Quantum Information and Computation, Vol. 5, No. 4&5 (2005) 380–395

BIDIRECTIONAL COHERENT CLASSICAL COMMUNICATION

ARAM W. HARROW
MIT Physics Dept., 77 Massachusetts Avenue
Cambridge, Massachusetts 02139, USA

DEBBIE W. LEUNG
MSC 107-81, IQI, Caltech
Pasadena, California 91125, USA

Received December 16, 2004
Revised May 11, 2005

A unitary interaction coupling two parties enables quantum or classical communication in both the forward and backward directions. Each communication capacity can be thought of as a tradeoff between the achievable rates of specific types of forward and backward communication. Our first result shows that for any bipartite unitary gate, bidirectional coherent classical communication is no more difficult than bidirectional classical communication — they have the same achievable rate regions. Previously this result was known only for the unidirectional capacities (i.e., the boundaries of the tradeoff). We then relate the tradeoff for two-way coherent communication to the tradeoff for two-way quantum communication and the tradeoff for coherent communication in one direction and quantum communication in the other.

Keywords: Bidirectional channels, channel capacities, coherent classical communication
Communicated by: H-K Lo

1. Introduction

Quantum communication theory typically studies channels which take an input quantum system from one party (call her Alice), act on it possibly with some noise (a trace preserving completely positive map[1]) and pass the system onto another party (call him Bob). A quantum channel can generate quantum or classical communication or entanglement at some rate. The maximum rate at which each task can be done with arbitrary precision and with an asymptotically large number of channel uses is called the *capacity*.

A bipartite unitary gate coupling Alice and Bob can achieve similar tasks, with either party (or both) in the role of sender or receiver. Early studies can be found in [2, 3, 4, 5], focusing on more specific systems and protocols. For example, a CNOT can send a classical bit from Alice to Bob, or from Bob to Alice or generate one EPR pair. *Asymptotic capacities* of a general bipartite unitary evolution to communicate and to generate entanglement were formalized in Ref. [6]. A general expression for the entanglement capacity was found in Refs. [7, 6] and that for entanglement-assisted one-way classical capacity was found in Ref. [6]. Expressions for various one-way quantum capacities were subsequently found in Ref. [8], by introducing

the concepts of *coherent classical communication* and entanglement recycling. (Their precise definitions, as well as concepts throughout the rest of this paragraph, will be clarified in Sec. 2). In particular, Ref. [8] showed that for any gate, the one-way classical capacity is equal to its one-way coherent capacity. This further provides an expression for the one-way classical capacity assisted by any linear amount of free entanglement, and allows the one-way quantum capacity and the *remote state preparation capacity* to be expressed in terms of this one-way classical capacity.

However, the core result for bipartite unitary evolution in Ref. [8], the equality of the one-way classical capacity and the coherent capacity, is left open for simultaneous two-way communication. Our main result is a proof of this equality in Sec. 3. For completeness, we also compare two-way classical communication and coherent classical communication in the regime of negative communication rates (i.e., consuming communication to help produce other resources). Following similar arguments as in Ref. [8], we list some corollaries. These are the two-way remote state preparation capacity and quantum capacity in terms of the classical capacity. Our main result is proved by using a coherent version of a one-time pad (analogous to that in Ref. [9]). The reason why a more direct extension of the proof from Ref. [8] fails is given in an appendix. A second appendix discusses the implications our results have on the definition of coherent classical communication.

2. Framework, definitions, and notations

Throughout the paper, we consider communication between two parties, Alice and Bob. Systems in their possession are denoted by respective subscripts A, $\mathrm{A}_{0,1,\ldots}$ and B, $\mathrm{B}_{0,1,\ldots}$. System labels are omitted when they are clear from the context. We also use superscripts (A) and (B) for *different* (but analogous) objects related to Alice and Bob (for example, their respective local operations). Exp and log are always base 2. We will primarily use the trace distance $\frac{1}{2}\|\rho - \sigma\|_1$ to quantify the proximity of any two states ρ and σ, where $\|X\|_1 := \mathrm{Tr}\sqrt{X^\dagger X}$. For two pure states $|\alpha\rangle, |\beta\rangle$, $\frac{1}{2}\| \, |\alpha\rangle\langle\alpha| - |\beta\rangle\langle\beta| \, \|_1 = \epsilon \;\Leftrightarrow\; |\langle\beta|\alpha\rangle|^2 = 1 - \epsilon^2$. We use $|\alpha\rangle \overset{\epsilon}{\approx} |\beta\rangle$ as a shorthand for $\frac{1}{2}\| \, |\alpha\rangle\langle\alpha| - |\beta\rangle\langle\beta| \, \|_1 \leq \epsilon$.

We now review some definitions and background results, mostly from Refs. [6, 8, 10]. Let $\{|x\rangle\}_{x=0,1}$ be a basis for $\mathbb{C}^2$. We first define various resources. Let an ebit denote a unit of shared quantum correlation, as quantified by an EPR pair $|\Phi\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{2}}\sum_{x=0}^{1} |x\rangle_{\mathrm{A}}|x\rangle_{\mathrm{B}}$. Throughout the paper, we omit the tensor product symbol, $\otimes$, if no confusion may arise. Following Ref. [8], we denote the ability to communicate a qubit in the forward direction (from Alice to Bob) as qubit($\rightarrow$), and mathematically, it corresponds to the isometry $|x\rangle_{\mathrm{A}} \rightarrow |x\rangle_{\mathrm{B}}$. Qubit communication in the opposite direction, the isometry $|x\rangle_{\mathrm{B}} \rightarrow |x\rangle_{\mathrm{A}}$, is denoted qubit($\leftarrow$). Nonunitary evolution can be viewed as a unitary evolution between all participating parties, together with an inaccessible one called the environment denoted by E. Then, the ability to communicate a classical bit in the forward direction, denoted as cbit($\rightarrow$), is given by the linear map $|x\rangle_{\mathrm{A}} \rightarrow |x\rangle_{\mathrm{B}}|x\rangle_{\mathrm{E}}$. In contrast, a cobit($\rightarrow$) is given by the map $|x\rangle_{\mathrm{A}} \rightarrow |x\rangle_{\mathrm{A}}|x\rangle_{\mathrm{B}}$. A cbit($\leftarrow$) and a cobit($\leftarrow$) are defined similarly. We call cobits *coherent* classical communication, and cbits *incoherent* classical communication or simply classical communication. One can view cobits as cbits in which Alice is given the environment E as

quantum feedback. The results of this paper imply that cobits may be equivalently defined as the ability to send cbits through unitary means. In Appendix B we will make this idea precise.

Communication theory is primarily concerned with converting available resources into desired ones. Roughly speaking, given two communication resources X and Y, we say that $X \geq rY$ if X can be transformed into Y asymptotically and approximately at rate r, i.e., $\forall\, \delta > 0$, $\exists N$ such that $\forall\, n \geq N$, n copies (or uses) of X can be transformed into $\geq n(r-\delta)$ copies (or uses) of Y, in an approximate manner to be defined. For example, Shannon's noisy coding theorem [11] for a classical channel (i.e. a stochastic map) T could be stated as $T \geq C(T)$ cbits, where $C(T) := \max_{P(\Xi)} [H(\Xi)+H(T(\Xi))-H(\Xi,T(\Xi))]$ is the classical capacity of the channel T, $H(\cdot)$ is the entropy of a random variable, and the maximization is over all distribution $P(\Xi)$ of the input alphabet Ξ. If $X \geq Y$ and $Y \geq X$, then we write that $X = Y$. For example, the reverse Shannon theorem [12] states that $C(T)$ cbits $\geq T$, so that $T_1 = \frac{C(T_1)}{C(T_2)}\, T_2$ for any two classical channels T_1, T_2 (in the presence of unlimited shared randomness). Another result [8] of this type, $2\,\text{cobits}(\rightarrow) = 1\,\text{ebit} + 1\,\text{qubit}(\rightarrow)$, will be used in Sec. 4 to relate the classical and quantum capacities of unitary gates.

The definition for $X \geq rY$ is only complete given an error definition, and a good one should ensure transitivity of resource inequalities: $X \geq rY$ and $Y \geq sZ$ implies $X \geq rsZ$. Operationally, the two corresponding resource transformations should be sufficiently accurate to be composable. Mathematically, we say that $X \geq rY$ if there exist *vanishing* sequences of nonnegative numbers, $\{\epsilon_n\}, \{\delta_n\}$, and protocols $\mathcal{P}_n$ each using X at most n times (and other allowed resources), such that $\mathcal{P}_n \overset{\epsilon_n}{\approx} Y^{\otimes(r-\delta_n)n}$. Here the notion of approximation $\overset{\epsilon_n}{\approx}$ is extended from states to operations as

$$\forall |\psi\rangle \quad \tfrac{1}{2} \parallel \mathcal{I}\otimes\mathcal{P}_n(|\psi\rangle) - \mathcal{I}\otimes Y^{\otimes(r-\delta_n)n}(|\psi\rangle) \parallel_1 \leq \epsilon_n\,, \tag{1}$$

where $\mathcal{I}$ denotes the identity operation on a *reference system* of dimension given by the input to $\mathcal{P}_n$. Including a reference system in Eq. (1) ensures that $\mathcal{P}_n$ and $Y^{\otimes(r-\delta_n)n}$ transform correlations similarly. Here, we use the symbol Y to denote the associated state transformation enabled by the resource (see Sec. 1 for examples). We will see examples of what the above means in the next section.

