# Remote Preparation of Quantum States

Charles H. Bennett, Patrick Hayden, Debbie W. Leung, Peter W. Shor, and Andreas Winter

***Abstract*—Remote state preparation is the variant of quantum state teleportation in which the sender knows the quantum state to be communicated. The original paper introducing teleportation established minimal requirements for classical communication and entanglement but the corresponding limits for remote state preparation have remained unknown until now: previous work has shown, however, that it not only requires less classical communication but also gives rise to a tradeoff between these two resources in the appropriate setting. We discuss this problem from first principles, including the various choices one may follow in the definitions of the actual resources.**

**Our main result is a general method of remote state preparation for arbitrary states of many qubits, at a cost of 1 bit of classical communication and 1 bit of entanglement per qubit sent. In this "universal" formulation, these ebit and cbit requirements are shown to be simultaneously optimal by exhibiting a dichotomy. Our protocol then yields the exact tradeoff curve for memoryless sources of pure states (including the case of incomplete knowledge of the ensemble probabilities), based on the recently established quantum-classical tradeoff for visible quantum data compression. A variation of that method allows us to solve the even more general problem of preparing entangled states between sender and receiver (i.e., purifications of mixed state ensembles).**

**The paper includes an extensive discussion of our results, including the impact of the choice of model on the resources, the topic of obliviousness, and an application to private quantum channels and quantum data hiding.**

***Index Terms*—Cryptography, entanglement, large deviations, teleportation, tradeoff.**

## I. Introduction

TELEPORTATION [5] implements the transmission of a quantum bit (1 qubit) by sending two classical bits (2 cbits), while using up quantum correlation amounting to one bit of entanglement (1 ebit)—although a description of this state would require an infinite number of cbits, even when assisted by unlimited classical correlation. What is more, in teleportation this description is not needed at all: both the Sender and the Receiver act physically on the state (i.e., by quantum operations: completely positive and trace-preserving linear maps), and the process can be used to transmit parts of entangled states faithfully. This and the phenomenon of dense coding [4] prove that one cannot do with less than these resources: both 2 cbits and 1 ebit are necessary.

However, allowing the Sender knowledge of the state to be communicated changes the task to what is now known as remote state preparation (r.s.p.) [36], [37], [43], and here two new phenomena occur: in [7] it is shown that at the cost of possibly spending more entanglement one can reduce the classical communication to 1 cbit per qubit in the asymptotics; and there is a tradeoff between the classical and the quantum resources needed, of which [7] and [22] provide bounds. In the present work, we put these results into their definite form by proving a formula for the exact tradeoff curve and by improving on the result of [7] to use only 1 cbit and 1 ebit per qubit.

By a protocol for r.s.p. we shall mean a procedure involving two parties, a Sender who is given a description of a state $\rho \in \mathcal{X} \subset \mathcal{S}(\mathcal{K})$ from a subset $\mathcal{X}$ of the state set $\mathcal{S}(\mathcal{K})$ of the Hilbert space $\mathcal{K}$ and a Receiver, who have access to a number of resources (both forward and backward classical communication, entanglement, shared randomness, or others). The protocol prescribes how to use these in a sequence of steps (based on the previous exchange of messages in the protocol, and on $\rho$ for the Sender), resulting in a state $\widetilde{\rho}$ held by the Receiver. The dimension $D = \dim \mathcal{K}$ will, in the entire following discussion be the principal asymptotic parameter (i.e., one should think of it as large).

We shall say that the protocol is *(deterministic) exact* if $\widetilde{\rho} = \rho$ for all choices of $\rho \in \mathcal{X}$.

It is said to have *fidelity* $F$ if for all $\rho \in \mathcal{X}$, $F(\widetilde{\rho}, \rho) \geq F$, with the mixed-state fidelity [32], [40]

$$F(\rho, \sigma) = \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1^2 = \mathrm{Tr} \left( \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right)^2 .$$

(Note that for $F = 1$ this is the same as an exact protocol.)

A notion in between these two is a *probabilistic exact* protocol *with error* $\epsilon$: this means that the protocol additionally produces a flag, accessible to both Sender and Receiver, which indicates "success" or "failure" such that for all $\rho \in \mathcal{X}$, $\Pr\{\text{"filure"}\} \leq \epsilon$ and $\widetilde{\rho} = \rho$ if the flag is "success"; $\widetilde{\rho}$ is arbitrary otherwise. (Note that such a protocol automatically has fidelity $\geq 1 - \epsilon$.)

Manuscript received August 22, 2003; revised June 7, 2004. The work of C. H. Bennett was supported by the U.S. National Security Agency and Advanced Research and Development Activity under Contracts DAAD19-01-1-06 and DAAD19-01-C-0056. The work of P. Hayden was supported by the Sherman Fairchild Foundation and the U.S. National Science Foundation under Grant EIA-0086038. The work of D. W. Leung was supported by the Richard C. Tolman Endowment Fund, the Croucher Foundation, and the U.S. National Science Foundation under Grant EIA-0086038. A. Winter was supported by the U.K. Engineering and Physical Sciences Research Council.

C. H. Bennett is with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: bennetc@watson.ibm.com).

P. Hayden and D. W. Leung are with the Institute for Quantum Information, Caltech 107-81, Pasadena, CA 91125 USA (e-mail: patrick@cs.caltech.edu; wcleung@cs.caltech.edu).

P. W. Shor was with AT&T Labs, Florham Park, NJ 07922 USA. He is now with the Department of Mathematics, the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: shor@math.mit.edu).

A. Winter is with the Department of Computer Science, University of Bristol, Bristol BS8 1UB, U.K (e-mail: a.j.winter@bris.ac.uk).

Communicated by E. Knill, Associate Editor for Quantum Information Theory.

Digital Object Identifier 10.1109/TIT.2004.839476

Sometimes we want to impose a probability distribution $P$ on $\mathcal{X}$ and we will also consider protocols which have *average fidelity* $\overline{F}$, meaning

$$\int \mathrm{d}P(\rho)F(\widetilde{\rho},\rho) \geq \overline{F}.$$

Varied as the parameters by which we judge the quality of a protocol are, so are the ways to account for the use of resources: we will come back to this issue later (Section VI-A), though the following example features not only various quality measures, but also some choices of resource accounting. For the moment, we think only about protocols which terminate at a certain prescribed point and the resources are those needed to get to this point in the worst case $\rho \in \mathcal{X}$.

*Example 1 (Column Method [7]):* The Sender is given an arbitrary pure state $\psi = |\psi\rangle\langle\psi|$ (note that we use *state* synonymous with *density operator*; if we want to denote a state vector it will be $|\psi\rangle$) on a $D$-dimensional space (in [7] $D = 2^n$, i.e., $n$ qubits), and that Sender and Receiver share sufficiently many maximally entangled states

$$|\Phi_D\rangle = \frac{1}{\sqrt{D}} \sum_{j=1}^{D} |j\rangle|j\rangle$$

of Schmidt rank $D$, labeled $1, 2, \ldots, K$.

The Sender performs the measurement

$$(A_0 = \overline{\psi}, A_1 = \mathbb{1} - \overline{\psi})$$

on each of the $K$ entangled states and records the outcome. Here $\bar{\cdot}$ denotes the complex conjugation with respect to the basis $\{|j\rangle\}$ used to define $\Phi_D$. The probability of a 0 clearly is $\frac{1}{D}$, the probability of a 1 is $1 - \frac{1}{D}$, hence, the probability of $K$ 1's in a row (this will be called "failure") is

$$\Pr\{\text{"failure"}\} = \left(1 - \frac{1}{D}\right)^K \leq \exp\left(-\frac{K}{D}\right).$$

Thus, if $\log K \geq \log D + \log\log\frac{1}{\epsilon}$, "failure" occurs with probability at most $\epsilon$. If this does not happen, there is at least one 0 in the measurement results, and it requires $\log K$ cbits to communicate the label of the entangled state where it occurred to the Receiver. For definiteness, let us say that the Sender selects one position of outcome 0 at random. Simple algebra shows that in this case, the Receiver's reduced state is just $\psi$.

This is an example of a probabilistic exact protocol with asymptotic cost of classical communication of 1 cbit per qubit and success probability $1 - \epsilon$. By ignoring the possibility of failure, it becomes a fidelity $1 - \epsilon$ protocol. The protocol requires $K \log D$ ebits, which is exponential in the number of qubits. Most of this, however, can be recovered ("recycled") using back communication after completion of the remote state preparation (see [7]) such that only $O(\log D)$ ebits are irrecoverably lost.

Clearly, to make this method deterministic exact, one must not put a limit on the number of trials $K$ (in which case the communication cost becomes infinite), or we must allow for a deterministic exact procedure in the case of "failure," e.g., teleportation. As this will increase the worst case communication cost to 2 cbits per qubit, we are motivated to also consider *expected* cbit cost, which in this example is $1 + 2\epsilon$ per qubit.

As an aside to this exposition, one can also consider making the task easier for the Receiver, by only requiring that he is able to simulate any measurement of which he is given a description, performed on the state of which the Sender is given a description: this is known as *classical teleportation* [14], and though it is related to our subject it lies outside the scope of the present paper.

The organization of the rest of the paper is as follows: in Section II, we present a general method of remote state preparation, which uses 1 cbit and 1 ebit per qubit asymptotically. It is based on an efficient state randomization method (see also [9]). In Section III, it is shown that any universal high-fidelity protocol has to use 1 cbit and 1 ebit per qubit, asymptotically. The cbit bound is true even if unlimited quantum back communication is allowed, and the ebit bound is proved even in the presence of shared randomness. We proceed to derive the exact tradeoff curve between ebits and cbits for an arbitrary ensemble of candidate states, in Section IV, using the recently established analogous but simpler tradeoff in quantum data compression between qubits and cbits [30]. Section V discusses the corresponding result if ensembles of pure entangled states between the Sender and the Receiver are to be prepared: again, we can prove the exact tradeoff between ebits and cbits.

We conclude with a discussion of our findings and open questions in Section VI: in particular, considerations of the issue of *obliviousness* (cf. [35]) and a discussion of the impact of certain slight changes in the model on our conclusions.

Several appendices contain separate or more technical issues: in Appendix A, facts about Gaussian distributed vectors are related; Appendix B contains the proofs for the central technical result, the state randomization; in Appendix C, it is shown that universal description of quantum states by qubits and cbits exhibits only a trivial tradeoff between the resources: there is a dichotomy between full quantum with no classical information and no quantum with infinite classical information. Facts about typical subspaces, used in various proofs, are collected in Appendix D. Appendix E contains thoughts on further operational links between the qubit/cbit and the ebit/cbit tradeoff, based on a conjecture on the compressibility of mixed-state sources. Finally, in Appendix F, miscellaneous proofs are collected.

Global notation conventions are as follows: we use $^*$ for the Hermitian adjoint, $^\top$ for the transpose (in some given basis); exp and log are to basis 2 (for the natural basis we use $e$, and the natural logarithm is denoted ln).

## II. Universal r.s.p.: 1 cbit + 1 ebit $\succ$ 1 qubit

We begin with a result on universal (approximate) state randomization by unitaries.

*Theorem 2:* For Hilbert space $\mathcal{H}$ of dimension $D$ and $\epsilon > 0$ there exist

$$K \leq \left(\frac{10}{\epsilon}\right)^2 D \log\left(\frac{20D}{\epsilon}\right)$$

unitaries $U_k$ on $\mathcal{H}$ such that for every state $\varphi$

$$\frac{1}{K}\sum_{k=1}^{K} U_k \varphi U_k^* \in \left[\frac{1-\epsilon}{D}\mathbb{1}; \frac{1+\epsilon}{D}\mathbb{1}\right] \tag{1}$$

where the closed interval to the right refers to the operator order.

*Proof:* Select the unitaries independently at random from the Haar measure on the unitary group. Observe that (1) says that for all pure states $\varphi$ and $\psi$

$$\left|\frac{1}{K}\sum_{k=1}^{K} \mathrm{Tr}\,(U_k \varphi U_k^* \psi) - \frac{1}{D}\right| \leq \frac{\epsilon}{D}.$$

Fix an $\frac{\epsilon}{4D}$-net $\mathcal{M}$, according to Lemma 4. Lemma 3 below allows us to bound

$$\Pr\left\{\exists \varphi,\psi \in \mathcal{M} \left|\frac{1}{K}\sum_{k=1}^{K} \mathrm{Tr}\,(U_k \varphi U_k^* \psi) - \frac{1}{D}\right| > \frac{\epsilon}{2D}\right\} \leq 2\left(\frac{20D}{\epsilon}\right)^{4D} \exp\left(-K\frac{\epsilon^2}{24}\right).$$

With triangle inequality for the trace norm we finally get

$$\Pr\left\{\exists \varphi,\psi \left|\frac{1}{K}\sum_{k=1}^{K} \mathrm{Tr}\,(U_k \varphi U_k^* \psi) - \frac{1}{D}\right| > \frac{\epsilon}{D}\right\} \leq 2\left(\frac{20D}{\epsilon}\right)^{4D} \exp\left(-K\frac{\epsilon^2}{24}\right),$$

so if $K$ is as large as stated in the theorem there exist $U_1,\ldots,U_K$ such that (1) is true. □

The probabilistic and geometrical facts used in the above proof are contained in the following lemmas. The first is applied in the above proof with $p = 1$ but the general version is used later on.

*Lemma 3:* Let $\varphi$ be a pure state, $P$ a rank-$p$ projector, and let $(U_k)_{k=1}^K$ be an independent and identically distributed (i.i.d.) sequence of $U(D)$-valued random variables, distributed according to Haar measure. Then, for $0 < \epsilon \leq 1$

$$\Pr\left\{\left|\frac{1}{K}\sum_{k=1}^{K} \mathrm{Tr}\,(U_k \varphi U_k^* P) - \frac{p}{D}\right| \geq \frac{\epsilon p}{D}\right\} \leq 2\exp\left(-Kp\frac{\epsilon^2}{6}\right).$$

*Proof:* In Appendix B. □

*Lemma 4:* Let $\mathcal{H}$ be a Hilbert space of dimension $D$. Then there exists, for every $\epsilon > 0$, a set $\mathcal{M}$ of pure state vectors in $\mathcal{H}$ of cardinality

$$|\mathcal{M}| \leq \left(\frac{5}{\epsilon}\right)^{2D}$$

such that for every state vector $|\varphi\rangle \in \mathcal{H}$ there exists a state vector $|\psi\rangle \in \mathcal{M}$ such that

$$\|\varphi - \psi\|_1 \leq 2\sqrt{1 - F(\varphi,\psi)} \leq 2\||\varphi\rangle - |\psi\rangle\|_2 \leq \epsilon.$$

Such a set $\mathcal{M}$ we call *$\epsilon$-net*.

