\beta\rangle^{A'}|\beta\rangle^{B'}|\beta\rangle^{C'},$$

and because the $J^{(\beta)}$ are all of the same type, they are permutations of each other, so the states $|\psi_{J^{(\beta)}}\rangle^{AB}$ can be taken to a standard state $|\psi_{\bar{J}}\rangle^{AB}$—say, the lexicographically first sequence of type $P$—by (controlled) permutations of the $n$ subsystems. So they arrive at the state

$$|Z\rangle^{AA'BB'C'} \approx |\psi_{\bar{J}}\rangle^{AB}\otimes\frac{1}{\sqrt{N}}\sum_{\beta=0}^{N-1}|\beta\rangle^{A'}|\beta\rangle^{B'}|\beta\rangle^{C'},$$

This concludes the proof, since the rate of $N$ is asymptotically $\chi_A$, and the rate of $E(\psi_{\bar{J}}^{AB})$ is asymptotically $\bar{E}$. □

*Remark*. We have presented the POVM's of theorems 1 and 2 in the simplest possible terms. One can also minimize these POVM's by not taking all sets $\mathcal{J}$. This can be done as shown in [18,27], yielding for theorem 1 a rank-1 measurement with $\approx 2^{nS(C)}$ elements; for theorem 2 the POVM has $\approx 2^{n[H(q)-\chi_A]}$ operators. □

Now we show that theorem 2 is in a certain sense optimal: namely, it gives the largest GHZ rate among all protocols which consist only of (i) a local operation with measurement at $C$, (ii) sending the classical information obtained in the measurement to $A$ and $B$, and (iii) local operations of $A$ and of $B$ depending on the message. In particular, we allow no feedback communication and no communication between Alice and Bob. These are severe restrictions, but at least the protocol from theorem 2 is of this type: we call it *one-way broadcast*.

*Theorem 9*. Under one-way broadcast protocols from $C$ to $AB$, the asymptotic GHZ rate from the state $\psi^{ABC}$ cannot exceed $\min\{S(A),S(B)\}-E_C(\psi^{AB})$.

*Proof*. We show actually a bit more: the rate of three-way *common randomness* distillable by such protocols is asymptotically bounded by the same number. This problem was studied in [27] for two players with one-way communication, and the relevant observation here is that with one-way broadcast, the task is equivalent to two simultaneous two-player common randomness distillations: from $C$ to $A$ and from $C$ to $B$.

The setup is the following: the sender $C$ and the receiver ($A$ or $B$) initially share a quantum state and by local operations and one-way classical communication want to distill a maximum amount of shared randomness, which, however, has to be independent of the communicated message(s).

A particular protocol for doing this is to distill GHZ states by a one-way broadcast protocol and then all three measure these states in the computational basis—by purity of the measured state, the resulting perfect shared randomness is independent of everything else in the protocol.

It was shown in [27] that the maximum rate achievable between $C$ and $A$ is the maximum of Eq. (3)—actually regularized for many copies of the state—and similarly between $C$ and $B$ the—regularized—maximum of Eq. (4). The smaller of these numbers clearly is just $\min\{S(A),S(B)\}-E_C(\psi^{AB})$. □

*Remark*. In general, the GHZ rate obtainable from a pure state $|\psi\rangle^{ABC}$ by general LOCC has the easy upper bound $\min\{S(A),S(B),S(C)\}$. Remarkably, our protocol of theorem 2 achieves this for a broad class of states—namely, when one of the reduced states $\psi^{AB}$, $\psi^{BC}$, or $\psi^{AC}$ is separable. □

*Example 10 (Groisman, Linden, and Popescu [4])*. Consider the family of states

$$|\Upsilon_\alpha\rangle^{ABC}=\alpha|0\rangle^A|\Phi^+\rangle^{BC}+\beta|1\rangle^A|\Phi^-\rangle^{BC},$$

with $0\leq\alpha\leq\beta$ and $\alpha^2+\beta^2=1$, interpolating between a state $|\Phi_2\rangle^{BC}(\alpha^2=0)$ and $|\Gamma\rangle^{ABC}$ ($\alpha^2=1/2$; up to local unitaries). Observe that it is certainly possible to obtain one EPR state between $B$ and $C$ from $\Upsilon_\alpha$.

