

# Capacity Theorems for Quantum Multiple-Access Channels: Classical-Quantum and Quantum-Quantum Capacity Regions

Jon Yard, *Member, IEEE*, Patrick Hayden, *Member, IEEE*, and Igor Devetak

**Abstract**—In this paper, we consider quantum channels with two senders and one receiver. For an arbitrary such channel, we give multiletter characterizations of two different two-dimensional capacity regions. The first region comprises the rates at which it is possible for one sender to send classical information, while the other sends quantum information. The second region consists of the rates at which each sender can send quantum information. For each region, we give an example of a channel for which the corresponding region has a single-letter description. One of our examples relies on a new result proved here, perhaps of independent interest, stating that the coherent information over any degradable channel is concave in the input density operator. We conclude with connections to other work and a discussion on generalizations where each user simultaneously sends classical and quantum information.

**Index Terms**—Multiple-access channels, quantum capacity, quantum information.

## I. INTRODUCTION

A CLASSICAL multiple-access channel with two senders and one receiver is described by a probability transition matrix  $p(z|x, y)$ . For the situation in which each sender wishes to send independent information, Ahlswede [1] and Liao [2]

Manuscript received August 4, 2005; revised October 15, 2007. The work of J. Yard was supported by the Army Research Office MURI under Contract DAAD-19-99-1-0215, by the National Science Foundation under Grant CCR-0311633, by the Stanford Networking Research Center under Grant 1059371-6-WAYTE, by CIFAR through McGill University, by the National Science Foundation (NSF) under Grant PHY-0456720 through Caltech, and by the Center for Nonlinear Studies (CNLS), the Quantum Institute, and the LDRD program of the U.S. Department of Energy through LANL. The work of P. Hayden was supported by CIFAR, by the Sherman Fairchild Foundation, and by the U.S. National Science Foundation under Grant EIA-0086038. Most of this research was done when J. Yard was with the Information Systems Laboratory, Stanford University, Stanford, CA.

J. Yard was with the School of Computer Science, McGill University, Montreal, QC H3A 2A7, Canada and Institute for Quantum Information, California Institute of Technology (Caltech), Pasadena, CA 91125 USA. He is now with the Los Alamos National Laboratory, Los Alamos, NM 87545 USA (e-mail: jtyard@gmail.com).

P. Hayden is with the School of Computer Science, McGill University, Montreal, QC H3A 2A7, Canada (e-mail: patrick@cs.mcgill.ca).

I. Devetak is with the Electrical Engineering Department, University of Southern California, Los Angeles, CA 90089-2565 USA (e-mail: devetak@usc.edu).

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

Color version of Figure 1 in this paper is available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2008.924665

showed that the capacity region  $\mathcal{C}$  admits a single-letter characterization, given by the convex hull of the closure of the set of rate pairs  $(r, s)$  satisfying

$$\begin{aligned} r &\leq I(X; Z|Y) \\ s &\leq I(Y; Z|X) \\ r + s &\leq I(XY; Z) \end{aligned}$$

for some  $p(x)p(y)$ . Further analysis by Cover, El Gamal, and Salehi [3] gives single-letter characterizations of a set of correlated sources that can be reliably transmitted over a multiple-access channel, generalizing the above, as well as Slepian–Wolf source coding and cooperative multiple-access channel capacity. They also give a multiletter expression for the capacity region, showing that an independent identically distributed (i.i.d.) source  $(U, V)$  can be reliably transmitted if and only if

$$\begin{aligned} H(U|V) &\leq \frac{1}{n} I(X^n; Z^n | U^n Y^n) \\ H(V|U) &\leq \frac{1}{n} I(Y^n; Z^n | V^n X^n) \\ H(U, V) &\leq \frac{1}{n} I(X^n Y^n; Z^n) \end{aligned}$$

for some  $n$  and  $p(x^n|u^n), p(y^n|v^n)$ , where  $x^n$  refers to the sequence of symbols  $(x_1, \dots, x_n)$ . A similar convention has been used for sequences of jointly distributed random variables, as  $X^n \equiv (X_1, \dots, X_n)$ . Such a characterization is of limited practical use, however, as it does not apparently lead to a finite computation for deciding if a source can be transmitted.

In quantum Shannon theory, various capacities of a single quantum channel are not currently known to be computable in general. For instance, the best known formula for the capacity of an arbitrary quantum channel for transmitting quantum information does not generally result in a single-letter optimization problem [4]. Furthermore, a general single-letter expression for the classical capacity of a general quantum channel is only known when the encoder is restricted to preparing product states.

Winter [5] has shown that the capacity region of a multiple-access channel with classical inputs and a quantum output for the transmission of independent classical messages admits a single-letter characterization, which is identical in form to that of  $\mathcal{C}$ . Results on the classical capacity region of quantum binary adder channels are contained in [6] and [7]. In what follows, we

will examine the capacity region of an arbitrary quantum multiple-access channel with quantum inputs and a quantum output, used in two distinct ways for the transmission of uncorrelated information from each terminal. Our first result describes the capacity region for the case in which one user sends quantum information, and the other classical. The second result characterizes the capacity region for the scenario in which each user wishes to send only quantum information.

This paper is organized as follows. Section II contains the relevant background material necessary to state and prove our main results. This includes mention of the notational conventions we will use throughout this paper, definitions of the distance measures for states we will use, as well as definitions of the information quantities that will characterize our rate regions. We also introduce two of the three equivalent information processing tasks that will be considered in this paper, entanglement transmission and entanglement generation. The constraints on successful transmission in these two scenarios is analogous to the *average* error criterion in classical information theory. Section III contains statements of Theorems 1 and 2, the main results of this paper. We collect various relationships between our distance measures, a number of lemmas, and statements of existing coding propositions in Section IV, which also contains the proofs of Theorems 1 and 2. In Section V, we will introduce a third information processing scenario, strong subspace transmission, whose success criterion is analogous to a *maximal* error constraint in classical information theory. All three scenarios will be proved equivalent in that section as well.

While the characterizations of the capacity regions given in Theorems 1 and 2 involve an infinite sequence of optimizations over many instances of the channel, Section VI provides examples of channels for which only a single optimization over one instance is required. These channels are related to a class of single-user channels whose quantum capacities have known single-letter descriptions—the so-called degradable channels [29]. We review these in Section VI-A, where we also prove a new concavity result for these channels. The additive examples for the classical-quantum (cq) capacity appear in Section VI-B, while those for the quantum-quantum (qq) capacity are in Section VI-C. We reflect on and relate our results to others in Section VII. Finally, the Appendix contains proofs of the convexity of our capacity regions (Appendix IX-A), and of the sufficiency of the bound on the cardinality of the set of classical message states for cq protocols (part B of the Appendix).

## II. BACKGROUND

We assume a basic understanding of the rudiments of quantum information theory—a good reference is [8]. Here, we review the notational conventions used in this paper. A typical quantum system will be labeled  $A$ . Its Hilbert space will be  $\mathcal{H}_A$ . The dimension of  $\mathcal{H}_A$  will be abbreviated as  $|A| = \dim \mathcal{H}_A$ . For convenience, the label  $A$  will often be shorthand for some collection of operators on  $\mathcal{H}_A$  when the context makes this apparent. For example, a density matrix  $\rho$  on the system  $A$  is a nonnegative operator on  $\mathcal{H}_A$  of unit trace. We will often abbreviate this by writing  $\rho^A$  to remind the reader of the system to which the state refers. A channel  $\mathcal{N}: A \rightarrow B$  is a linear map  $\mathcal{N}: \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$  that is

completely positive and trace preserving. Two systems  $A$  and  $B$  may be combined with a tensor product, resulting in the system  $AB$ , where  $\mathcal{H}_{AB} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$ . Given a multipartite state such as  $\rho^{ABC}$ , we abbreviate its partial traces by omitting superscripts, e.g.,  $\rho^{AB} \equiv \text{Tr}_C \rho^{ABC}$ . Where convenient, we occasionally refer to the global state (the one defined on the largest number of subsystems) by omitting all superscripts; i.e.,  $\rho \equiv \rho^{ABC}$  in this case. The system  $A^n$  has a Hilbert space  $\mathcal{H}_{A^n} \equiv \mathcal{H}_A^{\otimes n}$ , and the various operator algebras described by  $A^n$  will be appropriate subsets of  $\mathcal{B}(\mathcal{H}_A^{\otimes n})$ . We will freely identify  $\mathcal{N} \equiv \mathcal{N} \otimes \mathbb{1}^C$ , where  $C$  is any other system, in order to simplify long expressions. This procedure will always result in a unique completely positive map, since every channel in this paper will be completely positive. The maximally mixed state on a Hilbert space  $\mathcal{H}_A$  will always be written as  $\pi^A = \frac{1}{|A|} \mathbb{1}^A$ , and we reserve the symbol  $|\Phi\rangle$  for bipartite states that are maximally entangled. An exception to this convention will be made when, given a density matrix  $\rho^{A'}$ , we write  $|\Phi_\rho\rangle^{AA'}$  for a purification of  $\rho^{A'}$ . When we write the density matrix of a pure state  $|\psi\rangle$ , we will freely make the abbreviation  $\psi \equiv |\psi\rangle\langle\psi|$ . A measurement on a quantum system in the state  $\rho$  is modeled by a positive operator-valued measure (POVM) consisting of a set  $\{\Lambda_x\}$  of nonnegative definite matrices satisfying  $\sum_x \Lambda_x = \mathbb{1}$  yielding probabilities  $p(x) = \text{Tr} \Lambda_x \rho$ .

We will use the following conventions for distance measures between states. If  $\rho$  and  $\sigma$  are density matrices, we will write

$$F(\sigma, \rho) = \left( \text{Tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right)^2$$

for (the squared version of) the fidelity [9]. It is not hard to check that  $F$  is symmetric. For two pure states, this reduces to

$$F(|\phi\rangle, |\psi\rangle) = |\langle\phi|\psi\rangle|^2$$

while for a pure state and a mixed state

$$F(|\phi\rangle, \rho) = \langle\phi|\rho|\phi\rangle = \text{Tr} \phi \rho.$$

In this last case, we may interpret the fidelity as the success probability for a measurement that tests for the presence of the pure state  $\phi$ , when a physical system with density matrix  $\rho$  is presented. Indeed, for a POVM  $\{\phi, \mathbb{1} - \phi\}$

$$\text{Pr}\{\text{measure } \phi | \text{prepared } \rho\} = \text{Tr} \phi \rho = F(\phi, \rho).$$

The *trace norm* of an operator  $A \in \mathcal{B}(\mathcal{H})$  is defined as the sum of its singular values, and can be expressed as

$$\|A\|_1 = \text{Tr} \sqrt{A^\dagger A}.$$

This gives rise to another useful distance measure on states, the *trace distance*, defined as the trace norm of the difference between the states. It can be written explicitly as

$$\|\sigma - \rho\|_1 = \text{Tr} \sqrt{(\sigma - \rho)^2}$$

and carries a normalization which assigns a distance of 2 to states with orthogonal support.

