Time Reversal and Exchange Symmetries of Unitary Gate Capacities

Time Reversal and Exchange Symmetries of Unitary Gate Capacities

Aram W. Harrow and Peter W. Shor

***Abstract*—Unitary gates are interesting resources for quantum communication in part because they are always invertible and are intrinsically bidirectional. This paper explores these two symmetries: time-reversal and exchange of Alice and Bob. We will present examples of unitary gates that exhibit dramatic separations between forward and backward capacities (even when the back communication is assisted by free entanglement) and between entanglement-assisted and unassisted capacities, among many others. Along the way, we will give a general time-reversal rule for relating the capacities of a unitary gate and its inverse that will explain why previous attempts at finding asymmetric capacities failed. Finally, we will see how the ability to erase quantum information and destroy entanglement can be a valuable resource for quantum communication.**

***Index Terms*—Asymmetric, capacity, coherent, erase, quantum, reverse, unitary.**

I. Introduction: Communication Using Bipartite Unitary Gates

THIS paper investigates the asymptotic communication capacities of bipartite unitary quantum gates; for example, a controlled-NOT (CNOT) gate with control qubit held by Alice and target qubit held by Bob. For a review of this topic, see [7], [23] and references therein. The question of unitary gate capacity arises when studying our ability to communicate or generate entanglement using naturally occurring physical interactions; moreover, studying unitary gate capacity has often led to new ideas that are useful for other topics in quantum information theory[22].

In some ways unitary gates are like classical bidirectional channels or noisy quantum channels, but they are both more complex than one-way channels because of their intrinsic bidirectionality, and simpler than noisy channels because they involve no interactions with the environment. For example, any nonlocal unitary gate has nonzero capacities to send classical messages in either direction and to create entanglement [3], [7]. By contrast, bidirectional classical channels exist that have no capacity in either direction, but can be useful for nonlocal tasks like reducing communication complexity [15]. Even deterministic classical bidirectional channels, like the classical CNOT, can have capacities that are nonzero only in one direction. Another feature of unitary gates is that, unlike noisy quantum channels, knowing the classical capacity (as a function of the amount of entanglement assistance) of a unitary gate also determines its quantum capacity (again parameterized by the amount of entanglement assistance). Moreover, allowing free classical communication does not improve the entanglement capacity; on the other hand, the quantum capacity appears to no longer be simply equal to the entanglement generating capacity. In short, the usual questions (like additivity) about noisy channel capacities are replaced by an intriguingly different, yet perhaps related, set of questions about unitary gate capacities.

In this paper, we will investigate the questions of symmetry, both time-reversal and exchange of Alice and Bob, that arise in connection with unitary gate capacities. We will demonstrate the following.

- A general rule for relating capacity regions of a gate U to those of its inverse $U^\dagger$ (Section II). Along the way, we recast the main result of [25] as a sort of structure theorem for communication protocols based on unitary gates, which leads us to propose a new way to view the capacity region of unitary gates.
- A gate that exhibits nearly the strongest possible separation between forward and backward capacities, even when free entanglement is allowed for back communication (Section III).
- A gate with a nearly maximal separation between its ability to create and to destroy entanglement. This gate also exhibits a near-maximal improvement in communication capacity when assisted by entanglement (Section IV). (A variant of the former result was independently proved for a different gate in [38].)
- A quantum communication resource, "coherent erasure," that can be thought of as the time reversal of coherent classical communication (Section V).
- A more restricted type of resource inequality, which we call a "clean resource inequality" (described in more detail below).
- Alternate proofs for the two main unitary gate capacity theorems that are currently known (the Appendix). The proofs are simpler and establish slightly stronger versions of the capacity theorems; we also include them because they make this paper a self-contained summary of almost every result to date on asymptotic unitary gate capacities.

As we will see, a number of speculative claims about unitary gate capacities remain to be fully resolved, and more interestingly, we are only beginning to pose our questions about them

Manuscript received June 15, 2008; revised May 16, 2009. Current version published December 23, 2009. The work of A. W. Harrow is funded by the U.K. EPRSC under grant "QIP IRC" and the QAP project (under Contract IST-2005-15848). The work of P. W. Shor is funded by NSF under Grant CCF-0431787: "Quantum Channel Capacities and Quantum Complexity."

A. W. Harrow is with the Department of Mathematics, University of Bristol, Bristol, BS8 1TW, U.K. (e-mail: a.harrow@bris.ac.uk).

P. W. Shor is with the Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: shor@math.mit.edu).

Communicated by A. Winter, Associate Editor for Quantum Information Theory.

Digital Object Identifier 10.1109/TIT.2009.2034899

in the right way. We conclude in Section VI with some ideas about future research.

The remainder of this section reviews notation and some background results. Following [19], [23], we state our coding theorems in the language of asymptotic resource inequalities. The basic asymptotic resources are $[c \to c]$ (one use of a noiseless classical channel from Alice to Bob, also known as (a.k.a.) a cbit), $[qq]$ (the state $|\Phi\rangle^{AB} = \frac{1}{\sqrt{2}} \sum_{x=0}^{1} |x\rangle^A |x\rangle^B$, a.k.a. an ebit), and $[q \to q]$ (one use of a noiseless quantum channel, a.k.a. a qubit). Protocols transforming these resources into one another (e.g., teleportation) are expressed as asymptotic resource inequalities such as $2[c \to c] + [qq] \geq [q \to q]$. We will also make use of coherent bits, or cobits, which are denoted $[q \to qq]$ and correspond to the isometry $\sum_{x=0}^{1} |x\rangle^A |x\rangle^B \langle x|^A$. Coherent bits were introduced in [22] which proved that $2[q \to qq] = [q \to q] + [qq]$ (though only as an asymptotic relation; see [21] for a single-shot version). Since we are interested in two-way communication, define $[c \leftarrow c]$, $[q \leftarrow q]$, and $[qq \leftarrow q]$ to be cbits, qubits, and cobits, respectively, sent from Bob to Alice. These definitions are summarized in Section VII.

For a unitary gate U, let $\langle U \rangle$ denote the corresponding asymptotic resource; we can use it to state resource inequalities such as $\langle \text{CNOT} \rangle \geq [c \to c]$. Define $\text{CCE}(U)$ to be the three-dimensional capacity region of U to send cbits forward, send cbits backwards, and generate entanglement:

$$\begin{aligned}&\text{CCE}(U)\\ &:= \{(C_1, C_2, E) : \langle U \rangle \geq C_1[c \to c] + C_2[c \leftarrow c] + E[qq]\}. \end{aligned} \quad (1)$$

For example, if we define SWAP to exchange a qubit of Alice's with a qubit of Bob's, then $(1, 1, 0) \in \text{CCE}(\text{SWAP})$, since one use of SWAP can be used to send one bit forward and another bit backwards at the same time. When C_1, C_2, E take on negative values, we move the corresponding resources to the left-hand side of the resource inequality, e.g., $E < 0$ for entanglement-assisted communication. Continuing with the SWAP example, we could use superdense coding to consume 2 ebits and send 2 cbits in either direction: thus, $(2, 2, -2) \in \text{CCE}(\text{SWAP})$. We can define capacities in terms of $\text{CCE}(U)$ as extremal points of the region: the entanglement capacity $E(U) := \max\{E : (0, 0, E) \in \text{CCE}(U)\}$, the forward classical capacity $C_{\to}(U) := \max\{C : (C, 0, 0) \in \text{CCE}(U)\}$, the backwards capacity $C_{\leftarrow}(U) := \max\{C : (0, C, 0) \in \text{CCE}(U)\}$, the simultaneous capacity $C_{+}(U) := \max\{C_1 + C_2 : (C_1, C_2, 0) \in \text{CCE}(U)\}$, and entanglement-assisted versions $C_{\to}^{E}(U) := \max\{C : (C, 0, -\infty) \in \text{CCE}(U)\}$, $C_{\leftarrow}^{E}(U) := \max\{C : (0, C, -\infty) \in \text{CCE}(U)\}$ and $C_{+}^{E}(U) := \max\{C_1 + C_2 : (C_1, C_2, \infty) \in \text{CCE}(U)\}$. In [3], [7], it was shown that one of these capacities is nonzero if and only if all of them are nonzero. Various quantitative relations among these capacities were also shown, but they will be subsumed in what follows.

We can analogously define the capacity region of achievable rates of entanglement generation and *coherent* communication in both directions. In [25], this region was shown to coincide with $\text{CCE}(U)$ for the $C_1, C_2 \geq 0$ quadrant, and to be trivially related for other quadrants. Here we will present this result in a slightly stronger form. If $\alpha = (\alpha_n)_{n=1}^{\infty}$ and $\beta = (\beta_n)_{n=1}^{\infty}$ are two *pure*[1] asymptotic resources such that $\alpha \geq \beta$, then we say that this resource inequality is clean (denoted $\alpha \overset{\text{clean}}{\geq} \beta$) if α_n can be mapped to $\beta_{n(1-\delta_n)}$ using a protocol that discards only $n\delta'_n$ qubits, which (up to error ϵ_n) are all in the state $|0\rangle$, where $\epsilon_n, \delta_n, \delta'_n \to 0$ as $n \to \infty$.[2] We will also call protocols "semi-clean" when, at the end of the protocol, they discard arbitrary $n\delta'_n$-qubit states which depend only on n and not on any other inputs or outputs of the protocol. In most cases of interest, semi-clean protocols are also clean.

Implicit in the definition of a clean protocol is the idea that all local quantum operations are represented by isometries (perhaps increasing the dimension) followed by discarding some qubits. The advantage of this formulation is that apart from the discarding step, protocols can be easily reversed. On the other hand, requiring that resources be pure is quite a restrictive condition, and hopefully future work will able to fruitfully relax it. Many common resource inequalities, such as entanglement concentration/dilution, remote state preparation, channel coding, etc., ..., can be shown to admit "clean" versions, but other simple inequalities such as $2[qq] \geq [qq]$ do not. Now define

$$\begin{aligned}&\text{clCCE}(U)\\ &:= \{(C_1, C_2, E) : \langle U \rangle \overset{\text{clean}}{\geq} C_1[q \to qq] + C_2[qq \leftarrow q] + E[qq]\}. \end{aligned} \quad (2)$$

References [23], [25] considered the similar region $\text{CoCoE}(U)$ in which there was no requirement that the protocols be clean. Although $\text{clCCE}(U)$ is still convex, it is no longer monotone in the sense that throwing away resources does not always yield valid protocols; for example, while points like $(0, 0, -\infty)$ are in $\text{CCE}(U)$, one can show that

$$E(U^\dagger) = \max\{E : (0, 0, -E) \in \text{clCCE}(U)\}.$$

This result can be proven directly using the formula for $E(U)$ in [7], [37], and will also follow from a more general theorem relating $\text{clCCE}(U)$ to $\text{clCCE}(U^\dagger)$ that we prove in Section II.

The main use of clCCE in this paper will be the following strengthening of [25]'s main result.

Theorem 1 (Standard Form of Unitary Protocols): If

$$\langle U \rangle \geq C_1[c \to c] + C_2[c \leftarrow c] + E[qq]$$

and $C_1, C_2 \geq 0$, then there exists $E' \geq E$ such that

$$\langle U \rangle \overset{\text{clean}}{\geq} C_1[q \to qq] + C_2[qq \leftarrow q] + E'[qq].$$

In other words, there exists a series of protocols $(\mathcal{P}_n)_{n=1}^{\infty}$, with

$$\mathcal{P}_n = (A_n^{(n)} \otimes B_n^{(n)}) U \ldots U (A_1^{(n)} \otimes B_1^{(n)}) U (A_0^{(n)} \otimes B_0^{(n)}) \quad (3)$$

[1]Following [19] we say that a resource is pure if it is an isometry or a pure state.

[2]The idea that discarding information should be costly dates back to Szilard's interpretation in 1929 of Maxwell's demon[45] (see also [4]), and in the context of quantum Shannon theory has been discussed in [41].

for local isometries $A_0^{(n)}, B_0^{(n)}, \ldots, A_n^{(n)}, B_n^{(n)}$ (possibly adding ancillas), such that for all $x \in \{0,1\}^{n(C_1-\delta_n)}, y \in \{0,1\}^{n(C_2-\delta_n)}$, we have

$$\mathcal{P}_n |x\rangle^A |y\rangle^B \overset{\epsilon_n}{\approx} |x,y\rangle^A |x,y\rangle^B (|\Phi\rangle^{AB})^{\otimes n(E-\delta_n)} (|00\rangle^{AB})^{\otimes n\delta_n'}. \quad (4)$$

Here, $\epsilon_n, \delta_n, \delta_n' \to 0$ as $n \to \infty$ and $\rho \overset{\epsilon}{\approx} \sigma$ means that $\frac{1}{2}||\rho - \sigma||_1 \leq \epsilon$.