*Proof:* In Appendix B. □

A few words of interpretation: it is known [2], [13] that if $\epsilon = 0$, one needs $K \geq D^2$, and this is tight as the example of the generalized Pauli (sometimes called Weyl) operators shows. We call a selection of unitaries as in the theorem "randomising," because application of a randomly chosen $U_k$ results in an almost maximally mixed state. Clearly, this has cryptographic applications, an exploration of which is to be found in our separate paper [9].

Let us now show how to use this result to build a remote state preparation protocol: first of all, given a pure state $\psi$, one can write down the family of operators

$$A_k = \frac{D}{K(1+\epsilon)} U_k \overline{\psi} U_k^* \qquad (k = 1,\ldots,K)$$

$$A_{\text{failure}} = \mathbb{1} - \sum_{k=1}^{K} A_k.$$

This is a positive operator valued measure (POVM) by virtue of Theorem 2.

**Protocol Π (Description of $\psi$ at the Sender)**:

1) The Sender measures the POVM $(A_k)$ of the above description on her half of the entangled state $\Phi_D$ and announces the result (either "failure" or $k = 1,\ldots,K$).
2) If the message received is not "failure," say $k$, the Receiver applies the unitary $U_k^\top$ to his part of the state $\Phi_D$.

*Theorem 5:* The above protocol realizes remote state preparation for an arbitrary state $|\psi\rangle \in \mathcal{K}$ exactly with a probability of failure of exactly $\frac{\epsilon}{1+\epsilon} \leq \epsilon$.

In particular, exact probabilistic r.s.p. with error $\epsilon$ is possible using

$$\log D + 2\log\frac{10}{\epsilon} + \log\log\frac{20D}{\epsilon} \text{ cbits}$$

and

$$\log D \text{ ebits}.$$

*Proof:* It is straightforward to check that the protocol, in case it does not produce a failure, exactly prepares $|\psi\rangle$ at the Receiver.

For the probability assertions: the event $k$ of the POVM $(A_k)$ is triggered with probability exactly $\frac{1}{K(1+\epsilon)}$. Hence, the probability of failure is

$$1 - K\frac{1}{K(1+\epsilon)} = 1 - \frac{1}{1+\epsilon} = \frac{\epsilon}{1+\epsilon}.$$

The remaining claims are easy consequences of this. □

*Corollary 6:* Probabilistic exact remote state preparation is possible with 1 cbit and 1 ebit per qubit, asymptotically. □

## III. Optimality of cbit and ebit Resources

We will now show that both 1 ebit and 1 cbit per qubit are necessary asymptotically for universal r.s.p. protocols with high fidelity. More precisely, we assume a protocol like our protocol Π in Section II, which takes as input the description of an arbitrary state $\psi$ on a $D$-dimensional space $\mathcal{K}$, uses an entangled state of Schmidt rank $S$, forward communication of one out of $K$ messages, such that the output states $\widetilde{\rho}$ have fidelity $F$ to the ideal $\psi$.

Regarding the communication resources, causality shows that $K \geq FD$ is necessary, even if unlimited *quantum* back communication is allowed: this is because the mere capability to remotely prepare an orthogonal basis of states with fidelity $F$ clearly allows the Sender to transmit one out of $D$ classical messages with probability at least $F$ of correct decoding. Imagine now that Sender and Receiver follow the r.s.p. protocol with the

modification that each forward communication is skipped and replaced by the Receiver guessing it at random.

In this modification of the protocol, the probability of correct decoding clearly is $\geq \frac{F}{K}$, as the Receiver has only to guess the correct classical communication out of $K$. But the modified protocol involves no forward transmission at all, hence the probability of correctly identifying the Sender's message—1 out of $D$—is $\leq \frac{1}{D}$: this shows $\frac{1}{D} \geq \frac{F}{K}$.

We have thus proved the following.

*Theorem 7:* Any r.s.p. protocol with fidelity $F$ requires classical communication of

$$C = \log K \geq \log D + \log F \text{ cbits}$$

even if unlimited quantum back communication is allowed. □

Regarding the entanglement, we have the following result of an extremely strong dichotomy.

*Theorem 8:* Any r.s.p. protocol using an entangled state of Schmidt rank $S \leq qD$ $(q < F)$ requires classical communication of

$$C \geq \frac{q(1-q)}{6} D - O(\log D) \text{ cbits}$$

even if unlimited shared randomness is available.

On the other hand, there is a protocol with fidelity $F \geq 1-\epsilon$, which uses no entanglement at all (i.e., $S = 1$), and classical communication of

$$C \leq \left(4 + \log \frac{1}{\epsilon}\right) D \text{ cbits.}$$

Thus, in the asymptotic limit, and with normalized resources $E = \log S / \log D$ and $R = C / \log D$ for the entanglement and communication rates, the rate point $E = 1$ marks the threshold between two drastically different regimes: for $E \geq 1$, the classical communication rate $R = 1$ is sufficient by Corollary 6 and necessary by Theorem 7. For any entanglement rate $E < 1$, Theorem 8 shows that no finite classical communication rate is possible: $R \to \infty$ with $D \to \infty$. Thus, $E \geq 1$ and $R \geq 1$ hold simultaneously and both equalities can be achieved at the same time (Theorem 5), unless $R = \infty$ in which case $E = 0$, i.e., there is only a trivial tradeoff between ebits and cbits.

*Proof of Theorem 8:* Consider any protocol, using a shared random variable $\nu$, so that the output state $\widetilde{\rho}$ is the mixture of the output states for the various values of $\nu$. Such a protocol clearly has average fidelity $\overline{F} \geq F$, with respect to the uniform (i.e., unitarily invariant) distribution on the pure states

$$\overline{F} = \int \mathrm{d}|\psi\rangle F(\widetilde{\rho}, |\psi\rangle\langle\psi|).$$

Because of the linearity of the pure state fidelity in $\widetilde{\rho}$, $\overline{F}$ is the probabilistic average of the fidelities $\overline{F}_\nu$ of the protocol for the value $\nu$ of the shared random variable. Hence, there exists a $\nu$ such that $\overline{F}_\nu \geq F$, and we can consider a new protocol, without shared randomness, which has the same fidelity as the original.

Thus, without loss of generality (w.l.o.g.), we may assume a protocol of the form described in the first paragraph of this section, which uses only the entangled state $\Phi$ and forward classical communication. In general terms, it proceeds by the Sender performing a measurement on her half of $\Phi$ and communicating the outcome $m$ to the Receiver, who then applies a quantum operation $T_m$ to his half of $\Phi$. Observe that after the Sender's measurement, the state of the Receiver is collapsed to a state supported on the support of the restriction of $\Phi$, which is a space of dimension $S$. Thus, effectively, the Sender supplies the Receiver with a message $m$ and a state $\xi$ on an $S$-dimensional system, from the combination of which an approximation of $\psi$ is obtained: $\widetilde{\rho} = T_m(\xi)$. Once more, using bilinearity of the pure state fidelity, we may assume that the choice of the pair $(m, \xi)$ from $\psi$ is deterministic, and that $\xi$ is a pure state. (This no longer describes an r.s.p. protocol, where uncontrollable randomness due to measurements is the rule: what is important here is that this can only enhance the capabilities of the Sender.)

We now invoke Theorem 24 from Appendix C, which lower-bounds the classical communication cost of such a quantum-classical state description: we obtain

$$C \geq \frac{q(1-q)}{6} D - O(\log D)$$

which is our claim.

Conversely, in the situation with no entanglement, pick an $\sqrt{4\epsilon}$-net $\mathcal{M}$ of cardinality at most $(\frac{5}{2\sqrt{\epsilon}})^{2D}$, according to Lemma 4. Clearly, a valid protocol is one that follows.

Given a state description of $\psi$, the Sender picks a $|\phi\rangle \in \mathcal{M}$ with fidelity $1-\epsilon$ (because Lemma 4 is strong enough for that) to $\psi$, and sends the Receiver an identifier for $\phi$, which requires $\log |\mathcal{M}|$ cbits. □

## IV. Ensemble Tradeoff Curve

While in the previous sections we considered universal r.s.p. (even though asymptotic, allowing *any* input state), in the present section as well as in Section V we want to look at *ensemble asymptotics*: we consider an ensemble of quantum states $\mathcal{E} = \{|\psi_i\rangle, p_i\}$ on the Hilbert space $\mathcal{H}$ of dimension $d$, and are interested in r.s.p. of the ensemble $\{|\psi_I\rangle, p_I\}$ on $\mathcal{H}^{\otimes n}$, with states and probabilities

$$\begin{aligned} |\psi_I\rangle &= |\psi_{i_1}\rangle \otimes \cdots \otimes |\psi_{i_n}\rangle \\ p_I &= p_{i_1} \ldots p_{i_n} \\ I &= i_1 \ldots i_n \end{aligned}$$

and for large $n$. The notation for letters (lower case) and blocks (upper case) is used throughout this and in Section V.

Note that even in the case that the ensemble contains *all* pure states on $\mathcal{H}$, the asymptotics will capture only the product states in $\mathcal{K} = \mathcal{H}^{\otimes n}$, unlike the model of the previous sections.

We shall be interested in protocols which have average fidelity $\overline{F}$, i.e.,

$$\sum_I p_I \mathrm{Tr}(|\psi_I\rangle\langle\psi_I|\widetilde{\rho}_I) \geq \overline{F}. \tag{2}$$

By the monotonicity of the fidelity under partial traces, this implies the weaker condition

$$\sum_I p_I \frac{1}{n} \sum_{k=1}^{n} \mathrm{Tr}\left(|\psi_{i_k}\rangle\langle\psi_{i_k}| \mathrm{Tr}_{\neq k} \widetilde{\rho}_I\right) \geq \overline{F} \tag{3}$$

which we will find useful at times.

Note that by considering average fidelities as we do here, shared randomness becomes automatically useless, because we aim to prepare pure states with high fidelity (compare the proof of Theorem 8).

On block length $n$, a protocol for r.s.p. uses a maximally entangled state $\Phi_D$ of Schmidt rank $D$ shared between Sender (A) and Receiver (B). We consider here protocols which use only forward communication: their general form is described by a measurement POVM depending on $I$, $\boldsymbol{M}^I = \left(M_j^I\right)_j$ with $j$ running over a set $\{1,\dots,K\}$: after performing this POVM on her half of $\Phi_D$, the Sender communicates $j$, and the Receiver applies a quantum operation $T_j$ to his half of $\Phi_D$. We write $(\boldsymbol{M},T)$ to denote such a protocol, sometimes adding a subscript $n$ to indicate the block length.

The resources used are defined, in a way similar to [30], as the entanglement rate

$$\operatorname{esupp}(\boldsymbol{M},T) := \frac{1}{n}\log D,$$

and the communication rate

$$\operatorname{csupp}(\boldsymbol{M},T) := \frac{1}{n}\log K.$$

(The notation is meant to remind one of "support," since what we count here is the number of bits necessary to support the entanglement and the classical messages, respectively.) We say that a rate pair $(R,E)$ is *achievable* if for all $\epsilon,\delta>0$ there exists $n_0$ such that for all $n\geq n_0$ there are r.s.p. protocols $(\boldsymbol{M},T)_n$ with fidelity $1-\epsilon$ and resources

$$\begin{aligned}\operatorname{csupp}(\boldsymbol{M},T) &\leq R+\delta\\ \operatorname{esupp}(\boldsymbol{M},T) &\leq E+\delta.\end{aligned}$$

This allows us to rigorously define the tradeoff function $E^*$ by

$$E^*(R) = \min\{E|(R,E) \text{ is achievable}\}.$$

A similar tradeoff is studied in [30] between cbits and transmitted qubits instead of ebits, which is a visible coding generalization of the familiar Schumacher quantum data compression [34], [38]: such a protocol consists of a pair $(E_n,D_n)$ of encoding and decoding maps. The encoding takes $I$ to a combination of a quantum message supported on $n\operatorname{qsupp}(E_n,D_n)$ qubits and a classical message comprising $n\operatorname{csupp}(E_n,D_n)$ cbits, while the decoding is a quantum operation acting on these two, with the aim as before, to achieve a large average input–output fidelity.

Defining achievable rate pairs $(R,Q)$ analogous to the above, and letting

$$Q^*(R) = \min\{Q|(R,Q) \text{ is achievable}\}$$

we have the following single-letter formula for the quantum-classical tradeoff (q.c.t.) curve.

*Theorem 9 (Hayden, Jozsa, and Winter) [30]:*

$$Q^*(R) = M(\mathcal{E},R) := \min\{S(A:B|C)|S(A:C)\leq R\} \tag{4}$$

where the minimization is over all tripartite states

$$\omega = \sum_i p_i|i\rangle\langle i|^A \otimes \psi_i^B \otimes \sum_j p(j|i)|j\rangle\langle j|^C \tag{5}$$

for stochastic matrices $p(j|i)$; $j$ has a range of at most $m+1$ if the ensemble consists of $m$ states.

$$\begin{aligned}S(A:C) &= S(A)+S(C)-S(AC) \text{ and}\\ S(A:B|C) &= S(AC)+S(BC)-S(ABC)-S(C)\end{aligned}$$

are the (conditional) quantum mutual information, defined via the von Neumann entropy $S$, referring implicitely to the state $\omega$: $S(AC)$ is the von Neumann entropy of $\omega$ restricted to $AC$, etc. □

In brief, once an optimal channel $p(j|i)$ is chosen, the scheme essentially works as sending part of the classical encoding $J=j_1\dots j_n$ (only typical) using the Reverse Shannon Theorem [10], and then Schumacher-compressing the induced "conditional" ensemble

$$\begin{aligned}\{\psi_I, q(I|J) &= q(i_1|j_1)\dots q(i_n|j_n)\} \quad \text{with}\\ q(i|j) &= \left(\sum_i p_i p(j|i)\right)^{-1} p_i p(j|i)\end{aligned}$$

to its von Neumann entropy (note that the ensemble is a product of independent ensembles even though they are not all identical).

For each point $(R,Q)$ on the tradeoff curve for the ensemble $\mathcal{E}$ we can, with the method of the previous section, construct an asymptotic and approximate r.s.p. protocol using $C=R+Q$ cbits and $E=Q$ ebits: We only have to use Theorem 5 to remotely prepare the encoded state on $Q$ qubits, using $Q$ ebits and an additional $Q$ cbits, all per qubit.

We can summarize the finding as an upper bound on $E^*(R)$, in a strange implicit form

$$E^*(R+Q^*(R)) \leq Q^*(R). \tag{6}$$

*Remark 10:* Devetak and Berger [22] happened to parametrize the q.c.t. curve for the uniform qubit ensemble. Using teleportation instead of our Theorem 5 they obtained r.s.p. protocols using $C+2Q$ cbits and $Q$ ebits.