In order to introduce the information quantities that will be used to characterize our capacity regions, we first introduce the concept of a cq density matrix or state. Let  $X$  be

an  $\mathcal{X}$ -valued random variable distributed according to  $p(x)$ , where  $\mathcal{X}$  is a finite set. We can define a Hilbert space  $\mathcal{H}_X$  with a fixed orthonormal basis  $\{|x\rangle^X\}_{x \in \mathcal{X}}$ . This sets up an identification  $|\cdot\rangle^X: \mathcal{X} \rightarrow \mathcal{H}_X$  between the elements of  $\mathcal{X}$  and that particular basis. By this correspondence, the probability distribution  $p(x)$  can be mapped to a density matrix  $\rho^X = \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x|^X$ , which is diagonal in the basis  $\{|x\rangle^X\}_{x \in \mathcal{X}}$ . Further, to every subset  $S \subseteq \mathcal{X}$  corresponds a projection matrix  $\Pi_S = \sum_{x \in S} |x\rangle\langle x|^X$ , which commutes with  $\rho$ . This way, we can express concepts from classical probability theory in the language of quantum probability, such as the equivalence  $\Pr\{X \in S\} = \text{Tr}\rho^X \Pi_S$ . From the early development of quantum mechanics, noncommutativity has been seen to be the hallmark of quantum behavior. It is to be expected that classical probability, embedded in quantum theory's framework, is described entirely with commuting matrices.

Consider now a collection of density matrices  $\{\rho_x^A\}_{x \in \mathcal{X}}$ , indexed by the finite set  $\mathcal{X}$ . If those states occur according to the probability distribution  $p(x)$ , we may speak of an *ensemble*  $\{p(x), \rho_x^A\}$  of quantum states. In order to treat classical and quantum probabilities in the same framework, a joint density matrix can be constructed

$$\rho^{XA} = \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x|^X \otimes \rho_x^A.$$

This is known as a *cq state*, and describes the classical and quantum aspects of the ensemble on the extended Hilbert space  $\mathcal{H}_X \otimes \mathcal{H}_A$  [10]. The semiclassical nature of the ensemble is reflected in the embedding of a direct sum of Hilbert spaces  $\bigoplus_{x \in \mathcal{X}} \mathcal{H}_{A_x}$  into  $\mathcal{H}_X \otimes \mathcal{H}_A$ . This should be compared with the purely classical case, where a direct sum of one-dimensional vector spaces  $\bigoplus_{x \in \mathcal{X}} \mathbb{C}$  was embedded into  $\mathcal{H}_X$ . Just as our classical density matrix  $\rho^X$  was diagonal in a basis corresponding to elements of  $\mathcal{X}$ , the cq density matrix  $\rho^{XA}$  is *block diagonal*, where the diagonal block corresponding to  $x$  contains the non-normalized density matrix  $p(x)\rho_x^A$ . The classical state  $\rho^X$  is recovered as a partial trace of  $\rho^{XA}$ , while the average quantum state is  $\rho^A = \sum_{x \in \mathcal{X}} p(x)\rho_x^A$ . We will further speak of cq states, which consist of two quantum parts and one classical. When even more systems are involved, we will defer to the terminology cq to mean that some subsystems are classical, while some are quantum. Such states are not only of interest in their own right; information quantities evaluated on cq states play an important role in characterizing what is possible in quantum information theory. Now, let  $\rho^{ABX}$  be some cq state, in block-diagonal form

$$\rho^{ABX} = \sum_x p(x) |x\rangle\langle x|^X \otimes \rho_x^{AB}.$$

We write

$$H(A)_\rho = H(\rho^A) = -\text{Tr}\rho^A \log \rho^A$$

for the von Neumann entropy of the density matrix associated with  $A$ .  $H(AB)_\rho$  is defined analogously. We will omit sub-

scripts when the state under consideration is apparent. The *mutual information*, or *Holevo  $\chi$  quantity* [11], is defined as

$$I(X; B) \equiv H(X) + H(B) - H(XB).$$

Depending on the context, the *coherent information* [12] will be expressed in one of two ways. For a fixed joint state on  $AB$ , we write

$$I_c(A)B) \equiv H(B) - H(AB) = -H(A|B).$$

Otherwise, if we are given a density matrix  $\rho^{A'}$  and a channel  $\mathcal{N}: A' \rightarrow B$ , which give rise to a joint state  $(\mathbb{1}^A \otimes \mathcal{N})(\Phi_\rho^{AA'})$ , where  $|\Phi_\rho\rangle^{AA'}$  is any purification of  $\rho^{A'}$ , we will often use the notation

$$I_c(A)B) = I_c(\rho^{A'}, \mathcal{N}) = H(\mathcal{N}(\rho^{A'})) - H((\mathbb{1} \otimes \mathcal{N})(\Phi_\rho^{AA'})).$$

It can be shown that this latter expression is independent of the particular purification  $|\Phi_\rho\rangle^{AA'}$  that is chosen for  $\rho^{A'}$ .

Despite their distinct forms, the mutual and coherent informations do share a common feature. For a fixed input state, each is a convex function of the channel. We further remark that the quantity  $I_c(A)BX)$  can be considered as a conditional, or expected, coherent information, as

$$I_c(A)BX)_\sigma = \sum_x p(x) I(A)B)_{\sigma_x}.$$

A particular departure of this quantity from its classical analog, the conditional mutual information  $I(X; Y|Z)$ , is that the latter is only equal to  $I(X; YZ)$  when  $X$  and  $Z$  are independent, while the former always allows either interpretation, provided the conditioning variable is classical.

Conditional coherent information arises in another context; suppose that  $\mathcal{N}: A' \rightarrow XB$  is a *quantum instrument* [13], meaning that  $\mathcal{N}$  acts as

$$\mathcal{N}: \rho^{A'} \mapsto \sum_x |x\rangle\langle x|^X \otimes \mathcal{N}_x(\rho^{A'}).$$

The completely positive maps  $\{\mathcal{N}_x\}$  are the *components* of the instrument. While they are generally trace reducing, their sum  $\mathcal{N} = \sum_x \mathcal{N}_x$  is always trace preserving. It is not difficult to show that

$$I_c(\rho^{A'}, \mathcal{N}) = I_c(A)BX)$$

where the latter quantity is computed with respect to the state

$$\sum_x p(x) |x\rangle\langle x|^X \otimes (\mathbb{1}^A \otimes \mathcal{N}_x)(\Phi_\rho^{AA'}).$$

For us, a quantum multiple-access channel is a channel  $\mathcal{N}: A'B' \rightarrow C$  with two inputs and one output. We will assume that the inputs  $A'$  and  $B'$  are under the control of Alice and Bob, respectively, and that the output  $C$  is maintained by Charlie. We will present three different quantum information processing scenarios which, as we will see, lead to equivalent cq and qq capacity regions.

### A. Classical-Quantum (cq) Protocols

These protocols will be relevant to Theorem 1. Using a large number  $n$  of instances of  $\mathcal{N}$ , Alice tries to send classical information to Charlie at rate  $r$ , while Bob simultaneously attempts to convey *quantum information* at rate  $S$ . Alice's communication consists of sending Charlie one of  $2^{nr}$  classical messages. To this end, we allow her to prepare arbitrary pure states  $|\phi_m\rangle^{A^n}$  at her input  $A^n$  to the channel. It is assumed that neither Alice nor Bob shares any additional resources with Charlie or among themselves, such as entanglement or noiseless quantum channels. We consider three different information processing tasks which Bob can perform, introduced in order of apparently increasing strength. The first two, entanglement generation and entanglement transmission, are outlined below, as each plays an essential role in the proof of our main result. The third, strong subspace transmission, is described in Section V-A. While not essential for the understanding of our main results, we include it in this paper because the composability properties implied by its more stringent constraints on successful communication make it particularly attractive as a building block for creating more intricate protocols from simpler ones. That each of these aforementioned scenarios can justifiably be considered as "sending quantum information" to Charlie will be proved in Sections V-B and V-C, where we will show that each gives rise to the same collection of achievable rates.

*I-Entanglement Generation:* With the goal of eventually sharing near-maximal entanglement with Charlie, Bob begins by preparing a bipartite pure state  $|\Upsilon\rangle^{BB^n}$ , entangled between a physical system  $B$  located in his laboratory, and the  $B^n$  part of the inputs of  $\mathcal{N}^{\otimes n}$ . Charlie's postprocessing procedure will be modeled by a quantum instrument. While the outer bound provided by our converse theorem will apply to any decoding modeled by an instrument, our achievability proof will require a less general approach, consisting of the following steps.