We can now define the achievable classical rate region of a unitary gate U as the set of points (C_1, C_2, E) such that $U \geq C_1\,\text{cbits}(\rightarrow) + C_2\,\text{cbits}(\leftarrow) + E\,\text{ebits}$. When C_1, C_2, or E is negative, it means that the resource is being consumed; for example, if $E < 0$ and $C_1, C_2 \geq 0$, then $U + (-E)\,\text{ebits} \geq C_1\,\text{cbits}(\rightarrow) + C_2\,\text{cbits}(\leftarrow)$ represents entanglement-assisted communication. This paper is mostly concerned with $C_1, C_2 \geq 0$ and arbitrary E. Part of the (C_1, C_2, E) achievable region has been characterized, for the special cases of $C_1, C_2 \leq 0$ (entanglement capacity [6, 7] which is not increased by free classical communication), $C_2 = 0, E = -\infty$ (one-way classical communication with unlimited entanglement assistance [6], though the actual protocol requires only finite entanglement assistance) and $C_2 = 0$ (one-way classical communication with arbitrary entanglement assistance [8]). We can define the achievable coherent classical rate region of U analogously as the triples (C_1, C_2, E) so that $U \geq C_1\,\text{cobits}(\rightarrow) + C_2\,\text{cobits}(\leftarrow) + E\,\text{ebits}$.

Reference [8] showed that $U \geq C\,\text{cbits}(\rightarrow) + E\,\text{ebits}$ if and only if $U \geq C\,\text{cobits}(\rightarrow) + E\,\text{ebits}$,

i.e., the coherent and incoherent classical rate regions coincide on the planes $C_1 = 0$ and $C_2 = 0$. In the next section we prove that the coherent and incoherent rate regions are identical in the entire $C_1, C_2 \geq 0$ quadrant. Other quadrants will be considered for completeness – this amounts to understanding how to best use back classical communication. We will see that assistance by cobits only generates entanglement and that cbits are useless. We then apply the result to relate the capacity regions of different types of forward and backward communication.

3. Bidirectional coherent classical communication

Theorem 1 *For any bipartite unitary or isometry U and $C_1, C_2 \geq 0$,*

$$U \geqslant C_1 \text{ cbits}(\rightarrow) + C_2 \text{ cbits}(\leftarrow) + E \text{ ebits} \quad \text{iff} \tag{2}$$

$$U \geqslant C_1 \text{ cobits}(\rightarrow) + C_2 \text{ cobits}(\leftarrow) + E \text{ ebits} \tag{3}$$

Proof: Since 1 cobit $\geq$ 1 cbit, it suffices to prove the forward implication. In other words, given the existence of protocols achieving the resource transformation in Eq. (2), we will construct protocols that achieve the resource transformation in Eq. (3). We delay the discussion for $E \neq 0$ until the end of this section. For now, suppose $E = 0$.

• *The definition of $\mathcal{P}_n$*

Formally, Eq. (2) indicates the existence of sequences of nonnegative real numbers $\{\epsilon_n\}, \{\delta_n\}$ satisfying $\epsilon_n, \delta_n \to 0$ as $n \to \infty$; a sequence of protocols $\mathcal{P}_n = (V_n \otimes W_n)\, U \cdots U\, (V_1 \otimes W_1)\, U \times (V_0 \otimes W_0)$, where V_j, W_j are local isometries that may also act on extra local ancilla systems, and sequences of integers $C_1^{(n)}, C_2^{(n)}$ satisfying $nC_1 \geq C_1^{(n)} \geq n(C_1 - \delta_n)$, $nC_2 \geq C_2^{(n)} \geq n(C_2 - \delta_n)$, such that the following success criterion holds.

Let $a \in \{0,1\}^{C_1^{(n)}}$ and $b \in \{0,1\}^{C_2^{(n)}}$ be the respective messages of Alice and Bob. Let $|\varphi_{ab}\rangle := \mathcal{P}_n(|a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1})$. Note that $|\varphi_{ab}\rangle$ generally occupies a space of larger dimension than $\mathrm{A}_1 \otimes \mathrm{B}_1$ since $\mathcal{P}_n$ may add local ancillas. To say that $\mathcal{P}_n$ can transmit classical messages, we require that local measurements on $|\varphi_{ab}\rangle$ can generate messages b' for Alice and a' for Bob according to a distribution $\Pr(a'b'|ab)$ such that

$$\forall_{a,b} \quad \sum_{a',b'} \tfrac{1}{2} \left| \Pr(a'b'|ab) - \delta_{a,a'}\delta_{b,b'} \right| \leq \epsilon_n \tag{4}$$

where a', b' are summed over $\{0,1\}^{C_1^{(n)}}$ and $\{0,1\}^{C_2^{(n)}}$ respectively. Eq. (4) follows from applying Eq. (1) to classical communication, taking the final state to be the distribution of the output classical messages. Since any measurement can be implemented as a joint unitary on the system and an added ancilla, up to a redefinition of V_n, W_n, we can assume

$$|\varphi_{ab}\rangle := \mathcal{P}_n(|a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1}) = \sum_{a',b'} |b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}|\gamma^{a,b}_{a',b'}\rangle_{\mathrm{A}_2\,\mathrm{B}_2} \tag{5}$$

where the dimensions of A_1 and B_1 are interchanged by $\mathcal{P}_n$, and $|\gamma^{a,b}_{a',b'}\rangle$ are subnormalized states with $\Pr(a'b'|ab) := \langle\gamma^{a,b}_{a',b'}|\gamma^{a,b}_{a',b'}\rangle$ satisfying Eq. (4). Thus, for each a, b most of the

weight of $|\varphi_{ab}\rangle$ is contained in the $|\gamma_{a,b}^{a,b}\rangle$ term, corresponding to error-free transmission of the messages. See Fig. I(a).

• *The three main ideas for turning classical communication into coherent classical communication*

We first give an informal overview of the construction and the intuition behind it. For simplicity, consider the error-free term with $|\gamma_{a,b}^{a,b}\rangle$ in $\mathrm{A_2\,B_2}$. To see why classical communication via unitary means should be equivalent to coherent classical communication, consider the special case when $|\gamma_{a,b}^{a,b}\rangle_{\mathrm{A_2\,B_2}}$ is independent of a, b. In this case, copying a, b to local ancilla systems $\mathrm{A_0, B_0}$ before $\mathcal{P}_n$ and discarding $\mathrm{A_2\,B_2}$ after $\mathcal{P}_n$ leaves a state $\overset{\epsilon_n}{\approx} |b\rangle_{\mathrm{A_1}}|a\rangle_{\mathrm{A_0}}|a\rangle_{\mathrm{B_1}}|b\rangle_{\mathrm{B_0}}$—the desired coherent classical communication. See Fig. I(b). In general $|\gamma_{a,b}^{a,b}\rangle_{\mathrm{A_2\,B_2}}$ will carry information about a, b, so tracing $\mathrm{A_2\,B_2}$ will break the coherence of the classical communication. Moreover, if the Schmidt coefficients of $|\gamma_{a,b}^{a,b}\rangle_{\mathrm{A_2\,B_2}}$ depend on a, b, then knowing a, b is not sufficient to coherently eliminate $|\gamma_{a,b}^{a,b}\rangle_{\mathrm{A_2\,B_2}}$ without some additional communication. The remainder of our proof is built around the need to coherently eliminate this ancilla.

Our first strategy is to *encrypt* the classical messages a, b by a shared key, in a manner that preserves coherence (similar to that in Ref. [9]). The coherent version of a shared key is a maximally entangled state. Thus Alice and Bob (1) again copy their messages to $\mathrm{A_0, B_0}$, then (2) encrypt, (3) apply $\mathcal{P}_n$, and (4) decrypt. Encrypting the message makes it possible to (5) almost decouple the message from the combined "key-and-ancilla" system, which is approximately in a state $|\Gamma_{00}\rangle$ independent of a, b (exact definitions will follow later). (6) Tracing out $|\Gamma_{00}\rangle$ gives the desired coherent communication. Let $\mathcal{P}'_n$ denote steps (1)-(5) (see Fig. I(c)).