Using the chain rule $S(A:B|C)+S(A:C)=S(A:BC)$, we can put together Theorem 9 and (6) to obtain that

$$E^*(R) \leq \min\{S(A:B|C)|S(A:BC)\leq R\}.$$

In fact, we shall show in a moment that equality holds here.

*Theorem 11:*

$$E^*(R) = N(\mathcal{E},R) := \min\{S(A:B|C)|S(A:BC)\leq R\} \tag{7}$$

where the minimization is over all tripartite states $\omega$ as in (5).

Before we prove this, we state a little lemma collecting some properties of $N$.

*Lemma 12:* $N$ is convex, continuous and strictly decreasing in the interval where it takes finite positive values, which is

$[S(B); S(A)]$. It obeys the following additivity relation for ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$:

$$N(\mathcal{E}_1 \otimes \mathcal{E}_2, R) = \min\{N(\mathcal{E}_1 R_1) + N(\mathcal{E}_2, R_2) | R_1 + R_2 = R\}. \quad (8)$$

*Proof:* In Appendix F. □

*Proof of Theorem 11:* Only the direction "$\geq$" has to be proved: assume an r.s.p. protocol for block length $n$ and with average fidelity

$$\overline{F} = \sum_I p_I \mathrm{Tr}(|\psi_I\rangle\langle\psi_I|\widetilde{\rho}_I) \geq 1 - \epsilon.$$

Let us describe the protocol again: the Sender performs a measurement on her half of $n(E+\delta)$ Einstein–Podolsky–Rosen (EPR) pairs and sends the measurement result $j$ (obtained with probability $p(J|I)$ and collapsing the Receiver's state to $\sigma_{I,j}$) to the Receiver (using $n(R+\delta)$ classical bits), who performs a quantum operation $T_j$ on his half of the EPR pairs. The state thus produced is $\widetilde{\rho}_{I,j}$ and obviously $\widetilde{\rho}_I = \sum_j p(j|I)\widetilde{\rho}_{I,j}$.

Now, the post-measurement state, including a classical system $A$ to record $I$, can be written in the general form

$$\sigma = \sum_I p_I |I\rangle\langle I|^A \otimes \sum_j p(j|I)\sigma^B_{I,j} \otimes |j\rangle\langle j|^C$$

where $C$ is the classical system used for communicating $j$.

Entropic quantities of this state are related to the resources required by the protocol: first of all, $S_\sigma(A:BC) \leq n(R+\delta)$ because in total $n(R+\delta)$ bits are communicated, and their information cannot be exceeded by the information in what the receiver eventually gets, by causality. Similarly, because all the $\sigma^B_{I,j}$ are supported on the $nE$ qubits which form the Receiver's half of the EPR pairs, we get

$$n(E+\delta) \geq S_\sigma(B) \geq S_\sigma(B|C) \geq S_\sigma(A:B|C).$$

We may assume that the $T_j$ do not affect the system $C$, and because (conditional) mutual informations are nonincreasing under local quantum operations, we obtain that

$$n(R+\delta) \geq S_{\widetilde{\rho}}(A:BC) \quad (9)$$
$$n(E+\delta) \geq S_{\widetilde{\rho}}(A:B|C) \quad (10)$$

with the state

$$\widetilde{\rho} = \sum_I p_I |I\rangle\langle I|^A \otimes \sum_j p(j|I)\widetilde{\rho}^B_{I,j} \otimes |j\rangle\langle j|^C.$$

(Note that $\widetilde{\rho}_{I,j} = T_j(\sigma_{I,j})$ for all $I, j$.) Our goal is now to switch in the latter expression to the ideal states $|\psi_I\rangle\langle\psi_I|$, arguing that we retain high fidelity to $\widetilde{\rho}$, and then invoking general continuity bounds for the entropy.

More precisely, define

$$\Omega = \sum_I p_I |I\rangle\langle I|^A \otimes \sum_j p(j|I)|\psi_I\rangle\langle\psi_I|^B \otimes |j\rangle\langle j|^C.$$

Then we can estimate

$$\begin{aligned}\|\Omega - \widetilde{\rho}\|_1 &= \sum_I p_I \sum_j p(j|I)\, \||\psi_I\rangle\langle\psi_I| - \widetilde{\rho}_{I,j}\|_1 \\ &\leq \sum_I p_I \sum_j p(j|I) 2\sqrt{1 - \mathrm{Tr}(|\psi_I\rangle\langle\psi_I|\widetilde{\rho}_{I,j})} \\ &\leq \sum_I p_I 2\sqrt{1 - \mathrm{Tr}(|\psi_I\rangle\langle\psi_I|\widetilde{\rho}_I)} \\ &\leq 2\sqrt{1-\overline{F}} \leq 2\sqrt{\epsilon}\end{aligned}$$

where in the second line we have used the inequality $\frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1-F(\rho,\sigma)}$ for states $\rho$, $\sigma$[27], and then concavity of the square root function. Because for states $\rho$, $\sigma$ on a $D$-dimensional system, $\|\rho-\sigma\|_1 \leq \epsilon \leq \frac{1}{2}$, we have the Fannes inequality [25] $|S(\rho) - S(\sigma)| \leq -\epsilon \log \frac{\epsilon}{D}$, we obtain that there exists a function $f(\epsilon)$, vanishing as $\epsilon \to 0$, such that

$$n(R+\delta) \geq S_\Omega(A:BC) - nf(\epsilon) \quad (11)$$
$$n(E+\delta) \geq S_\Omega(A:B|C) - nf(\epsilon). \quad (12)$$

The reasoning is that the entropies of combinations of $A$, $B$, and $C$ relative to the states $\Omega$ and $\widetilde{\rho}$, see (9) and (10), can be estimated against each other by the Fannes inequality, observing that Hilbert space dimensions are of the form $X^n$ with a constant $X$.

Hence, we get (letting $\delta' = \delta + f(\epsilon)$)

$$\begin{aligned}n(E+\delta') &\geq \min_\Omega\{S(A:B|C) | S(A:BC) \leq n(R+\delta')\} \\ &= N(\mathcal{E}^{\otimes n}, n(R+\delta')).\end{aligned}$$

Now we invoke Lemma 12 to estimate further

$$\begin{aligned}E+\delta' &\geq \frac{1}{n} N(\mathcal{E}^{\otimes n}, n(R+\delta')) \\ &= \min\left\{ \frac{1}{n}\sum_{k=1}^n N(\mathcal{E}, R_k) \,\middle|\, \frac{1}{n}\sum_{k=1}^n R_k = R+\delta' \right\} \\ &\geq N(\mathcal{E}, R+\delta')\end{aligned}$$

the second line by (8), the third by convexity of $N$. Using the continuity of $N$ with $\delta' \to 0$ (which occurs with $\epsilon, \delta \to 0$), we arrive at $E \geq N(\mathcal{E}, R)$, as desired. □

Readers of [30] will notice the similarity of the proofs of the lower bounds in Theorems 9 and 11. Given that for the upper bound we use an operational transformation of a q.c.t. protocol into an r.s.p. protocol, one may wonder if there is not a proof of the optimality of this reduction by an inverse reduction of an r.s.p. protocol to a q.c.t. protocol. We relate one such attempt in Appendix E.

There are two generalizations of Theorem 9 which we can transport to obtain more general versions of Theorem 11: the first is to lift the restriction to discrete ensembles, which is not really necessary—it is shown in [30] by suitable approximation (using in fact the net Lemma 4) that Theorem 9 holds true for an arbitrary probability distribution $p$ on the pure states of $\mathcal{H}$. This shows automatically that Theorem 11 also holds in the same form for general ensembles (in general, with $\inf$ instead of $\min$).

The second concerns the so-called arbitrarily varying sources (AVS): an ensemble is generally taken to represent some partial knowledge about the states to be encountered, and this model allows us to fine-tune this to even less knowledge: an AVS is a family of probability distributions $p_s$, $s \in \mathcal{S}$, on the space of pure states, with the intention that at each time step each of

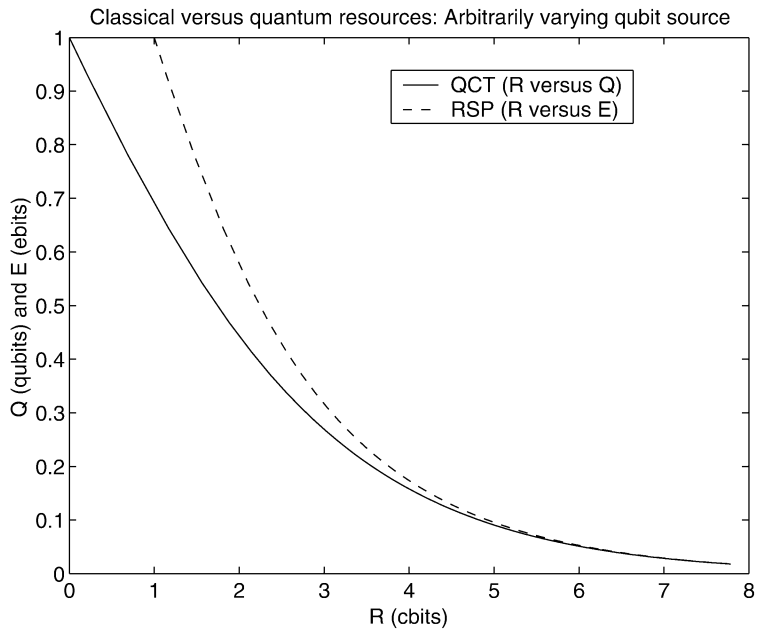


Fig. 1. The q.c.t. tradeoff curve for qubits versus cbits according to Devetak and Berger [22] (solid) and the implied r.s.p. tradeoff for ebits versus cbits (dashed).

the distributions $p_s$ can occur. One might want to think of an adversary choosing $s^n = s_1, \ldots, s_n$, thus presenting a given protocol with the distribution of states

$$p_{s^n} = p_{s_1} \otimes \cdots \otimes p_{s_n}.$$

A protocol (of either q.c.t. or r.s.p.) is said to have fidelity $\overline{F}$ if for all choices $s^n \in \mathcal{S}^n$

$$\int \mathrm{d}p_{s^n}(\psi) F(|\psi\rangle\langle\psi|, \widetilde{\rho}) \geq \overline{F}$$

where $\widetilde{\rho}$ is the output state on input $\psi$.

It turns out [30] that for q.c.t. there is still a tradeoff in this case, and that $Q^*(R)$ is given by the tradeoff for the worst case ensemble distribution from the convex hull $\boldsymbol{P} = \mathrm{conv}\{p_s | s \in \mathcal{S}\}$ of the $p_s$.

*Theorem 13:* For an AVS $\{p_s\}_{s\in\mathcal{S}}$, the q.c.t. tradeoff curve is given by

$$Q^*(R) = \sup_{p\in\boldsymbol{P}} Q^*(p, R)$$

where $Q^*(p, R)$ is the tradeoff of Theorem 9 as a function of cbit rate $R$ and the ensemble distribution $p$, made explicit. □

This immediately implies, by the same reasoning, the corresponding theorem for remote state preparation.

*Theorem 14:* For an AVS $\{p_s\}_{s\in\mathcal{S}}$, the r.s.p. tradeoff curve is given by

$$E^*(R) = \sup_{p\in\boldsymbol{P}} E^*(p, R)$$

where $E^*(p, R)$ is the tradeoff of Theorem 11 as a function of cbit rate $R$ and the ensemble distribution $p$, made explicit. □

In particular, dropping all restrictions, i.e., for the AVS with $\boldsymbol{P} = \{\text{all distributions}\}$ (which means that the adversary may pick an arbitrary product state for the protocol), we obtain the "ultimate" tradeoff functions $\boldsymbol{Q}^*$ and $\boldsymbol{E}^*$: these govern the asymptotic qubit/cbit and ebit/cbit cost of compressing and remotely preparing blocks of arbitrary states. Because we know that $Q^*(R)$ for the uniform distribution dominates all other curves with fixed input distribution ([30, Theorem 6.1 and Corollary 9.2]), we have $\boldsymbol{Q}^*(R) = Q^*(R, \text{uniform})$ and hence $\boldsymbol{E}^*(R) = E^*(R, \text{uniform})$. For qubits we thus can plot $\boldsymbol{E}^*$ thanks to the results of Devetak and Berger [22] (Fig. 1).

A word might be necessary to explain why there is no contradiction between this universal tradeoff curve (which evidently exists not just for qubits, but for any qu*d*its; to our knowlegde, however, it has not been worked out explicitly for $d > 2$), and the proof of the nonexistence of any finite tradeoff in Section III. This is because in the present section the task is much less ambitious: we only want to remotely prepare large blocks of (admittedly arbitrary) qubit states, i.e., a long product of pure states in small dimension. The set of product states however is much smaller than the set of all pure states on the large blocks. This fact is sufficient to allow an efficient tradeoff between ebits (or qubits) and cbits.

## V. Preparation of Entangled States

It is tempting to consider the generalization of the previous section to mixed-state sources. Observing however that our solution of the pure-state case rested on the quantum-classical tradeoff for pure state compression [30]—itself a generalization of Schumacher's source coding [38]—we might be discouraged by the corresponding mixed-state compression being far from resolved. A glimpse of this is provided in Appendix E, but see a more detailed discussion in [3], [42], and references therein.

Instead, we target a seemingly harder problem: the Sender (A) should remotely prepare an entangled state between the Receiver (B) and herself, drawn from an ensemble. Clearly, the Receiver in this way obtains the mixed-state ensemble of the reduced states.

In detail, assume an ensemble $\mathcal{E} = \{|\varphi_i\rangle^{AB}, p_i\}_{i=1}^m$ of pure entangled states generating the i.i.d. source

$$\begin{aligned} I &= i_1 \dots i_n \\ |\varphi_I\rangle &= |\varphi_{i_1}\rangle \otimes \cdots \otimes |\varphi_{i_n}\rangle \\ p_I &= p_{i_1} \dots p_{i_n}. \end{aligned}$$

The protocols we consider are of a general form very similar to those in Section IV: they allow both parties to use a maximally entangled state $\Phi_D$ of Schmidt rank $D$, and consist of a family of instruments [20] $\boldsymbol{M}^I = \left(M_j^I\right)_j$ $(j = 1, \dots, M)$ for the Sender, i.e., each $M_j^I$ is a completely positive map, and their sum (over $j$) is a trace-preserving map for every $I$—this conveniently captures the notion of a (partial) measurement with a post-measurement state. Furthermore, there are quantum operations $T_j$ for the Receiver. The states prepared in this way are

$$\widetilde{\rho}_I = \sum_j \left(M_j^I \otimes T_j\right) \Phi_D$$

and as before we demand that the fidelity $\overline{F} \geq 1 - \epsilon$, with

$$\sum_I p_I \mathrm{Tr}(\varphi_I \widetilde{\rho}_I) \geq \overline{F}.$$

And similarly, we call a rate pair $(R, E)$ *achievable* if for all $\epsilon, \delta > 0$ and sufficiently large $n$ there exist r.s.p. protocols with

$$\begin{aligned} \frac{1}{n} \log M &\leq R + \delta, \\ \frac{1}{n} \log D &\leq E + \delta. \end{aligned}$$

Define the tradeoff function for the ensemble $\mathcal{E}$

$$E^*(R) = \min\{E | (R, E) \text{ achievable}\}.$$

We start by describing a protocol to achieve the rate point with the smallest $R$ allowed by causality (a different proof for the achievability of the cbit rate can be found in [12] even though with a method that is very wasteful in terms of entanglement, much like the column method of Example 1).