In terms of CCE and clCCE, this means that for any $(C_1, C_2, E) \in$ CCE there exists $E' \geq E$ such that $(C_1, C_2, E') \in$ clCCE.

Reference [23] sketched how to extend the proof of [25] to obtain the above theorem, but we will make use of [24] to give a more rigorous derivation in the Appendix .

Finally, we state a single-shot expression for the tradeoff curve between ebits and cbits sent from Alice to Bob [22], [23]; call this tradeoff curve $\mathrm{CE}(U)$ and define it to be $\mathrm{CE}(U) := \{(C,E) : (C,0,E) \in \mathrm{CCE}(U)\}$. Similarly, define $\mathrm{clCE}(U) := \{(C,E) : (C,0,E) \in \mathrm{clCCE}(U)\}$. Before we can state our expression for $\mathrm{clCE}(U)$, we will need a few more definitions (following [19]). For a state $\psi^{ABC} = |\psi\rangle\langle\psi|^{ABC}$, recall the definition of the von Neumann entropy as $H(A) = H(A)_\psi = H(\psi^A) = -\mathrm{Tr}(\psi^A \log \psi^A)$, where $\psi^A = \mathrm{Tr}_{BC}\, \psi^{ABC}$. Similarly, the quantum mutual information is $I(A;B) = H(A) + H(B) - H(AB)$ and the conditional information is defined as $H(B|A) = H(AB) - H(A)$. Here, and elsewhere, we omit subscripts when the underlying state is obvious. We will denote an ensemble of pure states by

$$\mathcal{E}^{XABA'B'} = \sum_x p_x |x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^{ABA'B'}.$$

Here X is a classical label, the gate U acts on AB, and $A'B'$ are ancilla systems of arbitrary finite dimension. Let $U(\mathcal{E})$ stand for $(U^{AB} \otimes I^{XA'B'})(\mathcal{E})$. In terms of these ensembles we can define

$$\begin{aligned}\Delta_{I,E}(U) := \{(C,E) : \exists \mathcal{E} \text{ s.t.} \\ I(X;BB')_{U(\mathcal{E})} - I(X;BB')_{\mathcal{E}} \geq C \\ \text{and } H(BB'|X)_{U(\mathcal{E})} - H(BB'|X)_{\mathcal{E}} = E\}\end{aligned} \quad (5)$$

where $\mathcal{E}$ is an ensemble of bipartite pure states in $ABA'B'$ conditioned on a classical register X. This corresponds to the set of single-shot increases in mutual information ($I(X;BB')$) and average entanglement ($H(BB'|X)$) that are possible. It turns out that these increases are also achievable asymptotically, as expressed in the following theorem.

Theorem 2: $\mathrm{clCE}(U)$ is equal to the closure of $\Delta_{I,E}(U)$.

The direct coding theorem was proven in [22] and the converse in [23, Section 3.4.2]. In the Appendix , we will give a new, and more self-contained, proof of the coding theorem.

II. Reversing Unitary Communication Protocols

In this section, we present a general theorem for relating the capacity region of U with the capacity region of $U^\dagger$. Many of the key ideas are illustrated by the gate U_{XOXO}, which was conjectured in [7] to have asymmetric communication capacities. U_{XOXO} acts on a $2^m \times 2^m$-dimensional space, with m a parameter, and is defined as

$$\begin{aligned} U_{\mathrm{XOXO}} |x0\rangle &= |xx\rangle \qquad \forall x \in \{0,1\}^m \\ U_{\mathrm{XOXO}} |xx\rangle &= |x0\rangle \qquad \forall x \in \{0,1\}^m \\ U_{\mathrm{XOXO}} |xy\rangle &= |xy\rangle \qquad \forall x \neq y \in \{0,1\}^m. \end{aligned}$$

Clearly, $\langle U_{\mathrm{XOXO}} \rangle \geq m[q \to qq]$, but at first glance it appears that U_{XOXO} cannot easily be used for communication from Bob to Alice. Indeed, [23] proved that when starting without correlation or entanglement, a single use of U_{XOXO} could not send more than $0.7m + o(m)$ bits from Bob to Alice.

However, by consuming entanglement, U_{XOXO} can be used to send m bits from Bob to Alice. The protocol is as follows:

(1) Start with $2^{-m/2} \sum_{x \in \mathbb{Z}_2^m} |\vec{x}\rangle^A |\vec{x}\rangle^B$.

(2) To encode message $b_1, \ldots, b_m$, Bob applies $Z_1^{b_1} \cdots Z_m^{b_m}$ and obtains the state

$$2^{-m/2} \sum_{x \in \mathbb{Z}_2^m} (-1)^{\vec{b}\cdot\vec{x}} |\vec{x}\rangle^A |\vec{x}\rangle^B.$$

(3) U_{XOXO} is applied to yield the state

$$2^{-m/2} \sum_{x \in \mathbb{Z}_2^m} (-1)^{\vec{b}\cdot\vec{x}} |\vec{x}\rangle^A |0\rangle^B.$$

(4) Alice applies $H^{\otimes m}$ and obtains $\left|\vec{b}\right\rangle^A |0\rangle^B$.

Thus, $\langle U_{\mathrm{XOXO}} \rangle + m[qq] \geq m[qq \leftarrow q]$. As a corollary, $C_{\leftarrow}(U_{\mathrm{XOXO}}) \geq m/2$. If we could prove that this were roughly tight (say, that $C_{\leftarrow}(U_{\mathrm{XOXO}}) \leq m/2 + o(m)$), then we might conclude that forward and backward capacities can be separated by a constant factor, but that this separation vanishes when entanglement is allowed for free. We will later demonstrate unitary gates with much stronger separations, even between entanglement-assisted capacities.

First, we can generalize the backward communication protocol of U_{XOXO} to obtain the following result.

Theorem 3:

$$(C_1, C_2, E) \in \mathrm{clCCE}(U) \Leftrightarrow (C_2, C_1, -E - C_1 - C_2) \in \mathrm{clCCE}(U^\dagger).$$

Proof: The proof follows almost immediately from Theorem 1. Suppose $(C_1, C_2, E) \in \mathrm{clCCE}(U)$, so that for any $\delta, \epsilon > 0$ and all n sufficiently large there exists a protocol $\mathcal{P}_n$ of the form of (3) that satisfies (4). We can assume without loss of generality (WLOG) that the local isometries in (3) are in fact unitaries, with all the ancillas being added at the beginning. Then we take the complex conjugate of (3) to obtain

$$\mathcal{P}_n^\dagger = (A_0^{(n)} \otimes B_0^{(n)})^\dagger U^\dagger \ldots U^\dagger (A_n^{(n)} \otimes B_n^{(n)})^\dagger. \quad (6)$$

Observe that $\mathcal{P}_n^\dagger$ is a protocol that uses $U^\dagger$ n times together with local resources. Alice and Bob have only to create the $n\delta_n'$ copies of $|00\rangle$, which they can do using local isometries for free. Then they can use $U^\dagger$ n times to apply $\mathcal{P}_n^\dagger$. Up to error ϵ_n, this maps $|x\rangle^A |x\rangle^B \to |x\rangle^A$ for $x \in \{0,1\}^{n(C_1-\delta_n)}$ and $|y\rangle^A |y\rangle^B \to |y\rangle^B$ for $y \in \{0,1\}^{n(C_2-\delta_n)}$, while consuming

$n(E+\delta_n)$ ebits (or generating $-n(E+\delta_n)$ ebits). By applying the protocol outlined above for U_{XOXO}, this can be used to send $n(C_2-\delta_n)$ cobits from Alice to Bob and $n(C_1-\delta_n)$ cobits from Bob to Alice, while consuming $n(C_1+C_2+E-\delta_n)$ ebits (or generating $-n(C_1+C_2+E-\delta_n)$ ebits). □

From the definition of entanglement capacity we now obtain a statement claimed in the last section.

Corollary 1: For any unitary U, $E(U^\dagger) = \max\{E : (0,0,-E) \in \mathrm{clCCE}(U)\}$.

We can also obtain a few immediate corollaries for the case of free entanglement.

Corollary 2: For any unitary U

(1) $(C_1,C_2,-\infty) \in \mathrm{CCE}(U) \iff (C_2,C_1,-\infty) \in \mathrm{CCE}(U^\dagger)$.
(2) In particular, $C_{\rightarrow}^{E}(U) = C_{\leftarrow}^{E}(U^\dagger)$.
(3) If $U = U^\dagger$ then $C_{\leftarrow}^{E}(U) = C_{\rightarrow}^{E}(U)$.
(4) If $U = U^\dagger$ then $C_{\rightarrow}(U) \geq C_{\rightarrow}^{E}(U)/2$.

The only nontrivial claim here is (4). To prove it, first note that the entanglement-assisted capacity can be achieved with the assistance of $\leq E(U^\dagger)$ ebits. This is because $(C,0,-E) \in \mathrm{clCCE}(U)$ implies $(0,0,-E) \in \mathrm{clCCE}(U)$ (since cobits can always be discarded cleanly) and Theorem 3 implies that $E(U^\dagger) \geq E$. Since we have assumed $U = U^\dagger$, we have $E(U) = E(U^\dagger) \geq E$. Thus, two uses of U can send C cobits: the first use generates E ebits and the second use consumes them to send C cobits. (A similar result was proved in [12].)

Note that cases (3) and (4) of Corollary 2 apply to U_{XOXO}, and show us why we should not expect a dramatic separation of capacity for U_{XOXO}, or indeed any gate equal to its inverse. However, a straightforward modification of the argument in the next section can be used to prove [7]'s conjecture that $C_{\leftarrow}(U_{\mathrm{XOXO}}) \leq m/2 + o(m)$ for m large, which nearly saturates the bound $C_{\leftarrow}(U) \geq C_{\rightarrow}(U)/2$ that we now understand for gates satisfying $U = U^\dagger$.

An alternate proof of the reversal theorem can be obtained from the resource equality $2[q \rightarrow qq] \stackrel{\mathrm{clean}}{=} [q \rightarrow q] + [qq]$. Theorem 3 is then equivalent to the claim that

$$(Q_1,Q_2,E) \in \mathrm{clQQE}(U) \Leftrightarrow (Q_2,Q_1,-E) \in \mathrm{clQQE}(U^\dagger).$$

See [17] for similar examples of reversing quantum communication protocols.

III. V_m: A Gate With Asymmetric Capacities

Guided by Theorem 3, we will construct a gate that is quite different from its inverse. Again choosing a positive integer m as a parameter, define V_m on $\mathbb{C}^{2^m} \otimes \mathbb{C}^{2^m}$ by

$$\begin{aligned} V_m |x\rangle^A |0\rangle^B &= |x\rangle^A |x\rangle^B, && \forall x \\ V_m |x\rangle^A |y\rangle^B &= |x\rangle^A |y-1\rangle^B, && \text{for } 0 < y \leq x \\ V_m |x\rangle^A |y\rangle^B &= |x\rangle^A |y\rangle^B, && \text{for } y > x. \end{aligned}$$

The first line means that $\langle V_m \rangle \geq m[q \rightarrow qq]$. Though we will not need this fact, it turns out that $C_{\rightarrow}(V_m) = m$. The proof, following a similar argument for U_{XOXO} in [7] is as follows. Since $V_m = \sum_x |x\rangle\langle x| \otimes B_x$ for some operators B_x, we know that the Schmidt rank of V_m is $\leq 2^m$. Thus

$$m \leq C_{\rightarrow}(V_m) \leq E(V_m) \leq \log \mathrm{Sch}(V_m) \leq m$$

and each inequality must be an equality.

To bound the back communication of V_m, we now claim that V_m can be simulated to within an accuracy of ϵ by using $m[q \rightarrow qq]$ and $O(\log m/\epsilon)([q \rightarrow q] + [q \leftarrow q])$. We would like to say that this implies

$$\begin{aligned} m[q \rightarrow qq] + O(\log m)([q \rightarrow q] + [q \leftarrow q]) \\ \geq \langle V_m \rangle \geq m[q \rightarrow qq] \end{aligned} \quad (7)$$

but our tools are not strong enough to actually prove this. This is because simulating n copies of V_m to constant accuracy would require $O(n(m+\log nm))$ qubits of communication, which is superlinear in n for any fixed m. However, for now suppose that (7) were true. It would imply that

$$C_{\leftarrow}^{E}(V_m) \leq O(\log m) \ll m = C_{\rightarrow}(V_m), \quad (8)$$

a rather dramatic separation between forward and backward capacities, even when we allow free entanglement to assist the back communication. By using techniques specialized to unitary gates, we will give a proof of (8) later in this section; the proof is inspired by (7), but of course does not rely on it.