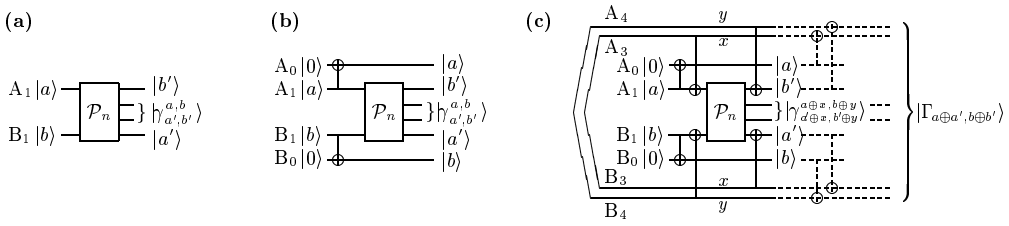

Fig. 1. Schematic diagrams for $\mathcal{P}_n$ and $\mathcal{P}'_n$. (a) A given protocol $\mathcal{P}_n$ for two-way classical communication. The output is a superposition (over all a', b') of the depicted states, with most of the weight in the $(a', b')=(a, b)$ term. The unlabeled output systems in the state $|\gamma_{a',b'}^{a,b}\rangle$ are $\mathrm{A_2, B_2}$. (b) The same protocol with the inputs copied to local ancillas $\mathrm{A_0, B_0}$ before $\mathcal{P}_n$. If $|\gamma_{a,b}^{a,b}\rangle$ is independent of a, b, two-way coherent classical communication is achieved. (c) The five steps of $\mathcal{P}'_n$. Steps (1)-(4) are shown in solid lines. Again, the inputs are copied to local ancillas, but $\mathcal{P}_n$ is used on messages encrypted by a coherent one-time-pad (the input $|a\rangle_{\mathrm{A_1}}$ is encrypted by the coherent version of the key $|x\rangle_{\mathrm{A_3}}$ and the output $|a' \oplus x\rangle_{\mathrm{B_1}}$ is decrypted by $|x\rangle_{\mathrm{B_3}}$; similarly, $|b\rangle_{\mathrm{B_1}}$ is encrypted by $|y\rangle_{\mathrm{B_4}}$ and $|b' \oplus y\rangle_{\mathrm{A_1}}$ decrypted by $|y\rangle_{\mathrm{A_4}}$. The intermediate state is shown in the diagram. Step (5), shown in dotted lines, decouples the messages in $\mathrm{A_{0,1}, B_{0,1}}$ from $\mathrm{A_{2,3,4}, B_{2,3,4}}$, which is in the joint state very close to $|\Gamma_{00}\rangle$.

If entanglement were free, then our proof of Theorem 1 would be finished. However, we have borrowed $C_1^{(n)}+C_2^{(n)}$ ebits as the encryption key and replaced it with $|\Gamma_{00}\rangle$. Though the

entropy of entanglement has not decreased (by any significant amount), $|\Gamma_{00}\rangle$ is not directly usable in subsequent runs of $\mathcal{P}'_n$. To address this problem, we use a second strategy of running k copies of $\mathcal{P}'_n$ in parallel and performing entanglement concentration of $|\Gamma_{00}\rangle^{\otimes k}$ using the techniques of [13]. For sufficiently large k, with high probability, we recover most of the starting ebits. The regenerated ebits can be used for more iterations of ${\mathcal{P}'_n}^{\otimes k}$ to offset the cost of making the initial $k\left(C_1^{(n)}+C_2^{(n)}\right)$ ebits, without the need of borrowing from anywhere.

However, a technical problem arises with simple repetition of $\mathcal{P}'_n$, which is that errors accumulate. In particular, a naïve application of the triangle inequality gives an error $k\epsilon_n$ but k, n are not independent. In fact, the entanglement concentration procedure of Ref. [13] requires $k \gg \mathrm{Sch}(|\Gamma_{00}\rangle) = \exp(O(n))$ and we cannot guarantee that $k\epsilon_n \to 0$ as $k, n \to \infty$. Our third strategy is to treat the k uses of $\mathcal{P}'_n$ as k uses of a slightly noisy channel, and encode only l messages (each having $C_1^{(n)}, C_2^{(n)}$ bits in the two directions) using classical error correcting codes. The error rate then vanishes with a negligible reduction in the communication rate and now making no assumption about how quickly ϵ_n approaches zero. We will see how related errors in decoupling and entanglement concentration are suppressed.

We now describe the construction and analyze the error in detail.

- *The definition of* $\mathcal{P}'_n$

  0. Alice and Bob begin with inputs $|a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1}$ and the entangled states $|\Phi\rangle_{\mathrm{A}_3\,\mathrm{B}_3}^{\otimes C_1^{(n)}}$ and $|\Phi\rangle_{\mathrm{A}_4\,\mathrm{B}_4}^{\otimes C_2^{(n)}}$. (Systems 3 and 4 hold the two separate keys for the two messages a and b.) The initial state can then be written as

  $$\frac{1}{\sqrt{N}}\sum_x |xx\rangle_{\mathrm{A}_3\,\mathrm{B}_3}\sum_y |yy\rangle_{\mathrm{A}_4\,\mathrm{B}_4}\ |a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1} \tag{6}$$

  where x and y are summed over $\{0,1\}^{C_1^{(n)}}$ and $\{0,1\}^{C_2^{(n)}}$, and $N = \exp\left(C_1^{(n)}+C_2^{(n)}\right)$.

  1. They coherently copy the messages to $\mathrm{A}_0, \mathrm{B}_0$.

  2. They encrypt the messages using the one-time-pad $|a\rangle_{\mathrm{A}_1}|x\rangle_{\mathrm{A}_3} \to |a\oplus x\rangle_{\mathrm{A}_1}|x\rangle_{\mathrm{A}_3}$ and $|b\rangle_{\mathrm{B}_1}|y\rangle_{\mathrm{B}_4} \to |b\oplus y\rangle_{\mathrm{B}_1}|y\rangle_{\mathrm{B}_4}$ coherently to obtain

  $$|a\rangle_{\mathrm{A}_0}|b\rangle_{\mathrm{B}_0}\ \frac{1}{\sqrt{N}}\sum_{xy}|x\rangle_{\mathrm{A}_3}|y\rangle_{\mathrm{A}_4}|x\rangle_{\mathrm{B}_3}|y\rangle_{\mathrm{B}_4}\ |a\oplus x\rangle_{\mathrm{A}_1}|b\oplus y\rangle_{\mathrm{B}_1}\,. \tag{7}$$

  3. Using U n times, they apply $\mathcal{P}_n$ to registers A_1 and B_1, obtaining an output state

  $$|a\rangle_{\mathrm{A}_0}|b\rangle_{\mathrm{B}_0}\frac{1}{\sqrt{N}}\sum_{xy}|x\rangle_{\mathrm{A}_3}|y\rangle_{\mathrm{A}_4}|x\rangle_{\mathrm{B}_3}|y\rangle_{\mathrm{B}_4}\sum_{a',b'}|b'\oplus y\rangle_{\mathrm{A}_1}|a'\oplus x\rangle_{\mathrm{B}_1}|\gamma^{a\oplus x,b\oplus y}_{a'\oplus x,b'\oplus y}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{8}$$

  4. Alice decrypts her message in A_1 using her key A_4 and Bob decrypts B_1 using B_3 coherently as $|b'\oplus y\rangle_{\mathrm{A}_1}|y\rangle_{\mathrm{A}_4} \to |b'\rangle_{\mathrm{A}_1}|y\rangle_{\mathrm{A}_4}$ and $|a'\oplus x\rangle_{\mathrm{B}_1}|x\rangle_{\mathrm{B}_3} \to |a'\rangle_{\mathrm{B}_1}|x\rangle_{\mathrm{B}_3}$ producing a state

  $$|a\rangle_{\mathrm{A}_0}|b\rangle_{\mathrm{B}_0}\frac{1}{\sqrt{N}}\sum_{xy}|x\rangle_{\mathrm{A}_3}|y\rangle_{\mathrm{A}_4}|x\rangle_{\mathrm{B}_3}|y\rangle_{\mathrm{B}_4}\sum_{a',b'}|b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}|\gamma^{a\oplus x,b\oplus y}_{a'\oplus x,b'\oplus y}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{9}$$

5. Further CNOTs $A_1 \to A_4$, $A_0 \to A_3$, $B_1 \to B_3$ and $B_0 \to B_4$ will leave $A_{2,3,4}$ and $B_{2,3,4}$ almost decoupled from the classical messages. To see this, the state has become

$$|a\rangle_{A_0}|b\rangle_{B_0}\sum_{a',b'}|b'\rangle_{A_1}|a'\rangle_{B_1}\frac{1}{\sqrt{N}}\sum_{xy}|a\oplus x\rangle_{A_3}|a'\oplus x\rangle_{B_3}|b'\oplus y\rangle_{A_4}|b\oplus y\rangle_{B_4}|\gamma^{a\oplus x,b\oplus y}_{a'\oplus x,b'\oplus y}\rangle_{A_2B_2}$$
$$= |a\rangle_{A_0}|b\rangle_{B_0}\sum_{a',b'}|b'\rangle_{A_1}|a'\rangle_{B_1}|\Gamma_{a\oplus a',b\oplus b'}\rangle_{A_{2,3,4}B_{2,3,4}}, \quad (10)$$

where

$$|\Gamma_{a\oplus a',b\oplus b'}\rangle_{A_{2,3,4}B_{2,3,4}} := \frac{1}{\sqrt{N}}\sum_{xy}|a\oplus x\rangle_{A_3}|a'\oplus x\rangle_{B_3}|b'\oplus y\rangle_{A_4}|b\oplus y\rangle_{B_4}|\gamma^{a\oplus x,b\oplus y}_{a'\oplus x,b'\oplus y}\rangle_{A_2B_2}. \quad (11)$$