In order to ascertain Alice's message  $M$ , Charlie first performs some measurement on  $C^n$  whose statistics are given by a POVM  $\{\Lambda_m\}_{m \in 2^{nr}}$ . We let the result of that measurement be denoted  $\hat{M}$ , his declaration of the message sent by Alice. Based on the result of that measurement, he will perform one of  $2^{nr}$  decoding operations  $\mathcal{D}'_m: C^n \rightarrow \hat{B}$ . These two steps can be mathematically combined to define a *quantum instrument*  $\mathcal{D}: C^n \rightarrow \hat{M}\hat{B}$  with (trace-reducing) components

$$\mathcal{D}_m: \tau^{C^n} \mapsto \mathcal{D}'_m(\sqrt{\Lambda_m} \tau^{C^n} \sqrt{\Lambda_m}).$$

The instrument acts as

$$\mathcal{D}: \tau^{C^n} \mapsto \sum_{m=1}^{2^{nr}} |m\rangle\langle m|^{\hat{M}} \otimes \mathcal{D}_m(\tau^{C^n})$$

and induces the trace-preserving map  $\mathcal{D}: C^n \rightarrow \hat{B}$ , acting according to

$$\mathcal{D}: \tau^{C^n} \mapsto \text{Tr}_{\hat{M}} \mathcal{D}(\tau^{C^n}) = \sum_{m=1}^{2^{nr}} \mathcal{D}_m(\tau^{C^n}).$$

We again remark that this is the most general decoding procedure required of Charlie. Any situation in which he were to

iterate the above steps by measuring, manipulating, measuring again, and so on, is equivalent to a single instance of the above protocol.  $(\{|\phi_m\rangle^{A^n}\}_{m \in 2^{nr}}, |\Upsilon\rangle^{BB^n}, \mathcal{D})$  will be called a  $(2^{nr}, 2^{nS}, n, \epsilon)$  cq entanglement generation code for the channel  $\mathcal{N}$  if

$$2^{-nr} \sum_{m=1}^{2^{nr}} P_s^I(m, |\Upsilon\rangle^{BB^n}) \geq 1 - \epsilon \quad (1)$$

where

$$P_s^I(m, |\Upsilon\rangle) = F\left(|m\rangle|\Phi\rangle^{\hat{B}\hat{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\phi_m^{A^n} \otimes \Upsilon^{BB^n})\right). \quad (2)$$

We will say that  $(r, S)$  is an achievable cq rate pair for entanglement generation if there exists a sequence of  $(2^{nr}, 2^{nS}, n, \epsilon_n)$  cq entanglement generation codes with  $\epsilon_n \rightarrow 0$ . The capacity region  $\mathcal{C}_{Q_I}(\mathcal{N})$  is defined to be the closure of the collection of all achievable cq rate pairs for entanglement generation.

*II-Entanglement Transmission:* In this scenario, rather than generating entanglement with Charlie, Bob will act to transmit *preexisting* entanglement to him. We assume that Bob is presented with the  $\hat{B}$  part of the maximally entangled state  $|\Phi\rangle^{\hat{B}\hat{B}}$ . It is assumed that he has complete control over  $\hat{B}$ , while he has no access to  $B$ . He will perform a physical operation in order to transfer the quantum information embodied in his system  $\hat{B}$  to the inputs  $B^n$  of the channel, modeled by an encoding operation  $\mathcal{E}: \hat{B} \rightarrow B^n$ . The goal of this encoding will be to make it possible for Charlie, via postprocessing of the information embodied in the system  $C^n$ , to hold the  $\hat{B}$  part of a state which is close to that which would have resulted if Bob had sent his system through a perfect quantum channel  $\text{id}: \hat{B} \rightarrow \hat{B}$ . Here, we imagine that  $\hat{B}$  and  $\hat{B}$  denote two distinct physical systems with the same number of quantum degrees of freedom. The role of the identity channel is to set up a unitary correspondence, or isomorphism, between the degrees of freedom of  $\hat{B}$  in Bob's laboratory and those of  $\hat{B}$  in Charlie's. We will often tacitly assume that such an identity map has been specified ahead of time in order to judge how successful an imperfect quantum transmission has been. This convention will be taken for granted many times throughout this paper, wherein specification of an arbitrary state  $|\Psi\rangle^{\hat{B}\hat{B}}$  will immediately imply specification of the state  $|\Psi\rangle^{B\hat{B}} = (\mathbb{1}^B \otimes \text{id})|\Psi\rangle^{\hat{B}\hat{B}}$ . Decoding is the same as it is for scenario I.

$(\{|\phi_m\rangle\}_{m \in 2^{nr}}, \mathcal{E}, \mathcal{D})$  will be called a  $(2^{nr}, 2^{nS}, n, \epsilon)$  cq entanglement transmission code for the channel  $\mathcal{N}$  if

$$2^{-nr} \sum_{m=1}^{2^{nr}} P_s^{II}(m) \geq 1 - \epsilon \quad (3)$$

where

$$P_s^{II}(m) = F\left(|m\rangle|\Phi\rangle^{\hat{B}\hat{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\phi_m^{A^n} \otimes \mathcal{E}(\Phi^{\hat{B}\hat{B}}))\right). \quad (4)$$

Achievable rate pairs and the capacity region  $\mathcal{C}_{Q_{II}}(\mathcal{N})$  are defined analogous to those for scenario I.

Scenario III will be introduced in Section V, where it will also be shown that all three scenarios gives rise to the same set of

achievable rates. For this reason, we will henceforth only speak of a single capacity region

$$\mathcal{CQ}(\mathcal{N}) = \mathcal{CQ}_I(\mathcal{N}) = \mathcal{CQ}_{II}(\mathcal{N}) = \mathcal{CQ}_{III}(\mathcal{N}).$$

### B. Quantum-Quantum (qq) Protocols

The subject of Theorem 2, these protocols concern the case in which Alice and Bob each wish to send only quantum information to Charlie at rates  $R$  and  $S$ , respectively. As in the cq case, we will initially describe two different senses in which such a task can be considered. Again, Section V will introduce a third scenario, which will be shown to be equivalent to the following two.

*I-Entanglement Generation:* For encoding, Alice and Bob, respectively, prepare the states  $|\Upsilon_1\rangle^{AA^n}$  and  $|\Upsilon_2\rangle^{BB^n}$ , entangled with the  $A^n$  and  $B^n$  parts of the inputs of  $\mathcal{N}^{\otimes n}$ . Their goal is to do this in such a way so that Charlie, after applying a suitable decoding operation  $\mathcal{D}: C^n \rightarrow \hat{A}\hat{B}$ , can hold the  $\hat{A}\hat{B}$  part of a state which is close to  $|\Phi_1\rangle^{A\hat{A}}|\Phi_2\rangle^{B\hat{B}}$ . Formally,  $(|\Upsilon_1\rangle^{AA^n}, |\Upsilon_2\rangle^{BB^n}, \mathcal{D})$  is a  $(2^{nR}, 2^{nS}, n, \epsilon)$  qq entanglement generation code for the channel  $\mathcal{N}$  if

$$F(|\Phi_1\rangle^{A\hat{A}}|\Phi_2\rangle^{B\hat{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\Upsilon_1^{AA^n} \otimes \Upsilon_2^{BB^n})) \geq 1 - \epsilon. \quad (5)$$

As before,  $(R, S)$  is an achievable qq rate pair for entanglement generation if there is a sequence of  $(2^{nR}, 2^{nS}, n, \epsilon_n)$  qq entanglement generation codes with  $\epsilon_n \rightarrow 0$ . The capacity region  $\mathcal{Q}_I(\mathcal{N})$  is the closure of the collection of all such achievable rates.

*II-Entanglement Transmission:* Alice and Bob each, respectively, have control over the  $\tilde{A}$  and  $\tilde{B}$  parts of the separate maximally entangled states  $|\Phi_1\rangle^{A\tilde{A}}, |\Phi_2\rangle^{B\tilde{B}}$ , while neither has access to  $A$  or  $B$ . Alice transfers the correlations in her system to the  $A^n$  parts of the inputs of  $\mathcal{N}^{\otimes n}$  with an encoding operation  $\mathcal{E}_1: \tilde{A} \rightarrow A^n$ . Bob acts similarly with  $\mathcal{E}_2: \tilde{B} \rightarrow B^n$ . Their goal is to preserve the respective correlations, so that Charlie can apply a decoding operation  $\mathcal{D}: C^n \rightarrow \hat{A}\hat{B}$ , in order to end up holding the  $\hat{A}\hat{B}$  part of a state which is close to  $|\Phi_1\rangle^{A\hat{A}}|\Phi_2\rangle^{B\hat{B}}$ . Formally,  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$  is a  $(2^{nR}, 2^{nS}, n, \epsilon)$  qq entanglement transmission code for the channel  $\mathcal{N}$  if

$$F(|\Phi_1\rangle^{A\hat{A}}|\Phi_2\rangle^{B\hat{B}}, \Omega^{A\hat{A}B\hat{B}}) \geq 1 - \epsilon \quad (6)$$

where

$$\Omega^{A\hat{A}B\hat{B}} = \mathcal{D} \circ \mathcal{N}^{\otimes n} \circ (\mathcal{E}_1 \otimes \mathcal{E}_2)(\Phi_1^{A\tilde{A}} \otimes \Phi_2^{B\tilde{B}}).$$

Achievable qq rate pairs for entanglement generation and the capacity region  $\mathcal{Q}_{II}(\mathcal{N})$  are defined as in the previous scenario. As in the cq case, we defer to Section V, the introduction of scenario III, as well as the proof that

$$\mathcal{Q}(\mathcal{N}) = \mathcal{Q}_I(\mathcal{N}) = \mathcal{Q}_{II}(\mathcal{N}) = \mathcal{Q}_{III}(\mathcal{N}).$$

## III. MAIN RESULTS

Our first theorem gives a characterization of  $\mathcal{CQ}(\mathcal{N})$  as a regularized union of rectangles.

*Theorem 1:* Given a quantum multiple-access channel  $\mathcal{N}: A'B' \rightarrow C$ , its cq capacity region  $\mathcal{CQ}(\mathcal{N})$  is given by the closure of

$$\bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{CQ}^{(1)}(\mathcal{N}^{\otimes k})$$

where  $\mathcal{CQ}^{(1)}(\mathcal{M})$  equals the pairs of nonnegative rates  $(r, S)$  satisfying

$$\begin{aligned} r &\leq I(X; C)_\sigma \\ S &\leq I_c(B)CX)_\sigma \end{aligned}$$

for some pure state ensemble  $\{p(x), |\phi_x\rangle^{A'}\}_{x \in \mathcal{X}}$  and a bipartite pure state  $|\Psi\rangle^{BB'}$  giving rise to

$$\sigma^{XBC} = \sum_x p(x) |x\rangle\langle x|^X \otimes \mathcal{M}(\phi_x^{A'} \otimes \Psi^{BB'}). \quad (7)$$

It is sufficient to consider  $|\mathcal{X}| \leq \min\{|A'|, |C|\}^2 + 1$  when computing  $\mathcal{CQ}^{(1)}$ .

The next theorem offers a characterization of  $\mathcal{Q}(\mathcal{N})$  as a regularized union of pentagons.