Our simulation also means that V_m could be thought of as almost equivalent, at least for large m, to the resource of coherent classical communication. This is interesting both because it is more natural to implement cobits as a unitary gate than as an isometry and because unitary gates, unlike isometries, are reversible. We will return to this second point in the next section when we discuss $V_m^\dagger$. However, we cannot state this as a more precise statement about asymptotic resources since the sequence $(V_m)_{m=1}^{\infty}$ does not fit the definition of an asymptotic resource given in [19].

A. A Simulation for V_m

In this subsection, we show how V_m can be simulated up to error ϵ by a protocol that uses $m[q \rightarrow qq]$ and $O(\log m/\epsilon)([q \rightarrow q] + [q \leftarrow q])$. A key subroutine used in the simulation is a classical communication protocol for distributed comparison. Suppose $x,y \in \{0,1\}^m$, Alice holds x and Bob holds y, which we interpret as integers between 0 and 2^m-1. Then for any error probability $\epsilon > 0$ they can probabilistically determine whether $x = y$, $x > y$, or $x < y$ using $O(\log m/\epsilon)$ bits of communication [40]. The comparison protocol is designed for classical information, but our simulation of V_m will run it coherently using quantum communication.

For the simulation for V_m, suppose Alice and Bob start with $|x\rangle^{A_1} |y\rangle^{B_1}$ for $x,y \in \{0,1\}^m$. Our protocol is as follows.

(1) Define the indicator variable w to be 1 if $y = 0$, 2 if $0 < y \leq x$, or 3 if $y > x$. Use $O(\log m/\epsilon)$ qubits of communication in either direction to coherently compute w. This leaves them (up to error ϵ) with the state

$$|x\rangle^{A_1} |y\rangle^{B_1} |w\rangle^{A_2} |w\rangle^{B_2} |f(x,y)\rangle^{A_3B_3}$$

and where $f(x,y)$ is the state of the ancillas produced by the comparison subprotocol. Using a standard procedure (compute $|w\rangle$, copy it to a new register and then uncompute the first copy of $|w\rangle$ along with the ancilla states produced along the way), Alice and Bob can eliminate the ancilla register to hold simply

$$|x\rangle^{A_1}|y\rangle^{B_1}|w\rangle^{A_2}|w\rangle^{B_2}$$

again up to error ϵ.

(2) Use $m[q \to qq]$ as follows:
- If $w = 1$, then Alice inputs $|x\rangle^{A_1}$, which maps to the state $|x\rangle^{A_1}|x\rangle^{B_3}$. Since $y = 0$, the B_1 register is in the $|0\rangle$ state. Bob swaps B_1 and B_3 to obtain $|x\rangle^{B_1}|0\rangle^{B_3}$.
- If $w = 2$ or 3, then Alice inputs $|0\rangle^{A_3}$, which maps to the state $|0\rangle^{A_3}|0\rangle^{B_3}$. Then she discards A_3.

In either case, Bob discards register B_3, which always contains the $|0\rangle$ state.

(3) If $w = 2$, then Bob maps $|y\rangle^{B_1}$ to $|y-1\rangle^{B_1}$.

(4) Alice and Bob use an additional $O(\log m/\epsilon)$ qubits of communication to uncompute $|w\rangle^{A_2}|w\rangle^{B_2}$. This point is slightly subtle, as the meaning of w as changed: now $w = 1$ means that $y = x$, $w = 2$ means that $y < x$, and $w = 3$ means that $y > x$. Thus, Alice and Bob will compute $|w'\rangle^{A_3}|w'\rangle^{B_3}$ corresponding to these new cases, and will each map $|w\rangle|w'\rangle$ to $|w \oplus w' \pmod 3\rangle|w'\rangle$. With probability $\geq 1 - 2\epsilon$, we have $w = w'$, so this operation effectively erases $|w\rangle$. Then they uncompute $|w'\rangle^{A_3}|w'\rangle^{B_3}$, along with all the ancilla produced along the way.

The entire procedure uses $m[q \to qq] + O(\log m/\epsilon)([q \to q] + [q \leftarrow q])$. To see that the protocol works, first observe that in an ideal protocol where equality testing was perfectly accurate we would obtain precisely V_m. Thus, when we replace equality testing with an approximate version that is accurate (in the sense of cb-norm[34]) to within ϵ, the overall protocol has error $\leq 2\epsilon$.

B. Bounding the Backwards Capacity of V_m

We now use our simulation for V_m to prove that $C_{\leftarrow}^{E}(V_m) \leq O(\log m)$. Since the cost of our simulation depends on the desired error rate, standard resource arguments are not enough to prove this claim. Instead, let V_m^ϵ denote the result of simulating V_m to accuracy ϵ using the above procedure. Since V_m^ϵ is built out of cobits and qubits, it will be an isometry rather than a unitary operator. This is not a serious problem: Theorem 2 still applies and by extending the space that V_m acts on, we have $\|V_m - V_m^\epsilon\|_\infty \leq \epsilon$, where

$$\|X\|_\infty := \max_{|\psi\rangle:\langle\psi|\psi\rangle=1} \sqrt{\langle\psi|X^\dagger X|\psi\rangle}.$$

Moreover, we can simulate V_m^ϵ *exactly* using $m[q \to qq] + O(\log m/\epsilon)([q \to q] + [q \leftarrow q])$. Thus, $C_{\leftarrow}^{E}(V_m^\epsilon) \leq O(\log m/\epsilon)$. Then we will conclude that $C_{\leftarrow}^{E}(V_m) \leq O(\log m)$ by choosing $\epsilon = 1/m$ and applying the following lemma.

Lemma 1 (Continuity of One-Way Capacity): If U and V are isometries with outputs in $\mathbb{C}^d \otimes \mathbb{C}^d$ such that $\|U - V\|_{\mathrm{cb}} \leq \epsilon$ then for all $(C,E) \in \mathrm{clCE}(U)$ there exists $(C',E') \in \mathrm{clCE}(V)$ such that $|C - C'| \leq \epsilon'$ and $|E - E'| \leq \epsilon'$, where $\epsilon' = 8\epsilon \log d + 4H_2(\epsilon)$ and $H_2(\epsilon) := -\epsilon\log\epsilon - (1-\epsilon)\log(1-\epsilon)$.

Proof: By Theorem 2, for any $\delta > 0$ there exists an ensemble of bipartite pure states $\mathcal{E}^{XABA'B'}$ such that

$$I(X;BB')_{U(\mathcal{E})} - I(X;BB')_{\mathcal{E}} \geq C - \delta \quad (9)$$

$$\left|H(BB'|X)_{U(\mathcal{E})} - H(BB'|X)_{\mathcal{E}} - E\right| \leq \delta \quad (10)$$

where U acts on the d-dimensional systems A and B, while X, A' and B' can have arbitrarily large dimension. Thus, we will need to use a recently proved variant of Fannes' inequality[2] which bounds the change in relative entropy as a function of the dimension of only the first system. Begin by using the chain rule[14] to express $H(BB'|X)$ as $H(B|B'X) + H(B'|X)$. Reference [2] states that if $\|U(\mathcal{E}) - V(\mathcal{E})\|_1 \leq \epsilon$ then

$$\left|H(B|B'X)_{U(\mathcal{E})} - H(B|B'X)_{V(\mathcal{E})}\right| \leq 2H_2(\epsilon) + 4\epsilon\log d$$

where $d = \dim B$. On the other hand, $\mathcal{E}^{B'X} = U(\mathcal{E})^{B'X} = V(\mathcal{E})^{B'X}$, so $H(B'|X)_{U(\mathcal{E})} = H(B'|X)_{V(\mathcal{E})}$ and

$$\left|H(BB'|X)_{U(\mathcal{E})} - H(BB'|X)_{V(\mathcal{E})}\right| \leq 2H_2(\epsilon) + 4\epsilon\log d.$$

Similarly

$$\begin{aligned} I(X;BB') &= H(BB') - H(BB'|X) \\ &= H(B') + H(B|B') - H(BB'|X) \end{aligned}$$

and $H(B')$ is unchanged by applying a unitary to AB, so

$$\left|I(X;BB')_{U(\mathcal{E})} - I(X;BB')_{V(\mathcal{E})}\right| \leq 4H_2(\epsilon) + 8\epsilon\log d.$$

If we now take $\delta \to 0$ and apply Theorem 2 again to relate $\Delta_{I,E}(V)$ to $\mathrm{clCE}(V)$, we obtain the proof of the lemma. □

Remark: We suspect that the entire two-way communication capacity region $\mathrm{CCE}(U)$ is similarly continuous. However, without a characterization of the two-capacity analogous to Theorem 2, the proof technique used in Lemma 1 will not work.

IV. More Asymmetry: The Capacity Region of $V_m^\dagger$

In this section, we will demonstrate nearly tight bounds for the capacity region of $V_m^\dagger$, just as we did with V_m. We will find large separations between entanglement-assisted and -unassisted capacities, as well as between entanglement-creation and -destruction capacities.

First use Theorem 3 to show that $\langle V_m^\dagger\rangle + m[qq] \geq m[qq \leftarrow q]$, or equivalently, that $\langle V_m^\dagger\rangle \geq m([qq \leftarrow q] - [qq])$. Next, we will present an approximate simulation for $V_m^\dagger$ that uses a nearly optimal amount of communication; i.e., barely more than $m([qq \leftarrow q] - [qq])$. Thus, by analogy with the nearly optimal simulation of V_m in (7), we will have

$$\begin{aligned} m[qq \leftarrow q] + O(\log m/\epsilon)([q \to q] + [q \leftarrow q]) \\ \gtrsim \langle V_m^\dagger\rangle + m[qq] \geq m[qq \leftarrow q]. \end{aligned} \quad (11)$$

Again, the first inequality is not really a resource inequality, since it does not give us a way of having the overall error vanish when simulating n copies of $V_m^\dagger$ using $(1+o(1))n$ times the resource cost. However, it will still be enough for us to give us nearly tight bounds on the capacity region of $V_m^\dagger$ for m large.

We begin by discussing this capacity region. Choosing $\epsilon = 1/m$ and again using the Continuity Lemma, we can show $C_{\rightarrow}^{E}(V_m^\dagger) \leq O(\log m) \ll m \leq C_{\leftarrow}^{E}(V_m^\dagger)$, a similar sort of capacity separation between forward and backward communication. However, $V_m^\dagger$ also exhibits a dramatic gap between entanglement-assisted and -unassisted capacity. Since $m[qq \leftarrow q] + O(\log m)([q \rightarrow q] + [q \leftarrow q])$ can create no more than $m + O(\log m)$ ebits, the Continuity Lemma implies that $\langle V_m^\dagger \rangle + m[qq]$ also must have entanglement capacity $\leq m + O(\log m)$. Finally, we use the fact that the entanglement capacity of isometries is additive (from [7], [37] as well as Theorem 2) to establish that $E(V_m^\dagger) \leq O(\log m)$.

This implies that *all* unassisted capacities are small; for example, $C_+(V_m^\dagger) \leq E(V_m^\dagger) \leq O(\log m)$. Thus, $V_m^\dagger$ is almost useless without entanglement; none of its unassisted capacities are greater than $O(\log m)$. On the other hand, its capacity (from Bob to Alice) rises when entanglement is supplied, at a rate of nearly one cbit per ebit. No such behavior is known for noisy quantum channels, although there are qudit channels with $O(d)$ multiplicative separations between entanglement-assisted and –unassisted capacities[9], [10], [32].[3] We also obtain a separation between entanglement-creating and -destroying capabilities: $E(V_m^\dagger) \leq O(\log m) \ll m = E(V_m)$. (An independently derived, and completely different, example of a gate with $E(U) \neq E(U^\dagger)$ is in [38].)

There are two ways we can derive the simulation of $V_m^\dagger$ posited in (11). The simplest is to time-reverse our simulation of V_m. First we will need to replace the $m[q \rightarrow qq]$ with $\frac{m}{2}([q \rightarrow q] + [qq])$. Reversing this will replace $[q \rightarrow q]$ with $[q \leftarrow q]$ and $[qq]$ with $-[qq]$. Together, this is $\frac{m}{2}([q \leftarrow q] - [qq]) = m([qq \leftarrow q] - [qq])$, plus of course the $O(\log m/\epsilon)$ terms.