*Proposition 15:* There exists an r.s.p. protocol which achieves the rate pair

$$\begin{aligned} R &= \chi\left(\left\{p_i, \varphi_i^B\right\}\right) \\ E &= S\left(\sum_i p_i \varphi_i^B\right) \end{aligned}$$

with the Holevo quantity $\chi$ of the Receiver's mixed-state ensemble $\left\{p_i, \varphi_i^B\right\}$.

*Proof:* Consider a string $I = i_1 \dots i_n$ of type (i.e., relative letter frequencies) $Q$—see Appendix D for details—and construct (with $\delta > 0$) the conditional typical projector $\Pi^n_{\varphi^B, \delta}(I)$ for $\varphi_I^B = \varphi_{i_1}^B \otimes \cdots \otimes \varphi_{i_n}^B$. By (D5), for sufficiently large $n$

$$\mathrm{Tr}\left(\varphi_I^B \Pi^n_{\varphi^B, \delta}(I)\right) \geq 1 - \epsilon.$$

Construct also the typical projector $\Pi := \Pi^n_{\rho, \delta}$ of the average state $\rho = \sum_i Q(i) \varphi_i^B$: by Lemma 26, for sufficiently large $n$

$$\mathrm{Tr}\left(\varphi_I^B \Pi^n_{\rho, \delta}\right) \geq 1 - \epsilon.$$

Hence, if we define (for all $I$ of type $Q$)

$$\pi_I := \Pi^n_{\rho, \delta} \Pi^n_{\varphi^B, \delta}(I) \varphi_I^B \Pi^n_{\varphi^B, \delta}(I) \Pi^n_{\rho, \delta}$$

these operators have the properties

$$\mathrm{Tr}\pi_I \geq 1 - 2\epsilon \tag{13}$$

$$\pi_I \leq \exp\left(-n\left(S(\varphi^B|Q) - \delta\right)\right) \Pi^n_{\rho, \delta}, \tag{14}$$

the latter is obtained by the definition of the conditional typical projector in Appendix D; here, $S(\varphi^B|Q) = \sum_i Q(i) S\left(\varphi_i^B\right)$.

Denoting the subspace onto which $\Pi^n_{\rho, \delta}$ projects by $\mathcal{T}$, its dimension, by (D3) is bounded

$$D := \dim \mathcal{T} \leq \exp(n(S(\rho) + \delta)). \tag{15}$$

Now, for the Haar measure $\mathrm{d}U$ on the unitaries on $\mathcal{T}$

$$\int \mathrm{d}U U \pi_I^\top U^* = (\mathrm{Tr}\pi_I) \frac{1}{D} \Pi.$$

Draw $U_1, \dots, U_K$ i.i.d. according to the Haar measure. Then, according to Lemma 16 stated later

$$\begin{aligned} \Pr\left\{\frac{1}{K} \sum_k \frac{U_k \pi_I^\top U_k^*}{\mathrm{Tr}\pi_I} \notin \frac{1}{D}[(1 \pm \epsilon)\Pi]\right\} \\ \leq 2D \exp\left(-K \exp(-n(\chi + 2\delta)) \frac{(1 - 2\epsilon)\epsilon^2}{2}\right) \end{aligned}$$

with $\chi = S(\rho) - S(\varphi^B|Q)$: because we can rescale the $\pi_I$ with the factor on the right-hand side of (14). Thus, by the union bound, there exist $U_1, \dots, U_K$ such that for all $I$ of type $Q$

$$\frac{1 - \epsilon}{D} \Pi \leq \frac{1}{K} \sum_k \frac{1}{\mathrm{Tr}\pi_I} U_k \pi_I^\top U_k^* \leq \frac{1 + \epsilon}{D} \Pi \tag{16}$$

if

$$K = (1 + n \log m + \log D) \frac{2}{(1 - 2\epsilon)\epsilon^2} \exp(n(\chi + 2\delta)).$$

The r.s.p. protocol now works as follows: the Sender, on getting $I$, determines its type $Q$ and sends it to the Receiver. If $\|p - Q\|_1 > \delta$, the protocol aborts here (this happens with probability $\leq \epsilon$ if $n$ is sufficiently large, by the law of large numbers). For type $Q$, they have agreed on a list of unitaries $U_1, \dots, U_K$ as in (16): the Sender can construct the measurement POVM

$$\begin{aligned} A_k &= \frac{D}{K(1 + \epsilon)\mathrm{Tr}\pi_I} U_k \pi_I^\top U_k^* \\ A_{\text{failure}} &= \Pi - \sum_k A_k \end{aligned}$$

and measures it (nondestructively) on the maximally entangled state $\Phi$ on $\mathcal{T}^A \otimes \mathcal{T}^B$. The outcome "failure" occurs with probability less than $\epsilon$, and in the case of outcome $k$ the Receiver, on learning the value $k$, can apply the unitary $U_k^\top$: it is straightforward to check that in this case he and the Sender share a purification of $\frac{1}{\mathrm{Tr}\pi_I}\pi_I$. Because of (13) and the gentle measurement Lemma 17 below, this state has high fidelity to $\varphi_I^B$, so by [32],

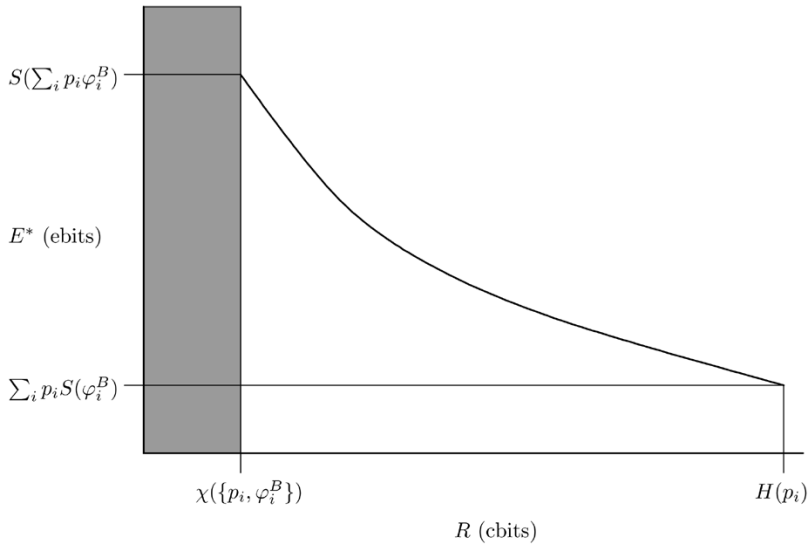


Fig. 2. Schematic of the tradeoff curve for an ensemble of entangled states. The shaded area is forbidden by causality and the curve begins at the point $\left(\chi\left(\{p_i,\varphi_i^B\}\right), S\left(\sum_i p_i\varphi_i^B\right)\right)$, due to the protocol of Proposition 15. It can never go below $E=\sum_i p_i S\left(\varphi_i^B\right)$, which is reached at cbit rate $R=H(p)$, as this is the very amount of entanglement in the ensemble.

[40] she can apply a unitary to her post-measurement state to obtain a high-fidelity approximation of $\varphi_I^{AB}$.

Clearly, this protocol has a high average fidelity. In terms of resources, it requires a logarithmic number of bits to communicate the type $Q$ and

$$\log K \leq n\left(\chi\left(\left\{Q(i),\varphi_i^B\right\}\right)+f(\delta)\right)$$

to communicate the result of the measurement described above, with a function $f$ which vanishes as $\delta\to 0$. By (15), it uses

$$\leq n\left(S\left(\sum_i Q(i)\varphi_i^B\right)+\delta\right)$$

ebits. With Fannes inequality [25] for $\|p-Q\|_1\leq\delta$, we obtain the claim. □

*Lemma 16 ("Operator Chernoff Bound" [1]):* Let $X_1,\ldots,X_M$ be i.i.d. random variables taking values in the operators $\mathcal{B}(\mathcal{H})$ on the $D$-dimensional Hilbert space $\mathcal{H}$, $0\leq X_j\leq\mathbb{1}$, with $A=\boldsymbol{E}X_j\geq\alpha\mathbb{1}$, and let $0<\eta\leq 1/2$. Then

$$\Pr\left\{\frac{1}{M}\sum_{j=1}^{M}X_j\notin[(1-\eta)A;(1+\eta)A]\right\}\leq 2D\exp\left(-M\frac{\alpha\eta^2}{2\ln 2}\right).\quad\square$$

*Lemma 17:* For a state $\rho$ and an operator $0\leq X\leq\mathbb{1}$, if $\mathrm{Tr}(\rho X)\geq 1-\epsilon$, then

$$\left\|\rho-\sqrt{X}\rho\sqrt{X}\right\|_1\leq\sqrt{8\epsilon}.$$

The main result of the present section is that this is essentially optimal.

*Theorem 18:* For the ensemble $\mathcal{E}=\{p_i,\varphi_i\}$ of pure bipartite states and $R\geq 0$

$$E^*(R)=N(\mathcal{E},R):=\min\{S(B|C)|S(X:BC)\leq R\}$$

where the entropic quantities are with respect to the state $\omega$, and minimization is over all 4-partite states $\omega$ as follows:

$$\omega=\sum_i p_i|i\rangle\langle i|^X\otimes\varphi_i^{AB}\otimes\sum_j p(j|i)|j\rangle\langle j|^C \tag{17}$$

with a classical channel $p(j|i)$.

This theorem should be compared to the unentangled case, Theorem 11, to which it provides a pleasingly direct generalization. We see that despite the fact that the theorem applies to ensembles of entangled states, register $A$ does not appear in any of the entropic quantities involved. The tradeoff curve is a function solely of the ensemble of mixed states at the Receiver. See Fig. 2 for a schematic view of the tradeoff curve.

Before giving the proof, we state a crucial lemma (compare Lemma 12), which we prove in Appendix F.

*Lemma 19:* $N$ is convex, continuous, and strictly decreasing in the interval $[S(X:B);S(X)]$. It obeys the following additivity relation for ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$:

$$N(\mathcal{E}_1\otimes\mathcal{E}_2,R)=\min\{N(\mathcal{E}_1,R_1)+N(\mathcal{E}_2,R_2)|R_1+R_2=R\}. \tag{18}$$

□

*Proof of Theorem 18:* First, to show that, for fixed $R$, the pair $(R,N(R))$ is achievable, consider any channel $q(j|i)$, and let Sender and Receiver perform the following procedure (where all information quantities we encounter refer to the state $\omega$).

In step one, the channel $q^n$ is simulated (using shared randomness) on the typical $I$ by the Reverse Shannon Theorem [10], [42], using $n(I(X:C)+\delta)$ of forward communication, within average total variational distance $\epsilon$ if $n$ is large enough.

Assuming that the channel $q^n$ is simulated ideally, we can proceed: with probability $1-\epsilon$, $J=j_1,\ldots,j_n$ is typical for the distribution $q_j=\sum_i p_i p(j|i)$, i.e., if $\mathcal{I}_j$ is the set of indices $i$ such that $j_i=j$, then

$$\forall\, j\quad\left||\mathcal{I}_j|-nq_j\right|\leq\delta.$$

Now Proposition 15 is used to remotely prepare the ensemble $\{q(i|j), \varphi_i\}$ on the block $\mathcal{I}_j$, with the conditional distribution

$$q(i|j) = \frac{1}{q_j} p_i p(j|i).$$

This requires

$$\begin{aligned} &\leq n(q_j + \delta)\left(\chi\left(\{q(i|j), \varphi_i^B\}\right) + \delta\right) \text{ cbits} \\ &\leq n(q_j + \delta)\left(S\left(\sum_i q(i|j)\varphi_i^B\right) + \delta\right) \text{ ebits.} \end{aligned}$$

In total, we use $n(S(X : C) + S(X : B|C) + f(\delta))$ cbits, and $n(S(B|C) + f(\delta))$ ebits, and the average fidelity can be made arbitrarily close to 1. Finally, the shared randomness can be disposed of, because the average fidelity is an average over it—hence, there exists a value of the shared random variable such that the average fidelity is even larger.

Now for the converse direction, that $N$ is a lower bound: if $(R, E)$ is achievable, then for sufficiently large $n$ there exist protocols which use $n(R+\delta)$ cbits and $n(E+\delta)$ ebits, of fidelity $1 - \epsilon$

$$\sum_I p_I \langle \varphi_I | \widetilde{\rho}_I | \varphi_I \rangle \geq 1 - \epsilon$$

where $\widetilde{\rho}_I$ is the output state for input $I$. Any protocol has the following form: the Sender performs a measurement on her half of $n(E+\delta)$ EPR pairs, and then sends $n(R+\delta)$ bits of classical message $j$ to the Receiver. Conditioned on the classical message $j$, he then performs a decoding operation $T_j$ on his system. The outcome is a state $\widetilde{\rho}_{I,j}$ such that

$$\widetilde{\rho}_I = \sum_j p(j|I)\widetilde{\rho}_{I,j}.$$

The post-measurement state, including a classical system recording $I$, can be written in the form

$$\omega = \sum_I p_I |I\rangle\langle I|^X \otimes \omega_{I,j}^{AB} \otimes \sum_j p(j|I)|j\rangle\langle j|^C$$

where system $C$ is communicated, and with $\widetilde{\rho}_{I,j} = T_j(\omega_{I,j})$. By causality, $S_\omega(X : BC) \leq n(R + \delta)$. Moreover, we can assume that the Receiver's operation $T_j$ does not damage the $C$ register since the contents of the register could be copied prior to the application of $T_j$. Since $T_j$ cannot increase $S(X : BC)$ by data processing, however, we find that for the state

$$\widetilde{\rho} = \sum_I p_I |I\rangle\langle I|^X \otimes \widetilde{\rho}_{I,j}^{AB} \otimes \sum_j p(j|I)|j\rangle\langle j|^C$$

the inequality

$$S_{\widetilde{\rho}}(X : BC) \leq n(R + \delta)$$

holds. Introducing

$$\Omega = \sum_I p_I |I\rangle\langle I|^X \otimes \varphi_I^{AB} \otimes \sum_j p(j|I)|j\rangle\langle j|^C$$

we conclude that

$$S_\Omega(X : BC) \leq n(R + \delta + f(\epsilon)) \tag{19}$$

with some universal function $f$ vanishing with $\epsilon$: this is because of our fidelity assumption on the protocol and the bilinearity of the pure state fidelity $F(\Omega, \widetilde{\rho}) \geq 1 - \epsilon$. (Compare the analogous computation in the proof of Theorem 11.)