In [4] it is observed that the local entropies of $\Upsilon_\alpha$ are consistent with the hypothetical existence of an asymptotically reversible transformation into

$$H_2(\alpha^2)\,|\Gamma\rangle^{ABC}\text{ and }\quad[1-H_2(\alpha^2)]\,|\Phi_2\rangle^C \tag{5}$$

per copy of the state, but the authors present heuristic arguments for its impossibility.

Let us see what our results tell us about the distillability of GHZ and EPR states: by applying theorem 2 with $B$ (or equivalently $C$) in the role of the helper, we obtain (since $\Upsilon^{AC}$ is separable) a GHZ rate of $H_2(\alpha^2)$, but no EPR states. The GHZ rate is evidently optimal under general LOCC protocols, as it coincides with Alice's entropy (i.e., her entanglement with the rest of the players). By applying theorem 2 with $A$ as the helper, we have to calculate the entanglement cost of the Bell mixture $\Upsilon^{BC}$, which happens to be known by [28,29]:

$$E_C(\Upsilon^{BC})=E_F(\Upsilon^{BC})=H_2\left(\frac{1}{2}-\alpha\beta\right).$$

Hence we get distillation of

$$\left[1-H_2\left(\frac{1}{2}-\alpha\beta\right)\right]|\Gamma\rangle^{ABC}\text{ and }\quad H_2\left(\frac{1}{2}-\alpha\beta\right)|\Phi_2\rangle^C$$

per copy of the state, and theorem 9 shows that this GHZ rate is optimal among all one-way broadcast protocols from $A$ to $BC$. Note that the GHZ rate is slightly worse than the one stated in Eq. (5).

*Example 11 (W-State)*. Another interesting example is provided by the $W$-state [8]

$$|W\rangle=\frac{1}{\sqrt{3}}(|001\rangle+|010\rangle+|100\rangle),$$

which is interesting because it cannot be converted to a GHZ state even probabilistically (on a single copy).

Theorem 1 tells us that any two parties can obtain a rate of $H_2(\frac{1}{3})\approx0.918$ EPR states, with assistance from the third. Since two EPR pairs between different players can be converted into a GHZ state, we can obtain a GHZ rate of at least $\frac{1}{2}H_2(\frac{1}{3})\approx0.459$.

However, using theorem 2, we can do a bit better: the entanglement of formation of any two-party reduced state is evaluated with the help of [29], and we get rates of $H_2(\frac{1}{3})-H_2((1-\sqrt{5/9})/2)\approx0.368$ for GHZ states and of $H_2((1-\sqrt{5/9})/2)\approx0.550$ for EPR states between any pair of players. Converting the EPR states to GHZ states as before, we arrive at an overall GHZ rate of $H_2(\frac{1}{3})-\frac{1}{2}H_2((1-\sqrt{5/9})/2)$, which is $\approx0.643$.

## VI. SINGLET DISTILLATION WITH THE HELP OF MANY (DISTANT) FRIENDS: ASYMPTOTIC LOCALIZABLE ENTANGLEMENT

Consider now the $m$-party generalization of scenario 1: distillation of EPR pairs between $A$ and $B$ with the help of $C_1,\ldots,C_{m-2}$ from an $m$-partite pure state, by LOCC.

In analogy to the upper bound, Eq. (2), we can easily obtain an upper bound on the achievable rate $R$ in this scenario: surely, the distillable entanglement can only go up if we allow Alice to team up with a subset $\mathcal{S}$ of the helpers $C_i$ and Bob with the complement $\overline{\mathcal{S}}=\{1,\ldots,m-2\}\backslash\mathcal{S}$, such that all collective operations on $A\mathcal{S}$ and on $B\overline{\mathcal{S}}$ are allowed, and LOCC between these two groups. Thus, $R\leq S(A\mathcal{S})$, and we get Eq. (1),

$$R\leq\min_{\mathcal{S}}S(A\mathcal{S})=\min_{\mathcal{S}}S(B\overline{\mathcal{S}}).$$

[Note that this reduces to the inequality (2) for $m=3$.]