It may be instructive to also consider a more explicit construction of the $V_m^\dagger$ simulation. Note that $V_m^\dagger$ acts on basis states $|x\rangle^A |y\rangle^B$ as follows:

$$\begin{aligned} V_m^\dagger |x\rangle^A |x\rangle^B &= |x\rangle^A |0\rangle^B, && \forall x \\ V_m^\dagger |x\rangle^A |y\rangle^B &= |x\rangle^A |y+1\rangle^B, && \text{for } 0 \leq y < x \\ V_m^\dagger |x\rangle^A |y\rangle^B &= |x\rangle^A |y\rangle^B, && \text{for } y > x. \end{aligned}$$

[3]Every example of such a channel has capacity much smaller than $\log d$; e.g., the channel that maps ρ to $\mathcal{N}(\rho) = \epsilon\rho + (1-\epsilon)I/d$. For concreteness, follow [32] and choose $\epsilon = -1/(d^2-1)$, so $\mathcal{N}(\rho) = (dI-\rho)/(d^2-1)$. Then the single-shot Holevo–Schumacher–Westmorelend (HSW) capacity $C^{(1)}(\mathcal{N})$ is $\Theta(d^{-3})$ and the entanglement-assisted capacity $C_E(\mathcal{N})$ is $\Theta(d^{-2})$. Achieving this entanglement-assisted rate using the protocols of [10], [32] requires $\log d$ ebits per use of $\mathcal{N}$, meaning that we consume many ebits to get a small enhancement in classical capacity. Communication protocols which used less entanglement were given in [19], [44], but here too we conjecture that $\Omega(\log d)$ ebits are necessary to raise the capacity of $\mathcal{N}$ from $O(d^{-3})$ to $\Omega(d^{-2})$, or even to $\Omega(d^{-3+\delta})$ for any $\delta > 0$.

Again, Alice and Bob can determine whether $y < x$, $y = x$, or $y > x$ to accuracy ϵ by exchanging $O(\log m/\epsilon)$ qubits, and by performing these calculations coherently, can uncompute this information at the end of the protocol. As with the V_m simulation, the three cases at the end of the protocol are different ($0 < y \leq x$, $y = 0$, and $y > x$), so it is important that they store no more information than which case holds.

The interesting case is when $x = y$ and Alice and Bob would like to map $|x\rangle^A |x\rangle^B \rightarrow |x\rangle^A |0\rangle^B$ for arbitrary values of x. As we have argued above, this can be simulated by reversing $m([q \rightarrow qq]) = \frac{m}{2}([q \rightarrow q] + [qq])$, which requires a resource cost of $m([qq \leftarrow q] - [qq]) = \frac{m}{2}([q \leftarrow q] - [qq])$. Let us now examine this reverse procedure in more detail. For simplicity, suppose $m = 2$. The procedure we would like to reverse is coherent superdense coding ($[q \rightarrow q] + [qq] \geq 2[q \rightarrow qq]$), which maps $|x_1\rangle^{A_1} |x_2\rangle^{A_2}$ to $|x_1\rangle^{A_1} |x_2\rangle^{A_2} |x_1\rangle^{B_1} |x_2\rangle^{B_2}$ as follows: First Alice and Bob add a maximally entangled state $|\Phi\rangle^{A_3 B_1}$. Then Alice applies the Pauli operator $X^{x_1} Z^{x_2}$ to the A_3 system, leaving the state

$$\begin{aligned} |x_1\rangle^{A_1} |x_2\rangle^{A_2} (X^{x_1} Z^{x_2} \otimes I) |\Phi\rangle^{A_3 B_1} \\ =: |x_1, x_2\rangle^{A_1 A_2} |\Phi_{x_1,x_2}\rangle^{A_3 B_1} \end{aligned} \quad (12)$$

where $|\Phi_{x_1,x_2}\rangle := (X^{x_1} Z^{x_2} \otimes I) |\Phi\rangle$. Note that the four $|\Phi_x\rangle$ form an orthonormal basis, and thus we can define the unitary map $U_{\text{SD}} := \sum_{x_1=0}^{1} \sum_{x_2=0}^{1} |x_1, x_2\rangle \langle \Phi_{x_1,x_2}|$. Superdense coding proceeds by Alice sending her half of $|\Phi_{x_1,x_2}\rangle$ to Bob, who applies U_{SD} to yield the state $|x_1, x_2\rangle^A |x_1, x_2\rangle^B$.

Now we explain how this protocol can be reversed to map $|x_1, x_2\rangle^A |x_1, x_2\rangle^B$ to $|x_1, x_2\rangle^A$. Bob first applies $U_{\text{SD}}^\dagger$ to $|x_1, x_2\rangle^B$ and obtains $|\Phi_x\rangle^{B_1 B_2}$. Using $[q \leftarrow q]$, Bob sends half of $|\Phi_x\rangle$ to Alice, so the joint state becomes $|x_1, x_2\rangle^A |\Phi_{x_1,x_2}\rangle^{AB}$. Since Alice has a copy of x, she can apply $(X^{x_1} Z^{x_2})^\dagger$ to her half of $|\Phi_{x_1,x_2}\rangle^{AB}$ and transform it to $|\Phi_{0,0}\rangle = |\Phi\rangle$. Thus, not only has Bob's copy of $|x_1, x_2\rangle$ been erased, but Alice and Bob are left sharing the state $|\Phi\rangle$. This means that two bits on Bob's side can be coherently erased by using the resource $[q \leftarrow q] - [qq] = 2([qq \leftarrow q] - [qq])$.

V. Coherent Erasure and Time Reversal

By now we have seen many examples of how unitary communication protocols can be reversed to give new protocols. In this section, we summarize these reversal rules and, inspired by our observations about $V_m^\dagger$ in the last section, introduce the new communication resource of "coherent erasure," which is the time reversal of coherent classical communication. Time-reversal symmetry was also discussed in [17], which explained how the quantum reverse Shannon theorem[6] is the time-reversal of the quantum Slepian–Wolf theorem (a.k.a. state merging[33] or the mother[18], [19]), once all the protocols are made fully coherent.

For the standard resources of $[q \rightarrow q]$ and $[qq]$, time-reversal is quite simple. If we let † denote the time-reverse of a resource, then $[q \rightarrow q]^\dagger = [q \leftarrow q]$ and $[qq]^\dagger = -[qq]$. This means that sending qubits in one direction is reversed by sending qubits in the other direction, and that creating entanglement is reversed by destroying entanglement. As we have argued above,

$[q \to qq] = ([q \to q] + [qq])/2$ so the time-reversal of $[q \to qq]$ is $[q \to qq]^\dagger = ([q \leftarrow q] - [qq])/2$. This resource corresponds to the map $|x\rangle^A |x\rangle^B \to |x\rangle^A$, so we call it "coherent erasure" and label it $[q \leftarrow qq]$. Since one bit of coherent classical communication is a cobit, we (following a suggestion of Charlie Bennett's) call $[q \leftarrow qq]$ a co-cobit from Bob to Alice, where the first "co" stands for "complementary" and the second "co" stands for "coherent." The co-cobit from Alice to Bob is denoted $[qq \to q]$ and corresponds to the map $|x\rangle^A |x\rangle^B \to |x\rangle^B$.

Of course, the map $|x\rangle^A |x\rangle^B \to |x\rangle^A$ is not defined for all inputs (what if Alice and Bob do not input the same state?), but in this section we will explain how coherent erasure nevertheless makes sense as a communication resource. First in Section V-A we will explain how coherent erasure can be derived from other resources and then in Section V-B we will describe some uses of coherent erasure.

We stress at the outset that coherent erasure is equivalent to standard resources, and there is no need to introduce new concepts such as "erasure capacity" and the like. However, it may prove a useful metaphor in analyzing other communication protocols.

A. Producing Coherent Erasure

We have already seen three ways of producing coherent erasure, which we briefly review here.

1) Reversing Clean Protocols With $[q \to qq]$: If a clean resource inequality involves $[q \to qq]$, then in the time-reversed version of the resource inequality $[q \to qq]$ is replaced with $[q \to qq]^\dagger = [q \leftarrow qq]$. This was an implicit part of the argument of Theorem 3, which was based on reversing $\langle U \rangle \overset{\text{clean}}{\geq} C[q \to qq]$ to obtain $\langle U^\dagger \rangle \overset{\text{clean}}{\geq} C[q \leftarrow qq]$.

2) Reversing Super-Dense Coding: This was explained in Section IV, and is basically a special case of the last point: time-reversing $[q \to q] + [qq] \geq 2[q \to qq]$ yields

$$[q \leftarrow q] - [qq] \geq 2[q \leftarrow qq]. \tag{13}$$

3) The Gate $V_m^\dagger$: Just as V_m is equivalent to m cobits up to small errors and inefficiencies, $V_m^\dagger$ is roughly equivalent to m co-cobits. Of course $(V_m^\dagger)_{m=1}^\infty$ is not a proper asymptotic resource, but it is still useful as a concrete way to imagine implementing coherent erasure.

B. Using Coherent Erasure

1) Entanglement-Assisted Communication: Theorem 3 (or more precisely, the protocol sketched in Section II for entanglement-assisted back communication using U_{XOXO}) explained how

$$[q \leftarrow qq] + [qq] \geq [qq \leftarrow q] \tag{14}$$

by giving an explicit protocol. Another way to derive this resource inequality is reversing coherent teleportation

$$2[q \to qq] \geq [q \to q] + [qq]$$

to obtain

$$2[q \leftarrow qq] \geq [q \leftarrow q] - [qq].$$

Substituting $[q \leftarrow q] = 2[qq \leftarrow q] - [qq]$ then gives the desired result.

Interestingly, coherent erasure is capable of no communication on its own, but can convert one ebit into one cobit. This is a sharper version of the separation we observed between the entanglement-assisted and -unassisted capacities of $V_m^\dagger$.

2) State Merging and Partial Quantum Communication: First note that (13) and (14) can be combined to obtain the equality

$$\begin{aligned}[q \leftarrow qq] &= [qq \leftarrow q] - [qq] \\ &= \frac{[q \leftarrow q] - [qq]}{2} \\ &= [q \leftarrow q] - [qq \leftarrow q].\end{aligned} \tag{15}$$

This last point means that $[q \to q] = [q \to qq] + [qq \to q]$, so the task of sending a qubit can be split into the tasks of sending one cobit and coherently erasing one bit. There is a direct protocol which performs this. Suppose Alice would like to send the state $\sum_x \alpha_x |x\rangle^A$ to Bob. If she applies $[q \to qq]$ then they will obtain the entangled state $\sum_x \alpha_x |x\rangle^A |x\rangle^B$. Finally, applying $[qq \to q]$ will erase Alice's state and leave Bob with $\sum_x \alpha_x |x\rangle^B$.

Of course this also works for coherent superpositions of messages. If Alice would like to send her half of $\sum_x \alpha_x |x\rangle^R |x\rangle^A$ to Bob then she can first apply $[q \to qq]$ to obtain

$$\sum_x \alpha_x |x\rangle^R |x\rangle^A |x\rangle^B$$

and then $[qq \to q]$ will again erase Alice's state to leave $\sum_x \alpha_x |x\rangle^R |x\rangle^B$.

The general problem here is *state merging*[33], in which Alice gives Bob her piece of a tripartite state $|\psi\rangle^{ABR}$, perhaps generating or consuming entanglement in the process. Reference [17] argued that cobits are the canonical example of feedback channels, which are isometries that map from A to AB. Likewise, we claim that coherent erasure is the canonical example of state merging.

To justify this interpretation, we will now show how to generalize the decomposition $[q \to q] = [q \to qq] + [qq \to q]$ to a decomposition of perfect quantum communication from $A \to B$ into an isometry from $A \to AB$ followed by merging AB into B. The isometry from $A \to AB$ can be simulated by $I(R;B)[q \to qq] + I(B\rangle A)[qq]$ using the quantum reverse Shannon theorem[6], [17], where the coherent information $I(B\rangle A)$ is defined as

$$I(B\rangle A) := -H(B|A) = H(A) - H(AB) = H(A) - H(R).$$

The $I(R;B)[q \to qq]$ cost represents the difficulty of creating the desired correlations between Bob and the reference system, while the $I(B\rangle A)[qq]$ cost is necessary because back communication would allow Bob to distill that much entanglement with Alice while preserving his correlations with R. Then the tripartite state can be mapped (using state merging) to one where Bob holds Alice's part using $I(R;A)[qq \to q] - I(A\rangle B)[qq]$. Here the

erasure cost measures the amount of correlation with the reference system that Alice has and needs to give up, while $I(A\rangle B)$ is the amount of entanglement that is recovered by the procedure once Bob has the entire purification of Alice's state. Indeed, this version of state merging amounts to a coherent version of entanglement distillation ($I(A;E)[c\to c]+\langle\psi^{ABE}\rangle \geq I(A\rangle B)[qq]$) in which $[c\to c]$ is replaced with $[qq\to q]$ and as a result Bob holds the purification of the environment at the end of the protocol. Finally, the total resource cost $I(R;B)[q\to qq]+I(B\rangle A)[qq]+I(R;A)[qq\to q]-I(A\rangle B)[qq]$ is simply equal to $H(R)[q\to q]$, the cost of sending the reference system directly to Bob.