To bound the entanglement, observe that because the Sender's measurement was on her half of $n(E + \delta)$ EPR pairs that the state $\omega_j^B = \sum_I q(I|j)\omega_{I,j}^B$ has support no larger than $2^{n(E+\delta)}$. (Note that $p_I p(j|I)$ defines a joint distribution on $I$ and $j$. We use $q(I|j)$ and $q_j$ to denote the associated conditional and marginal distributions.) Therefore, for the state $\omega_{XABC}$

$$S_\omega(B|C) = \sum_j q_j S\left(\omega_j^B\right) \leq n(E + \delta).$$

If Bob's decoding operation $T_j$ were guaranteed to be unitary we could conclude $S_{\widetilde{\rho}}(B|C) \leq n(E + \delta)$. More generally, $T_j$ can be decomposed into three steps: adjoining an ancilla, applying a unitary and then tracing over the ancilla system. The first two steps leave the entropy invariant so without loss of generality, assume that conditioned on $j$, Sender and Receiver share a state $\omega_{I,j}^{ABB'}$ and that $T_j = \mathrm{Tr}_{B'}$. Our strategy will be to use the fact that the states $\varphi_I^{AB}$ are pure to argue that the partial trace should not increase the entropy.

First, we now have $\omega_{I,j}^{AB} = \widetilde{\rho}_{I,j}^{AB}$. Let $\langle \varphi_I | \widetilde{\rho}_{I,j} | \varphi_I \rangle = 1 - \epsilon_{I,j}$. We can choose an extension $\varphi_I^{ABB'}$ of $\varphi_I$ such that

$$F\left(\varphi_I^{ABB'}, \omega_{I,j}^{ABB'}\right) = 1 - \epsilon_{I,j}$$

[32], [40]. By the concavity of the fidelity, we then conclude that for

$$\begin{aligned} \varphi_j &:= \sum_I q(I|j)\varphi_I^{ABB'} \\ \omega_j &:= \sum_I q(I|j)\omega_{I,j}^{ABB'} \end{aligned}$$

we have

$$F(\varphi_j, \omega_j) \geq \sum_I q(I|j)(1 - \epsilon_{I,j}) =: 1 - \epsilon_j.$$

Now, because $\varphi_I^{AB}$ is pure, the state $\varphi_j$ must be separable across the $AB - B'$ cut. Therefore, $S(\varphi_j^{BB'}) \geq S(\varphi_j^B)$. On the other hand, using the Fannes inequality and the concavity of its bound, we obtain

$$\begin{aligned} n(E + \delta) &\geq \sum_j q_j S\left(\omega_j^{BB'}\right) \\ &\geq \sum_j q_j \left[S\left(\varphi_j^{BB'}\right) - nf(\epsilon_j)\right] \\ &\geq \sum_j q_j S\left(\varphi_j^B\right) - nf(\epsilon) \end{aligned}$$

for some universal function $f$ vanishing with $\epsilon$. Hence,

$$S_\Omega(B|C) \leq n(E + \delta + f(\epsilon)). \tag{20}$$

Putting this together with (19), we get, with $\delta' = \delta + f(\epsilon)$ and the definition of $N$

$$n(E + \delta') \geq N(\mathcal{E}^{\otimes n}, n(R + \delta')).$$

Now we can invoke Lemma 19, and obtain

$$\begin{aligned} E + \delta' &\geq \frac{1}{n} N(\mathcal{E}^{\otimes n}, n(R + \delta')) \\ &= \min\left\{\frac{1}{n}\sum_{k=1}^n N(\mathcal{E}, R_k) \,\middle|\, \frac{1}{n}\sum_{k=1}^n R_k = R + \delta'\right\} \\ &\geq N(\mathcal{E}, R + \delta'). \end{aligned}$$

TABLE I

| | Worst Case | Expected |
|---|---|---|
| Det. exact | ?$\geq$ 1 ebit, 2 $\geq$?$\geq$ 1 cbits | 1 ebit, 1 cbit |
| Prob. exact | 1 ebit, 1 cbit | |
| High fidelity | 1 ebit, 1 cbit | |
| Oblivious | 1 ebit, 2 cbits | |
| Approx. obl. | 1 ebit, 1 cbit | |

Finally, using continuity of $N$ in $R$, we obtain the result

$$E \geq N(\mathcal{E}, R). \qquad \square$$

## VI. Discussion

In Sections VI-A–C we want to review what we have achieved, while pointing out open questions.

### A. Models and Resources

In the Introduction, we have mentioned various subtly different ways to define remote state preparation (deterministic exact, probabilistic exact, high fidelity; see Section VI-B for *oblivious*), as well as ways to account for the resources used (worst case and expected cost).

Subsequently, we have concentrated on probabilistic and high-fidelity asymptotic protocols (for which worst case and expected cost coincide, as one can easily see). The justification of this choice is that it seems to be the one best suited to the asymptotic considerations at our focus.

However, as shown in Table I, our conclusions are for the most part independent of the particulars of the model.

The entries "1 ebit, 1 cbit" derive their achievability from our protocol $\Pi$ (Theorem 5)—directly in the cases "Probabilistic exact," "High fidelity," and "Approximately oblivious" (see Section VI-B), and augmented by teleportation in the failure event for "Deterministic exact, Expected cost." The upper bound "1 ebit, 2 cbits" is, of course, teleportation, which indeed is oblivious (see Section VI-B); that in the oblivious case 2 cbits are indeed necessary was shown in [35].

So, only the entry in the field "Deterministic exact, Worst case" is not entirely understood: in [29] it is shown that an exact r.s.p. protocol for a single qubit requires at least 1 ebit and 2 cbits, just like teleportation. Whether the analogous statement for higher dimensions is true is unknown.

### B. Approximate Obliviousness

An r.s.p. protocol is called oblivious to the Sender [35] if, like teleportation, it can be made into a quantum operation for her, which she can execute without knowing classically what state she is attempting to prepare. A protocol is called oblivious to the Receiver [35] if, again like teleportation, it leaks no information about the state being prepared beyond giving him a single specimen of it. In [35] it was shown that if a deterministic exact protocol for preparing states in dimension $D$ is oblivious to the Receiver, then it must be oblivious to the Sender also, and must, therefore, like teleportation, use at least $\log D$ ebits and $2 \log D$ cbits.

A similar penalty for receiver obliviousness exists even in a purely classical analog of r.s.p., namely, the simulation of a noisy classical channel by noiseless forward classical communication (cbits) and shared randomness (rbits) between Sender and Receiver. The classical Reverse Shannon Theorem [10] gives a deterministic exact protocol for this task at an expected cbit cost approaching the simulated channel's classical capacity $C$ in the limit of large block size, but it is not hard to show that for some channels any such exact efficient simulation must 1) have a worst case cost exceeding its expected cost, and 2) must be nonoblivious to the Receiver. For example, consider a binary-symmetric channel with crossover probability $p$ and capacity $C = 1 + p \log p + (1-p)\log(1-p)$. Note that such a channel, given a block of $n$ inputs, has probability $P_0 = (1-p)^n$ of transmitting the whole block exactly, without crossovers, and of course any exact simulation of the channel must simulate this rare event with the correct probability. But to avoid a violation of causality, the expected cost of the simulation, in instances where no crossover occurs in a block of size $n$, must be at least $n - \log(1/P_0)$; otherwise, as in the column method, the Sender could use $\log(1/P_0) + O(1)$ cbits of additional classical communication to designate a no-crossover instance within a general simulation, thereby communicating $n$ cbits about the input in less than $n$ cbits of forward communication. For $0 < p < 1/2$, the causality-imposed cost $n - \log(1/P_0) = n(1 + \log(1-p))$ exceeds the expected cost $nC$ of an efficient simulation according to the Reverse Shannon Theorem; therefore, in any efficient exact simulation, 1) the worst case cost must be at least $n - \log(1/P_0)$; and 2) the occurrence of a cost exceeding the expected cost $nC$ must be negatively correlated with the number of crossovers, leaking extra information about the channel input besides that contained in the correctly simulated output.

Resuming our discussion of obliviousness in r.s.p., we observe that the previously studied notions of obliviousness to the Receiver are exact, requiring that the protocol leak no information whatever about the input. In the present paper's main context of approximate simulations it is more appropriate to use a more robust notion of approximate obliviousness.

*Definition 20:* An r.s.p. protocol for a set $\boldsymbol{X}$ of states on $\mathcal{K}$ is said to be approximate and approximately oblivious with parameters $(\epsilon, \delta)$ if

1) For all $\sigma \in \boldsymbol{X}$, if the Receiver's output state is denoted $\rho^B$: $\frac{1}{2}\|\sigma - \rho^B\|_1 \leq \epsilon$.
2) There exists a completely positive and trace preserving (c.p.t.p.) map $T$ on the Receiver's system that maps his output state $\rho^B$ to a close approximation of the whole of what he gets from the protocol: the pre-image of $\rho^B$ (under his decoding operation), possible residual quantum states, and the classical messages, i.e.,

$$\frac{1}{2}\|\{\text{Receiver's record}\} - T(\rho^B)\|_1 \leq \delta.$$

Note that our notion of "approximate obliviousness" does not arise from some *a priori* concept of what the Receiver must not learn. It is rather modeled after "zero-knowledge" in zero-knowledge proofs: the verifier gets nothing that he could not have simulated himself (see [28] and subsequent literature).

Note that for $\epsilon = \delta = 0$ we recover the definition of [35] of a deterministic exact and exactly oblivious protocol. It would be natural to conjecture that a robust version of the main result of [35] should hold.

For an approximate and approximately oblivious r.s.p. protocol with parameters $(\epsilon, \delta)$ (for the set $\boldsymbol{X} = \mathcal{P}(\mathcal{K})$ of all pure states on $\mathcal{K}$), the communication cost is $\geq 2 - f(\epsilon, \delta)$ cbits per qubit, and it has to use $\geq 1 - f(\epsilon, \delta)$ ebits per qubit. There, $f$ is a function that vanishes with $\epsilon, \delta \to 0$.

Instead, it turns out that our protocol $\Pi$ is indeed approximate and approximately oblivious in the sense of Definition 20, with parameters $(\epsilon, \epsilon)$:

Clearly, part 1) of the definition is satisfied (we remove the failure event by having the Sender choose one uniformly distributed from the "good" messages in the case of a "failure"). Part 2) also is easily seen to be true: the simulating map is simply

$$T : |\psi\rangle\langle\psi| \longmapsto \sum_{k=1}^{K} \frac{1}{K} |k\rangle\langle k| \otimes \overline{U_k} |\psi\rangle\langle\psi| U_k^{\top}.$$

As an aside, we may return to the column method, presented in Example 1 (without recycling of entanglement): it is not hard to see that in fact also this procedure is approximate and approximately oblivious. Indeed, to simulate the Receiver's view of the protocol, he only has to create an arbitrary state, say $\left(\frac{1}{D}\mathbb{1}\right)^{\otimes K}$ and an arbitrary classical message (say, uniformly distributed) with probability $\epsilon$: this is to simulate the failure. With probability $\frac{1-\epsilon}{K}$ each, he generates the states

$$\left(\frac{1}{D}\mathbb{1}\right)^{\otimes(k-1)} \otimes |\psi\rangle\langle\psi| \otimes \left(\frac{1}{D}\mathbb{1}\right)^{\otimes(K-k)}$$

and the classical message $k = 1, \ldots, K$. It is easily seen that this is $\epsilon$-close to the Receiver's actual view.

### C. Further Applications of Randomization and Tradeoff r.s.p.

The remote state preparation of state ensembles turns out to have applications to other problems, which we simply list here for reference.

1) The protocol we described in Section V for optimal preparation of pure entangled states produces, when one ignores the Sender's half of the state, mixed states at the Receiver's system with a classical communication cost exactly equal to the Holevo quantity of his ensemble. This result is in fact the Quantum Reverse Shannon Theorem [6] for the so-called cq-channels (mapping a discrete set of *classical* inputs $x$ to *quantum* states $\varphi_x$), and follows also from the alternative protocol described in [12].

2) Optimal remote state preparation of entangled states (Section V) is invoked to prove capacity formulas and bounds for the classical communication capacity of bipartite unitaries assisted by unlimited or bounded entanglement [8], [31].

3) At the heart of our r.s.p. protocol is the state randomization by relatively few unitaries (Theorem 2). In fact, similar to previously considered private quantum channels [2], [13] we obtain a private channel scheme, but with halved key length! By applying the randomization to half of an entangled state, one even obtains very efficient schemes for data hiding in bipartite quantum states [23], [24]. Our separate paper [9] is devoted to an exploration of these applications.

## APPENDIX A
## GAUSSIAN DISTRIBUTED VECTORS

This appendix is largely a compilation of known facts about the distribution of random vectors following a Gaussian law, and of some of their moments: we freely use textbook knowledge of probability theory (see, e.g., [26]), as well as parts of the treatment of large deviation theory by Dembo and Zeitouni [21].

Recall that the Gaussian (or normal) distribution on the reals with mean $\mu$ and variance $\sigma^2$, denoted $N(\mu, \sigma^2)$, is defined by the density

$$N(\mu, \sigma^2)\{\mathrm{d}t\} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t.$$

We shall phrase most of the following in terms of random variables. That a random variable $X$ is distributed according to some Gaussian is denoted $X \sim N(\mu, \sigma^2)$.

*Definition 21:* A Gaussian complex number with mean $\mu \in \boldsymbol{C}$ and variance $\sigma^2 > 0$ is a random variable $\gamma = X + iY$, where $X$ and $Y$ are independent real random variables with

$$X \sim N\left(\mathrm{Re}\mu, \frac{\sigma^2}{2}\right) \quad \text{and} \quad Y \sim N\left(\mathrm{Im}\mu, \frac{\sigma^2}{2}\right).$$

Its distribution is denoted $N_{\boldsymbol{C}}(\mu, \sigma^2)$.

Note that in this definition we insist that real and imaginary variance are equal, in contrast to the most general Gaussian distribution in $\boldsymbol{R}^2$.