*Theorem 2:* Given a quantum multiple-access channel  $\mathcal{N}: A'B' \rightarrow C$ , its qq capacity region  $\mathcal{Q}(\mathcal{N})$  is given by the closure of

$$\bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{Q}^{(1)}(\mathcal{N}^{\otimes k})$$

where  $\mathcal{Q}^{(1)}(\mathcal{M})$  equals the pairs of nonnegative rates  $(R, S)$  satisfying

$$\begin{aligned} R &\leq I_c(A)BC)_\sigma \\ S &\leq I_c(B)AC)_\sigma \\ R + S &\leq I_c(AB)C)_\sigma \end{aligned}$$

for some bipartite pure states  $|\Psi_1\rangle^{AA'}$  and  $|\Psi_2\rangle^{BB'}$  giving rise to

$$\sigma^{ABC} = (\mathbb{1}^{AB} \otimes \mathcal{M})(\Psi_1^{AA'} \otimes \Psi_2^{BB'}). \quad (8)$$

We remark here that there does not appear to be any obstacle preventing application of the methods used in this paper to prove many-sender generalizations of the above theorems. For simplicity, we have focused on the situations with two senders. In the Appendix, we prove that the multiletter nature of the regions of Theorems 1 and 2 implies that they do not require convexification. This is in contrast to the general single-letter result in the classical case for which it is indeed necessary. The characterizations given in each of the above theorems do not apparently lead to a finite computation for determining the capacity

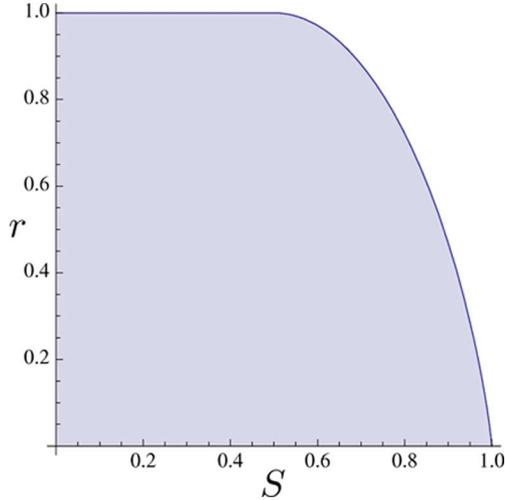


Fig. 1.  $\mathcal{CQ}$  (erasure channel) with  $d = 2$ .

regions, as neither admits a single-letter characterization in general. However, we prove in Section VI-B (Theorem 3) that for a class of quantum multiple-access channels, one has that  $\mathcal{CQ}$  is equal to the convex hull of  $\mathcal{CQ}^{(1)}$ . As a concrete application of this theorem, consider a channel  $\mathcal{N}: A'B' \rightarrow C$  with  $|A'| = 2$ ,  $|B'| = d$ , and  $|C| = d + 1$ . This channel first measures  $A'$  in the computational basis. If the outcome is  $|1\rangle^{A'}$ , Charlie receives Bob's input without error, while if the outcome is  $|2\rangle^{A'}$ , Bob's input is replaced by some fixed pure state. In addition, Charlie always knows whether an error has occurred. We show there that the cq region of this channel is equal to the set of all pairs of nonnegative cq rates  $(r, S)$  with outer boundary given by the parameterized curve

$$(r, S) = \begin{cases} (1, q \log d), & 0 \leq q \leq 1/2 \\ (H(q), q \log d), & 1/2 \leq q \leq 1. \end{cases}$$

This region is pictured in Fig. 1 for  $d = 2$ .

In Section VI-C, we define a family of “collective phase flip channels” which, with probability  $p$ , flip the phases of both input qubits. We show there that the qq capacity region of such a channel is given by a single pentagon of nonnegative qq rates  $(R, S)$  satisfying

$$\begin{aligned} R &\leq 1 \\ S &\leq 1 \\ R + S &\leq 2 - H(p). \end{aligned}$$

The characterizations given in Theorems 1 and 2 are not the only possible ways to describe the corresponding regions. It is possible to prove coding theorems and converses for regularizations of distinct single-letter regions for each of  $\mathcal{CQ}(\mathcal{N})$  and  $\mathcal{Q}(\mathcal{N})$ . Although the characterizations given above are the best which we are aware of, we refer the reader to Section VII for further discussion regarding other characterizations.

#### IV. PROOFS OF THEOREMS 1 AND 2

We first collect some relevant results that will be used in what follows, starting with some relationships between our distance

measures. If  $\rho$  and  $\sigma$  are density matrices defined on the same Hilbert space, set

$$F = F(\rho, \sigma) \quad \text{and} \quad T = \|\rho - \sigma\|_1.$$

Then, the following inequalities hold (see, e.g., [8]):

$$1 - \sqrt{F} \leq T/2 \leq \sqrt{1 - F} \quad (9)$$

$$1 - T \leq F \leq 1 - T^2/4. \quad (10)$$

From these inequalities, we can derive the following more useful relationships:

$$F \geq 1 - \epsilon \quad \Rightarrow \quad T \leq 2\sqrt{\epsilon} \quad (11)$$

$$T \leq \epsilon \quad \Rightarrow \quad F \geq 1 - \epsilon \quad (12)$$

which are valid for  $0 \leq \epsilon \leq 1$ . Uhlmann [9] has given the following characterization of fidelity:

$$F(\rho, \sigma) = \max_{|\Psi_\rho\rangle, |\Phi_\sigma\rangle} |\langle \Psi_\rho | \Phi_\sigma \rangle|^2 = \max_{|\Psi_\rho\rangle} |\langle \Psi_\rho | \Phi_\sigma \rangle|^2$$

where the first maximization is over all purifications of each state, and the second maximization holds for any fixed purification  $|\Phi_\sigma\rangle$  of  $\sigma$ . This characterization is useful in two different ways. First, for any two states, it guarantees the existence of purifications of those states whose squared inner product equals the fidelity. Second, one can derive from that characterization the following monotonicity property [14] associated with an arbitrary trace-preserving channel  $\mathcal{N}$ :

$$F(\rho, \sigma) \leq F(\mathcal{N}(\rho), \mathcal{N}(\sigma)). \quad (13)$$

An analogous property is shared by the trace distance [15]

$$\|\rho - \sigma\|_1 \geq \|\mathcal{N}(\rho) - \mathcal{N}(\sigma)\|_1 \quad (14)$$

which holds even if  $\mathcal{N}$  is trace reducing. A simple proof for the trace-preserving case can be found in [8]. These inequalities reflect the fact that completely positive maps are *contractive* and cannot improve the distinguishability of quantum states; the closer states are to each other, the harder it is to tell them apart. We will often refer to either of these two properties as just “monotonicity,” as the particular one to be used will always be clear from the context. Another useful property will be the multiplicativity of fidelities under tensor products

$$F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = F(\rho_1, \sigma_1)F(\rho_2, \sigma_2). \quad (15)$$

Since the trace distance is a norm, it satisfies the triangle inequality. The fidelity is not a norm, but it is possible to derive the following analog by applying (9) and (10) to the triangle inequality for the trace distance:

$$F(\rho_1, \rho_3) \geq 1 - 2\sqrt{1 - F(\rho_1, \rho_2)} - 2\sqrt{1 - F(\rho_2, \rho_3)}. \quad (16)$$

It will be possible to obtain a sharper triangle-like inequality as a consequence of the following lemma, which states that if a measurement succeeds with high probability on a state, it will also do so on a state that is close to that state in trace distance.

*Lemma 1:* Suppose  $\rho, \sigma, \Lambda \in \mathcal{B}(\mathcal{H})$ , where  $\rho$  and  $\sigma$  are density matrices, and  $0 \leq \Lambda \leq \mathbb{1}$ . Then,  $\text{Tr}\Lambda\sigma \geq \text{Tr}\Lambda\rho - \|\rho - \sigma\|_1$ .

*Proof:*

$$\begin{aligned} \text{Tr}\lambda\sigma &= \text{Tr}\lambda\rho - \text{Tr}\lambda(\rho - \sigma) \\ &\geq \text{Tr}\lambda\rho - \max_{0 \leq \lambda \leq \mathbb{1}} \text{Tr}\lambda(\rho - \sigma) \\ &= \text{Tr}\lambda\rho - \frac{1}{2}\|\rho - \sigma\|_1 \end{aligned}$$

where the last equality follows from a characterization of trace distance given in [8].  $\square$

Since  $F(\phi, \rho) = \text{Tr}\phi\rho$  when  $\phi$  is a pure state, a corollary of Lemma 1 is that

$$F(\phi, \sigma) \geq F(\phi, \rho) - \frac{1}{2}\|\rho - \sigma\|_1$$

a fact we will refer to as the ‘‘special triangle inequality.’’

The following lemma can be thought of either as a type of transitivity property inherent to any bipartite state with a component near a pure state, or as a partial converse to the monotonicity of fidelity.

*Lemma 2:* For finite-dimensional Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , let a pure state  $|\phi\rangle^A$  and two density matrices  $\rho^B$  and  $\Omega^{AB}$  be given. Then

$$F(\phi^A \otimes \rho^B, \Omega^{AB}) \geq 1 - \|\rho^B - \Omega^B\|_1 - 3(1 - F(\phi^A, \Omega^A)).$$

*Proof:* We begin by defining the subnormalized density matrix  $\tilde{\omega}^B$  via the equation

$$(\phi^A \otimes \mathbb{1}^B)\Omega^{AB}(\phi^A \otimes \mathbb{1}^B) = \phi^A \otimes \tilde{\omega}^B \quad (17)$$

so that  $\tilde{\omega}^B$  is the upper-left block of  $\Omega^{AB}$  when the basis for  $\mathcal{H}_A$  is chosen so that  $|\phi\rangle^A = (1, 0, \dots, 0)^T$ . Notice that  $F(\phi^A, \Omega^A) = \text{Tr}\tilde{\omega} \equiv f$ . The normalized state  $\omega^B = \tilde{\omega}^B/f$  satisfies

$$\|\omega^B - \tilde{\omega}^B\|_1 = (1 - f)\|\omega^B\|_1 = 1 - f. \quad (18)$$

Omitting superscripts for clarity, we write

$$\begin{aligned} \sqrt{F(\phi \otimes \rho, \Omega)} &= \text{Tr}\sqrt{\sqrt{(\phi \otimes \rho)\Omega}(\phi \otimes \rho)} \\ &= \text{Tr}\sqrt{(\mathbb{1} \otimes \sqrt{\rho})(\phi \otimes \mathbb{1})\Omega(\phi \otimes \mathbb{1})(\mathbb{1} \otimes \sqrt{\rho})} \\ &= \text{Tr}\sqrt{(\mathbb{1} \otimes \sqrt{\rho})(\phi \otimes \tilde{\omega})(\mathbb{1} \otimes \sqrt{\rho})} \\ &= \text{Tr}\sqrt{\phi \otimes (\sqrt{\rho}\tilde{\omega}\sqrt{\rho})} \\ &= \text{Tr}\sqrt{\sqrt{\rho}\tilde{\omega}\sqrt{\rho}} \\ &= \sqrt{F(\tilde{\omega}, \rho)} \\ &= \sqrt{fF(\omega, \rho)} \\ &\geq \sqrt{f(1 - \|\omega - \rho\|_1)}. \end{aligned} \quad (19)$$

The first line is the definition of fidelity and the third follows from (17). The last equality relies on the fact that the fidelity, as we have defined it, is linear in either of its two inputs, while the inequality follows from (10).