3) Rule I: Coherently Decoupled Input Cbits: Suppose $\alpha+C[c\to c]\geq\beta$ is a resource inequality in which the classical message sent is nearly independent of all residual quantum systems, including the environment. In this case, we say that the input cbits are *coherent decoupled* and "Rule I" of [18], [19], [23] proves that they can be replaced by $C([q\to qq]-[qq])$. Equivalently, we can replace them by $C[qq\to q]$. Thus, coherent erasure can be used whenever we need to send a classical message whose contents can be guaranteed to be almost completely independent of the remaining quantum systems.

The simplest example of a coherently decoupled input is of course teleportation, which quickly leads us to the familiar resource inequality $2[qq\to q]+[qq]\geq[q\to q]$. A slightly more nontrivial example is remote state preparation[8], in which $\log d$ ebits and $\log n := \log d+\log\log d+2\log 1/\epsilon+O(1)$ cbits are used for Alice to prepare an arbitrary d-dimensional state in Bob's lab (i.e., $\log d$ "remote qubits"). Since the input cbits are coherently decoupled, they can be replaced by $(\log n)([q\to qq]-[qq])$. Asymptotically, this means that one coherent bit is at least as strong as one remote qubit, which is an interesting statement because cobits and remote qubits both lie somewhere in between cbits and qubits, and there does not appear to be any trivial protocol relating the two. Another way to interpret remote state preparation is that $[qq\to q]+[qq]$ yield one remote qubit; this, by contrast, can be implemented by a relatively straightforward single-shot protocol. Suppose the state that Alice wishes to prepare for Bob is $|\alpha\rangle=\sum_{x=1}^d\alpha_x|x\rangle$. A particularly easy case is when $|\alpha_x|^2=1/d$ for all x. Then Alice could locally map $|\Phi_d\rangle := \frac{1}{\sqrt{d}}\sum_{x=1}^d|x\rangle^A|x\rangle^B$ to $\sum_{x=1}^d\alpha_x|x\rangle^A|x\rangle^B$ and then use $\log d[qq\to q]$ to leave Bob with the state $\sum_{x=1}^d\alpha_x|x\rangle^B$.

In general, $|\alpha_x|$ will not always be equal to $1/\sqrt{d}$, and blithely applying the above method for a general state will only achieve a fidelity of $F(\alpha):=\sum_x|\alpha_x|/\sqrt{d}$. Moreover, even a randomly chosen state will usually have F close to $\sqrt{\pi}/2$ when d is large (as can be seen using $\mathbb{E}|\alpha_x|=\sqrt{\pi/4d}$ and the central limit theorem). This means that multiplying $|\alpha\rangle$ by a fixed random unitary will not be sufficient to obtain high fidelity. Instead, we will use a small sequence of unitaries $\{U_1,\ldots,U_\kappa\}$, along with a κ-dimensional ancilla register that controls which unitary is applied. To decode, Bob will need this ancilla register, which we will require $\log\kappa$ extra qubits of communication from Alice to Bob. Fortunately, we will see that the error shrinks rapidly with κ for a variety of choices of $\{U_1,\ldots,U_\kappa\}$.

The procedure for Alice to remotely prepare $|\alpha\rangle$ in Bob's lab is as follows. Let

$$|\beta\rangle := \frac{1}{\kappa}\sum_{k=1}^{\kappa}|k\rangle U_k|\alpha\rangle=\sum_{x=1}^d\beta_x|b_x\rangle|x\rangle$$

where in the last expression we have introduced states $|b_x\rangle$ for each x. Assume for now that $F(\beta)=\sum_x|\beta_x|/\sqrt{d}$ is close to one. Then, starting with the shared state $|\Phi_d\rangle$, Alice can create

$$\frac{1}{\sqrt{d}}\sum_x|x\rangle^A|x\rangle^B|b_x\rangle^{A'}.$$

If she then applies $\log d$ co-cobits to AB and sends A' to Bob using $\log\kappa$ qubits then Bob will have a fidelity-$F(\beta)$ approximation to $|\beta\rangle$, from which he can obtain $|\alpha\rangle$.

We can summarize the correctness of the protocol as follows.

Theorem 4: There exists a subspace $V\subset\mathbb{C}^d$ with $d' := \dim V = d\epsilon^2/32\log(5/\epsilon)$ such that the remote state preparation protocol above can prepare any state $|\alpha\rangle\in V$ in Bob's lab using $\log d[qq\to q]+\log d[qq]+\log\kappa[q\to q]$ and with fidelity $1-\frac{1}{2\kappa}-2\epsilon$.

This translates into remotely preparing a state of $\log d' = \log d-2\log 1/\epsilon-O(\log\log 1/\epsilon)$ qubits, which is similar to the performance of [8].

Proof: To analyze the protocol, we first seek to lower-bound $\mathbb{E}_\alpha F(\beta)$, where α is uniformly randomly chosen. We will choose U_k to be $\sum_x|x+k\rangle\langle x|$ for $k\in\{1,\ldots,\kappa\}$ and where addition is $\mathrm{mod}d$. We could also choose the U_k to be (approximately) mutually unbiased bases (meaning that $|\langle x|U_k^\dagger U_{k'}|x\rangle|^2$ is (approximately) equal to $1/d$), but we omit the analysis here. To evaluate the expected fidelity, we use linearity of expectation

$$\begin{aligned}\mathbb{E}_\alpha F(\beta)&=\frac{1}{\sqrt{d}}\sum_{x=1}^d\mathbb{E}_\alpha|\beta_x|\\&=\frac{1}{\sqrt{d\kappa}}\sum_{x=1}^d\mathbb{E}_\alpha\sqrt{\sum_{k=1}^{\kappa}|\langle x+k|\alpha\rangle|^2}.\end{aligned}\tag{16}$$

Using the fact that the distribution of $|\alpha\rangle$ is rotationally invariant, we find that

$$\mathbb{E}_\alpha F(\beta)=\sqrt{\frac{d}{\kappa}}\mathbb{E}_\alpha\sqrt{\mathrm{Tr}\,P\alpha}$$

where P is a rank-κ projector (assuming that $\kappa\leq d$). To estimate this last term, we will use the inequality $\mathbb{E}|X|\geq(\mathbb{E}X^2)^{\frac{3}{2}}/(\mathbb{E}X^4)^{\frac{1}{2}}$ which holds for any random variable [11]. Next, we calculate $\mathbb{E}\,\mathrm{Tr}\,P\alpha=\kappa/d$ and

$$\begin{aligned}\mathbb{E}(\mathrm{Tr}\,P\alpha)^2&=\mathrm{Tr}(P\otimes P)\cdot\mathbb{E}(\alpha\otimes\alpha)\\&=\mathrm{Tr}(P\otimes P)\cdot(I+\ \mathrm{SWAP})/d(d+1)\\&=\kappa(\kappa+1)/d(d+1).\end{aligned}$$

Putting this together we find that

$$\mathbb{E}_\alpha F(\beta) \geq \sqrt{\frac{1+\frac{1}{d}}{1+\frac{1}{\kappa}}} \geq 1 - \frac{1}{2\kappa}.$$

The remaining steps are quite similar to the arguments in [1], [29]. We will argue that not only is F close to its expectation for most values of α, but in fact if we choose a random subspace V of dimension $d' = O(d\epsilon^2/32\log(5/\epsilon))$ then with nonzero probability every vector $|\alpha\rangle \in V$ will have $F(\beta) \geq 1 - \frac{1}{2\kappa} - \epsilon$. Observe that the Lipschitz constant of $F(\beta)$ (defined to be $\max_{|\alpha\rangle} \sum_x (\partial F/\partial_{\alpha_x})^2$) is constant. In fact, it is 1. Then Levy's Lemma[35] states that

$$\Pr_\alpha \left[F(\alpha) \leq 1 - \frac{1}{2\kappa} - \epsilon \right] \leq 2\exp\left(-\frac{d\epsilon^2}{31}\right) \qquad (17)$$

for any $\epsilon > 0$. Now consider a random d'-dimensional subspace V. According to [29, Lemma III.6], we can choose a mesh of $(5/\epsilon)^{d'} = \exp(d\epsilon^2/32)$ points such that any point in V is within ϵ (in trace distance) of some point in the mesh. Applying the union bound and the triangle inequality to (17), we find that there exists a subspace V such that $F(\alpha) \geq 1 - \frac{1}{2\kappa} - 2\epsilon$ for all $|\alpha\rangle \in V$. □

Thus, coherent erasure is an alternate, and arguably more direct, way to think about remote state preparation. It is perhaps interesting that our protocol is nontrivially different from the comparably efficient protocol that could be obtained from applying "Rule I" of [18], [19], [23] to [8]. While Rule I guarantees that in general co-cobits can be used in place of coherently decoupled input bits, the proof is indirect and involves catalysis. It would be interesting to know whether there is a more direct and natural use of coherent erasure, such as the one we showed for remote state preparation, in any protocol with coherently decoupled input cbits.

VI. Conclusion

The results of this paper help resolve many questions surrounding unitary gate capacities. We now understand that, while all the capacities of any nonlocal gate are nonzero, there can be asymptotically large separations between these capacities. The main separation left unproven by this paper is finding a gate with $E(U) \gg C_+(U)$. An early version of this paper proposed the gate on $\mathbb{C}^d \otimes \mathbb{C}^d$ which exchanges $|01\rangle$ and $|\Phi_d\rangle$, while leaving the other states unchanged: $U = I - |01\rangle\langle 01| - |\Phi_d\rangle\langle\Phi_d| + |01\rangle\langle\Phi_d| + |\Phi_d\rangle\langle 01|$. Since U requires $\log d$ cbits to simulate, even using unlimited EPR pairs (as we will prove later in this section), the simulation techniques in this paper will need to be modified. Since the first version of this paper appeared, [26] established the conjectured separation (showing that $C_+^E(U) \leq O(\log\log d) \ll \log d \leq E(U)$) by using non-maximally entangled states for the simulation.

Another limitation of our work is that the separations we have found are between $o(\log d)$ and $\approx \log d$; on the other hand, [38] has proven that if $E(U) = 2\log d$ then $E(U^\dagger) = 2\log d$ as well. It would be interesting to see which capacity separations are possible for gates with $E(U)$ between $\log d$ and $2\log d$.

TABLE I
Definitions of Communication Resources

abbr.	name	formula
$[qq]$	ebit	$\frac{1}{\sqrt{2}}(\|00\rangle^{AB} + \|11\rangle^{AB})$
$[c \to c]$	cbit	$\sum_{x=0}^{1} \|x\rangle^B\|x\rangle^E\langle x\|^A$
$[q \to q]$	qubit	$\sum_{x=0}^{1} \|x\rangle^B\langle x\|^A$
$[q \to qq]$	cobit	$\sum_{x=0}^{1} \|x\rangle^A\|x\rangle^B\langle x\|^A$
$[q \leftarrow qq]$	co-cobit	$\sum_{x=0}^{1} \|x\rangle^A\langle x\|^A\langle x\|^B$
$\langle U\rangle$	unitary	U

Of course, separating communication capacities is mostly intended as a step towards better understanding quantum communication using unitary gates as well as other resources. For example, it has led us to the resource of coherent erasure, which hopefully will turn out to be a useful concept the way coherent classical communication has.

Entanglement destruction is another resource in quantum information theory that seems to be worth exploring. Once we restrict protocols to be clean, destroying entanglement is a nonlocal task. Equivalently, creating coherent superpositions of states with varying amount of entanglement requires communication, even if the two parties are allowed unlimited numbers of maximally entangled states[30]. This task comes up in entanglement dilution[27], [30], in its generalization, remote preparation of known entangled states (though this is not explicitly acknowledged in [8]), and in the quantum reverse Shannon theorem[6], which may be thought of as a further generalization of remote state preparation. In each case, the resource of "entanglement spread"—meaning the ability to generate superpositions of states with varying amounts of entanglement—appears to be necessary. Entanglement spread can be generated (using, e.g., remote state preparation) by sending cbits in either direction, or even without any communication at all, if Alice and Bob can make catalytic use of an embezzling state[28]. It appears that a unitary gate U has "spread capacity" equal to $E(U) + E(U^\dagger)$; for example, a simple application of [30] can prove that $C_1[c \to c] + C_2[c \leftarrow c] + \infty[qq] \leq \langle U\rangle$ implies that $C_1 + C_2 \geq E(U) + E(U^\dagger)$, a result previously proved only for two-qubit gates, using very different arguments [12]. Of course, without a precise definition of entanglement spread it will be difficult to formalize these arguments. Reference [30] is a promising first step towards defining spread as a resource, though the unusual scaling of embezzling states and of the cost of entanglement dilution suggest that the independent and identically distributed (i.i.d.) resource model of [19], [23] might not fit well.

VII. Notation

In this section, we collect some of the notation used in the rest of the paper.