Now let $\mathcal{H}$ be a complex Hilbert space (of finite dimension $d$). In general, a Gaussian distributed vector is a sum of the form $|\Gamma\rangle = \sum_j \gamma_j |v_j\rangle$, with an orthonormal basis $\{|v_j\rangle\}$ and independent Gaussian complex numbers $\gamma_j \sim N_{\boldsymbol{C}}\left(\mu_j, \sigma_j^2\right)$. Its distribution is uniquely determined by the mean $|\mu\rangle = \boldsymbol{E}|\Gamma\rangle$ and the covariance operator $S^2 = \boldsymbol{E}|\Gamma\rangle\langle\Gamma| \geq 0$: the density is given by

$$\Pr\left\{|\Gamma\rangle - |\mu\rangle \in |v\rangle + \mathrm{d}^{2d}|w\rangle\right\} = \frac{1}{\pi^d \det(S^2)} e^{-\langle v|S^{-2}|v\rangle} \mathrm{d}^{2d}|w\rangle$$

with the unitarily and translationally invariant normalized volume element $\mathrm{d}^{2d}w$ in $\mathcal{H} \simeq \boldsymbol{R}^{2d}$ (i.e., standard Lebesgue measure).

However, we shall be interested only in the special case that all means $\mu_j = 0$ and all $\sigma_j$ are equal.

*Definition 22:* A symmetric Gaussian vector with variance $\sigma^2$ is a randomly distributed $|\Gamma\rangle \in \mathcal{H}$ such that in one orthonormal basis $\{|v_j\rangle\}$

$$|\Gamma\rangle = \sum_j \gamma_j |v_j\rangle$$

with independent $\gamma_j \sim N_{\boldsymbol{C}}\left(\mu_j, \frac{\sigma^2}{d}\right)$.

Equivalently, we could also define it by its covariance operator being $\boldsymbol{E}|\Gamma\rangle\langle\Gamma| = \frac{\sigma^2}{d}\mathbb{1}$. From this it follows that the distribution of $\Gamma$ is unitarily invariant, hence, in Definition 22 we can allow any orthonormal basis, a fact we shall make frequent use

of. Note that $\sigma^2 = \boldsymbol{E}\langle\Gamma|\Gamma\rangle$. This distribution on $\mathcal{H}$ is denoted $N_{\mathcal{H}}(0,\sigma^2)$.

According to Cramér's theorem [18] (see [21] for its derivation in the present context: it requires only the "Bernstein trick" and Markov inequality), for i.i.d. real random variables $X, X_1, \ldots, X_N$

$$\Pr\left\{\frac{1}{N}\sum_{i=1}^{N} X_i \geq a\right\} \leq \exp\left(-N\frac{1}{\ln 2}\inf_{x\geq a}\Lambda^*(x)\right)$$
$$\Pr\left\{\frac{1}{N}\sum_{i=1}^{N} X_i \leq a\right\} \leq \exp\left(-N\frac{1}{\ln 2}\inf_{x\leq a}\Lambda^*(x)\right) \quad \text{(A1)}$$

with the rate function

$$\Lambda^*(x) = \sup_{y\in\boldsymbol{R}}\left[yx - \ln \boldsymbol{E}e^{yX}\right].$$

For a squared Gaussian this can be evaluated explicitly.

*Lemma 23:* For $X = Y^2$, with a Gaussian variable $Y \sim N(0,\sigma^2)$, the rate function evaluates to

$$\Lambda^*(x) = \begin{cases} \frac{1}{2}\left[\frac{x}{\sigma^2} - 1 - \ln\left(\frac{x}{\sigma^2}\right)\right], & x > 0 \\ \infty, & x \leq 0. \end{cases}$$

*Proof:* First we calculate $\Lambda(y) = \ln \boldsymbol{E}e^{yX}$

$$\begin{aligned}\boldsymbol{E}e^{yX} &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{yt^2}e^{-\frac{t^2}{2\sigma^2}}\mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{(y-\frac{1}{2\sigma^2})t^2}\mathrm{d}t \\ &= \frac{1}{\sqrt{1-2y\sigma^2}}\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{\tau^2}{2\sigma^2}}\mathrm{d}\tau \\ &= \frac{1}{\sqrt{1-2y\sigma^2}}.\end{aligned}$$

Hence,

$$\Lambda(y) = \begin{cases} -\frac{1}{2}\ln(1-2y\sigma^2), & y < \frac{1}{2\sigma^2} \\ \infty, & y \geq \frac{1}{2\sigma^2}. \end{cases}$$

Differentiation reveals one extremum of $yx - \Lambda(y)$ at $y = \frac{1}{2\sigma^2} - \frac{1}{2x}$, which must be the maximum because $yx - \Lambda(y)$ is upper-bounded for $x > 0$ and $-\infty$ at both ends of the permissible interval of $y$. This yields the claim. □

Observe in particular, that $\boldsymbol{E}X = \sigma^2$, so that we get for $a = (1+\epsilon)\sigma^2$ and $a = (1-\epsilon)\sigma^2$ $(\epsilon \geq 0)$ in (A1)

$$\Pr\left\{\frac{1}{N}\sum_{i=1}^{N} X_i > (1+\epsilon)\sigma^2\right\} \leq \exp\left(-N\frac{\epsilon - \ln(1+\epsilon)}{2\ln 2}\right)$$
$$\Pr\left\{\frac{1}{N}\sum_{i=1}^{N} X_i < (1-\epsilon)\sigma^2\right\} \leq \exp\left(-N\frac{-\epsilon - \ln(1-\epsilon)}{2\ln 2}\right). \quad \text{(A2)}$$

We shall make use of the following lower bound:

$$\text{For all } -1 \leq \xi \leq 1, \quad \frac{1}{2\ln 2}(\xi - \ln(1+\xi)) \geq \frac{\xi^2}{12\ln 2}. \quad \text{(A3)}$$

Proof is by Taylor expansion: for $|\xi| = 1$ it is obviously true, and for $|\xi| < 1$ we have

$$\begin{aligned}\xi - \ln(1+\xi) &= \xi - \left(\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\xi^n}{n}\right) \\ &= \sum_{n=2}^{\infty}(-1)^n\frac{\xi^n}{n} \\ &= \sum_{k=1}^{\infty}\left[\frac{\xi^{2k}}{2k} - \frac{\xi^{2k+1}}{2k+1}\right] \\ &\geq \sum_{k=1}^{\infty}\left[\frac{\xi^{2k}}{2k} - \frac{\xi^{2k}}{2k+1}\right] \geq \frac{\xi^2}{6}.\end{aligned}$$

## Appendix B
## State Randomization

### A. Proof of Lemma 3

Since the Haar measure is left and right invariant, we may assume that $\varphi = |e_1\rangle\langle e_1|$ and $P = \sum_{i=1}^{p}|e_i\rangle\langle e_i|$ for some fixed orthonormal basis $\{|e_i\rangle\}$. Let $|\Gamma_j\rangle = \sum_{i=1}^{D} g_{ij}|e_i\rangle$, where $g_{ij} \sim N_{\boldsymbol{C}}(0,1)$ are i.i.d. (see Appendix A). The distribution of $|\Gamma_j\rangle$ is the same as the distribution for $\||\Gamma_j\|_2 U|e_1\rangle$ if $U$ is chosen using the Haar measure.

For fixed $U = U_j$ and $|\Gamma\rangle = |\Gamma_j\rangle$, the convexity of exp implies that

$$\begin{aligned}&\boldsymbol{E}_\Gamma \exp\left(\frac{y}{D}\sum_{i=1}^{p}|\langle e_i|\Gamma\rangle|^2\right) \\ &\quad= \boldsymbol{E}_U\boldsymbol{E}_\Gamma \exp\left(\frac{y\||g\|_2^2}{D}\sum_{i=1}^{p}|\langle e_i|U|e_1\rangle|^2\right) \\ &\quad\geq \boldsymbol{E}_U \exp\left(\boldsymbol{E}_\Gamma\frac{y\||g\|_2^2}{D}\sum_{i=1}^{p}|\langle e_i|U|e_1\rangle|^2\right) \\ &\quad= \boldsymbol{E}_U \exp\left(y\sum_{i=1}^{p}|\langle e_i|U|e_1\rangle|^2\right) \\ &\quad= \boldsymbol{E}_U \exp(y\mathrm{Tr}(U\varphi U^*P)).\end{aligned}$$

Invoking Cramér's theorem, this inequality between the moment generating functions establishes that

$$\frac{1}{n}\sum_{j=1}^{n}\mathrm{Tr}(U\varphi U^*P)$$

converges to its mean value

$$\boldsymbol{E}_U\mathrm{Tr}(U\varphi U^*P) = \boldsymbol{E}_\Gamma\frac{1}{D}\sum_{i=1}^{p}|\langle e_i|\Gamma\rangle|^2 = \frac{p}{D}$$

at least as quickly as

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{p}\frac{1}{D}|\langle e_i|\Gamma_j\rangle|^2 = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{p}\frac{1}{D}|g_{ij}|^2.$$

That is, the exponential rate function $\Lambda_U^*$ controlling large deviations of $\mathrm{Tr}(U\varphi U^*P)$ is at least as large as the corresponding function $\Lambda_\Gamma^*$ for $\frac{1}{D}\sum_{i=1}^{p}|\langle e_i|\Gamma\rangle|^2$.

The latter we have evaluated and estimated in section A: if $|\epsilon| \leq 1$, $\Lambda_\Gamma^*(1+\epsilon) \geq \frac{1}{6}\epsilon^2 p$ and the result follows by an application of the union bound. □

### B. Proof of Lemma 4

We begin by relating the trace norm to the Hilbert space norm

$$\begin{aligned} \||\psi\rangle - |\varphi\rangle\|_2^2 &= 2 - 2\mathrm{Re}\langle\psi|\varphi\rangle \\ &\geq 2 - 2|\langle\psi|\varphi\rangle| \\ &= 2\left(1 - \sqrt{F(\psi,\varphi)}\right) \\ &\geq 1 - F(\psi,\varphi) \\ &\geq \left(\frac{1}{2}\|\psi - \varphi\|_1\right)^2 \end{aligned}$$

where the last line is a well-known relation between fidelity and trace norm distance [27]. Thus, it will be sufficient to find an $\epsilon/2$-net for the Hilbert space norm. Let $\mathcal{M} = \{|\varphi_i\rangle : 1 \leq i \leq m\}$ be a maximal set of pure states satisfying $\||\varphi_i\rangle - |\varphi_j\rangle\|_2 \geq \epsilon/2$ for all $i$ and $j$. By definition, $\mathcal{M}$ is an $\epsilon/2$-net for $\|\cdot\|_2$. We can estimate $m$ by a volume argument, however. As subsets of $\boldsymbol{R}^{2D}$, the open balls of radius $\epsilon/4$ about each $|\varphi_i\rangle$ are pairwise disjoint and all contained in the ball of radius $1+\epsilon/4$ centered at the origin. Therefore,

$$m(\epsilon/4)^{2D} \leq (1+\epsilon/4)^{2D}$$

and we are done. □

## Appendix C
## Universal Quantum-Classical State Description

In Section III, we reduced universal r.s.p. with little entanglement resources to universal visible quantum data compression with the same amount of qubit resources. Here we study the latter question.

For a Hilbert space $\mathcal{K}$ of dimension $D$ a (universal) quantum-classical state compression (or quantum-classical state description) of fidelity $F$ consists of the following: first, a map

$$E : \psi \longmapsto E(\psi) = (\xi(\psi), m(\psi))$$

mapping every pure state vector $|\psi\rangle \in \mathcal{K}$ to a pair $(\xi, m)$, where $|\xi\rangle \in \mathcal{C}$ is a state vector in the (quantum) code space and $m$ is a classical message from the set $\mathcal{M}$. Second, a family of completely positive and trace-preserving linear maps

$$D_m : \mathcal{B}(\mathcal{C}) \longrightarrow \mathcal{B}(\mathcal{K})$$

such that

$$\forall\, |\psi\rangle \in \mathcal{K} \quad F(\psi, D_{m(\psi)}(\xi(\psi))) \geq F.$$

We call such a compression/description "universal" because it has to have high fidelity for every possible input pure state. Note that both the quantum and classical parts of the state description are of fixed size, in contrast to variable-length coding schemes existing in classical and quantum data compression, for which the qualifier "universal" has a quite different meaning: there it means that the encoding of a state has the minimal possible length according to some standard. Here we are interested in how the two resources we have trade against each other, in a "universal" way.

There are two extreme examples. One is "no classical message, "i.e., $|\mathcal{M}| = 1$ and a $D$-dimensional $\mathcal{C} \simeq \mathcal{K}$: for this the Sender simply prepares the desired state $\psi$ in $\mathcal{C}$. On the other end, $\dim \mathcal{C} = 1$ (i.e., no quantum message), in which case one can achieve fidelity $1-\epsilon$ by identifying an element of an $\epsilon$-net $\mathcal{M}$ in $\mathcal{K}$: by Lemma 4, this requires $\left(4 + \log \frac{1}{\epsilon}\right) D$ cbits.

The following theorem says that there occurs a jump in going from one extreme to the other, in the sense that as soon as the quantum resources are less than $\log D$ qubits, an exponential number of classical bits are needed:

*Theorem 24:* A quantum-classical state compression with average fidelity $F$,

$$\int \mathrm{d}|\psi\rangle F(\psi, D_{m(\psi)}(\xi(\psi))) \geq F,$$

which uses a code space $\mathcal{C}$ of dimension $S \leq qD$ $(q < F)$, requires exponential classical resources

$$\log|\mathcal{M}| \geq \frac{q(1-q)}{6} D - 2\log D + \log\left(1 - \sqrt{\frac{1-F}{1-q}}\right).$$

*Proof:* Write the fidelity $F = 1-\epsilon$ and define

$$\mathcal{S}_m := \{|\psi\rangle \in \mathcal{K} | \exists |\phi\rangle \in \mathcal{C} F(T_m(\phi), \psi) \geq 1 - \vartheta\},$$

the set of pure states which can be reached to fidelity $1-\vartheta$ using the message $m$ and some quantum code state. Clearly, $\Sigma := \bigcup_{m\in\mathcal{M}} \mathcal{S}_m$ is the set of all states which can be decoded with fidelity $1-\vartheta$. By Markov's inequality

$$\lambda(\Sigma) \geq 1 - \frac{\epsilon}{\vartheta} = 1 - \frac{1}{\sqrt{t}}$$

where $\lambda$ is the unique $\mathcal{U}(D)$-invariant measure on pure states, normalized to 1 (i.e., a probability measure), and with $t := \frac{1-q}{1-F} > 1$ and $\vartheta = \sqrt{t}\epsilon$.

Hence, to prove a lower bound on $|\mathcal{M}|$, it will be sufficient to prove an upper bound on the volume $\lambda(\mathcal{S}_m)$ of the sets $\mathcal{S}_m$.