The fact $|\Gamma_{a\oplus a',b\oplus b'}\rangle$ depends only on $a\oplus a'$ and $b\oplus b'$, without any other dependence on a and b, can be easily seen by replacing x, y with $a\oplus x, b\oplus y$ in $\sum_{xy}$ in the RHS of the above. Note that $\langle\Gamma_{a\oplus a',b\oplus b'}|\Gamma_{a\oplus a',b\oplus b'}\rangle = \frac{1}{N}\sum_{xy}\Pr(a'\oplus x, b'\oplus y\,|\,a\oplus x, b\oplus y)$, so in particular for the state corresponding to the error-free term, we have $\langle\Gamma_{00}|\Gamma_{00}\rangle = \frac{1}{N}\sum_{xy}\Pr(xy|xy) := 1-\bar{\epsilon}_n \geq 1-\epsilon_n$ [14].

Suppose that Alice and Bob could project onto the space where $a'=a$ and $b'=b$, and tell each other they have succeeded (by using a little extra communication); then the resulting ancilla state $\frac{1}{\sqrt{1-\bar{\epsilon}_n}}|\Gamma_{00}\rangle$ has at least $C_1^{(n)}+C_2^{(n)}+\log(1-\epsilon_n)$ ebits, since its largest Schmidt coefficient is $\leq \left[\,\exp(C_1^{(n)}+C_2^{(n)})(1-\bar{\epsilon}_n)\,\right]^{-1/2}$ and $\bar{\epsilon}_n \leq \epsilon_n$. (A similar state was studied in Ref. [6] in the proof that the entanglement capacity of a unitary gate was at least as large as its classical communication capacity.) Furthermore, $|\Gamma_{00}\rangle$ is manifestly independent of a, b. We will see how to improve the probability of successful projection onto the error free subspace by using block codes for error correction, and how correct copies of $|\Gamma_{00}\rangle$ can be identified if Alice and Bob can exchange a small amount of information.

• *Main idea on how to perform error correction*

As discussed before, $|\Gamma_{00}\rangle$ cannot be used directly as an encryption key – our use of entanglement in $\mathcal{P}'_n$ is not catalytic. Entanglement concentration of many copies of $|\Gamma_{00}\rangle$ obtained from many runs of $\mathcal{P}'_n$ will make the entanglement overhead for the one-time-pad negligible, but errors will accumulate. The idea is to suppress the errors in many uses of $\mathcal{P}'_n$ by error correction. This has to be done with care, since we need to simultaneously ensure low enough error rates in both the classical message and the state to be concentrated, as well as sufficient decoupling of the classical messages from other systems.

Our error-corrected scheme will have k parallel uses of $\mathcal{P}'_n$, but the k inputs are chosen to be a valid codeword of an error correcting code. Furthermore, for each use of $\mathcal{P}'_n$, the state in $A_{2,3,4}B_{2,3,4}$ will only be collected for entanglement concentration if the error syndrome is trivial for that use of $\mathcal{P}'_n$. We use the fact that errors occur rarely (at a rate of ϵ_n, which goes to zero as $n\to\infty$) to show that (1) most states are still used for concentration, and (2) communicating the indices of the states with non trivial error syndrome requires a negligible amount of communication.

• *Definition of $\mathcal{P}''_{nk}$: error corrected version of $(\mathcal{P}'_n)^{\otimes k}$ with entanglement concentration*

We construct two codes, one used by Alice to signal to Bob and one from Bob to Alice. We consider high distance codes. The distance of a code is the minimum Hamming distance between any two codewords, i.e. the number of positions in which they are different.

First consider the code used by Alice. Let $N_1 = 2^{C_1^{(n)}}$. Alice is coding for a channel that takes input symbols from $[N_1] := \{1, \ldots, N_1\}$ and has probability $\leq \epsilon_n$ of error on any input (the error rate depends on both a and b). We would like to encode $[N_1]^l$ in $[N_1]^k$ using a code with distance $2k\alpha_n$, where α_n is a parameter that will be chosen later. Such a code can correct up to any $\lfloor k\alpha_n - \frac{1}{2} \rfloor$ errors (without causing much problem, we just say that the code corrects $k\alpha_n$ errors). Using standard arguments [18], we can construct such a code with $l \geq k\left[\, 1 - 2\alpha_n - H_2(2\alpha_n)/C_1^{(n)} \,\right]$, where $H_2(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy. The code used by Bob is chosen similarly, with $N_2 = 2^{C_2^{(n)}}$ input symbols to each use of $\mathcal{P}'_n$. For simplicity, Alice's and Bob's codes share the same values of l, k and α_n. We choose $\alpha_n \geq \max(1/C_1^{(n)}, 1/C_2^{(n)})$ so that $l \geq k(1-3\alpha_n)$.

Furthermore, we want the probability of having $\geq k\alpha_n$ errors to be vanishingly small. This probability is $\leq \exp(-kD(\alpha_n \| \epsilon_n)) \leq \exp(k + k\alpha_n \log \epsilon_n)$ (using arguments from [19]) $\leq \exp(-k)$ if $\alpha_n \geq -2/\log \epsilon_n$.

Using these codes, Alice and Bob construct $\mathcal{P}''_{nk}$ as follows (with steps 1-3 performed coherently).

0. Let $(a_1^o, \cdots, a_l^o)$ be a vector of l messages each of $C_1^{(n)}$ bits, and $(b_1^o, \cdots, b_l^o)$ be l messages each of $C_2^{(n)}$ bits.

1. Using her error correcting code, Alice encodes $(a_1^o, \cdots, a_l^o)$ in a valid codeword $\vec{a} = (a_1, \cdots, a_k)$ which is a k-vector. Similarly, Bob generates a valid codeword $\vec{b} = (b_1, \cdots, b_k)$ using his code.

2. Let $\vec{\mathrm{A}}_1 := \mathrm{A}_1^{\otimes k}$ denote a tensor product of k input spaces each of $C_1^{(n)}$ qubits. Similarly, $\vec{\mathrm{B}}_1 := \mathrm{B}_1^{\otimes k}$. (We will also denote k copies of $\mathrm{A}_{0,2,3,4}$, and $\mathrm{B}_{0,2,3,4}$ by adding the vector symbol.) Alice and Bob apply $(\mathcal{P}'_n)^{\otimes k}$ to $|\vec{a}\rangle_{\vec{\mathrm{A}}_1}|\vec{b}\rangle_{\vec{\mathrm{B}}_1}$; that is, in parallel, they apply $\mathcal{P}'_n$ to each pair of inputs (a_j, b_j). The resulting state is a tensor product of states of the form given by Eq. (10):

$$\bigotimes_{j=1}^{k} \left[\, |a_j\rangle_{\mathrm{A}_0}|b_j\rangle_{\mathrm{B}_0} \sum_{a'_j, b'_j} |b'_j\rangle_{\mathrm{A}_1}|a'_j\rangle_{\mathrm{B}_1} \, |\Gamma_{a_j \oplus a'_j, b_j \oplus b'_j}\rangle_{\mathrm{A}_{2,3,4}\,\mathrm{B}_{2,3,4}} \,\right]. \tag{12}$$

Define $|\Gamma_{\vec{a}\oplus\vec{a}', \vec{b}\oplus\vec{b}'}\rangle_{\vec{\mathrm{A}}_{234}\vec{\mathrm{B}}_{234}} := \bigotimes_{j=1}^{k} |\Gamma_{a_j\oplus a'_j, b_j \oplus b'_j}\rangle_{\mathrm{A}_{2,3,4}\,\mathrm{B}_{2,3,4}}$. Then, Eq. (12) can be written more succinctly as

$$|\vec{a}\rangle_{\vec{\mathrm{A}}_0}|\vec{b}\rangle_{\vec{\mathrm{B}}_0} \sum_{\vec{a}', \vec{b}'} |\vec{b}'\rangle_{\vec{\mathrm{A}}_1}|\vec{a}'\rangle_{\vec{\mathrm{B}}_1}|\Gamma_{\vec{a}\oplus\vec{a}', \vec{b}\oplus\vec{b}'}\rangle_{\vec{\mathrm{A}}_{234}\vec{\mathrm{B}}_{234}}. \tag{13}$$

3. Alice performs the error correction step on $\vec{\mathrm{A}}_1$ and Bob does the same on $\vec{\mathrm{B}}_1$. According to our code constructions, this (joint) step fails with probability $p_{\text{fail}} \leq 2 \cdot 2^{-k}$. (We will see below why p_{fail} is independent of $\vec{a}$ and $\vec{b}$.)