Cobits were introduced in [22] and co-cbits were introduced in Section V. Note that co-cobits are only defined when Alice and Bob's joint state is constrained to lie in the subspace spanned by $|00\rangle$ and $|11\rangle$.

TABLE II
EXCHANGE AND TIME-REVERSAL SYMMETRIES

resource	exchange	reverse
$[qq]$	$[qq]$	$-[qq]$
$[c \to c]$	$[c \leftarrow c]$	undefined
$[q \to q]$	$[q \leftarrow q]$	$[q \leftarrow q]$
$[q \to qq]$	$[qq \leftarrow q]$	$[q \leftarrow qq]$
$[q \leftarrow qq]$	$[qq \to q]$	$[q \to qq]$
$\langle U \rangle$	$\langle \mathcal{F}U\mathcal{F} \rangle$	$\langle U^\dagger \rangle$

TABLE III
TRANSFORMATIONS BETWEEN STANDARD RESOURCES

$2[c \to c] + [qq] \overset{clean}{\geq} [q \to q]$	teleportation
$[q \to q] + [qq] \overset{clean}{\geq} 2[c \to c]$	super-dense coding
$[q \to q] + [qq] \overset{clean}{=} 2[q \to qq]$	from [22]
$[q \to q] - [qq] \overset{clean}{=} 2[q \leftarrow qq]$	from Sec. V
$[q \to qq] + [q \leftarrow qq] \overset{clean}{=} [q \to q]$	from last two lines

To understand the relations between the resources in Table I, we describe the effects of exchanging Alice and Bob and of running a protocol backwards. These are listed in the "exchange" and "reverse" columns of the of Table II. For example, $[q \to q]$ means sending a qubit from Alice to Bob, so either exchanging Alice and Bob or reversing time transforms $[q \to q]$ to $[q \leftarrow q]$. However, given a cobit from Alice to Bob, exchange and time-reversal do not act the same way: exchanging Alice and Bob yields a cobit from Bob to Alice while time-reversal yields a co-cobit from Bob to Alice.

In the first line of Table II, we can consider $[qq]$ to be the action of creating an ebit. Thus, the time-reversal of $[qq]$ corresponds to destroying an ebit, which, if done coherently, is a nontrivial resource. In the last line, $\mathcal{F}$ denotes the unitary operator that exchanges Alice and Bob's systems. The time-reversal relations for cobits and co-cobits are explained in Section V.

Next, we summarize some of the basic transformations of the resources in Table III that are possible. We have indicated where the protocols can be done cleanly by using $\overset{\text{clean}}{\geq}$ or $\overset{\text{clean}}{=}$ instead of $\geq$ or $=$.

Now define U to be a bipartite unitary gate. We have defined various capacity regions in Section I, which are summarized in Table IV.

Finally, we will explain how the results of the paper relate to the terms in Table IV. First, some of the theorems relate as follows:

- Theorem 1 shows that the capacity regions CCE and clCCE are nearly equivalent, other than the fact that entanglement can be thrown away in non-clean protocols. As a corollary, Theorem 1 also relates CE to clCE the same way.
- Theorem 2 and Lemma 1 concern the region $\text{clCE}(U)$, giving a single-letter formula for it, and proving its continuity (in terms of U), respectively.
- Theorem 3 shows that $\text{clCCE}(U^\dagger)$ can be completely determined from $\text{clCCE}(U)$ (and *vice versa*, of course).

Finally, Table V summarizes the separations in capacities proved in Sections III and IV.

TABLE IV
CAPACITIES AND CAPACITY REGIONS OF UNITARY GATES

$$\begin{aligned}
\text{CCE}(U) &= \{(C_1, C_2, E) : U \geq C_1[c \to c] + C_2[c \leftarrow c] + E[qq]\} \\
\text{clCCE}(U) &= \{(C_1, C_2, E) : U \overset{clean}{\geq} C_1[c \to c] + C_2[c \leftarrow c] + E[qq]\} \\
\text{CE}(U) &= \{(C, E) : U \geq C[c \to c] + E[qq]\} \\
\text{clCE}(U) &= \{(C, E) : U \overset{clean}{\geq} C[c \to c] + E[qq]\} \\
C^E_{\to}(U) &= \max\{C : U + \infty[qq] \geq C[c \to c]\} \\
C_{\to}(U) &= \max\{C : U \geq C[c \to c]\} \\
C^E_{\leftarrow}(U) &= \max\{C : U + \infty[qq] \geq C[c \leftarrow c]\} \\
C_{\leftarrow}(U) &= \max\{C : U \geq C[c \leftarrow c]\} \\
C_{+}(U) &= \max\{C_1 + C_2 : U \geq C_1[c \to c] + C_2[c \leftarrow c]\} \\
C^E_{+}(U) &= \max\{C_1 + C_2 : U + \infty[qq] \geq C_1[c \to c] + C_2[c \leftarrow c]\} \\
E(U) &= \max\{E : U \geq E[qq]\}
\end{aligned}$$

TABLE V
EXCHANGE AND TIME-REVERSAL SYMMETRIES

f. vs. b. comm.	$C^E_{\leftarrow}(V_m) \ll C_{\to}(V_m)$
ent. create/destroy	$E(V_m^\dagger) \ll E(V_m)$
ent.-assted f. vs. b. comm.	$C^E_{\to}(V_m^\dagger) \ll C^E_{\leftarrow}(V_m^\dagger)$
comm. w/ or w/o ent.-asst.	$C_{+}(V_m^\dagger) \ll C^E_{+}(V_m)$

Here f. means forward, b. means backwards, comm. means communication (which can be equivalently taken to be cbits, cobits or qubits), ent. means entanglement, and asst/assted means assistance/assisted.

APPENDIX
NEW PROOFS OF THEOREMS 1 AND 2

In this appendix, we sketch alternate proofs for Theorems 1 and 2. Both proofs we give are slightly simpler than previous versions and have slightly better convergence properties. Moreover, by including them, this paper can be more self-contained, especially given that the previous statements of Theorem 1 in [23], [25] were slightly weaker.

Proof of Theorem 1: We begin by following the approach of [25], where this was first proved. Suppose $\langle U \rangle \geq C_1[c \to c] + C_2[c \leftarrow c] + E[qq]$ for some $E > 0$ (similar arguments apply for $E \leq 0$). Then, if Alice and Bob copy their inputs before sending and refrain from performing their final von Neumann measurements, we obtain a sequence of protocols $\mathcal{P}_n$ such that for all $x \in \{0,1\}^{C_1^{(n)}}$ and $y \in \{0,1\}^{C_2^{(n)}}$ (with $C_j^{(n)} := n(C_j - \delta_n)$ for $j = 1, 2$)

$$\mathcal{P}_n |x\rangle^A |y\rangle^B \overset{\epsilon_n}{\approx} |x,y\rangle^A |x,y\rangle^B (|\Phi\rangle^{AB})^{\otimes n(E-\delta_n)} |\varphi_{x,y}\rangle^{AB}. \quad (18)$$

The main difference between (18) and our goal ((4)) is the presence of the ancilla $|\varphi_{x,y}\rangle^{AB}$ with its arbitrary depends on x and y rather than a string of zeroes that depends only on n. Simply discarding $|\varphi_{x,y}\rangle^{AB}$ will in general break superpositions between different values of x and y. Also the ancilla is not guaranteed to fit in $o(n)$ qubits.

Reference [25] made a series of modifications in order to obtain a clean protocol. In fact, [25] obtained a slightly weaker result than (4), in which Alice and Bob are left with an ancilla $|\varphi_n\rangle^{AB}$ which depends only on n and (it can be shown) can be stored in $\leq n\delta_n'$ qubits. We call protocols of this form "semi-clean," but when working with unitary gates this implies

the protocol can be made clean at an asymptotically negligible additional cost. This is due to [24], which proved that any nonlocal gate U can exactly generate any other fixed gate, such as SWAP, with a constant number of applications interspersed with local unitaries. Thus, $O(n\delta'_n)$ applications of U can exactly map $|\varphi'_n\rangle$ to $|00\rangle^{n\delta'_n}$.

We now review the steps of [25] in obtaining a semi-clean protocol before we describe our alternate approach. First they used entanglement as a sort of coherent one-time-pad, so that the ancillas would become correlated with the one-time-pad and not the message. Then a classical error-correcting code was applied to reduce the errors, and finally entanglement concentration[5] was used on the error-free blocks to recover the entanglement used for the one-time-pad. We use an approach that is only slightly different: first, using block coding to reduce the error to a nearly exponentially small amount, then using a coherent one-time-pad to decouple the ancillas, and finally recovering entanglement using an approximate form of entanglement concentration that needs far fewer states.

We now explain these components in more detail. First, absorb all of the output into the ancilla, except for the locally copied inputs, so that

$$\mathcal{P}_n |x\rangle^A |y\rangle^B = |x\rangle^{A_1} |y\rangle^{B_1} |\varphi_{x,y}\rangle^{A_2A_3B_2B_3} \tag{19}$$

with $\mathrm{Tr}(|y\rangle\langle y|^{A_2} \otimes |x\rangle\langle x|^{B_2} \otimes I^{A_3B_3})\varphi_{x,y} \geq 1 - \epsilon_n$ and with entanglement $E(|\varphi_{x,y}\rangle) := H(\varphi_{x,y}^{A_2A_3}) \geq n(E - \delta_n)$, for an appropriate redefinition of δ_n. This is clearly an equivalent formulation, but it allows us to speak more easily about the exact output of the protocol.

Now we will use classical bidirectional block-coding[43] to control how quickly ϵ_n vanishes as a function of n. By applying $\mathcal{P}_{n_1}$ n_2 times and slightly reducing the rate, it is possible to send $n_1n_2(C_1 - \delta)$ cbits forward and $n_1n_2(C_2 - \delta)$ cbits backwards with n_1n_2 uses of U and average error $\leq \exp(n_2(1+\alpha \log \epsilon_n))$ as long as $\delta \geq \delta_{n_1} + \alpha + H_2(\alpha)/(n_1(\min(C_1, C_2) - \delta_{n_1}))$. This can be simplified by choosing $\alpha = 2/\log(1/\epsilon_n)$, so the error probability is $\leq \exp(-n_2)$ and we still have $\delta \to 0$. Finally, for any $0 < \gamma < 1$, choose $n_2 = n_1^{\frac{1-\gamma}{\gamma}}$, so if $n := n_1n_2$, then $n_2 = n^{1-\gamma}$. Thus, we can WLOG assume that we have a sequence of protocols $(\mathcal{P}_n)_{n=1}^{\infty}$ with inefficiency δ_n and error ϵ_n satisfying $\lim_{n\to\infty} \delta_n = 0$ and $\epsilon_n \leq \exp(-n^{1-\gamma})$ (in fact, we could choose $\epsilon_n \leq \exp(-f(n))$ for any $f(n) = o(n)$ that we like).[4] Now we see the reason for using the form of (19); our block codes still satisfy (19), but not necessarily (18), since there may be arbitrary errors in the entangled states.