In [30] it was shown that whenever the right-hand side in the above equation is nonzero, then one of Alice and Bob can, with LOCC help from the other parties, distill EPR pairs at nonzero rate.

It turns out, however, that the right-hand side is achievable for any $m$, and we show here how to do it in the case of $m=4$ (the general case requires different arguments and is solved in [13]).

*Proof of theorem 3 for* $m=4$. Only the achievability of the minimum cut entanglement is left to be proved. For $m=4$—i.e., two helpers Charlie and Debbie—this means we are looking at $R=\min\{S(A),S(B),S(AC),S(BC)\}$.

Our goal will be to construct a measurement on $D^n$ such that for nearly all projected states $|\vartheta\rangle^{A^nB^nC^n}$,

$$\min\{S(\vartheta^{A^n}),S(\vartheta^{B^n})\}\geq n(R-\delta),$$

with arbitrary $\delta>0$ and sufficiently large $n$. I.e., we want to preserve (up to a small loss) the minimum cut entanglement, while disengaging Debbie. If we succeed doing this, we can invoke theorem 1 for the residual tripartite state.

Pick any basis of $D$, so that we can write $|\psi\rangle^{ABCD}=\sum_j\sqrt{q_j}|\psi_j\rangle^{ABC}|j\rangle^D$. As in the previous proofs, we have reduced state ensembles with Holevo information $\chi_A$, $\chi_B$, $\chi_{AC}$, and $\chi_{BC}$. By possibly swapping $A$ and $B$, we may assume that $\chi_A\leq\chi_B$. Invoking monotonicity of the Holevo information under partial trace, $\chi_A\leq\chi_{AC}$ and $\chi_B\leq\chi_{BC}$, we are left with one of the following orderings of the four quantities:

$$\chi_A\leq\chi_B\leq\chi_{AC}\leq\chi_{BC},$$

or

$$\chi_A \leqslant \chi_{AC} \leqslant \chi_B \leqslant \chi_{BC}.$$

Define $\chi_0 := \min\{\chi_B, \chi_{AC}\}$ and consider random codes $\mathcal{J}$ of rate $\chi_0 - \delta$ where, in a slight variation of the proof of theorem 1, the codewords are drawn from the distribution $q^{\otimes n}$—this is the original form of the HSW theorem [17], and the conclusion of lemma 6 holds true. Now construct a rank-1 measurement on $D$, in the same way as we did there. What can we say about the projected state

$$|\vartheta\rangle^{ABC} = \frac{1}{\sqrt{N}}\sum_\beta e^{-2\pi i\alpha\beta/N}|\psi_{J(\beta)}\rangle^{ABC}?$$

For the bipartition $B|AC$, essentially the same argument from that proof shows that with high probability, the entanglement of $\vartheta$ is

$$E(\vartheta^{B|AC}) \geqslant n(\min\{S(B),S(AC)\} - \delta) \geqslant n(R-\delta).$$

For the bipartition $A|BC$ this only works when $\chi_A = \chi_0$, to make the rate of the codes smaller than either Holevo information. So let us assume $\chi_A < \chi_0$ and $\delta$ so small that $\chi_A + \delta \leqslant \chi_0 - \delta$. The rate of the code is still smaller than $\chi_{BC}$, so there exists a (hypothetical) "local" decoding of $\beta$ from the register $BC$. I.e., with respect to the bipartite cut $A|BC$, the state $|\vartheta\rangle$ is equivalent to

$$|\tilde{\vartheta}\rangle^{ABC} \approx \frac{1}{\sqrt{N}}\sum_\beta e^{-2\pi i\alpha\beta/N}|\psi_{J(\beta)}\rangle^{ABC}|\beta\rangle^{\widetilde{BC}},$$

where the approximation has the same quality as in the proof of theorem 1. But then, we have

$$\tilde{\vartheta}^A \approx \frac{1}{N}\sum_\beta \psi_{J(\beta)}.$$

Now we can conclude the proof by invoking lemma 12 below, which states that for a random code (which is what the POVM will select) the average on the right-hand side is $\approx(\psi^A)^{\otimes n}$. Hence and using Fannes' inequality once more,