In order to describe the residual state, we now introduce $\mathcal{G}_{\rm A} = \{\vec{x} \in [N_1]^k : |\vec{x}| \leq k\alpha_n\}$ and $\mathcal{G}_{\rm B} = \{\vec{x} \in [N_2]^k : |\vec{x}| \leq k\alpha_n\}$, where $|\vec{x}| := |\{j : x_j \neq 0\}|$ denotes the Hamming weight of $\vec{x}$. Thus $\mathcal{G}_{\rm A,B}$ are sets of correctable (good) errors, in the sense that there exist local decoding isometries $\mathcal{D}_{\rm A}, \mathcal{D}_{\rm B}$ such that for any code word $\vec{a} \in [N_1]^k$ we have $\forall \vec{a}' \in \vec{a} \oplus \mathcal{G}_{\rm A}, \mathcal{D}_{\rm A}|\vec{a}'\rangle = |\vec{a}\rangle|\vec{a} \oplus \vec{a}'\rangle$ (and similarly, if $\vec{b} \in [N_2]^k$ is a codeword, then $\forall \vec{b}' \in \vec{b} \oplus \mathcal{G}_{\rm B}, \mathcal{D}_{\rm B}|\vec{b}'\rangle = |\vec{b}\rangle|\vec{b} \oplus \vec{b}'\rangle$). For concreteness, let the decoding maps take $\vec{\rm A}_1$ to $\vec{\rm A}_1\vec{\rm A}_5$ and $\vec{\rm B}_1$ to $\vec{\rm B}_1\vec{\rm B}_5$.

Conditioned on success, Alice and Bob are left with

$$\frac{1}{\sqrt{1-p_{\rm fail}}} |\vec{a},\vec{b}\rangle_{\vec{\rm A}_{0,1}} |\vec{a},\vec{b}\rangle_{\vec{\rm B}_{0,1}} \sum_{\vec{a}' \in \vec{a} \oplus \mathcal{G}_{\rm A}} \sum_{\vec{b}' \in \vec{b} \oplus \mathcal{G}_{\rm B}} |\vec{b} \oplus \vec{b}'\rangle_{\vec{\rm A}_5} |\vec{a} \oplus \vec{a}'\rangle_{\vec{\rm B}_5} |\Gamma_{\vec{a}\oplus\vec{a}', \vec{b}\oplus\vec{b}'}\rangle_{\vec{\rm A}_{234}\vec{\rm B}_{234}} \quad (14)$$

$$:= \frac{1}{\sqrt{1-p_{\rm fail}}} |\vec{a},\vec{b}\rangle_{\vec{\rm A}_{0,1}} |\vec{a},\vec{b}\rangle_{\vec{\rm B}_{0,1}} \sum_{\vec{a}'' \in \mathcal{G}_{\rm A}} \sum_{\vec{b}'' \in \mathcal{G}_{\rm B}} |\vec{b}''\rangle_{\vec{\rm A}_5} |\vec{a}''\rangle_{\vec{\rm B}_5} |\Gamma_{\vec{a}'', \vec{b}''}\rangle_{\vec{\rm A}_{234}\vec{\rm B}_{234}}, \quad (15)$$

where we have defined $\vec{a}'' := \vec{a} \oplus \vec{a}'$ and $\vec{b}'' := \vec{b} \oplus \vec{b}'$. Note that $2^{-k+1} \geq p_{\rm fail} = \sum_{(\vec{a}'', \vec{b}'') \notin \mathcal{G}_{\rm A} \times \mathcal{G}_{\rm B}} \langle \Gamma_{\vec{a}'', \vec{b}''} | \Gamma_{\vec{a}'', \vec{b}''} \rangle$, which is manifestly independent of $\vec{a}, \vec{b}$. The ancilla is now *completely* decoupled from the message, resulting in coherent classical communication. The only remaining issue is recovering entanglement from the ancilla, so for the remainder of the protocol we ignore the now decoupled states $|\vec{a},\vec{b}\rangle_{\vec{\rm A}_{0,1}} |\vec{a},\vec{b}\rangle_{\vec{\rm B}_{0,1}}$.

4. For any $\vec{x}$, define $S(\vec{x}) := \{j : x_j \neq 0\}$ to be set of positions where $\vec{x}$ is nonzero. If $\vec{x} \in \mathcal{G}_{\rm A}$ (or $\mathcal{G}_{\rm B}$), then $|S(\vec{x})| \leq k\alpha_n$. Thus, $S(\vec{x})$ can be written using $\leq \log \sum_{j \leq k\alpha_n} \binom{k}{j} \leq \log \binom{k}{k\alpha_n} + \log(k\alpha_n) \leq kH_2(\alpha_n) + \log(k\alpha_n)$ bits.

   The next step is for Alice to compute $|S(\vec{b}'')\rangle$ from $|\vec{b}''\rangle$ and communicate it to Bob using $\big(kH_2(\alpha_n) + \log(k\alpha_n)\big)$ cbits$(\rightarrow)$. Similarly, Bob sends $|S(\vec{a}'')\rangle$ to Alice using $\big(kH_2(\alpha_n) + \log(k\alpha_n)\big)$ cbits$(\leftarrow)$. Here we need to assume that some (possibly inefficient) protocol to send $O(k)$ bits in either direction with error $\exp(-k-1)$ (chosen for convenience) and with Rk uses of U for some constant R. Such a protocol was shown in Ref. [6] and the bound on the error can be obtained from the HSW theorem [16].

   Alice and Bob now have the state

$$\frac{1}{\sqrt{1-p_{\rm fail}}} \sum_{\vec{a}'' \in \mathcal{G}_{\rm A}} \sum_{\vec{b}'' \in \mathcal{G}_{\rm B}} |S(\vec{a}'')S(\vec{b}'')\rangle_{\vec{\rm A}_6} |\vec{b}''\rangle_{\vec{\rm A}_5} |S(\vec{a}'')S(\vec{b}'')\rangle_{\vec{\rm B}_6} |\vec{a}''\rangle_{\vec{\rm B}_5} |\Gamma_{\vec{a}'', \vec{b}''}\rangle_{\vec{\rm A}_{234}\vec{\rm B}_{234}}. \quad (16)$$

   Conditioning on their knowledge of $S(\vec{a}''), S(\vec{b}'')$, Alice and Bob can now identify $k' \geq k(1 - 2\alpha_n)$ positions where $a''_j = b''_j = 0$, and extract k' copies of $\frac{1}{\sqrt{1-p_{\rm fail}}}|\Gamma_{00}\rangle$. Note that leaking $S(\vec{a}''), S(\vec{b}'')$ to the environment will not affect the extraction procedure, therefore, coherent computation and communication of $S(\vec{a}''), S(\vec{b}'')$ is unnecessary. (We have not explicitly included the environment's copy of $|S(\vec{a}'')S(\vec{b}'')\rangle$ in the equations to minimize clutter.) After extracting k' copies of $\frac{1}{\sqrt{1-p_{\rm fail}}}|\Gamma_{00}\rangle$, we can safely discard the remainder of the state, which is now completely decoupled from both $\big[\frac{1}{\sqrt{1-p_{\rm fail}}}|\Gamma_{00}\rangle\big]^{\otimes k'}$ and the message $|\vec{a}\rangle_{\rm A_0}|\vec{b}\rangle_{\rm A_1}|\vec{b}\rangle_{\rm B_0}|\vec{a}\rangle_{\rm B_1}$.

5. Alice and Bob perform entanglement concentration $\mathcal{E}_{\rm conc}$ (using the techniques of [13]) on $\left[\frac{1}{\sqrt{1-p_{\rm fail}}}|\Gamma_{00}\rangle\right]^{\otimes k'}$. Note that since $\frac{1}{\sqrt{1-p_{\rm fail}}}|\Gamma_{00}\rangle$ can be created using U n times and then using classical communication and postselection, it must have Schmidt rank $\leq \mathrm{Sch}(U)^n$, where $\mathrm{Sch}(U)$ is the Schmidt number of the gate U [20]. Also recall that $E\left[\frac{1}{\sqrt{1-p_{\rm fail}}}|\Gamma_{00}\rangle\right] \geq C_1^{(n)} + C_2^{(n)} + \log(1-\epsilon_n)$. According to Ref. [13], $\mathcal{E}_{\rm conc}$ requires no communication and with probability $\geq 1-\exp\left[-\mathrm{Sch}(U)^n\left(\sqrt{k'}-\log(k'+1)\right)\right]$ produces at least $k'\left[C_1^{(n)}+C_2^{(n)}+\log(1-\epsilon_n)\right] - \mathrm{Sch}(U)^n\left[\sqrt{k'}-\log(k'+1)\right]$ ebits.