We concentrate on a particular message $m$ for the time being, so we drop the subscript $m$ in the sequel. The decoding operation $T : \mathcal{B}(\mathcal{C}) \to \mathcal{B}(\mathcal{K})$ can be written, by a result of Choi [16], as

$$T(\phi) = \sum_{i=1}^{D^2} A_i \phi A_i^*$$

with linear operators $A_i : \mathcal{C} \to \mathcal{K}$. Hence, we can write

$$T(\phi) = \sum_{i=1}^{D^2} p_i \phi_i$$

with probabilities $p_i$ and pure state vectors $|\phi_i\rangle \in A_i\mathcal{C} =: \mathcal{W}_i$, the latter an (at most) $S$-dimensional subspace of $\mathcal{K}$. But if $F(T(\phi),\psi) \geq 1-\epsilon$, there must exist $i$ such that $F(\phi_i,\psi) \geq 1-\epsilon$, by bilinearity of the pure state fidelity.

Hence,

$$\mathcal{S} \subset \bigcup_{i=1}^{D^2} B_\epsilon(\mathcal{W}_i) \tag{C1}$$

with

$$B_\epsilon(\mathcal{W}) := \left\{|\psi\rangle | \exists |\phi\rangle \in \mathcal{W} \, |\langle\psi|\phi\rangle|^2 \geq 1-\epsilon\right\}$$

and it will be sufficient to bound the volume of $B_\epsilon(\mathcal{W})$ for an arbitrary $S$-dimensional subspace $\mathcal{W}$.

Denoting the orthogonal projector onto $\mathcal{W}$ by $P$, we can rewrite $B_\epsilon(\mathcal{W})$ as

$$B_\epsilon(\mathcal{W}) = \left\{|\psi\rangle | \mathrm{Tr}(|\psi\rangle\langle\psi|P) \geq 1-\epsilon\right\}.$$

Also, since the volume $\lambda$ is a probability measure, we have

$$\begin{aligned}\lambda(B_\epsilon(\mathcal{W})) &= \Pr\left\{|\psi\rangle | \mathrm{Tr}(|\psi\rangle\langle\psi|P) \geq 1-\epsilon\right\} \\ &= \Pr\left\{U | \mathrm{Tr}(U|0\rangle\langle 0|U^* P) \geq 1-\epsilon\right\}\end{aligned}$$

with $\mathcal{U}(D)$-uniformly distributed unit vector $|\psi\rangle$ and a unitary $U$ distributed according to Haar measure. Observing that the expectation of the overlap $\mathrm{Tr}(|\psi\rangle\langle\psi|P)$ above is $q$, and defining

$$\eta := \min\left\{\frac{1-\vartheta}{q} - 1, 1\right\} \geq \sqrt{t\epsilon}$$

we can use Lemma 3 to bound this probability by

$$\exp\left(-qD\frac{\eta^2}{6}\right)$$

so using the union bound in (C1) we have

$$\lambda(\mathcal{S}) \leq D^2 \exp\left(-qD\frac{\eta^2}{6}\right)$$

which implies what we wanted

$$\log|\mathcal{M}| \geq \frac{q\eta^2}{6}D - 2\log D + \log\left(1 - \sqrt{\frac{1-F}{1-q}}\right). \quad \square$$

*Remark 25:* There exists a universal quantum-classical state compression with fidelity $\geq (1-\epsilon)^2$, which uses a code space $\mathcal{C}$ of dimension $S = \lceil (1-\frac{\epsilon}{2})D \rceil$ and classical communication of $\lceil \epsilon^{-1} \rceil$ cbits.

This works as follows: decompose $\mathcal{K}$ into orthogonal subspaces $\mathcal{H}_k$ $(k = 0, \ldots, K = \lceil \epsilon^{-1} \rceil)$, such that

$$\dim\mathcal{H}_0 < \dim\mathcal{H}_1 = \cdots = \dim\mathcal{H}_K = \left\lfloor \frac{D}{K} \right\rfloor.$$

Write $P_k$ for the projectors onto the orthogonal complement of $\mathcal{H}_k$: then

$$\frac{1}{K}\sum_{k=1}^{K} P_k \geq \left(1 - \frac{1}{K}\right)\mathbb{1} \geq (1-\epsilon)\mathbb{1}$$

which means that for every state vector $\psi$ the Sender can find $1 \leq k \leq K$ such that $\mathrm{Tr}(|\psi\rangle\langle\psi|P_k) \geq 1-\epsilon$. The Sender simply transmits the projected quantum state and $k$, from which the Receiver can reconstruct $\psi$ to the desired fidelity. The rank of the $P_k$ determines $S$, which is easily estimated. $\square$

This result is in contrast to the findings of [30], where for the asymptotic compression of longer and longer products of qubits (or qu-$d$-its in general) a rate tradeoff between qubits and cbits was exhibited. In the light of the present theorem we can understand how that comes about: the model of [30] admits only *product states* in larger and larger spaces. The tradeoff curve then quantifies how efficiently the manifold of product states can be covered by (neighborhoods of) small subspaces.

Once we admit all states in dimension $D$, this covering, instead of using polynomially (in $D$) many subspaces, requires exponentially (in $D$) many!

## Appendix D
## Typical Subspaces

The following material can be found in most textbooks on information theory, e.g., [17], [19], or in the original literature on quantum information theory [34], [38], [39], [41].

For strings of length $n$ from a finite alphabet $\mathcal{X}$, which we generically denote $x^n = x_1 \ldots x_n \in \mathcal{X}^n$, we define the *type of* $x^n$ as the empirical distribution of letters in $x^n$: i.e., $P$ is the type of $x^n$ if

$$\forall\, x \in \mathcal{X} \quad P(x) = \frac{1}{n}|\{k : x_k = x\}|.$$

It is easy to see that the total number of types is upper-bounded by $(n+1)^{|\mathcal{X}|}$.

The type class of $P$, denoted $\mathcal{T}_P^n$, is defined as all strings of length $n$ of type $P$. Obviously, the type class is obtained by taking all permutations of an arbitrary string of type $P$.

The following is an elementary property of the type class:

$$(n+1)^{-|\mathcal{X}|}\exp(nH(P)) \leq |\mathcal{T}_P^n| \leq \exp(nH(P)) \tag{D1}$$

with the (Shannon) entropy $H(P)$.

For $\delta > 0$, and for an arbitrary probability distribution $P$, define the set of *$P$-typical sequences* as

$$\mathcal{T}_{P,\delta}^n := \left\{x^n : \left| -\frac{1}{n}\log P^{\otimes n}(x^n) - H(P)\right| \leq \delta\right\}.$$

By the law of large numbers, for every $\epsilon > 0$ and sufficiently large $n$

$$P^{\otimes n}\left(\mathcal{T}_{P,\delta}^n\right) \geq 1-\epsilon. \tag{D2}$$

Furthermore

$$\left|\mathcal{T}_{P,\delta}^n\right| \leq \exp(n(H(P)+\delta)) \tag{D3}$$

$$\left|\mathcal{T}_{P,\delta}^n\right| \geq (1-\epsilon)\exp(n(H(P)-\delta)). \tag{D4}$$

For a (classical) channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$ (i.e., a stochastic map, taking $x \in \mathcal{X}$ to a probability distribution $W_x$ on $\mathcal{Y}$) and a string $x^n \in \mathcal{X}^n$ of type $P$ we denote the output distribution of $x^n$ in $n$ independent uses of the channel by

$$W_{x^n}^n = W_{x_1} \otimes \cdots \otimes W_{x_n}.$$

Let $\delta > 0$, and define the set of conditional $W$-typical sequences as

$$\mathcal{T}^n_{W,\delta}(x^n) := \left\{ y^n : \left| -\frac{1}{n}\log W^n_{x^n}(y^n) - H(W|P) \right| \leq \delta \right\}$$

where $H(W|P) = \sum_x P(x)H(W_x)$ is the conditional entropy.

Once more by the law of large numbers, for every $\epsilon$ and sufficiently large $n$

$$W^n_{x^n}\left(\mathcal{T}^n_{W,\delta}(x^n)\right) \geq 1-\epsilon. \tag{D5}$$

Furthermore

$$\left|\mathcal{T}^n_{W,\delta}(x^n)\right| \leq \exp(n(H(W|P)+\delta)) \tag{D6}$$

$$\left|\mathcal{T}^n_{W,\delta}(x^n)\right| \geq (1-\epsilon)\exp(n(H(W|P)-\delta)). \tag{D7}$$

All of these concepts and formulas have analogues as "typical projectors" $\Pi$ for quantum state: by virtue of the spectral decomposition, the eigenvalues of a density operator can be interpreted as a probability distribution over eigenstates. The subspaces spanned by the typical eigenstates are the "typical subspaces." The trace of a density operator with one of its typical projectors is then the probability of the corresponding set of typical sequences.

Notations like $\Pi^n_{\rho,\delta}$, $\Pi^n_{\varphi,\delta}(x^n)$, etc., for a state $\rho$ and a cq-channel $\varphi : x \longmapsto \varphi_x$ should be clear from this.

There is only one such statement for density operators that we shall use, which is not of this form.

*Lemma 26 (Operator Law of Large Numbers):* Let $x^n \in \mathcal{X}^n$ be of type $P$, and let $W : \mathcal{X} \longrightarrow \mathcal{S}(\mathcal{H})$ be a cq-channel. Denote the average output state of $W$ under $P$ as

$$\rho = \sum_x P(x)W_x.$$

Then, for every $\epsilon > 0$ and sufficiently large $n$

$$\mathrm{Tr}\left(W^n_{x^n}\Pi^n_{\rho,\delta}\right) \geq 1-\epsilon.$$

*Proof:* See [41], Lemma 6. □

## Appendix E
## A Possible Operational Reduction of R.S.P. to Q.C.T.

Our protocol in Section IV for (asymptotic) remote state preparation of ensembles reduces the problem to the quantum-classical tradeoff in visible source coding [30] by an operational reduction: we simply add our universal r.s.p. protocol, Theorem 5, on top of the q.c.t. coding, Theorem 9. The optimality proof, though modeled closely along the lines of the corresponding proof in [30], is however completely independent. It would be desirable to have a closer connection between the trading of qubits versus cbits and of ebits versus cbits, and in this appendix we describe an operational link going the other way, from r.s.p. to q.c.t., resting on an (as yet unproven) conjecture on mixed-state compression.

More precisely, given an r.s.p. protocol (asymptotic and approximate) of cbit rate $C$ and ebit rate $E$ construct a q.c.t. scheme with cbit rate $R = C - E$ and qubit rate $Q = E$. This would exactly revert the construction of Section IV.

We will prove that this is possible, assuming the following conjecture (see [3] and [42]).

*Conjecture 27:* Given an i.i.d. source $\mathcal{F} = \{p_i, \rho_i\}$ of *mixed states* it is possible to visibly compress the source asymptotically and approximately, using shared randomness between Sender and Receiver, and communicating qubits at rate

$$\chi(\{p_i;\rho_i\}) = S\left(\sum_i p_i\rho_i\right) - \sum_i p_i S(\rho_i).$$

Note that this is true if the ensemble consists of pure states, by Schumacher's quantum data compression [38]. Also observe that the conjecture certainly is true for commuting mixed states: this is essentially the content of the Reverse Shannon Theorem [10], see also [42].

Note (as we have observed earlier) that shared randomness can safely be assumed free, because we are considering an average pure state fidelity as quality measure of the protocol.

We assume the following general form of our r.s.p. protocol: it uses a standard maximally entangled state $|\Phi\rangle$ on $\mathcal{K}_A \otimes \mathcal{K}_B$, with $\dim \mathcal{K} \leq 2^{n(E+\delta)}$. Depending on $I = i_1 \ldots i_n$ the Sender makes a measurement on $\mathcal{K}_A$, described by a POVM $\boldsymbol{A}^{(I)} = (A^{(I)}_m)$, where $m$ is the message she subsequently sends to the Receiver, chosen from a set of $M \leq 2^{n(R+\delta)}$. Of course, as $n$ tends to infinity, $\delta$ will tend to zero. For each of the messages $m$, the Receiver can execute an operation $T_m$ on $\mathcal{K}_B$, acting on the state induced by the entanglement and the measurement, together with the outcome, denoted $\rho_{m|I}$. Denote the induced probability of the message $m$ (given $I$) as $q(m|I)$. We shall only assume the "local" fidelity condition, (3), not the stronger "global" one, (2).

Our goal is to re-enact the creation of the post-measurement state and the transmission of the classical message using only cbit and qubit communications. The key idea comes from the observation that there is noise in the system due to the uncontrollable randomness of the POVMs. We want to transfer the generation of this noise to the shared randomness.

We shall now look at blocks formed from the $n$-blocks given by the assumed r.s.p. protocol. We use the previous notation $I = i_1 \ldots i_n$ for an $n$-block, and introduce $I^N = I_1 \ldots I_N$ for such a block of blocks. By the Reverse Shannon Theorem (in the formulation of [42]) we can visibly encode the distribution $q(\cdot|I^N)$, at least for typical $I^N$, using shared randomness and communicating

$$\begin{aligned} \mathrm{I}(I:m) &= H(m) - \sum_I p_I H(q(\cdot|I)) \\ &\leq n(R+\delta) - \sum_I p_I H(q(\cdot|I)) \end{aligned}$$

cbits per $n$-block, where we treat $I$ and $m$ as jointly distributed random variables

$$\Pr\{I,m\} = p_I q(m|I)$$

with $H$ as the usual Shannon entropy, and $I$ the Shannon mutual information.

A feature of the Reverse Shannon Theorem that was noted earlier is that the Sender gets full feedback, i.e., she obtains the very (random) message $m$ the Receiver gets out. With the help of this feedback, she just prepares the post-measurement state on $\mathcal{K}$ that otherwise the Receiver would have found on his half of the entanglement, and sends it. Then, obviously, the Receiver

can proceed as in the r.s.p. protocol. It is clear, that we end up with a procedure having high fidelity according to the "local" fidelity criterion (3), now over a block of length $Nn$.

How does this behave in terms of resources? Clearly, we now use only qubits and cbits. Inspection of the above formulas reveals that all is fine if

$$\sum_I p_I H(q(\cdot|I)) \geq n(E-\delta') \tag{E1}$$

with $\delta' = o(1)$ as $n \to \infty$. Because then we have a q.c.t. scheme (satisfying (3)) that uses $Nn(E+\delta)$ qubits and $Nn(R-E+\delta+\delta')$ cbits. This is exactly the reduction we wanted: since in [30] the tradeoff curve was (implicitly) proved for the criterion (3), we obtain the desired bounds on $E$ and $R$.

We are left with proving that assuming the negation of (E1) leads to a contradiction: so, introducing the tripartite state

$$\omega = \sum_I p_I |I\rangle\langle I|^A \otimes \sum_m q(m|I)\rho^B_{m|I} \otimes |m\rangle\langle m|$$

for notational convenience, assume that there exists $\Delta E > 0$ such for all large $n$

$$S(B:C|A) \leq S(C|A) \leq n(E-\Delta E). \tag{E2}$$

The right inequality is the negation of (E1), and the left is by data processing: for each value of $I$ in $A$ the information between $B$ and $C$ (which is the Holevo quantity of the ensemble $\{q(\cdot|I), \rho_{\cdot|I}\}$) is upper-bounded by the entropy of $C$, i.e., $H(q(\cdot|I))$.