Noting that  $\Omega^B \geq \tilde{\omega}^B$ , we define another positive operator  $\omega' = \Omega^B - \tilde{\omega}^B$ , which satisfies  $\text{Tr}\omega'^B = 1 - f$  and can be interpreted as the sum of the rest of the diagonal blocks of  $\Omega^{AB}$ . The trace distance in the last line above can be bounded via double application of the triangle inequality as

$$\begin{aligned} \|\rho - \omega\|_1 &\leq \|\rho - (\rho - \omega')\|_1 + \|(\rho - \omega') - \tilde{\omega}\|_1 + \|\tilde{\omega} - \omega\|_1 \\ &= \text{Tr}\omega'^B + \|\rho^B - \Omega^B\|_1 + 1 - f \\ &= \|\rho^B - \Omega^B\|_1 + 2(1 - f) \end{aligned} \quad (20)$$

where the second line follows from (18). Combining (19) with (20), we obtain

$$\begin{aligned} F(\phi^A \otimes \rho^B, \Omega^{AB}) &\geq f(1 - \|\rho^B - \Omega^B\|_1 - 2(1 - f)) \\ &\geq 1 - \|\rho^B - \Omega^B\|_1 - 3(1 - f) \end{aligned}$$

as required.  $\square$

The following continuity lemma from [16] shows that if two bipartite states are close to each other, the difference between their associated coherent informations is small.

*Lemma 3 (Continuity of Coherent Information):* Let  $\rho^{QR}$  and  $\sigma^{QR}$  be two states of a finite-dimensional bipartite system  $QR$  satisfying  $\|\rho - \sigma\|_1 \leq \epsilon$ . Then

$$|I_c(Q)R_\rho - I_c(Q)R_\sigma| \leq 2H(\epsilon) + 4\log|Q|\epsilon$$

where  $H(\epsilon)$  is the binary entropy function.

Next is Winter’s ‘‘gentle measurement’’ lemma [17], which implies that a measurement which is likely to be successful in identifying a state tends not to significantly disturb that state.

*Lemma 4 (Gentle Measurement):* Let  $\mathcal{H}$  be a finite-dimensional Hilbert space. If  $\rho \in \mathcal{B}(\mathcal{H})$  is a density matrix and  $\Lambda \in \mathcal{B}(\mathcal{H})$  is nonnegative with spectrum bounded above by 1, then

$$\text{Tr}\rho\Lambda \geq 1 - \epsilon$$

implies

$$\|\sqrt{\Lambda}\rho\sqrt{\Lambda} - \rho\|_1 \leq \sqrt{8\epsilon}.$$

Our coding theorems for multiple-access channels will make use of existing coding theorems for single-user channels. Given a channel  $\mathcal{N}: A' \rightarrow B$ , we say that the pair of channels  $(\mathcal{E}, \mathcal{D})$ , where  $\mathcal{E}: \hat{A} \rightarrow A'^n$  and  $\mathcal{D}: B^n \rightarrow \hat{A}$ , comprise a  $(2^{nR}, n, \epsilon)$  entanglement transmission code for the channel  $\mathcal{N}$  if, for the  $2^{nR} \times 2^{nR}$  maximally entangled state  $|\Phi\rangle^{A\hat{A}}$ , we have

$$F(|\Phi\rangle^{A\hat{A}}, \mathcal{D} \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}(\Phi^{A\hat{A}})) \geq 1 - \epsilon.$$

A  $(2^{nR}, n, \epsilon)$  random entanglement transmission code consists of an collection of deterministic  $(2^{nR}, n, \epsilon)$  entanglement transmission codes  $(\mathcal{E}_\beta, \mathcal{D}_\beta, P_\beta)$ , where the  $P_\beta$  are probabilities corresponding to a source of shared common randomness available to both sender and receiver. We will often omit the subscript once the randomness of the code has been clarified, and it will be understood that  $\mathcal{E}$  and  $\mathcal{D}$  constitute a pair of correlated random maps.

Associated to a random code is its expected, or average code density operator  $\mathbb{E}\mathcal{E}(\pi^{\hat{A}})$ , which is the expectation, over the shared randomness, of the image of the maximally mixed state on  $\hat{A}$ . Our reason for using random quantum codes will be to ensure that, on average, the input to  $\mathcal{N}^{\otimes n}$  is at least close to an  $n$ -fold product state.

The proof of the existence of quantum codes achieving the coherent information bound is attributed to Lloyd [18], Shor [19], and Devetak [20]. The following quantum coding proposition for single-user channels is proved in [20] and concerns the existence of random entanglement transmission codes whose average code density matrix can be made arbitrarily close to a product state.

*Proposition 1:* Given is a channel  $\mathcal{N}: A' \rightarrow B$ , a density matrix  $\rho^{A'}$ , and a number  $0 \leq R < I_c(\rho^{A'}, \mathcal{N})$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that there is a  $(2^{nR}, n, \epsilon)$  random entanglement transmission code  $(\mathcal{E}, \mathcal{D})$  for  $\mathcal{N}$  with an isometric encoder  $\mathcal{E}$  whose average code density operator  $\varrho^{A'^n} = \mathbb{E}\mathcal{E}(\pi^{A'})$  satisfies

$$\|\varrho^{A'^n} - (\rho^{A'})^{\otimes n}\|_1 \leq \epsilon.$$

Furthermore, given any particular isometric extension  $\mathcal{U}_{\mathcal{N}}: A' \rightarrow BE$  of  $\mathcal{N}$ , it is possible to choose isometric extensions  $\mathcal{U}_{\mathcal{D}_\beta}: B^n \rightarrow \hat{A}F$  of the deterministic decoders so that

$$F(|\Phi\rangle^{A\hat{A}}|\lambda\rangle^{E^n F}, \mathcal{U}_{\mathcal{D}_\beta} \circ \mathcal{U}_{\mathcal{N}}^{\otimes n} \circ \mathcal{E}_\beta|\Phi\rangle^{A\hat{A}}) \geq 1 - \epsilon \quad (21)$$

for some fixed pure state  $|\lambda\rangle^{E^n F}$ .

Next, we state an average error version of the Holevo–Schumacher–Westmoreland (HSW) theorem for cq codes with codewords chosen i.i.d. according to a product distribution [21], [22].

*Proposition 2 (HSW Theorem):* Given is a cq state

$$\sigma^{XQ} = \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^Q$$

and a number  $0 \leq R < I(X; Q)_\sigma$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that if  $2^{nR}$  codewords  $\mathcal{C} = \{X^n(m)\}$  are chosen i.i.d. according to  $p(x^n) = \prod_{i=1}^n p(x_i)$ , corresponding to input preparations

$$\rho_{x^n}^{Q^n} = \rho_{x_1}^{Q_1} \otimes \cdots \otimes \rho_{x_n}^{Q_n}$$

there exists a decoding POVM  $\{\Lambda_m\}$  on  $Q^n$ , depending on the random choice of codebook  $\mathcal{C}$ , which correctly identifies the index  $m$  with average probability of error less than  $\epsilon$ , in the sense that

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m=1}^{2^{nR}} \text{Tr} \rho_{X^n(m)}^{Q^n} \Lambda_m \geq 1 - \epsilon. \quad (22)$$

Due to the symmetry of the distribution of  $\mathcal{C}$  under codeword permutations, it is clear that the expectations of each term in the above sum are equal. In other words

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m=1}^{2^{nR}} \text{Tr} \rho_{X^n(m)}^{Q^n} \Lambda_m = \mathbb{E}_{\mathcal{C}} \text{Tr} \rho_{X^n(1)}^{Q^n} \Lambda_1 \quad (23)$$

so we will later, without loss of generality, make the assumption that Alice sends codeword  $M = 1$  during our analysis (see [23] for a detailed discussion in the classical case).

#### A. Proof of Theorem 1 (Converse)

We prove in Section V that any rate pair that is achievable for entanglement transmission is also achievable for entanglement generation. For this reason, we use the latter scenario to prove the converse part of Theorem 1. It should be noted that the reverse implication, namely, that entanglement generation implies entanglement transmission, follows from the fact the outer bound to be proved next coincides with the inner bound obtained by the coding theorem below.