• *Error and resource accounting*

$\mathcal{P}''_{nk}$ consumes a total of

(0) nk uses of U (in the k executions of $\mathcal{P}'_n$)
(1) Rk uses of U (for communicating $S(\vec{a}''), S(\vec{b}'')$)
(2) $k\left[C_1^{(n)}+C_2^{(n)}\right]$ ebits (for the encryption of classical messages).

$\mathcal{P}''_{nk}$ produces, with probability and fidelity $\geq 1-2\cdot 2^{-(k-1)} - \exp\left[-\mathrm{Sch}(U)^n\left(\sqrt{k'}-\log(k'+1)\right)\right]$, at least

(1) $l\,C_1^{(n)}\mathrm{cobits}(\rightarrow) + l\,C_2^{(n)}\mathrm{cobits}(\leftarrow)$
(2) $k'\left(C_1^{(n)}+C_2^{(n)}+\log(1-\epsilon_n)\right) - \mathrm{Sch}(U)^n\left(\sqrt{k'}-\log(k'+1)\right)$ ebits.

We restate the constraints on the above parameters: $\epsilon_n, \delta_n \to 0$ as $n\to\infty$; $C_1^{(n)} \geq n(C_1-\delta_n)$, $C_2^{(n)} \geq n(C_2-\delta_n)$; $\alpha_n \geq \max(1/C_1^{(n)}, 1/C_2^{(n)}, -2/\log\epsilon_n)$; $k' \geq k(1-2\alpha_n)$; $l \geq k(1-3\alpha_n)$.

We define "error" to include both infidelity and the probability of failure. To leading orders of k, n, this is equal to $2^{-(k-2)} + \exp\left[-\sqrt{k}\ \mathrm{Sch}(U)^n\right]$. We define "inefficiency" to include extra uses of U, net consumption of entanglement, and the amount by which the coherent classical communication rates fall short of the classical capacities. To leading order of k, n, these are respectively Rk, $2\alpha_n k(C_1^{(n)}+C_2^{(n)}) + \sqrt{k}\,\mathrm{Sch}(U)^n \approx 2\alpha_n kn(C_1+C_2) + \sqrt{k}\,\mathrm{Sch}(U)^n$, and $nk(C_1+C_2) - l(C_1^{(n)}+C_2^{(n)}) \leq nk(3\alpha_n(C_1+C_2)+2\delta_n)$. We would like the error to vanish, as well as the fractional inefficiency, defined as the inefficiency divided by kn, the number of uses of U. Equivalently, we can define $f(k,n)$ to be the *sum* of the error and the fractional inefficiency, and require that $f(k,n)\to 0$ as $nk\to\infty$. By the above arguments,

$$f(k,n) \leq 2^{-(k-2)} + \exp(-\sqrt{k}\ \mathrm{Sch}(U)^n) + 2\alpha_n(C_1+C_2) + \tfrac{1}{n\sqrt{k}}\ \mathrm{Sch}(U)^n + \frac{R}{n} + 3\alpha_n(C_1+C_2) + 2\delta_n\,. \qquad (17)$$

Note that for any fixed value of n, $\lim_{k\to\infty} f(k,n) = 5\alpha_n(C_1+C_2)+2\delta_n+R/n$. (This requires k to be sufficiently large and also $k \gg \mathrm{Sch}(U)^{2n}$.) Now, allowing n to grow, we have

$$\lim_{n\to\infty}\lim_{k\to\infty} f(k,n) = 0. \qquad (18)$$

The order of limits in this equation is crucial due to the dependence of k on n.

The only remaining problem is our catalytic use of $O(nk)$ ebits. In order to construct a protocol that uses only U, we need to first use U $O(nk)$ times to generate the starting entanglement. Then we repeat $\mathcal{P}''_n$ m times, reusing the same entanglement. The catalyst results in an additional fractional inefficiency of c/m (for some constant c depending only

on U) and the errors and inefficiencies of $\mathcal{P}_n''$ add up to no more than $mf(k,n)$. Choosing $m = \lfloor 1/\sqrt{f(k,n)} \rfloor$ will cause all of these errors and inefficiencies to simultaneously vanish. More generally,

$$\lim_{m\to\infty} \lim_{n\to\infty} \lim_{k\to\infty} \; mf(k,n) + \frac{c}{m} \; = \; 0\,. \tag{19}$$

This proves the resource inequality

$$U \geq C_1 \text{ cobits}(\rightarrow) + C_2 \text{ cobits}(\leftarrow). \tag{20}$$

• *The $E < 0$ and $E > 0$ cases*

If $E < 0$ then entanglement is consumed in $\mathcal{P}_n$, so there exists a sequence of integers $E^{(n)} \leq n(E + \delta_n)$ such that

$$\mathcal{P}_n\Big(|a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1}|\Phi\rangle^{E^{(n)}}_{\mathrm{A}_5\,\mathrm{B}_5}\Big) = \sum_{a',b'} |b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}|\gamma^{a,b}_{a',b'}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{21}$$

In this case, the analysis for $E^{(n)} = 0$ goes through, only with additional entanglement consumed. Almost all equations are the same, except now the Schmidt rank for $|\Gamma_{00}\rangle$ is upper-bounded by $\big[\mathrm{Sch}(U)2^{E+\delta_n}\big]^n$ instead of $\mathrm{Sch}(U)^n$. In particular, previous arguments still give Eq. (18) from the modified Eq. (17).

If instead $E > 0$, entanglement is created, so for some $E^{(n)} \geq n(E - \delta_n)$ we have

$$\mathcal{P}_n\big(|a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1}\big) = \sum_{a',b'} |b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}|\gamma^{a,b}_{a',b'}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{22}$$

for $E(|\gamma^{a,b}_{a,b}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}) \geq E^{(n)}$. Again, the previous construction and analysis go through, with an extra $E^{(n)}$ ebits of entanglement of entropy in $|\Gamma_{00}\rangle$, and thus an extra fractional efficiency of $\leq 2\alpha_n E$ in Eq. (17). The Schmidt rank of $|\Gamma_{00}\rangle$ is still upper bounded by $\mathrm{Sch}(U)^n$ in this case. □

So far, we have focused on the $C_1, C_2 \geq 0$ quadrant. The following theorem will relate the achievable regions for coherent and incoherent classical communication when $C_1 \leq 0$ or $C_2 \leq 0$.

Theorem 2 *For any bipartite unitary or isometry U and $C_1, C_2 \geq 0$,*

$$C_2 \text{ cbits}(\leftarrow) + U \;\geqslant\; C_1 \text{ cbits}(\rightarrow) + E \text{ ebits} \quad \text{iff} \tag{23}$$

$$U \;\geqslant\; C_1 \text{ cbits}(\rightarrow) + E \text{ ebits} \quad \text{iff} \tag{24}$$

$$U \;\geqslant\; C_1 \text{ cobits}(\rightarrow) + E \text{ ebits} \quad \text{iff} \tag{25}$$

$$C_2 \text{ cobits}(\leftarrow) + U \;\geqslant\; C_1 \text{ cobits}(\rightarrow) + (E{+}C_2) \text{ ebits} \tag{26}$$

and

$$C_1 \text{ cbits}(\rightarrow) + C_2 \text{ cbits}(\leftarrow) + U \;\geqslant\; E \text{ ebits} \quad \text{iff} \tag{27}$$

$$U \;\geqslant\; E \text{ ebits} \quad \text{iff} \tag{28}$$

$$C_1 \text{ cobits}(\rightarrow) + C_2 \text{ cobits}(\leftarrow) + U \;\geqslant\; (E{+}C_1{+}C_2) \text{ ebits} \tag{29}$$

In essence, the rates of unidirectional classical communication with arbitrary amount of entanglement assistance (or generation) are not increased by (in)coherent classical communication in the opposite direction, except for a trivial gain of entanglement when the assisting classical communication is coherent.

Proof: Using superdense coding to send 2 cobits and supplying the required 1 qubit of quantum communication by teleportation (using 2 cbits +1 ebit), we have

$$1\,\text{cbit} + 1\,\text{ebit} \geqslant 1\,\text{cobit}\ . \tag{30}$$

The above resource transformation is exact and does not require large blocks. Thus, composing it with other protocols poses no extra problem.

For the first part of the theorem, Eq. (23) $\Rightarrow$ Eq. (24) follows from how Ref. [8] characterizes the set of (C_1, E) that satisfies Eq. (23). Although the proof in Ref. [8] did not mention back communication, it can be easily modified to show that free classical communication from Bob to Alice does not change the capacity. In essence, the optimal tradeoff curve between C_1 and E has an upper bound that remains valid in the presence of back classical communication, and the same bound is achieved by a protocol that uses no back classical communication. A complete proof of this fact will also appear in Ref. [22].