$$E(\vartheta^{A|BC}) \geqslant n(S(A)-\delta) \geqslant n(R-\delta),$$

and we are done. □

*Lemma 12 (Density sampling [31]).* Consider the ensemble $\{(q_j,\rho_j)\}$ of states on a $d$-dimensional Hilbert space, with average density operator $\rho$ and Holevo information $\chi$. Let independent and identically distributed random variables $X_1,\ldots,X_N$, drawn from the states $\rho_J = \rho_{j_1}\otimes\cdots\otimes\rho_{j_n}$ with probability $q_J = q_{j_1}\cdots q_{j_n}$. Then, for every $\epsilon,\delta>0$, $N \geqslant 2^{n(\chi+\delta)}$, and sufficiently large $n$,

$$\left\|\frac{1}{N}\sum_k^N X_k - \rho^{\otimes n}\right\|_1 \leqslant \epsilon,$$

with probability $\geqslant 1-\epsilon$. □

Theorem 3 yields an exact expression for the asymptotic localizable entanglement [10] (except for the technical issue that there one has an infinite number of parties, whereas here we considered only finite $m$). The concept of localizable entanglement was introduced in the context of quantum spin systems and allows the definition of a notion of entanglement length when these spins are part of a lattice with a given geometry. More precisely, consider the maximal bipartite entanglement that can be localized between two blocks of spins as a function of the distance between the blocks; typically this function is decaying exponentially with the distance, $\exp(-L/\xi)$, and the entanglement length is defined as the constant $\xi$ in this exponent. Theorem 3 gives the exact expression for the localizable entanglement between two blocks if asymptotically many realizations of these systems are available and *joint local* operations can be performed. If furthermore we are considering a state with infinitely many particles and translational symmetry (which is the usual case in condensed matter systems), then the strong subadditivity property of the von Neumann entropy enforces the entropy of a block of spins to grow when more spins are included in the block. It follows that the asymptotic localizable entanglement in such systems between two blocks is exactly given by the minimal entropy of these blocks, which proves that the upper bound given in [10] is actually the exact value for the localizable entanglement in the asymptotic limit. This is very surprising: the power of doing local asymptotic operations allows the distillation of entanglement between two blocks that are arbitrarily far from one another, and the rate at which this can be done is independent of the distance. This implies that any nontrivial ground state can be used as a perfect quantum repeater if many copies are available in parallel. The amount of entanglement that can be localized over these arbitrary distances is solely related to the entropy of a block of spins and not dependent on the distance. It is interesting to contrast the translationally invariant case to the one with random bond interactions [32]; in the latter case, the minimal entropy over all bipartite cuts will decrease algebraically with the distance between the blocks. This indicates that the entanglement in the case of random systems is essentially different than in the case of translationally invariant ones, something that is not revealed by looking at the entropy of a block of spins.

The problem of calculating the entropy of a block of spins has recently attracted a lot of attention in condensed matter physics [33], where it was shown that this entropy, in the case of ground states of one-dimensional systems, saturates to a finite value or increases logarithmically as a function of the size of the block, depending on whether the system is critical or not. The present work provides an operational meaning to these calculations in the sense of entanglement theory: this entropy quantifies the amount of entanglement that can be created at arbitrary distances if this ground state would be used as a quantum repeater. In higher-dimensional systems, the entropy of a block of spins grows as the boundary of that block, and therefore there is no bound on the amount of EPR pairs that could be localized between two far-away regions by doing joint local measurements on all the other spins; this is again the consequence of the fact that the asymptotic operations allow for perfect entanglement swapping in multipartite states.

## VII. ASYMPTOTIC NORMAL FORMS OF UNITAL AND GENERAL QUANTUM CHANNELS

Based on the well-known linear isomorphism between completely positive and trace-preserving maps and a set of quantum states [34]:

$$T:A \to B \Leftrightarrow \rho^{A'B} = \rho_T = (\mathrm{id} \otimes T)\phi^{A'A},$$

with a pure state $|\phi\rangle^{A'A}$ of Schmidt rank $d_A = \dim \mathcal{H}_A$ and the identity map id, we can interpret our findings in theorem 1 as statements on quantum channels. Note that $T$ is an isometry or unitary, if and only if the state $\rho_T$ is pure and the corresponding states of different isometries are equivalent to each other up to unitaries on the system $B$. We shall use this isomorphism in the following with a maximally entangled state $\Phi^{A'A}$ of Schmidt rank $d_A$, unless specified otherwise.