Suppose there exists a sequence of  $(2^{nr}, 2^{nS}, n, \epsilon_n)$  entanglement generation codes with  $\epsilon_n \rightarrow 0$ . Fixing a blocklength  $n$ , let  $\{|\phi_m\rangle\}_{m \in 2^{nr}}$ ,  $|\Upsilon\rangle^{BB'^n}$ , and  $\mathcal{D}: C^n \rightarrow \hat{B}\hat{M}$  comprise the corresponding cq entanglement generation code. The state induced by the encoding is

$$\omega^{MBC^n} = 2^{-nr} \sum_{m=1}^{2^{nr}} |m\rangle\langle m|^M \otimes (\mathbb{1}^B \otimes \mathcal{N}^{\otimes n})(\phi_m^{A'^n} \otimes \Upsilon^{BB'^n}).$$

After applying the decoding instrument, the state is

$$\Omega^{M\hat{M}B\hat{B}} = (\mathbb{1}^{MB} \otimes \mathcal{D})(\omega^{MBC^n}).$$

We upper bound the classical rate of the code as follows:

$$\begin{aligned} nr &= H(M)_\Omega \\ &= I(M; \hat{M})_\Omega + H(M|\hat{M})_\Omega \\ &\leq I(M; \hat{M})_\Omega + H(\epsilon_n) + nr\epsilon_n \\ &\leq I(M; C^m)_\omega + n\epsilon'_n. \end{aligned}$$

The first inequality follows from Fano's inequality (see, e.g., [23]), while in the second, we use the Holevo bound [11] and define  $\epsilon'_n = \frac{1}{n} + r\epsilon_n$ . The quantum rate of the code is upper bounded as

$$\begin{aligned} I_c(B)C^m M)_\omega &\geq I_c(B)\hat{B}M)_\Omega \\ &\geq I_c(B)\hat{B})_\Omega \\ &\geq I_c(B)\hat{B})_\Phi - 2H(\epsilon_n) - 8nS\sqrt{\epsilon_n} \\ &= nS - n\epsilon''_n. \end{aligned}$$

In the above, the inequalities are consequences of the data processing inequality [12], the fact that conditioning cannot increase entropy (and thus cannot decrease coherent information) [8], a combination of Lemma 3 and (11), and the definition  $\epsilon''_n = \frac{2}{n} + nS\sqrt{\epsilon_n}$ . The second justification can be considered as an alternative statement of the well-known strong subadditivity

inequality [24], of which a recent simple proof can be found in [25]. Setting  $X = M$ , we have thus proven that

$$r \leq \frac{1}{n}I(X; C^n) + \epsilon'_n, \quad S \leq \frac{1}{n}I_c(B)C^n X + \epsilon''_n$$

whenever  $(r, S)$  is an achievable cq rate pair for entanglement generation, where  $\epsilon'_n, \epsilon''_n \rightarrow 0$ . It follows that for any achievable rate pair  $(r, S)$  and any  $\delta > 0$ , we have

$$(r - \delta, S - \delta) \in \frac{1}{n}\mathcal{CQ}^{(1)}(\mathcal{N}^{\otimes n}) \subseteq \mathcal{CQ}(\mathcal{N}).$$

Since  $\mathcal{CQ}(\mathcal{N})$  is closed by definition, this completes the proof.  $\square$

*B. Proof of Theorem 1 (Achievability)*

Our method of proof for the coding theorem will work as follows. We will employ random HSW codes and random entanglement transmission codes to ensure that the average state at the input of  $\mathcal{N}^{\otimes n}$  is close to a product state. Each sender will utilize a code designed for the product channel induced by the other’s random input, whereby existing coding theorems for product channels will be invoked. The quantum code used will be one that achieves the capacity of a modified channel, in which the classical input is copied, without error, to the output of the channel. As the random HSW codes will exactly induce a product state input, the existence of these quantum codes will follow directly from Proposition 1.

The random HSW codes will be those which exist for product channels. As random entanglement transmission codes exist with average code density matrix arbitrarily close to a product state, this will ensure that the resulting output states are distinguishable with high probability. Furthermore, obtaining the classical information will be shown to cause but a small disturbance in the overall joint quantum state of the system. As we will show, it is possible to mimic the channel for which the quantum code is designed by placing the identities of the estimated classical message states into registers appended to the outputs of each channel in the product.

The decoder for the modified channel will then be shown to define a quantum instrument which satisfies the success condition for a cq entanglement transmission code, on average. This feature will then be used to infer the existence of a particular, deterministic code which meets the same requirement.

Fix a pure state ensemble  $\{p(x), |\phi_x\rangle^{A'}\}$  and a bipartite pure state  $|\Psi\rangle^{B''B'}$ , which give rise to the cq state

$$\omega^{XB''C} = \sum_x p(x)|x\rangle\langle x|^X \otimes (\mathbf{1}^B \otimes \mathcal{N})(\phi_x^{A'} \otimes \Psi^{B''B'})$$

having the form of (7). Define the states

$$\rho_1^{A'} = \sum_x p(x)\phi_x^{A'} \quad \rho_2^{B'} = \text{Tr}_{B''}\Psi^{B''B'}$$

We will demonstrate the achievability of the corner point  $(I(X; C), I_c(B'')CX)_\omega$  by showing that for every  $\epsilon, \delta > 0$ , if

$$r = I(X; C)_\omega - \delta \quad S = I_c(B'')CX_\omega - \delta$$

there exists a  $(2^{nr}, 2^{nS}, n, \epsilon)$  cq entanglement transmission code for the channel  $\mathcal{N}$ , provided that  $n$  is sufficiently large and that  $S > 0$ . The rest of the region will follow by timesharing.

For encoding, Alice chooses  $2^{nr}$  sequences  $X^n(m)$ , i.i.d. according to the product distribution  $p(x^n) = \prod_{i=1}^n p(x_i)$ . As each sequence corresponds to a preparation of channel inputs

$$|\phi_m\rangle^{A^{1n}} = |\phi_{X_1(m)}\rangle^{A'_1} \otimes \cdots \otimes |\phi_{X_n(m)}\rangle^{A'_n}$$

the expected average density operator associated with Alice’s input to the channel is precisely

$$\begin{aligned} \mathbb{E}_{\mathcal{C}} 2^{-nr} \sum_{m=1}^{2^{nr}} |\phi_m\rangle\langle\phi_m|^{A^{1n}} &= \sum_{x^n} p(x^n) |\phi_{x^n}\rangle\langle\phi_{x^n}|^{A^{1n}} \\ &= (\rho_1^{A^{1n}})^{\otimes n}. \end{aligned}$$

Define a new channel  $\mathcal{N}_2: B' \rightarrow C\hat{X}$  (which is also an instrument) by

$$\mathcal{N}_2: \rho^{B'} \mapsto \sum_x p(x)\mathcal{N}(\phi_x^{A'} \otimes \rho^{B'}) \otimes |x\rangle\langle x|^{\hat{X}}.$$

This can be interpreted as a channel that reveals the identity of Alice’s input state to Charlie, with the added assumption that Alice chooses her inputs at random. Alternatively, one can view this as a channel with state information available to the receiver, where nature is randomly choosing the “state”  $x$  at Alice’s input. Observe that  $I_c(\rho_2^{B'}, \mathcal{N}_2) = I_c(B'')CX$ . By Proposition 1, there exists a  $(2^{n\tilde{S}}, n, \epsilon)$  random entanglement transmission code  $(\mathcal{E}_\beta, \mathcal{D}_\beta, Q_\beta)$  for the channel  $\mathcal{N}_2$ , where  $\mathcal{E}_\beta: \tilde{B} \rightarrow B'^n$  and  $\mathcal{D}_\beta: C^n \hat{X}^n \rightarrow \tilde{B}$ , such that the average code density operator  $\varrho^{B'^n} = \mathbb{E}_\beta \mathcal{E}(\pi^{\tilde{B}})$  satisfies

$$\|\varrho^{B'^n} - (\rho_2^{B'})^{\otimes n}\|_1 \leq \epsilon.$$

When Alice sends the message  $M = m$ , the resulting joint Bob–Charlie state is

$$\xi_m^{BC^n} = \mathcal{N}^{\otimes n}(\phi_m^{A^{1n}} \otimes \mathcal{E}(\Phi^{B\tilde{B}}))$$

where  $|\Phi\rangle^{B\tilde{B}}$  is the maximally entangled state which Bob is required to transmit. We will show that Charlie can distinguish the states  $\xi_m^{C^n}$  because they are close to a related set of states. By Proposition 2, for the channel  $\mathcal{N}_1: A' \rightarrow C$  given by

$$\mathcal{N}_1: \rho^{A'} \mapsto \mathcal{N}(\rho^{A'} \otimes \rho_2^{B'})$$

which would result if Bob’s average code density operator were exactly equal to  $(\rho_2^{B'})^{\otimes n}$ , there exists a decoding POVM  $\{\Lambda_m\}_{m \in 2^{nr}}$  on  $C^n$  that would identify Alice’s index  $m$  with expected average probability of error less than  $\epsilon$ , in the sense that

$$\mathbb{E}_{\mathcal{C}} 2^{-nr} \sum_{m=1}^{2^{nr}} \text{Tr} \Lambda_m \xi_m^{C^n} \geq 1 - \epsilon$$

where

$$\xi_m^{C^n} = \mathcal{N}^{\otimes n}(\phi_m^{A^{1n}} \otimes (\rho_2^{B'})^{\otimes n}).$$

By the symmetry of the random code construction, we utilize (23) to write this as

$$\mathbb{E}_C \text{Tr} \Lambda_1 \xi_1^{C^n} \geq 1 - \epsilon.$$

Note that

$$\mathbb{E}_\beta \xi_m^{C^n} = \mathcal{N}^{\otimes n}(\phi_m^{A^n} \otimes \varrho^{B^n}).$$

It follows from monotonicity of trace distance that

$$\|\mathbb{E}_\beta \xi_1^{C^n} - \xi_1^{C^n}\|_1 \leq \epsilon$$

which, together with Lemma 1, implies that

$$\mathbb{E}_C 2^{-nr} \sum_{m=1}^{2^{nr}} \text{Tr} \Lambda_m \mathbb{E}_\beta \xi_m^{C^n} = \mathbb{E}_{C\beta} \text{Tr} \Lambda_1 \xi_1^{C^n} \geq 1 - 2\epsilon.$$

This allows us to bound the expected probability of correctly decoding Alice's message as

$$\mathbb{E}_{C\beta} \text{Tr}(\mathbf{1} \otimes \Lambda_1) \xi_1^{BC^n} \geq 1 - 2\epsilon. \quad (24)$$

In order to decode, Charlie begins by performing the measurement  $\{\Lambda_m\}_{m \in 2^{nr}}$  on  $C^n$ . He declares Alice's message to be  $\widehat{M} = m$  if measurement result  $m$  is obtained. Charlie will then attempt to simulate the channel  $\mathcal{N}_2^{\otimes n}$  by associating a separate classical register  $\widehat{X}_i$  to each channel  $\mathcal{N}: A_i \rightarrow C_i$  in the product, preparing the states  $|X_i(m)\rangle_{\widehat{X}_i}$ , for  $1 \leq i \leq n$ . Additionally, he stores the result of the measurement in the system  $\widehat{M}$ , his declaration of the message intended by Alice. This procedure results in the global state

$$\Gamma^{BC^n \widehat{X}^n \widehat{M}} = \sum_{m=1}^{2^{nr}} \Gamma_m^{BC^n} \otimes |X^n(m)\rangle \langle X^n(m)|_{\widehat{X}^n} \otimes |m\rangle \langle m|_{\widehat{M}}$$

where we define the subnormalized states

$$\Gamma_m^{BC^n} = (\mathbf{1}^B \otimes \sqrt{\Lambda_m}) \xi_1^{BC^n} (\mathbf{1}^B \otimes \sqrt{\Lambda_m}).$$

If Charlie were able to perfectly reconstruct Alice's classical message, the global state would instead be

$$\Gamma'^{BC^n \widehat{X}^n \widehat{M}} = \xi_1^{BC^n} \otimes |X^n(1)\rangle \langle X^n(1)|_{\widehat{X}^n} \otimes |1\rangle \langle 1|_{\widehat{M}}.$$

When averaged over Alice's random choice of HSW code, the reduced state  $\Gamma'^{BC^n \widehat{X}^n}$  is precisely equal to the state that would arise via the action of the modified channel  $\mathcal{N}_2$ . This is because

$$\begin{aligned} \mathbb{E}_C \Gamma'^{BC^n \widehat{X}^n} &= \sum_{x^n} p(x^n) \xi_{x^n}^{BC^n} \otimes |x^n\rangle \langle x^n|_{\widehat{X}^n} \\ &= \mathcal{N}_2^{\otimes n} \circ \mathcal{E}(\Phi^{B\widehat{B}}) \end{aligned} \quad (25)$$

where we have written the state that results when Alice prepares  $\phi_{x^n}^{A^n}$  as

$$\xi_{x^n}^{BC^n} = \mathcal{N}^{\otimes n}(\phi_{x^n}^{A^n} \otimes \mathcal{E}(\Phi^{B\widehat{B}})).$$

However, our choice of a good HSW code ensures that he can almost perfectly reconstruct Alice's message. A consequence of

this will be that the two states  $\Gamma^{BC^n \widehat{X}^n}$  and  $\Gamma'^{BC^n \widehat{X}^n}$  are almost the same, as we will now demonstrate.