Ref. [8] also proved that Eq. (24) $\Leftrightarrow$ Eq. (25), and it is trivial that Eq. (25) $\Rightarrow$ Eq. (26). Finally, Eq. (26) $\Rightarrow$ Eq. (23) because of Eq. (30).

For the second part of the theorem, Ref. [6] proved that Eq. (27) $\Rightarrow$ Eq. (28). It is trivial that Eq. (28) $\Rightarrow$ Eq. (29). Using Eq. (30), Eq. (29) $\Rightarrow$ Eq. (27). □

4. Achievable regions for bidirectional communication

Bipartite unitary gates can be used for several inequivalent purposes simultaneously, including some (possibly different) forms of forward and backward communications and entanglement generation. It is thus natural to define their capacities in terms of achievable rate regions (in 3-dimensional space) and trade-off surfaces.

For example, let CCE be the achievable rate region $\{(C_1, C_2, E)\ :\ U \geqslant C_1\,\text{cbits}(\rightarrow) + C_2\,\text{cbits}(\leftarrow) + E\,\text{ebits}\}$, and $\text{C}_\text{o}\text{C}_\text{o}\text{E}$ be the achievable rate region $\{(C_1, C_2, E) : U \geqslant C_1\,\text{cobits}(\rightarrow) + C_2\,\text{cobits}(\leftarrow) + E\,\text{ebits}\}$. Theorems 1 and 2 provide a mapping between CCC and $\text{C}_\text{o}\text{C}_\text{o}\text{E}$:

$$(C_1, C_2, E) \in \text{CCE} \iff (C_1, C_2, E - \min(C_1, 0) - \min(C_2, 0)) \in \text{C}_\text{o}\text{C}_\text{o}\text{E}\,. \tag{31}$$

Finding relations between different capacity regions will simplify our study of capacities of bipartite unitary gates and elicit their nonlocal properties.

As a second example of relation of achievable regions, consider remote state preparation, which is the ability to prepare a quantum state $|\psi\rangle$ in the laboratory of the receiver, assuming that the sender has a classical description of $|\psi\rangle$ (assuming pure states for simplicity). We claim that the achievable region RRE for two-way (but independent forward and backward) remote state preparation is the same as CCE. To prove this, first note that ∞ cbits $\geq$ n remote qubits $\geq n$ cbits, where n remote qubits denotes the ability to remotely prepare an n-qubit state. Combining this with the fact that even unlimited back-communication does not

improve classical capacity implies that RRE $\subset$ CCE. On the other hand, Ref. [8] showed that n coherent bits $\geq n$ remote qubits. Thus the first quadrants ($C_1, C_2 \geq 0$) of RRE and C$_o$C$_o$E (and thus CCE) are the same, and the other quadrants of RRE are related to C$_o$C$_o$E the same way that CCE is: backwards cobits can be used to generate entanglement, but free backwards remote qubits do not improve the forward capacity. This means that RRE = CCE.

Similarly, define QQE to be the region $\{(Q_1, Q_2, E) : U \geqslant Q_1 \text{ qubits}(\rightarrow) + Q_2 \text{ qubits}(\leftarrow) + E \text{ ebits}\}$, corresponding to two-way quantum communication. We can also consider coherent classical communication in one direction and quantum communication in the other; let QC$_o$E be the region $\{(Q_1, C_2, E) : U \geqslant Q_1 \text{ qubits}(\rightarrow) + C_2 \text{ cobits}(\leftarrow) + E \text{ ebits}\}$ and define C$_o$QE similarly.

Ref. [8] related the one-way tradeoff curves C$_o$E and QE, defined as C$_o$E $= \{(C, E) : (C, 0, E) \in$ C$_o$C$_o$E$\}$ and QE $= \{(Q, E) : (Q, 0, E) \in$ QQE$\}$. There it was claimed that

$$(Q, E) \in \text{QE} \Leftrightarrow (2Q, E - Q) \in \text{C}_o\text{E}\,. \tag{32}$$

We now rephrase the proof of Eq. (32) in a form that readily extends to a relation between entire achievable rate regions (for different types of bidirectional communication). Eq. (32) is due to the equivalence 2 cobits = 1 qubit +1 ebit. Note that this equivalence involves resource transformations that are exact and do not require large blocks. Thus, composing these transformations with other protocols poses no extra problem, and the equivalence can be used "freely." To prove Eq. (32), choose any $(Q, E) \in$ QE. Then $U \geq Q$ qubits $+ E$ ebits $= 2Q$ cobits $+ (E - Q)$ ebits, so $(2Q, E - Q) \in$ C$_o$E. Conversely, if $(2Q, E - Q) \in$ C$_o$E, then $U \geq 2Q$ cobits $+ (E - Q)$ ebits $= Q$ qubits $+ E$ ebits, so $(Q, E) \in$ QE.

Note that the above argument still works if we replace U with a different resource, such as $U - Q_2$ qubits($\leftarrow$). Therefore, the same argument that proved Eq. (32) also establishes the following equivalences for bidirectional rate regions:

$$\begin{array}{ccc} (Q_1, Q_2, E) \in \text{QQE} & \Longleftrightarrow & (2Q_1, Q_2, E - Q_1) \in \text{C}_o\text{QE} \\ \Updownarrow & & \Updownarrow \\ (Q_1, 2Q_2, E - Q_2) \in \text{QC}_o\text{E} & \Longleftrightarrow & (2Q_1, 2Q_2, E - Q_1 - Q_2) \in \text{C}_o\text{C}_o\text{E} \end{array}\,. \tag{33}$$

Finally, Eq. (31) further relates QQE, QCE, CQE, CCE, where QCE and CQE are defined similarly to QC$_o$E and C$_o$QE but with incoherent classical communication instead.

Thus once one of the capacity region (say C$_o$C$_o$E) is determined, all other capacity regions discussed above are determined.

Acknowledgements

We are grateful to the Perimeter Institute for their hospitality while we did this work. Feedback from the anonymous referees was much appreciated. Thanks to Igor Devetak, Andreas Winter, and Jon Yard for useful discussions, especially on the relation between the worst-case and the average-case errors and on the significance of [23]. AWH acknowledges partial support from the NSA and ARDA under ARO contract DAAD19-01-1-06. DWL acknowledges sup-

port from the Tolman Endowment Fund, the Croucher Foundation, and the US NSF under grant no. EIA-0086038.

References

1. M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum information*, Cambridge University Press (Cambridge, U.K., 2000).
2. C.H. Bennett, S. Braunstein, I.L. Chuang, D.P. DiVincenzo, D. Gottesman, J.A. Smolin, B.M. Terhal, W. K. Wootters, unpublished discussion during the sixth quantum computation workshop, ISI, Torino, Italy (1998). C.H. Bennett and D. Gottesman, email communication (1998).
3. J. Eisert, K. Jacobs, P. Papadopoulos, and M.B. Plenio, "Optimal local implementation of non-local quantum gates," Phys. Rev. A **62** (2000) 052317, quant-ph/0005101v1.
4. D. Collins, N. Linden, and S. Popescu, "The non-local content of quantum operations," quant-ph/0005102v1
5. W. Dür, G. Vidal, J.I. Cirac, N. Linden, S. Popescu, "Entanglement capabilities of non-local Hamiltonians," quant-ph/0006034; Phys. Rev. Lett. **87**, 137901 (2001).
6. C.H. Bennett, A.W. Harrow, D.W. Leung and J.A. Smolin, "On the capacities of bipartite Hamiltonians and unitary gates," quant-ph/0205057; IEEE Trans. Inf. Theory **49**, 1895 (2003).
7. M. Leifer, L. Henderson, and N. Linden, "Optimal entanglement generation from quantum operations," quant-ph/0205055; Phys. Rev. A **67**, 012306 (2003).
8. A.W. Harrow, "Coherent communication of classical messages," quant-ph/0307091; Phys. Rev. Lett. **92**, 097902 (2004).
9. D.W. Leung, "Quantum Vernam Cipher," quant-ph/0012077; Quant. Inf. Comp. **2**, no. 1, 14-34 (2002).
10. I. Devetak, A.W. Harrow, A. Winter, "Quantum Shannon theory, resource inequalities, and optimal tradeoffs for a family of quantum protocols," in preparation.
11. C.E. Shannon. "A mathematical theory of communication." Bell Sys. Tech. Journal, **27** 379-423, 623-656 (1948).
12. C.H. Bennett, P.W. Shor, and J.A. Smolin, and A.V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," quant-ph/0106052; IEEE Trans. Inf. Th. **48** 2637 (2002).
13. C.H. Bennett, H.J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating Partial Entanglement by Local Operations," quant-ph/9511030; Phys. Rev. A, **53** 2046 (1996).
14. Thus it turns out that Eq. (4) was more than we needed; the *average* error (over all a, b) would have been sufficient. In general, this argument shows that using shared entanglement (or randomness in the case of classical communication) can convert an average error condition into a maximum error condition, and will be further developed in [15].
15. I. Devetak and A. Winter, "Maximal and average error capacity regions coincide—under randomised encodings," in preparation (2005).
16. A.S. Holevo, IEEE Trans. Inform. Theory **44**, 269 (1998); B. Schumacher and M.D. Westmoreland, Phys. Rev. A **56**, 131 (1997).
17. D.W. Berry and B.C. Sanders, "Relation between classical communication capacity and entanglement capability for two-qubit unitary operations," quant-ph/0207065; Phys. Rev. A. 68, 032312 (2003).
18. We show the existence of a maximal code by repeatedly adding new codewords that have distance $\geq 2k\alpha_n$ from all other chosen codewords. This gives at least $N^k / \operatorname{Vol}(N, 2k\alpha_n, k)$ codewords, where $\operatorname{Vol}(N, k\delta, k)$ is the number of words in $[N]^k$ within a distance $k\delta$ of a fixed codeword. But $\operatorname{Vol}(N, k\delta, k) \leq \binom{k}{k\delta} N^{k\delta} \leq 2^{kH_2(\delta)} N^{k\delta}$. (See [19] for a derivation of $\binom{k}{k\delta} \leq 2^{kH_2(\delta)}$, or simply consider k i.i.d. tosses of a coin each with probability δ of coming up heads. Prob($k\delta$ heads) $= \binom{k}{k\delta} \delta^{k\delta} (1-\delta)^{k(1-\delta)} = \binom{k}{k\delta} 2^{-kH_2(\delta)} \leq 1$.) Altogether, the number of codewords