Let $T$ be a unital quantum channel on a system—i.e., mapping the identity on $A$ to the identity on $B$ (and assume input and output system to be of the same dimension $d = d_A = d_B$ for the moment). The corresponding state has the properties $\rho^B = \mathrm{Tr}_{A'} \rho_T = (1/d)\mathbb{1}$ and $\rho^{A'} = \mathrm{Tr}_B \rho_T = (1/d)\mathbb{1}$. Thanks to theorem 1, we know that in the asymptotic scenario the entanglement of assistance of $\rho_T$ is given by $\log d$.

Clearly, the unital channels (and, equally, the states with maximally mixed marginals as above) form a convex set, and the question of determining its extremal points has attracted quite some attention [35]. The classical analog of this problem is about doubly stochastic maps which, thanks to Birkhoff's theorem, are known to be exactly the convex combinations of permutations. For quantum doubly stochastic maps (another popular name for unital trace-preserving channels) the "obvious" generalization is wrong: there exist unital channels which are *not* convex combinations of unitaries. Under the Jamiołkowski isomorphism, this means that the state $\rho_T$ is not a convex combination of maximally entangled states; a specimen of this type we have actually studied in example 4.

However, theorem 1 points a way to resolving this unsatisfactory state of affairs in the asymptotic limit: since the asymptotic entanglement of assistance of $\rho_T$ is $\log d$, we can say that $\rho_T^{\otimes n}$ is well approximated by a convex combination of "almost" maximally entangled states in the sense that their entropies of entanglement are $n(\log d - \delta)$ for arbitrarily small $\delta > 0$ and sufficiently large $n$. We would like to deduce from this that $T^{\otimes n}$ is well approximated by a convex combination of unitaries (in the appropriate norm), but unfortunately the latter is really a stronger statement since it would give an approximation of $\rho_T^{\otimes n}$ by a convex combination of states that have high fidelity to some maximally entangled state. And that is not even mentioning the issues of the different norms to be used for comparing states and for channels.

Similarly, for a general channel and general $\phi^{A'A}$, the state $\rho_T^{\otimes n}$ can be restricted to the typical subspaces [36] of $(\rho_T^{A'})^{\otimes n}$ on Alice's side and of $(\rho_T^B)^{\otimes n}$ on Bob's side, without changing the state very much. This projected state resides in a $(D_A \times D_B)$-dimensional system, with $D_A \approx 2^{nS(A)}$ and $D_B \approx 2^{nS(B)}$. By theorem 1 it is well approximated by a convex combination of pure states with entanglement $n(\min\{S(A), S(B)\} - \delta)$, which again is too weak to say that the components have high fidelity with maximally entangled states.

Nevertheless, we take these observations as positive evidence for the following conjecture

*Conjecture 13*. Let $T$ be a unital quantum channel on a system or, more generally, a map $T:A \to B$ such that for all input states $\rho^A$, $S(\rho) \leq S(T(\rho))$. Then, for sufficiently large $n$, $T^{\otimes n}$ is arbitrarily well approximated by mixtures of isometries (unitaries in the unital case).

In general, $T^{\otimes n}$ is arbitrarily well approximated by mixtures of partial isometries between $A^n$ and $B^n$ (i.e., unitary transformations between subspaces of systems $A^n$ and $B^n$).

The appropriate distance measure for quantum channels $T$ and $T'$ to be used here is

$$\|T - T'\|_{\mathrm{cb}} = \max_\phi \|(\mathrm{id} \otimes T)\phi - (\mathrm{id} \otimes T')\phi\|_1,$$

the completely bounded norm (cb norm) [37].