Next, use a coherent one-time-pad in the same way as [25]. Alice and Bob start with $C_1^{(n)} + C_2^{(n)}$ ebits. Together with their initial messages $|x\rangle^A |y\rangle^B$, their state is

$$\frac{1}{\sqrt{N}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} |a, b, x\rangle^A |a, b, y\rangle^B$$

[4]This step closely resembles the code construction in [25], which contains a slightly more detailed proof. The basic tools for the proof can also be found in [14], [43]; note that they (especially [43]) only bound the *average* decoding error, rather than the maximum error (similarly in [12], which introduced the idea of double-blocking unitary communication protocols). An average error can easily be turned into a maximum error[20], though in our case it is not necessary, or rather, our entire protocol can be thought of as a coherent version of [20].

where $\mathcal{A} := \{0,1\}^{C_1^{(n)}}$, $\mathcal{B} := \{0,1\}^{C_2^{(n)}}$, and $N := |\mathcal{A}| \cdot |\mathcal{B}| = \exp(C_1^{(n)} + C_2^{(n)})$. This can be locally mapped to

$$\frac{1}{\sqrt{N}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} |a, b, a \oplus x\rangle^A |a, b, b \oplus y\rangle^B .$$

Relabelling the sum over a, b shows that this is equivalent to

$$\frac{1}{\sqrt{N}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} |a \oplus x, b \oplus y, a\rangle^A |a \oplus x, b \oplus y, b\rangle^B .$$

Now $\mathcal{P}_n$ is applied to $|a\rangle^A |b\rangle^B$, obtaining

$$\frac{1}{\sqrt{N}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} |a \oplus x, b \oplus y, a\rangle^A |a \oplus x, b \oplus y, b\rangle^B |\varphi_{a,b}\rangle .$$

Since $\mathrm{Tr}(|b\rangle\langle b|^A \otimes |a\rangle\langle a|^B \otimes I^{AB})\varphi_{a,b} \geq 1 - \epsilon_n$, we can extract $|b\rangle^A |a\rangle^B$ from $|\varphi_{a,b}\rangle$ while causing $O(\sqrt{\epsilon_n})$ disturbance. This yields a state within $O(\sqrt{\epsilon_n})$ of

$$\frac{1}{\sqrt{N}} \sum_{\substack{a\in\mathcal{A} \\ b\in\mathcal{B}}} |a \oplus x, b \oplus y, a, b\rangle^A |a \oplus x, b \oplus y, b, a\rangle^B |\varphi_{a,b}\rangle$$

which can be locally mapped to

$$|x, y\rangle^A |x, y\rangle^B \frac{1}{\sqrt{N}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} |a, b\rangle^A |a, b\rangle^B |\varphi_{a,b}\rangle =: |x, y\rangle^A |x, y\rangle^B |\bar{\varphi}\rangle^{AB} .$$

Thus, we have performed the desired coherent communication (up to error $O(\sqrt{\epsilon_n}) = O(\exp(-n^{1-\gamma}/2))$) and converted $n(C_1 + C_2 - 2\delta_n)$ ebits into $|\bar{\varphi}\rangle^{AB}$, which is the fixed pure state

$$\frac{1}{\sqrt{N}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} |a, b\rangle^A |a, b\rangle^B |\varphi_{a,b}\rangle .$$

The final step is to recover entanglement from $|\bar{\varphi}\rangle^{AB}$. Since $E(\varphi_{a,b}) \geq n(E - \delta_n)$, we can bound

$$E_0 := E(\bar{\varphi}) \geq n(C_1 + C_2 + E - 3\delta_n).$$

Also, $|\bar{\varphi}\rangle^{AB}$ can be obtained from $O(n)[qq] + O(n)\langle U\rangle$, so it must have Schmidt rank $\leq \exp(O(n))$. We would like to repeat the entire protocol k times in parallel, and then apply entanglement concentration to $|\bar{\varphi}\rangle^{\otimes k}$ in order to recover standard EPR pairs. This will inevitably increase our error; we get up to $k \cdot O(\sqrt{\epsilon_n})$ from repeating the protocol, and some additional errors from the entanglement concentration. However, as long as our final error ϵ' is $o(1)$ then we can use the catalytic entanglement safely; we merely repeat the protocol $\lceil 1/\epsilon' \rceil$ times for a total error of $\sqrt{\epsilon'}$ and an additional fractional inefficiency of $\sqrt{\epsilon'}$, both of which are still $o(1)$.

Unfortunately, the original entanglement concentration protocol of [5] will not suffice for this purpose (without using a more elaborate procedure, as in [25]). This is because [5] requires $k/\log k$ to grow faster than the Schmidt rank of $|\bar{\varphi}\rangle$, meaning that $k = \exp(\Omega(n))$. Since $\epsilon_n = \exp(-o(n))$, we would be unable to guarantee that $k\epsilon_n = o(1)$.

Instead, we will use an approximate version of entanglement concentration (due to Andreas Winter[47]) which only requires $k = \text{poly}\log(\text{Sch}(\bar{\varphi})) = \text{poly}(n)$ to achieve vanishing error and inefficiency. Since $\text{poly}(n)\epsilon_n = o(1)$, this will complete our proof. □

It now remains only to describe Winter's new version of entanglement concentration[47]. Since this result may be more broadly useful, we rename the variables from the proof of Theorem 1 to more conventional notation, and state a slightly more general result than we need above.

Theorem 5 (Due to A. Winter): Let $|\psi_1\rangle^{AB}, \ldots, |\psi_n\rangle^{AB}$ be bipartite states each with Schmidt rank $\leq d$, and with total entanglement $E := \sum_{i=1}^n H(\psi_i^A)$. Then for any $\delta > 0$ such that $n \geq \max(\delta^{-2}3(\log d)^3, 20\delta^{-1}\log n\delta)$, Alice and Bob can extract $E - n\delta$ ebits from $|\psi_1\rangle^{AB} \otimes \cdots \otimes |\psi_n\rangle^{AB}$ with error $O(\exp(-(n\delta^2)^{1/3}/2\ln 2))$ using no communication. Up to the above error, their residual state has Schmidt rank $\leq \exp(2n\delta)$.

In particular, $E(1 - o(1))$ ebits can be extracted with error $o(1)$ while creating a sublinear-size garbage state if $n = \text{poly}(\log d)$. This suffices to complete the above proof of Theorem 1.

Proof of Theorem 5: First we describe the protocol, then analyze its correctness. Alice and Bob begin by using local unitaries to rotate the Schmidt basis of each $|\psi_i\rangle$ into a standard basis. This leaves them with the state

$$\sum_{j_1=1}^{d} \sqrt{p_{j_1}^1}\, |j_1\rangle^A |j_1\rangle^B \otimes \cdots \otimes \sum_{j_n=1}^{d} \sqrt{p_{j_n}^n}\, |j_n\rangle^A |j_n\rangle^B$$

where $(p_1^i, \ldots, p_d^i)$ are the Schmidt coefficients of $|\psi_i\rangle$ (possibly not all nonzero).

Next they each project onto the subspace where the Schmidt coefficients are in the range $\exp(-E \pm n\delta/2)$; that is spanned by $|j_1, \ldots, j_n\rangle \otimes |j_1, \ldots, j_n\rangle$ for those values of $j_1, \ldots, j_n$ satisfying $|\sum_i \log p_{j_i}^i + E| \leq n\delta/2$. We will later argue that this projection almost always succeeds, and thus causes very little disturbance.

Now they divide the interval $[\exp(-E - n\delta/2), \exp(-E + n\delta/2)]$ into m bins with geometrically spaced boundaries, for m a parameter we will pick later. That is, for $k = 1, \ldots, m$, bin k is the interval

$$\left[\exp\left(-E - \frac{n\delta}{2} + \frac{j-1}{m}n\delta\right), \exp\left(-E - \frac{n\delta}{2} + \frac{j}{m}n\delta\right)\right].$$

Still without using any communication, Alice and Bob each perform a projective measurement onto the different bins. By this we mean that the measurement operators are projectors onto subspaces spanned by strings $|j_1, \ldots, j_n\rangle$ whose Schmidt coefficients fall entirely into one of the above intervals. We claim that (a) that with high probability they will find a bin that contains many eigenstates, and (b) the resulting state will have high fidelity with a maximally entangled state. It remains only to quantify the various errors and inefficiencies we have encountered along the way.

Start with the last step. The probability that the bin measurement yields a bin with weight $\leq \epsilon$ (for ϵ a parameter we will set later) is $\leq m\epsilon$. Each Schmidt coefficient that we have kept is $\leq \exp(-E + n\delta/2)$, and thus any bin with weight $\geq \epsilon$ must contain $\geq \epsilon \exp(E - n\delta/2)$ eigenvalues. Choosing $\epsilon = 2^{-n\delta/2}$, we obtain a state that approximates $\geq \log 1/\epsilon + E - n\delta/2 = E - n\delta$ EPR pairs. To assess the fidelity of this approximation, note that all of the Schmidt coefficients of the projected and rescaled state are in a band between λ and $\lambda \exp(-n\delta/m) \geq \lambda(1 - n\delta(\ln 2)/m)$, for some normalization factor λ. Thus, this state has fidelity $1 - O(n\delta/m)$ with a maximally entangled state. Now choose $m = 2^{n\delta/4}$, so that both the $m\epsilon$ failure probability and the $O(n\delta/m)$ error are $\leq \exp(-n\delta/5)$.

All that remains for the error analysis is to assess the damage from projecting onto the Schmidt coefficients in the interval $-E \pm n\delta/2$. The key tool here is the Chernoff bound[13], which states that if $X_1, \ldots, X_n$ are independent (not necessarily independent) random variables satisfying $0 \leq X_i \leq \gamma$ then $X := \sum_{i=1}^n X_i$ satisfies

$$P\left(|X - \mathbb{E}X| \geq n\frac{\delta}{2}\right) \leq 2\exp\left(-\frac{n\delta^2}{\gamma^2 2\ln 2}\right). \tag{20}$$

We would like to apply this with X_i defined by $P(X_i = \log 1/p_j^i) = p_j^i$, so that X is likely to be close to $\mathbb{E}X = E$. Unfortunately, p_j^i can be arbitrarily close to zero, and thus $\log 1/p_j^i$ can be arbitrarily large, so we cannot immediately establish any upper bound on γ.[5] To do so, we will discard the Schmidt coefficients that are smaller than $2^{-\gamma}$, which automatically means that $0 \leq X_i \leq \gamma$. This causes $\leq d2^{-\gamma}$ damage for each $|\psi_i\rangle$, or $\leq nd2^{-\gamma}$ overall. Combining with (20), we find the total error is $nd2^{-\gamma} + e^{\frac{-n\delta^2}{2\gamma^2}}$. Finally, we set $\gamma = (n\delta^2)^{1/3}$ to obtain an overall error of $O(\exp(-(n\delta^2)^{1/3}/2\ln 2))$, which dominates the $\exp(-n\delta/5)$ error from the first part.

Now we explain how the protocol can be made semi-clean. The "which bin" measurement should instead be performed coherently, with the entanglement extraction proceeding conditioned on the quantum register storing the superposition of measurement outcomes. Since the different outcomes remain locally orthogonal, the overall Schmidt rank is equal to the sum over k of the rank of the state conditioned on obtaining bin k. After the initial projection onto the typical subspace succeeds, each Schmidt coefficient is $\geq \exp(-E - n\delta/2)$, and so each bin has rank $\leq \exp(E + n\delta/2)$. Since we are extracting $E - n\delta$ ebits from each bin, the rank of the residual state, conditioned on k, should be $\leq \exp\left(\frac{3}{2}n\delta\right)$. Once we sum over $m = 2^{n\delta/4}$ bins, we have an overall Schmidt rank of $\leq \exp(\frac{7}{4}n\delta) \leq \exp(2n\delta)$. □

Proof of Coding Theorem for Theorem 2: Suppose there exists an ensemble $\mathcal{E}^{XAA'BB'} = \sum_x p_x |x\rangle\langle x|^X \otimes \psi_x^{AA'BB'}$ such that

$$\begin{aligned} I(X;BB')_{U(\mathcal{E})} - I(X;BB')_{\mathcal{E}} &= C \\ \text{and } H(BB'|X)_{U(\mathcal{E})} - H(BB'|X)_{\mathcal{E}} &= E \end{aligned} \tag{21}$$

The idea is to use HSW coding[31], [42] for $U(\mathcal{E})^{BB'}$. For a string $\vec{x} = (x_1, \ldots, x_n)$, let $|\psi_{\vec{x}}\rangle := |\psi_{x_1}\rangle \otimes \cdots |\psi_{x_n}\rangle$. The HSW theorem states that for $\epsilon, \delta > 0$ and for n sufficiently large, choosing $N := \exp(n(I(X;BB')_{U(\mathcal{E})} - \delta))$ random codewords $U^{\otimes n}(\psi_{\vec{x}_1})^{BB'}, \ldots, U^{\otimes n}(\psi_{\vec{x}_N})^{BB'}$ according to the distribution $\vec{p}(\vec{x}) := p_{x_1}p_{x_2}\ldots p_{x_n}$ will result in a code with average error $\leq \epsilon$.

[5]The Chebyshev bound would avoid these difficulties, but at the cost of losing the exponential bounds on error probability. Nevertheless, for small values of n, it may be preferable. Here we can use the fact that $\text{Var}(X) \leq n\log^2 d$ to find that $P(|X - \mathbb{E}X| \geq n\delta/2) \leq 4(\log d)^2/n\delta^2$. Other than the revised error bound, the rest of the proof would be the same.

On the other hand, the operator Chernoff bound[46] states that a collection of $M := \exp(n(I(X;BB')_{\mathcal{E}} + \delta))$ random codewords $\psi_{\vec{x}_1}, \ldots, \psi_{\vec{x}_M}$ will have average state on Bob's side quite close to their expectation

$$\theta^{BB'} := \mathbb{E}_{\vec{x}} \psi_{\vec{x}}^{BB'} = \left(\sum_x p_x \psi_x^{BB'} \right)^{\otimes n}. \tag{22}$$

Choose an arbitrary purification $|\theta\rangle^{ABB'}$. Also let $N = LM$, so

$$\begin{aligned} L &= \exp(n(I(X;BB')_{U(\mathcal{E})} - I(X;BB')_{\mathcal{E}} - 2\delta)) \\ &= \exp(n(C - 2\delta)). \end{aligned}$$

Our strategy for the rest of the proof is for Alice and Bob to start with a state where Bob's part always looks like $\theta^{BB'}$, but Alice can reliably send one of L different messages by performing a local unitary and then applying U to the joint state.