$:= N^l \geq N^k/(2^{kH_2(2\alpha_n)}N^{2k\alpha_n})$, thus $l \geq k\left[\, 1 - 2\alpha_n - \frac{H_2(2\alpha_n)}{\log N}\,\right]$.

19. T. Cover and J. Thomas, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).
20. M. A. Nielsen, "Quantum information theory," PhD thesis, University of New Mexico, Albuquerque, NM, 1998.
21. A.W. Harrow, P. Hayden and D.W. Leung, "Super-dense coding of quantum states," quant-ph/0307221; Phys. Rev. Lett. **92**, 187901 (2003).
22. A.W. Harrow, "Applications of coherent classical communication and the Schur transform to quantum information theory," PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2005.
23. G. Dueck, "Maximal error capacity regions are smaller than average error capacity regions for multi-user channels." (English. Russian summary) Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform. 7 (1), 11 (1978).

Appendix A Why we cannot use the techniques in Ref. [8]

In this appendix, we review the proof of Prop. 1 in Ref. [8] (the unidirectional communication analogue of Theorem 1) and show how it breaks down when applied to two-way communication.

We first review HSW coding [16], since the proof of Prop. 1 in [8] is based on it. Given a channel which maps a classical input i to a quantum state ρ_i, the HSW theorem states that its classical capacity is $C := \max_p S(\sum_i p_i\rho_i) - \sum_i p_i S(\rho_i)$, where the maximization is over probability distributions p and $S(\rho) := -\operatorname{Tr}\rho\log\rho$ is the von Neumann entropy. The HSW theorem can be proved by random coding followed by expurgation. That is, we choose $2^{n(C-\delta_n)}$ length n codewords according to the product distribution $p^n(i_1,\ldots,i_n) = p(i_1)\cdots p(i_n)$ (with $\delta_n \to 0$ as $n \to \infty$). Then with high probability the codewords will on average be almost perfectly distinguishable from one another. We then discard (or "expurgate") the worst half of the codewords in order to signal with asymptotically vanishing maximum error at a rate approaching C.

Instead of choosing codewords according to p^n, we could instead randomly choose typical sequences (meaning that the frequency of a letter i is $np_i \pm O(\sqrt{n})$). In fact, since there are only $\operatorname{poly}(n)$ different type classes, we can choose all our codewords to be the same type and still achieve capacity C asymptotically. (The "type" of a string denotes the number of times each letter appears in the string.)

Now we review the application of the HSW theorem to coherent communication in Prop. 1 of [8]. Given a gate U such that $U \geq C$ cbits($\rightarrow$), we know (similar to Eq. (5)) that there exists a sequence of unitary protocols $\mathcal{P}_n$, each can communicate a bit string of length $\approx n(C-\delta_n)$ bits up to an error of ϵ_n for $\delta_n \to 0$, $\epsilon_n \to 0$. $\mathcal{P}_n$ can be viewed as a channel with HSW capacity $\approx nC$, i.e., by HSW coding, $\mathcal{P}_n$ can be used k times, sending $\approx nkC$ bits with *overall* error rate vanishing as $k \to \infty$. (This idea was used in [17] to bound the size of the ancilla systems used in unitary gate communication.)

Let p be the distribution that almost achieve the HSW capacity. Let $\vec{a} = (a_1, \cdots, a_k)$ be any HSW codeword. Running $\mathcal{P}_n$ k times produces the state $|\varphi\rangle = \bigotimes_{i=1}^k \mathcal{P}_n|a_i\rangle_{\mathrm{A}_1}$. Alice could have copied the input before the protocol, and by the construction of the HSW code, Bob can

extract $\vec{a}$ with negligible error and disturbance to $|\varphi\rangle$, and Alice and Bob will have possession of a state which is $k\epsilon_n$ close to $|\vec{a}\rangle_{A_0}|\vec{a}\rangle_{B_1} \bigotimes_{i=1}^{k}(\mathcal{P}_n|a_i\rangle)_{A_2 B_2}$. The state $|\vec{a}\rangle$ in A_0 and B_1 will allow Alice and Bob to coherently reorder the k copies of $\mathcal{P}_n|a_i\rangle$ (with preagreed total order of the set of all nC-bit words). The reordered state has no information on $\vec{a}$ except for the letter frequency. Thus, when all $\vec{a} = (a_1, \cdots, a_k)$ are of the same type, the reordered state becomes independent of $\vec{a}$ and can be discarded without breaking coherence of the communication of $|\vec{a}\rangle$. Or when all $\vec{a}$ are typical sequences, the small information on $\vec{a}$ can be removed with $O(\sqrt{k})$ qubits of communication. Here, k and n are independent, so that indeed $k\epsilon_n \to 0$.

(The original form of the HSW theorem in which we simply choose random codewords according to p^n and expurgate causes a problem in this application. With high probability, the codewords are typical, but some codewords can be highly nontypical, with corresponding ancilla that cannot be made identical to a "typical ancilla" using negligible resources.)

The same-type HSW coding technique cannot be easily applied in the two-way case. Even if Alice only uses HSW codewords $|\vec{a}\rangle$ of the same type and similarly for codewords $|\vec{b}\rangle$ of Bob, the joint string $(\vec{a}, \vec{b}) := ((a_1, b_1), \ldots (a_k, b_k))$ need not have the same type. With high probability $(\vec{a}, \vec{b})$ will be typical, but some are far from typical. Worst still, these are composite codewords that depend *jointly* on $\vec{a}$ and $\vec{b}$ and cannot be expurgated by independent expurgation of individual codewords used by Alice and Bob.

Thus we obtain the strange situation where the average error is small, but we cannot make the maximum error small because expurgation requires a linear amount of communication. A similar problem was found in bidirectional classical channels, where the achievable capacity regions are different depending on whether average or maximum error is considered [23]. Classically, this separation between achievable average and maximum error occurs only when we restrict to deterministic encodings; Ref. [15] points out that the capacity regions for maximum and average error are the same when we let randomness be introduced into the encodings. The main result of our paper can thus be thought of as a coherent version of Ref. [15].

Appendix B Implications on the definition of coherent classical communication

There are two ways to define a cbit. One is in terms of an abstract operation $|x\rangle_A \to |x\rangle_B|x\rangle_E$ for $x \in \{0, 1\}$. Another is more operational, that some sequence of operations $\mathcal{P}_n$ can send n cbits with error $\epsilon_n \to 0$ if $\mathcal{P}_n(|x\rangle_A) \overset{\epsilon_n}{\approx} |x\rangle_B$, for x an n-bit string. The fact that the operational and abstract definitions are equivalent allows us to think about classical communication in both ways interchangeably.

Similarly we can define a cobit either as an abstract operation $|x\rangle_A \to |x\rangle_A|x\rangle_B$ for $x \in \{0, 1\}$, or by saying that $\mathcal{P}_n$ can send n cobits with error $\epsilon_n \to 0$ if $\mathcal{P}_n$ can send n cbits with error ϵ_n *and* $\mathcal{P}_n$ is an isometry. By Prop 1 of [8], these definitions are equivalent for one-way communication. Thm 1 of this paper shows that these definitions are now equivalent for two-way communication. This justifies the name "coherent classical communication"; a cobit really is no more and no less than a cbit sent through coherent means (i.e. a unitary gate or isometry).