In further support of this conjecture, we now outline a proof for a weaker version of it, where the comparison of $T^{\otimes n}$ and the mixture $T'$ of unitaries is done not in the worst case over all input states, but with respect to a single state $\phi^{\otimes n}$: we want $\|\rho_T^{\otimes n} - (\mathrm{id} \otimes T')\phi^{\otimes n}\|_1 \leq \epsilon$. The significance of such a statement is that if $\phi$ is a purification of a mixed state $\sigma$ on $A$ and $\{p_k, \phi_k\}$ is any source ensemble on $A^n$ with average $\sigma^{\otimes n}$, then the average error $\Sigma_k p_k \|T^{\otimes n}(\phi_k) - T'(\phi_k)\|_1$ is also bounded by $\epsilon$.

For simplicity, we assume the $S(\sigma)$ is strictly smaller than $S(T(\sigma))$. Note that one could always modify the channel trivially by padding the output with a sufficiently maximally mixed state, to enforce this condition.

We can write down a purification of $\rho_T$ in Schmidt form

$$|\psi\rangle^{A'BC} = \sum_j \sqrt{q_j} |\psi_j\rangle^{A'B} |j\rangle^C,$$

with orthogonal states $\{|j\rangle\}_j$ and $\{|\psi_j\rangle\}_j$. By assumption,

$$\chi_{A'} \coloneqq S(\sigma) - \sum_j q_j S(\psi_j^{A'}) < S(T(\sigma)) - \sum_j q_j S(\psi_j^B) \eqqcolon \chi_B,$$

so we can choose a number $R$ between these two values. Now we go through the random coding argument in the proof of theorem 1, but actually in the form of the second case considered in the proof of theorem 3. Since here we assume that we have uniform distribution on the $j$, there is no need to restrict to the set of typical sequences.

What we get are random codes of $N = 2^{nR}$ sequences $J = j_1, \ldots, j_n$, such that the corresponding states $\psi_{J(\beta)}^B$ (dropping superscript $n$ as before) form a good code for Bob. That means that for the superpositions

$$|\vartheta\rangle^{A'B} = \frac{1}{\sqrt{N}} \sum_\beta e^{-2\pi i \alpha\beta} |\psi_{J(\beta)}\rangle^B$$

(resulting from projecting the system $C$ onto a vector $|t_\vartheta\rangle$), we obtain, as at the end of the proof of theorem 3, that $\vartheta^{A'} \approx (1/N)\Sigma_\beta \psi_{J(\beta)}^{A'}$. And exactly as there, we can use lemma 12 to conclude that $\vartheta^{A'} \approx \sigma^{\otimes n}$ with respect to the trace distance. Both approximations in fact hold with high probability over the choice of the code. That means that there is a purification $|\zeta_\vartheta\rangle^{A'B}$ of $\sigma^{\otimes n}$ such that $\vartheta^{A'B} \approx \zeta_\vartheta^{A'B}$.

The connection to channels is now made by using the Jamiołkowski isomorphism in the other direction: for the well-behaved $\vartheta$ as above, there exists an isometry

$V_\vartheta : A \to B$ such that $|\zeta_\vartheta\rangle = (\mathbb{1} \otimes V_\vartheta)|\phi^{\otimes n}\rangle$, and hence our candidate mixture of unitaries is

$$T'(\varphi) = \sum_{\vartheta \text{ well behaved}} w_\vartheta V_\vartheta \varphi V_\vartheta^\dagger,$$

where the $w_\vartheta$ are probability weights. They are obtained as essentially $|{}^C\langle t_\vartheta | \psi^{\otimes n}\rangle^{A'BC}|^2$, normalized to the probability of the well-behaved set. It is then straightforward to verify that indeed $\|\rho_T^{\otimes n} - (\mathrm{id} \otimes T')\phi^{\otimes n}\|_1$ is small.