With this in mind, we now rephrase the random codes described above. Draw $\{\vec{x}_{i,j}\}_{i \in [L], j \in [M]}$ from the distribution $\vec{p}$, and let $e_{i,j}$ be the probability of error when Bob attempts to decode $U^{\otimes n} |\psi_{\vec{x}_{i,j}}\rangle$. The HSW theorem states that with high probability the average error is low, i.e.,

$$\frac{1}{LM} \sum_{i=1}^{L} \sum_{j=1}^{M} e_{i,j} \leq \epsilon. \tag{23}$$

On the other hand, the operator Chernoff bound states that with high probability

$$\frac{1}{M} \sum_{j=1}^{M} \psi_{\vec{x}_{i,j}}^{BB'} \approx_{\epsilon^2/2} \theta^{BB'} \tag{24}$$

for all $i \in [L]$ (the reason to demand error $\epsilon^2/2$ will later be apparent). Using the union bound (see, e.g., the proof of Theorem 1 of [16] for detailed calculations), one can show that in fact with high probability both (23) and (24) hold simultaneously, and in particular that there exists a set of $\{\vec{x}_{i,j}\}$ for which this is true. Fix this set for the rest of the proof.

For each i, let $e_i := \sum_j e_{i,j}/M$ and define the set of good codewords to be $G = \{i : e_i \leq 2\epsilon\}$. By Markov's inequality, $|G| \geq L/2$.[6] The communication protocol proceeds as follows:

(1) Alice and Bob start with the state $|\theta\rangle^{ABB'}$.

(2) To send the message $i \in G$, Alice will perform a local unitary operation so that the overall state is within ϵ of

$$\frac{1}{\sqrt{M}} |i\rangle^A \sum_{j=1}^{M} |j\rangle^A \left|\psi_{\vec{x}_{i,j}}\right\rangle^{AA'BB'}.$$

This is possible because of (24), Uhlmann's theorem[36], and the fact that two mixed states with trace distance $\leq \epsilon^2/2$ have purifications with trace distance $\leq \epsilon$ [19, Lemma 2.2].

(3) Apply $(U^{AB})^{\otimes n}$ so that the two parties share a state within ϵ of

$$\frac{1}{\sqrt{M}} |i\rangle^A \sum_{j=1}^{M} |j\rangle^A (U^{AB} \otimes I)^{\otimes n} \left|\psi_{\vec{x}_{i,j}}\right\rangle^{AA'BB'}.$$

(4) Bob decodes coherently, to obtain a state within $\epsilon + e_i \leq 3\epsilon$ of

$$\frac{1}{\sqrt{M}} |ii\rangle^{AB} \sum_{j=1}^{M} |jj\rangle^{AB} (U^{AB} \otimes I)^{\otimes n} \left|\psi_{\vec{x}_{i,j}}\right\rangle^{AA'BB'}.$$

(5) Conditioned on i, j, Alice and Bob concentrate $\approx nH(BB'|X)_{U(\mathcal{E})}$ ebits from $(U^{AB} \otimes I)^{\otimes n} \left|\psi_{\vec{x}_{i,j}}\right\rangle^{AA'BB'}$. Since the dimension of the states is fixed and n can be made arbitrarily large, the entanglement concentration technique of [5] will suffice. Moreover, entanglement concentration can be performed cleanly, so a sublinear amount of additional communication will leave them with the state $M^{-1/2} \sum_{j=1}^{M} |j\rangle^A |j\rangle^B$, which is of course equivalent to $n(I(X;BB')_{\mathcal{E}} + \delta)$ ebits.

Alice and Bob have used U $n + o(n)$ times and sent $\log L = n(C - 2\delta)$ cobits. They started with the state $|\theta\rangle^{ABB'}$, which can be prepared with entanglement dilution using $nH(BB')_{\mathcal{E}} + o(n)$ ebits and $o(n)$ cbits[39], and end with $n(I(X;BB')_{\mathcal{E}} + H(BB'|X)_{U(\mathcal{E})}) \pm o(n)$ ebits, for a net change of

$$n(H(BB'|X)_{U(\mathcal{E})} - H(BB'|X)_{\mathcal{E}}) \pm o(n) = n(E \pm o(1))$$

as desired. □

Note that unlike the proof of Theorem 1 (or indeed the proof in [22] of the present result), no double-blocking is necessary here, except perhaps to deal with the sublinear communication used for entanglement dilution and to erase the ancilla states left by entanglement concentration.

[6]Note that unlike in standard HSW coding, we cannot simply throw out the worst half of all codewords, since then the $\vec{x}_{i,j}$ would no longer be independent and (24) would no longer necessarily hold.

ACKNOWLEDGMENT

The authors would like to thank Andreas Winter for allowing them to include Theorem 5, Harry Buhrman for teling them about [40], and Noah Linden for telling them about [38]. A. W. Harrow would also like to thank Charlie Bennett, Debbie Leung, and John Smolin for suggesting the problem of asymmetric unitary gate capacities and for many interesting discussions on the subject.

REFERENCES

[1] A. Abeyesinghe, P. Hayden, G. Smith, and A. Winter, "Optimal superdense coding of entangled states," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3635–3641, Aug. 2006.

[2] R. Alicki and M. Fannes, "Continuity of quantum conditional information," *J. Phys. A*, vol. 37, pp. L55–L57, 2004.

[3] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, "Causal and localizable quantum operations," *Phys. Rev. A*, vol. 64, p. 052309, 2001, Available at quant-ph/0102043.

[4] C. H. Bennett, "The thermodynamics of computation—A review," *Int. J. Theor. Phys.*, vol. 21, no. 12, pp. 905–940.

[5] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A*, vol. 53, pp. 2046–2052, 1996.

[6] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The Quantum Reverse Shannon Theorem," Tech. Rep., to be published.

[7] C. H. Bennett, A. W. Harrow, D. W. Leung, and J. A. Smolin, "On the capacities of bipartite Hamiltonians and unitary gates," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1895–1911, Aug. 2003.

[8] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. Winter, "Remote preparation of quantum states," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 56–74, Jan. 2005.
[9] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. Thapliyal, "Entanglement-assisted classical capacity of noisy quantum channels," *Phys. Rev. Lett.*, vol. 83, pp. 3081–3084, 1999.
[10] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. Thapliyal, "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.
[11] B. Berger, "The fourth moment method," in *Proc. 2nd Annu. ACM-SIAM Sympo. Discrete Algorithms (SODA '910*, Philadelphia, PA, 1991, pp. 373–383.
[12] D. W. Berry and B. C. Sanders, "Relation between classical communication capacity and entanglement capability for two-qubit unitary operations," *Phys. Rev. A*, vol. 68, p. 032312, 2003.
[13] H. Chernoff, "A measure of the asymptotic efficiency of tests of a hypothesis based on a sum of observations," *Ann. Math. Statist.*, vol. 23, pp. 493–507, 1952.
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Series in Telecommunication. New York: Wiley , 1991.
[15] W. van Dam, "Implausible Consequences of Superstrong Nonlocality," Tech. Rep., 2005 [Online]. Available: arXiv:quant-ph/0501159
[16] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 44–55, Jan. 2005.
[17] I. Devetak, "A triangle of dualities: reversibly decomposable quantum channels, source-channel duality, and time reversal," *Phys. Rev. Lett.*, vol. 97, p. 140503.
[18] I. Devetak, A. W. Harrow, and A. Winter, "A family of quantum protocols," *Phys. Rev. Lett.*, vol. 93, p. 239503, 2004.
[19] I. Devetak, A. W. Harrow, and A. Winter, "A resource framework for quantum Shannon theory," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, Oct. 2008, "," 2008, Tech. Rep. 10.
[20] I. Devetak and A. Winter, "Maximal and Average Error Capacity Regions Coincide—Under Randomised Encodings," Tech. Rep., 2005, unpublished.
[21] S. van Enk, "Quantifying the resource of sharing a reference frame," *Phys. Rev. A*, vol. 71, p. 032339, 2005.
[22] A. W. Harrow, "Coherent communication of classical messages," *Phys. Rev. Lett.*, vol. 92, p. 097902, 2004.
[23] A. W. Harrow, "Applications of Coherent Classical Communication and Schur Duality to Quantum Information Theory," Ph.D. dissertation, MIT, Cambridge, MA, 2005.
[24] A. W. Harrow, "Exact universality from any entangling gate without inverses," *Quart. Inf. Comp.*, vol. 8, no. 8&9, pp. 715–721, 2008, "," Tech. Rep., 2008.
[25] A. W. Harrow and D. W. Leung, "Bidirectional coherent classical communication," *Quantum Inf. Comp.*, vol. 5, no. 4–5, pp. 380–395.
[26] A. W. Harrow and D. W. Leung, "An Exponential Separation Between the Entanglement and Communication Capacities of a Bipartite Unitary Interaction," Tech. Rep., 2008.
[27] A. W. Harrow and H. K. Lo, "A tight lower bound on the classical communication cost of entanglement dilution," *IEEE Trans. Inf. Theory*, vol. 50, no. 2, pp. 319–327, Feb. 2004.
[28] P. Hayden and W. van Dam, "Universal entanglement transformations without communication," *Phys. Rev. A*, vol. 67, p. 060302(R), 2003.
[29] P. Hayden, D. W. Leung, and A. Winter, "Aspects of generic entanglement," *Commun. Math. Phys.*, vol. 265, p. 95, 2006.
[30] P. Hayden and A. Winter, "On the communication cost of entanglement transformations," *Phys. Rev. A*, vol. 67, p. 012306, 2003.
[31] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 269–273, Jan. 1998.
[32] A. S. Holevo, "On entanglement assisted classical capacity," *J. Math. Phys.*, vol. 43, no. 9, pp. 4326–4333, 2002.
[33] M. Horodecki, J. Oppenheim, and A. Winter, "Quantum information can be negative," *Nature*, vol. 436, pp. 673–676, 2005.
[34] D. Kretschmann and R. F. Werner, "Tema Con Variazioni: Quantum channel capacity," *New J. Phys.*, vol. 6, p. 26, 2004.
[35] M. Ledoux, *The Concentration of Measure Phenomenon*. Providence, RI: Amer. Math. Soc., vol. 89, AMS Mathematical Surveys and Monographs.
[36] M. A. Nielsen and I. A. Chuang, *Quantum Computation and Quantum Information*. New York: Cambridge Univ. Press, 2000.
[37] M. S. Leifer, L. Henderson, and N. Linden, "Optimal entanglement generation from quantum operations," *Phys. Rev. A*, vol. 67, p. 012306, 2003.
[38] N. Linden, J. A. Smolin, and A. Winter, "The Entangling and Disentangling Power of Unitary Transformations are Unequal," Tech. Rep., 2005.
[39] H.-K. Lo and S. Popescu, "The classical communication cost of entanglement manipulation: Is entanglement an inter-convertible resource?," *Phys. Rev. Lett.*, vol. 83, pp. 1459–1462, 1999.
[40] N. Nisan, "The communication complexity of threshold gates," in *Combinatorics, Paul Erdös is Eighty*, V. S. D. Mikl'os and T. Szonyi, Eds. Budapest, Hungary: Jànos Bolyai Math. Soc., vol. I., pp. 301–315.
[41] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, "A thermodynamical approach to quantifying quantum correlations," *Phys. Rev. Lett.*, vol. 89, p. 180402, 2002.
[42] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, pp. 131–138, 1997.
[43] C. E. Shannon, *Two-Way Communication Channels*. Berkeley, CA: Univ. California Press, 1961.
[44] P. W. Shor, "The classical capacity achievable by a quantum channel assisted by limited entanglement," *Quant. Inf. Comp.*, vol. 4, no. 6&7, Dec. 2004.
[45] L. Szilard, "Über die entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen," *Zeits. Physik*, vol. 53, pp. 840–856.
[46] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2481–2485, Nov. 1999.
[47] A. Winter, Personal Communication.

Aram W. Harrow received the Ph.D. degree in physics from the Massachusetts Institute of Technology (MIT), Cambridge, in 2005.

Since then he has worked as a Lecturer at the University of Bristol, Bristol, U.K., first in the Department of Computer Science, and then since 2008, the Department of Mathematics. His research interests include quantum information theory and quantum algorithms.

Peter W. Shor received the Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1985.

He has worked as a researcher at Bell Labs and since 2003 has been the Morss Professor of Applied Mathematics at MIT.

Prof. Shor invented the quantum algorithm for prime factorization in 1994, and has been awarded the Nevanlinna Prize (1998), MacArthur Fellowship (1999), the Gödel Prize (1999), and the King Faisal International Prize in Science (2002).