Note further that, because $\sum_m q(m|I)\rho_{m|I}$ equals the maximally mixed state for all $I$, we have $S(A:B)=0$, hence, by the chain rule for quantum mutual information

$$S(B:C|A) = S(AC:B).$$

Thus, for large enough $N$, we can, by Conjecture 27, encode $N$-blocks of the $\rho_{m|I}$ using shared randomness and sending

$$\begin{aligned} NS(AC:B) + o(N) &= NS(B:C|A) + o(N) \\ &\leq Nn(E-\Delta E) + o(N) \end{aligned}$$

qubits: the conjecture is applied to the ensemble $\{p_I q(m|I); \rho_{m|I}\}$, which partly is given (the input, $I$) and which partly is obtained by simulating the noisy classical channel $q(\cdot|\cdot)$ (the variable $m$). Observe that $m$ is by this method generated simultaneously at the Sender and at the Receiver.

Switching back to r.s.p. via (6) we end up with a protocol on $Nn$-blocks using only $Nn(E-\Delta E)+o(N)$ ebits and

$$\begin{aligned} &NS(A:C) + NS(B:C|A) + o(N) \\ &\qquad = NS(AB:C) + o(N) \leq Nn(R+o(1)) \end{aligned}$$

cbits: the first term is due to the communication cost of the Reverse Shannon Theorem, and the second is the cost overhead to remotely prepare the $NS(B:C|A)+o(N)$ qubits of the compressed mixed states. In the limit, this leads to a rate pair $(R, E-\Delta E)$, contradicting the optimality of $(R,E)$. □

## APPENDIX F
## MISCELLANEOUS PROOFS

### A. Proof of Lemma 12

For finiteness of the values of $N$ we have to have

$$R \geq S(A:B) = S(B)$$

which is clearly sufficient. For $N(\mathcal{E}, R) = 0$, on the other hand, one has to have a state with

$$0 = S(A:B|C) = S(B|C).$$

But then

$$\begin{aligned} R &\geq S(A:BC) \\ &= S(A:C) + S(A:B|C) = S(A:C). \end{aligned}$$

However, $S(B|C)=0$ says that $B$ is in a pure state given $C$, which is only possible if $S(A|C)=0$. Hence, $R \geq S(A)$, which clearly is sufficient, too.

For convexity, let $\omega_1$ be optimal for $R_1$, $\omega_2$ optimal for $R_2$, i.e.,

$$\begin{aligned} S_{\omega_k}(A:B|C) &= N(R_k) \\ S_{\omega_k}(A:BC) &\leq R_k \end{aligned}$$

$k=1,2$. Furthermore, let $0 \leq \lambda \leq 1$. Then form the state

$$\omega = \lambda\omega_1 \otimes |1\rangle\langle 1|^{C'} + (1-\lambda)\omega_2 \otimes |2\rangle\langle 2|^{C'}.$$

By definition (with $\widetilde{C} = CC'$)

$$\begin{aligned} S_\omega(A:B\widetilde{C}) &\leq \lambda R_1 + (1-\lambda)R_2 \\ S_\omega(A:B|\widetilde{C}) &= \lambda N(R_1) + (1-\lambda)N(R_2) \end{aligned}$$

and thus, the minimization yields

$$N(\lambda R_1 + (1-\lambda)R_2) \leq \lambda N(R_1) + (1-\lambda)N(R_2).$$

Taking $R_1 = S(B)$ and $R_2 = S(A)$, we obtain that in the interval $[S(B); S(A)]$ is strictly decreasing and continuous—otherwise, there would be a contradiction to convexity. (Note that $N(R_2)=0$!)

Finally, for the additivity relation (8), observe that "$\leq$" is almost obvious: if $\omega_k$ are optimal for $(\mathcal{E}_k, R_k)$, $k=1,2$, it is immediate to check that $\omega = \omega_1 \otimes \omega_2$ is feasible for $(\mathcal{E}_1 \otimes \mathcal{E}_2, R = R_1 + R_2)$, implying an upper bound of $N(\mathcal{E}_1, R_1) + N(\mathcal{E}_2, R_2)$ for $N(\mathcal{E}_1 \otimes \mathcal{E}_2, R)$.

In the other direction, let $\omega$ be optimal for $(\mathcal{E}_1 \otimes \mathcal{E}_2, R)$

$$\omega = \sum_{i,i'} p_i p'_{i'} |i\rangle\langle i|^{A_1} \otimes |i'\rangle\langle i'|^{A_2} \pi_i^{B_1} \otimes \pi_{i'}^{B_2} \otimes \sum_j p(j|ii')|j\rangle\langle j|^C.$$

First, by the chain rule and data processing

$$\begin{aligned} R &\geq S(A_1A_2 : B_2B_2C) \\ &= S(A_1 : B_1B_2C) + S(A_2 : B_1B_2C|A_1) \\ &\geq S(A_1 : B_1C) + S(A_2 : B_2C|A_1). \end{aligned}$$

Thus, we can write $R = R_1 + R_2$ such that

$$S(A_1 : B_1C) \leq R_1, \quad S(A_2 : B_2C|A_1) \leq R_2. \tag{F1}$$

Second, by a similar reasoning

$$\begin{aligned} N(\mathcal{E}_1 \otimes \mathcal{E}_2, R) &= S(A_1A_2 : B_2B_2|C) \\ &= S(A_1 : B_1B_2|C) + S(A_2 : B_1B_2|CA_1) \\ &\geq S(A_1 : B_1|C) + S(A_2 : B_2|CA_1). \end{aligned}$$

Here, the first term is $\geq N(\mathcal{E}_1, R_1)$ by definition, using (F1). The second term is similarly $\geq N(\mathcal{E}_2, R_2)$, using additionally the convexity of $N$. □

### *B. Proof of Lemma 19*

Monotonicity follows directly from the definition.

For finite values we obviously have to have

$$R \geq S(X : BC) \geq S(X : B).$$

Also always (using that the conditional entropy can only increase under quantum operations—a consequence of strong subadditivity)

$$E \geq S(B|C) \geq S(B|X)$$

with equality when $C$ contains a copy of $X$.

Convexity is proved exactly as in the proof of Lemma 12. From this continuity in the domain of finite values follows, as well as strict monotonicity as long as $N(R) > S(B|X)$.

It remains to prove the additivity relation (18): "$\leq$" is the trivial inequality, after the pattern of the proof of Lemma 12. As for "$\geq$," consider an optimal state $\omega$ for $\mathcal{E}_1 \otimes \mathcal{E}_2$ and rate $R$

$$\omega = \sum_{i,i'} p_i p_{i'} |i\rangle\langle i|^{X_1} \otimes |i'\rangle\langle i'|^{X_2} \otimes \varphi_i^{A_1B_1} \otimes \varphi_{i'}^{A_2B_2} \otimes \sum_j p(j|ii')|j\rangle\langle j|^C.$$

Then, using the chain rule, data processing, and the independence of $X_1$ and $X_2$

$$\begin{aligned} R &\geq S(X_1X_2 : B_1B_2C) \\ &= S(X_1 : B_1B_2C) + S(X_2 : B_1B_2C|X_1) \\ &\geq S(X_1 : B_1C) + S(X_2 : B_2C|X_1) \\ &= S(X_1 : B_1C) + S(X_2 : B_2CX_1) \end{aligned}$$

so we can find $R_1$ and $R_2$ such that $R_1 + R_2 = R$ and

$$S(X_1 : B_1C) \leq R_1, \quad S(X_2 : B_2CX_1) \leq R_2.$$

On the other hand

$$\begin{aligned} N(\mathcal{E}_1 \otimes \mathcal{E}_2) &= S(B_1B_2|C) \\ &= S(B_1|C) + S(B_2|CB_1) \\ &\geq S(B_1|C) + S(B_2|CX_1) \\ &\geq N(\mathcal{E}_1, R_1) + N(\mathcal{E}_2, R_2) \end{aligned}$$

where in the third line we have used that conditional entropy can only increase under quantum operations, a consequence of strong subadditivity. The last line follows because with our choice of $R_1$ and $R_2$, $C$ and $CX_1$ are permitted in the definition of $N(\mathcal{E}_1, R_1)$ and $N(\mathcal{E}_2, R_2)$, respectively. □

## ACKNOWLEDGMENT

The authors wish to thank Anura Abeyesinghe, Igor Devetak, Chris Fuchs, Aram Harrow, Daniel Gottesman, and John Smolin for interesting and helpful conversations. P. Hayden, D. W. Leung, and A. Winter gratefully acknowledge the hospitality and support of the Mathematical Sciences Research Institute, Berkeley, during part of the autumn term of 2002.

## REFERENCES

[1] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 569–579, Mar. 2002.

[2] A. Ambainis, M. Mosca, A. Tapp, and R. de Wolf, "Private quantum channels," in *Proc. 41st Annu. Symp. Foundations of Computer Science (FOCS)*, Redondo Beach, CA, Nov. 12–14, 2000, pp. 547–553.

[3] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. W. Schumacher, "On quantum coding for ensembles of mixed states," *J. Phys. A: Math. and Gen.*, vol. 34, no. 35, pp. 6767–6785, 2001.

[4] C. H. Bennett and S. Wiesner, "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states," *Phys. Rev. Lett.*, vol. 69, pp. 2881–2884, 1992.

[5] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Phys. Rev. Lett.*, vol. 70, no. 13, pp. 1895–1899, 1993.

[6] C. H. Bennett, I. Devetak, A. Harrow, P. W. Shor, and A. Winter, "The Quantum Reverse Shannon Theorem," in preparation.

[7] C. H. Bennett, D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, and W. K. Wootters, "Remote state preparation," *Phys. Rev. Lett.*, vol. 87, no. 077902, 2001. "Erratum," *Phys. Rev. Lett.*, vol. 88, 099902, 2002.

[8] C. H. Bennett, A. W. Harrow, D. W. Leung, and J. A. Smolin, "On the Capacities of Bipartite Hamiltonians and Unitary Gates," e-print, quant-ph/0205057, 2002.

[9] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. Winter, "Randomizing quantum states: constructions and applications," *Commun. Math. Phys.*, vol. 250, pp. 371–391, 2004.

[10] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted classical capacity of noisy quantum channels," *Phys. Rev. Lett.*, vol. 83, no. 15, pp. 3081–3084, 1999.

[11] ——, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002.

[12] D. W. Berry and B. C. Sanders, "Optimal remote state preparation," *Phys. Rev. Lett.*, vol. 90, no. 057901, 2003.

[13] P. O. Boykin and V. Roychowdhury, "Optimal encryption of quantum bits," *Phys. Rev. A*, vol. 67, no. 042317, 2003.

[14] N. Cerf, N. Gisin, and S. Massar, "Classical teleportation of a quantum bit," *Phys. Rev. Lett.*, vol. 84, no. 11, pp. 2521–2524, 2000.

[15] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, pp. 493–507, 1952.

[16] M.-D. Choi, "Completely positive linear maps on complex matrices," *Lin. Algebra and Appl.*, vol. 10, pp. 285–290, 1975.

[17] T. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[18] H. Cramér, "Sur un nouveau théorème-limite de la theorie des probabilités," in *Actualités Scientifiques et Industrielles (Colloque consacré à la theorie des probabilités no. 736)*. Paris, France: Hermann, 1938, pp. 5–23.

[19] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[20] E. B. Davies and J. T. Lewis, "An operational approach to quantum probability," *Commun. Math. Phys.*, vol. 17, pp. 239–260, 1970.

[21] A. Dembo and O. Zeitouni, *Large Deviations: Techniques and Applications*, 2nd ed, ser. Applications of Mathematics 38. New York: Springer-Verlag, 1998.
[22] I. Devetak and T. Berger, "Low-entanglement remote state preparation," *Phys. Rev. Lett.*, vol. 87, no. 197901, 2001.
[23] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal, "Quantum data hiding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 580–599, Mar. 2002.
[24] D. P. DiVincenzo, P. Hayden, and B. M. Terhal, "Hiding Quantum Data," e-print, quant-ph/0207147, 2002.
[25] M. Fannes, "A continuity property of the entropy density for spin lattice systems," *Commun. Math. Phys.*, vol. 31, pp. 291–294, 1973.
[26] W. Feller, *An Introduction to Probability Theory and its Applications*, 3rd ed. New York: Wiley, 1968, vol. I.
[27] C. A. Fuchs and J. van de Graaf, "Cryptographic distinguishability measures for quantum-mechanical states," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1216–1227, May 1999.
[28] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof systems," *SIAM J. Comput.*, vol. 18, no. 1, pp. 186–208, 1989.
[29] A. Hayashi, T. Hashimoto, and M. Horibe, "Remote state preparation without oblivious condition," *Phys. Rev. A*, vol. 67, no. 5, 052302, 2003.
[30] P. Hayden, R. Jozsa, and A. Winter, "Trading quantum for classical resources in quantum data compression," *J. Math. Phys.*, vol. 43, no. 9, pp. 4404–4444, 2002.
[31] A. Harrow, "Coherent classical communication," *Phys. Rev. Lett.*, vol. 92, no. 9, 097902, Mar. 2004.
[32] R. Jozsa, "Fidelity for mixed quantum states," *J. Mod. Opt.*, vol. 41, no. 12, pp. 2315–2323, 1994.
[33] E. Kushilevitz and N. Nisan, *Communication Complexity*. Cambridge, U.K.: Cambridge Univ. Press, 1996.
[34] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Mod. Opt.*, vol. 41, no. 12, pp. 2343–2349, 1994.
[35] D. W. Leung and P. W. Shor, "Oblivious remote state preparation," *Phys. Rev. Lett.*, vol. 90, no. 12, 127905, 2003.
[36] H.-K. Lo, "Classical communication cost in distributed quantum information processing—A generalization of quantum communication complexity," *Phys. Rev. A*, vol. 62, no. 012313, 2000.
[37] A. K. Pati, "Minimum classical bit for remote preparation and measurement of a qubit," *Phys. Rev. A*, vol. 63, no. 014302, 2001.
[38] B. W. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, no. 4, pp. 2738–2747, 1995.
[39] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, pp. 131–138, 1997.
[40] A. Uhlmann, "The 'transition probability' in the state space of a $^*$-algebra," *Rep. Math. Phys.*, vol. 9, no. 2, pp. 273–279, 1976.
[41] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2481–2485, Nov. 1999.
[42] ——, "Compression of Sources of Probability Distributions and Density Operators," e-print, quant-ph/0208131, 2002.
[43] B. Zeng and P. Zhang, "Remote-state preparation in higher dimension and the parallelizable manifold $S^{n-1}$," *Phys. Rev. A*, vol. 65, no. 2, 022316, 2002.