In what follows, we will need to explicitly keep track of the randomness in our codes by means of extra subscripts that are to be interpreted as indexing the corresponding deterministic states, which occur with the probabilities  $P_C$  and  $Q_\beta$ . Rewriting (24) as

$$\sum_{C\beta} P_C Q_\beta \text{Tr}(\mathbf{1}^B \otimes \Lambda_{1C}) \xi_{1C\beta}^{BC^n} \geq 1 - 2\epsilon$$

it is clear that we may write

$$\text{Tr}(\mathbf{1}^B \otimes \Lambda_1^C) \xi_{1C\beta}^{BC^n} \geq 1 - \epsilon_{C\beta}$$

for positive numbers  $\{\epsilon_{C\beta}\}$  chosen to satisfy

$$\sum_{C\beta} P_C Q_\beta \epsilon_{C\beta} = 2\epsilon.$$

By the gentle measurement lemma

$$\left\| (\mathbf{1}^B \otimes \sqrt{\Lambda_{1C}}) \xi_{1C\beta}^{BC^n} (\mathbf{1}^B \otimes \sqrt{\Lambda_{1C}}) - \xi_{1C\beta}^{BC^n} \right\|_1 \leq \sqrt{8\epsilon_{C\beta}}$$

and thus, by the concavity of the square root function

$$\begin{aligned} \mathbb{E}_{C\beta} \left\| (\mathbf{1}^B \otimes \sqrt{\Lambda_1}) \xi_1^{BC^n} (\mathbf{1}^B \otimes \sqrt{\Lambda_1}) - \xi_1^{BC^n} \right\|_1 \\ = \sum_{C\beta} P_C Q_\beta \left\| (\mathbf{1}^B \otimes \sqrt{\Lambda_{1C}}) \xi_{1C\beta}^{BC^n} (\mathbf{1}^B \otimes \sqrt{\Lambda_{1C}}) - \xi_{1C\beta}^{BC^n} \right\|_1 \\ \leq 4\sqrt{\epsilon}. \end{aligned}$$

Along with (24) and monotonicity with respect to  $\text{Tr}_{\widehat{M}}$ , this estimate allows us to express

$$\begin{aligned} \mathbb{E}_{C\beta} \|\Gamma^{BC^n \widehat{X}^n} - \Gamma'^{BC^n \widehat{X}^n}\|_1 \\ \leq \mathbb{E}_{C\beta} \|\Gamma^{BC^n \widehat{X}^n \widehat{M}} - \Gamma'^{BC^n \widehat{X}^n \widehat{M}}\|_1 \\ = \mathbb{E}_{C\beta} \left\| (\mathbf{1}^B \otimes \sqrt{\Lambda_1}) \xi_1^{BC^n} (\mathbf{1}^B \otimes \sqrt{\Lambda_1}) - \xi_1^{BC^n} \right\|_1 \\ + \mathbb{E}_{C\beta} \sum_{m=2}^{2^{nr}} \left\| (\mathbf{1}^B \otimes \sqrt{\Lambda_m}) \xi_1^{BC^n} (\mathbf{1}^B \otimes \sqrt{\Lambda_m}) \right\|_1 \\ = \mathbb{E}_{C\beta} \left\| (\mathbf{1}^B \otimes \sqrt{\Lambda_1}) \xi_1^{BC^n} (\mathbf{1}^B \otimes \sqrt{\Lambda_1}) - \xi_1^{BC^n} \right\|_1 \\ + \mathbb{E}_{C\beta} \sum_{m=2}^{2^{nr}} \text{Tr}(\mathbf{1}^B \otimes \Lambda_m) \xi_1^{BC^n} \\ \leq 4\sqrt{\epsilon} + 2\epsilon \\ \leq 5\sqrt{\epsilon} \end{aligned} \quad (26)$$

provided that  $\epsilon \leq \frac{1}{2}$ . Since the entanglement fidelity is linear in  $\mathcal{D}(\Gamma^{BC^n \widehat{X}^n})$ , which is itself linear in  $\Gamma^{BC^n \widehat{X}^n}$ , it follows that

$$F(|\Phi\rangle^{B\widehat{B}}, \mathcal{D}(\mathbb{E}_{C\beta} \Gamma^{BC^n \widehat{X}^n})) = F(|\Phi\rangle^{B\widehat{B}}, \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_C \Gamma^{BC^n \widehat{X}^n})).$$

The special triangle inequality can now be used to lower bound the right-hand side of this expression by

$$\begin{aligned} F(|\Phi\rangle^{B\widehat{B}}, \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_C \Gamma^{BC^n \widehat{X}^n})) \\ - \|\mathbb{E}_\beta \mathcal{D}(\mathbb{E}_C \Gamma'^{BC^n \widehat{X}^n}) - \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_C \Gamma^{BC^n \widehat{X}^n})\|_1. \end{aligned}$$

Using our earlier observation from (25) and the definition of a  $(2^{nS}, n, \epsilon)$  entanglement transmission code, we can bound the first term as

$$F(|\Phi\rangle^{B\hat{B}}, \mathcal{D}(\mathbb{E}_C \Gamma^{BC^n} \hat{X}^n)) = F(|\Phi\rangle^{B\hat{B}}, \mathcal{D} \circ \mathcal{N}_2^{\otimes n} \circ \mathcal{E}(\Phi^{B\hat{B}})) \geq 1 - \epsilon.$$

An estimate on the second term is obtained via

$$\begin{aligned} & \left\| \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_C \Gamma^{BC^n} \hat{X}^n) - \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_C \Gamma'^{BC^n} \hat{X}^n) \right\|_1 \\ & \leq \mathbb{E}_\beta \left\| \mathcal{D}(\mathbb{E}_C \Gamma^{BC^n} \hat{X}^n) - \mathcal{D}(\mathbb{E}_C \Gamma'^{BC^n} \hat{X}^n) \right\|_1 \\ & \leq \mathbb{E}_\beta \left\| \mathbb{E}_C \Gamma^{BC^n} \hat{X}^n - \mathbb{E}_C \Gamma'^{BC^n} \hat{X}^n \right\|_1 \\ & \leq \mathbb{E}_{C\beta} \left\| \Gamma^{BC^n} \hat{X}^n - \Gamma'^{BC^n} \hat{X}^n \right\|_1 \\ & \leq 5\sqrt{\epsilon} \end{aligned}$$

where first three lines are by convexity, monotonicity, and convexity once again of the trace norm. The last inequality follows from (26). Putting these together gives

$$\begin{aligned} \mathbb{E}_{C\beta} F(|\Phi\rangle^{B\hat{B}}, \mathcal{D}(\Gamma^{B\hat{X}^n C^n})) & \geq 1 - \epsilon - 5\sqrt{\epsilon} \\ & \geq 1 - 6\sqrt{\epsilon}. \end{aligned} \quad (27)$$

At last, observe that the final decoded state  $\Omega^{B\hat{B}M}$  (which still depends on both sources of randomness  $\mathcal{C}$  and  $\beta$ ) is equal to

$$\Omega^{B\hat{B}M} = \mathcal{D}(\Gamma^{BC^n} \hat{X}^n \hat{M}) \equiv \mathcal{D}(\xi_1^{BC^n}).$$

This implicitly defines the desired decoding instrument  $\mathcal{D}: C^n \rightarrow \hat{B}\hat{M}$ . The expectation of (3) can now be bounded as

$$\begin{aligned} \mathbb{E}_{C\beta} 2^{-nr} \sum_{m=1}^{2^{nr}} P_s^\Pi(m) & = \mathbb{E}_{C\beta} P_s^\Pi(1) \\ & = F(|1\rangle^M |\Phi\rangle^{B\hat{B}}, \mathbb{E}_{C\beta} \Omega^{B\hat{B}M}) \\ & \geq 1 - \left\| \mathbb{E}_{C\beta} \Gamma^M - |1\rangle\langle 1|^M \right\|_1 \\ & \quad - 3 \left( 1 - F(|\Phi\rangle^{B\hat{B}}, \mathcal{D}(\Gamma^{B\hat{X}^n C^n})) \right) \\ & \geq 1 - 2\sqrt{2\epsilon} - 18\sqrt{\epsilon} \\ & \geq 1 - 21\sqrt{\epsilon}. \end{aligned}$$

The third line above is by Lemma 2, the first estimate in the fourth line follows from (24), and the second estimate is by (27), together with (11). We may now conclude that there are particular values of the randomness indices  $\beta$  and  $\mathcal{C}$  such that the same bound is satisfied for a deterministic code. We have thus proven that  $(\{|\phi_m\rangle^{A^n}\}_{m \in 2^{nr}}, \mathcal{E}, \mathcal{D})$  comprises a  $(2^{nr}, 2^{nS}, n, 21\sqrt{\epsilon})$  entanglement transmission code. This concludes the coding theorem.  $\square$