We want to close this section with a few comments on the difficulties encountered in the attempt to extend this argument to a proof of our conjecture. Clearly, the vectors $|t_\vartheta\rangle$ determine Kraus operators $D_\vartheta$ for $T^{\otimes n}$ via

$$^C\langle t_\vartheta | \psi^{\otimes n}\rangle^{A'BC} = (\mathbb{1} \otimes D_\vartheta)|\phi^{\otimes n}\rangle.$$

Because the cb-norm difference of $T^{\otimes n}$ and $T'$ can be upper bounded by $\Sigma_\vartheta \|D_\vartheta - V_\vartheta\|$ (with the operator norm $\|\cdot\|$), it is tempting to aim at making the latter quantity small. But relative to the fixed source $\sigma$, we can hope to make a statement about the difference $D_\vartheta - V_\vartheta$ only on the typical subspace of the source. However, even there our technique gives an approximation of the channel output only on the average—it is conceivable, and consistent with our result, that the *operator norms* of the $D_\vartheta - V_\vartheta$, when restricted to the typical subspace, are all large (which would say that the Kraus operators act very differently in the worst case). In addition, in the above proof we can make our statements about $\vartheta$ only "with high probability" and the probability distribution is also determined by the source $\sigma$.

## VIII. DISCUSSION

We have presented a class of very general procedures to distill singlets and quantum superposition ("cat") states, in both tripartite and multipartite settings. These procedures give universally the largest EPR rate distillable between any pair of parties in a multipartite state, when the other players cooperate. For three parties, this problem and its solution is equivalent to the previously considered entanglement of assistance. We have shown how GHZ (and higher cat state) distillation protocols can be constructed from common randomness distillation schemes by "coherification." It should be clear that a good number of variations of what we have shown here can be done. As a consequence, we could solve the problem of quantum channel coding with maximal classical help from the environment.

We stress that even though we look here at pure state transformations, we did not attempt "entanglement concentration," which is meant to generalize the asymptotic theory of bipartite pure states: there we have asymptotic reversibility, and demanding this leads to the hard MREGS problems. Instead, we do a "distillation" of specific states, embracing the possibility of irreversibility, but going for the maximum rate. We think that understanding these problems will remain central even assuming the availability of a complete MREGS.

Thus, starting from the strange entanglement of the assistance problem, we discovered a great number of highly interesting results of multiparty entanglement processing. These also shed some light on issues like the entanglement length in spin chains. Perhaps even more important are the conceptual insights regarding possible asymptotic normal forms of quantum channels as mixtures of partial isometries. Finally, we want to mention a spin-off in quite another direction: based on the techniques of Sec. VI and developing them further, the problem of *distributed quantum data compression* (with unlimited classical side communication) could be solved in [13]. The methods of that paper also simplify some of our arguments regarding EPR distillation (they do not apply to GHZ distillation, however) and allow us to prove the equality in theorem 3 for all $m$.

## ACKNOWLEDGMENTS

We wish to thank Charles H. Bennett, Ignacio Cirac, Igor Devetak, Mark Fannes, Michał Horodecki, Debbie Leung, Jonathan Oppenheim, and Tobias Osborne for interesting discussions on the entanglement of assistance and unital channels and especially Berry Groisman, Noah Linden, and Sandu Popescu for sharing their results in [4] prior to publication.

J.A.S. acknowledges the support of the NSA and ARO under Contract No. DAAD19-01-C-0056. A.W. is supported by the EU project RESQ (Contract no. IST-2001-37559) and by the U.K. Engineering and Physical Sciences Research Council's "IRC QIP." The hospitality of the Isaac Newton Institute of Mathematical Sciences, Cambridge, during the topical semester on Quantum Information Sciences (16/08-17/12 2004) is gratefully acknowledged by J.A.S. and A.W. F.V. acknowledges support by the Gordon and Betty Moore Foundation (the Information Science and Technology Initiative, Caltech).

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

[2] C. H. Bennett, S. Popescu, D. Rohrlich, J. A. Smolin, and A. V. Thapliyal, Phys. Rev. A **63**, 012307 (2000).

[3] N. Linden, S. Popescu, B. Schumacher, and M. Westmoreland, e-print quant-ph/9912039.

[4] B. Groisman, N. Linden, and S. Popescu, e-print quant-ph/0508001, Phys. Rev. A (to be published).

[5] O. Cohen, Phys. Rev. Lett. **80**, 2493 (1998).

[6] D. P. DiVincenzo, C. A. Fuchs, H. Mabuchi, J. A. Smolin, A. V. Thapliyal, and A. Uhlmann, in *Proceedings of Quantum Computing and Quantum Communications: First NASA International Conference, Palm Springs, 1998*, *Springer Lecture Notes in Computer Science*, Vol. 1509 (Springer, Heidelberg, 1999), pp. 247–257.

[7] T. Laustsen, F. Verstraete, and S. J. van Enk, Quantum Inf. Comput. **3**, 64 (2003).

[8] A. Acín, E. Jané, W. Dür, and G. Vidal, Phys. Rev. Lett. **85**,

4811 (2000); W. Dür, G. Vidal, and J. I. Cirac, Phys. Rev. A **62**, 062314 (2000).
[9] O. Cohen and T. A. Brun, Phys. Rev. Lett. **84**, 5908 (2000); T. A. Brun and O. Cohen, Phys. Lett. A **281**, 88 (2001).
[10] F. Verstraete, M. Popp, and J. I. Cirac, Phys. Rev. Lett. **92**, 027901 (2004); F. Verstraete, M. A. Martin-Delgado, and J. I. Cirac, *ibid.* **92**, 087201 (2004).
[11] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[12] P. Hayden, M. Horodecki, and B. M. Terhal, J. Phys. A **34**, 6891 (2001).
[13] M. Horodecki, J. Oppenheim, and A. Winter, Nature (London) **436**, 673 (2005).
[14] E. Schrödinger, Proc. Cambridge Philos. Soc. **32**, 446 (1936); L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A **183**, 14 (1993).
[15] F. Yura, J. Phys. A **36**, L237 (2003).
[16] M. Fannes, Commun. Math. Phys. **31**, 291 (1973).
[17] A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998); B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997).
[18] I. Devetak and A. Winter, Proc. R. Soc. London, Ser. A **461**, 207 (2005).
[19] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).
[20] M. Gregoratti and R. F. Werner, J. Mod. Opt. **50**, 913 (2003).
[21] P. Hayden and C. King, e-print quant-ph/0409026.
[22] B. Schumacher, Phys. Rev. A **54**, 2614 (1996); B. Schumacher and M. A. Nielsen, *ibid.* **54**, 2629 (1996).
[23] I. Devetak, IEEE Trans. Inf. Theory **51**, 44 (2005).
[24] S. Lloyd, Phys. Rev. A **55**, 1613 (1997).
[25] P. W. Shor (unpublished). Online at http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/; MSRI Workshop on Quantum Information, Berkeley, 2002.
[26] I. Devetak and A. Winter, Phys. Rev. Lett. **93**, 080501 (2004).
[27] I. Devetak and A. Winter, IEEE Trans. Inf. Theory **50**, 3183 (2004).
[28] G. Vidal, W. Dür, and J. I. Cirac, Phys. Rev. Lett. **89**, 027901 (2002).
[29] W. K. Wootters, Phys. Rev. Lett. **80**, 2245 (1998).
[30] A. V. Thapliyal and J. A. Smolin, Phys. Rev. A **68**, 062324 (2003).
[31] R. Ahlswede and A. Winter, IEEE Trans. Inf. Theory **48**, 569 (2002).
[32] G. Refae and J. E. Moore, Phys. Rev. Lett. **93**, 260602 (2004).
[33] C. Holzhey, F. Larsen, and F. Wilczek, Nucl. Phys. B **424**, 443 (1994); I. Peschel, M. Kaulke, and O. Legeza, Ann. Phys. **8**, 153 (1999); G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, Phys. Rev. Lett. **90**, 227902 (2003); B.-Q. Jin and V. E. Korepin, J. Stat. Phys. **116**, 79 (2004); P. Calabrese and J. Cardy, J. Struct. Mech. P06002 (2004).
[34] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).
[35] L. J. Landau and R. F. Streater, Linear Algebr. Appl. **193**, 107 (1993); K. R. Parthasarathy, e-print quant-ph/0307182; O. Rudolph, J. Math. Phys. **45**, 4035 (2004).
[36] B. Schumacher, Phys. Rev. A **51**, 2738 (1995); R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[37] V. I. Paulsen, *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, Cambridge, England, 2002).