### C. Proof of Theorem 2 (Converse)

Suppose that  $(R, S)$  is an achievable qq rate pair for entanglement generation. By definition, this means that there must exist a sequence of  $(2^{nR}, 2^{nS}, n, \epsilon_n)$  entanglement generation codes with  $\epsilon_n \rightarrow 0$ . Fixing a blocklength  $n$ , let  $|\Upsilon_1\rangle^{AA^n}, |\Upsilon_2\rangle^{BB^n}$

and  $\mathcal{D}: C^n \rightarrow \hat{A}\hat{B}$  comprise the corresponding encodings and decodings. Define

$$\omega^{ABC^n} = (\mathbb{1}^{AB} \otimes \mathcal{N}^{\otimes n})(\Upsilon_1^{AA^n} \otimes \Upsilon_2^{BB^n})$$

to be the result of sending the respective  $A^n$  and  $B^n$  parts of  $|\Upsilon_1\rangle^{AA^n}$  and  $|\Upsilon_2\rangle^{BB^n}$  through the channel  $\mathcal{N}^{\otimes n}$ . Further defining

$$\Omega^{AB\hat{A}\hat{B}} = (\mathbb{1}^{AB} \otimes \mathcal{D})(\omega^{ABC^n})$$

as the corresponding state after decoding, the entanglement fidelity of the code is given by

$$F_{AB} = F(|\Phi_1\rangle^{A\hat{A}} |\Phi_2\rangle^{B\hat{B}}, \Omega^{AB\hat{A}\hat{B}}) \geq 1 - \epsilon_n \quad (28)$$

where  $|\Phi_1\rangle^{A\hat{A}}$  and  $|\Phi_2\rangle^{B\hat{B}}$  are the maximally entangled target states. The sum rate can be bounded as

$$\begin{aligned} I_c(AB)C^n)_\omega & \geq I_c(AB) \hat{A}\hat{B})_\Omega \\ & \geq I_c(AB) \hat{A}\hat{B})_{\Phi_1 \otimes \Phi_2} \\ & \quad - 2H(\epsilon_n) - 8n(R+S)\sqrt{\epsilon_n} \\ & \geq n(R+S) - n\epsilon'_n. \end{aligned}$$

The first step is by the data processing inequality. The second step uses Lemma 3 and (11), along with monotonicity applied to (28). The last step has defined  $\epsilon'_n = \frac{2}{n} - 8(R+S)\sqrt{\epsilon_n}$  and holds because the binary entropy  $H(\cdot)$  is upper bounded by 1. We can bound Alice's rate  $R$  by writing

$$\begin{aligned} I_c(A)BC^n)_\omega & \geq I_c(A)C^n)_\omega \\ & \geq I_c(A) \hat{A}\hat{B})_\Omega \\ & \geq I_c(A) \hat{A})_\Omega \\ & \geq I_c(A) \hat{A})_{\Phi_1} - 2H(\epsilon_n) - 8nR\sqrt{\epsilon_n} \\ & \geq nR - n\epsilon'_n. \end{aligned}$$

The first three steps above are by data processing [12]. The remaining steps hold for the same reasons as in the previous chain of inequalities. Similarly, Bob's rate also must satisfy

$$nS \leq I_c(B)AC^n)_\omega + n\epsilon'_n.$$

Since  $\epsilon_n \rightarrow 0$  implies  $\epsilon'_n \rightarrow 0$ , this means that for every  $\delta > 0$ , any achievable qq rate pair  $(R, S)$  must satisfy

$$(R - \delta, S - \delta) \in \frac{1}{n} \mathcal{Q}^{(1)}(\mathcal{N}^{\otimes n}) \subseteq \mathcal{Q}(\mathcal{N}).$$

Since  $\mathcal{Q}(\mathcal{N})$  is closed by definition, this completes the proof.  $\square$

*Remark:* Strictly speaking, the pair of nonnegative rates  $(R, S)$  needs to be contained in some pentagon whose corner points  $(\frac{1}{k} I_c(A)C^k)_\sigma, \frac{1}{k} I_c(B)AC^k)_\sigma$  and  $(\frac{1}{k} I_c(A)BC^k)_\sigma, \frac{1}{k} I_c(B)C^k)_\sigma$  are located in the upper right quadrant of  $\mathbb{R}^2$ , where  $\sigma^{ABC^k}$  is some state of the form (8). For large enough  $n$ , the induced states  $\omega^{ABC^n}$  in the above proof fulfill this role. To see this, note that an artifact of the steps that

upper bound Alice's rate  $R$  is that  $\frac{1}{n}I_c(A)BC^m) \geq R - \epsilon_n$  and  $I_c(A)C^m) \geq R - \epsilon_n$ . Since  $\epsilon_n \rightarrow 0$ , the right-hand sides are eventually positive whenever  $R > 0$ . The similar steps that bound Bob's rate complete the argument.

#### D. Proof of Theorem 2 (Achievability)

Fix bipartite pure states  $|\Psi_1\rangle^{A''A'}$  and  $|\Psi_2\rangle^{B''B'}$  that give rise to the state

$$\omega^{A''B''C} = (\mathbb{1}^{A''B''} \otimes \mathcal{N}^{\otimes n})(\Psi_1^{A''A'} \otimes \Psi_2^{B''B'})$$

and define

$$\rho_1^{A'} = \text{Tr}_{A''}\Psi_1^{A''A'} \quad \rho_2^{B'} = \text{Tr}_{B''}\Psi_2^{B''B'}$$

Letting  $\epsilon, \delta > 0$  be arbitrary, we will show that there exists a  $(2^{nR}, 2^{nS}, n, \epsilon)$  qq entanglement transmission code where

$$R = I_c(A''\rangle C)_\omega - \delta \quad \text{and} \quad S = I_c(B''\rangle A''C)_\omega - \delta$$

provided that  $R, S \geq 0$ . Note that the rates in Theorem 2 will be implied by taking the channel to be  $\mathcal{N}^{\otimes k}$ , with  $\omega^{ABC^k}$  defined similarly.

We first choose an isometric extension  $\mathcal{U}_\mathcal{N} : A'B' \rightarrow CE$  of  $\mathcal{N}$ . Define the ideal channel  $\mathcal{N}_1 : A' \rightarrow C$ , which would effectively be seen by Alice were Bob's average code density operator exactly equal to  $(\rho_2^{B'})^{\otimes n}$  as

$$\mathcal{N}_1 : \tau^{A'} \mapsto \mathcal{N}(\tau^{A'} \otimes \rho_2^{B'}).$$

We now use  $\mathcal{U}_\mathcal{N}$  to define a particular isometric extension  $\mathcal{U}_{\mathcal{N}_1} : A' \rightarrow CE'$  of  $\mathcal{N}_1$ , where  $E' = B''E$ , as

$$\mathcal{U}_{\mathcal{N}_1}|\psi\rangle^{A'} = \mathcal{U}_\mathcal{N}|\psi\rangle^{A'}|\Psi_2\rangle^{B''B'}.$$

Observe that Bob's fake input  $B''$  is treated as part of the environment of Alice's ideal induced channel. We then further define the channel  $\mathcal{N}_2 : B' \rightarrow A''C$  by

$$\mathcal{N}_2 : \tau^{B'} \mapsto \mathcal{N}(\Psi_1^{A''A'} \otimes \tau^{B'}).$$

In contrast to the interpretation of  $\mathcal{N}_1$ , this may be viewed as the channel that would be seen by Bob if Alice were to input the  $A'$  part of the purification  $|\Psi_2\rangle^{A''A'}$  of  $\rho_2^{A'}$  to her input of the channel and then send the  $A''$  system to Charlie via a noiseless quantum channel. A pictorial depiction of the maps  $\mathcal{N}_1$  and  $\mathcal{N}_2$  is given in Fig. 2. As in the proof of Theorem 1, Charlie will first decode Alice's information, after which he will attempt to simulate the channel  $\mathcal{N}_2$ , allowing a higher transmission rate for Bob than if Alice's information was treated as noise. Since quantum information cannot be copied, showing that this is indeed possible will require different techniques than were utilized in the previous coding theorem. Although ensembles of random codes will be used in this proof, we introduce the technique of *coherent coding*, in which we pretend that the common randomness is purified. The main advantage of this approach will be that working with states in the enlarged Hilbert space allows monotonicity to be easily exploited in order to provide the estimates we require. Additionally, before we derandomize at the end of the proof, it will ultimately be only Bob who is using a random code. Alice

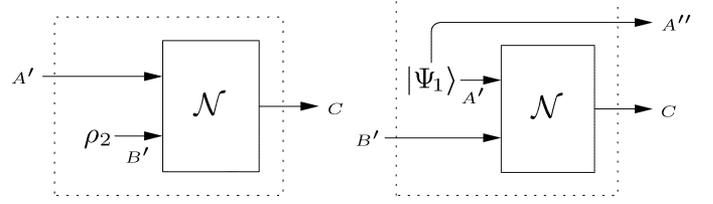


Fig. 2. Structure of  $\mathcal{N}_1$  (left) and  $\mathcal{N}_2$  (right).

will be able to use any deterministic code from her random ensemble, as Charlie will implement a decoding procedure that produces a global state which is close to that which would have been created had Alice coded with the coherent randomness. To show this, we will first analyze the state that would result if both senders used their full ensembles of codes. Then, we show that if Alice uses any code from her ensemble, Charlie can create the proper global state himself, allowing him to effectively simulate  $\mathcal{N}_2$  and ultimately decode both states at the desired rates. The overall structure of the decoder we will construct is illustrated in Fig. 3.

By Proposition 1, for large enough  $n$ , there exists a  $(2^{nR}, n, \epsilon)$  random entanglement transmission code  $(p_\ell, \mathcal{E}_1^\ell, \mathcal{D}_1^\ell)$  for the channel  $\mathcal{N}_1$ , where

$$R = I_c(\rho_1^{A'}, \mathcal{N}_1) - \delta = I_c(A''\rangle C) - \delta.$$

There similarly exists a  $(2^{nS}, n, \epsilon)$  random entanglement transmission code  $(q_m, \mathcal{E}_2^m, \mathcal{D}_2^m)$  for  $\mathcal{N}_2$ , with

$$S = I_c(\rho_2^{B'}, \mathcal{N}_2) - \delta = I_c(B''\rangle A''C) - \delta.$$

Proposition 1 further guarantees that these codes can be chosen so that their respective average code density operators

$$\varrho_1^{A'^n} = \sum_\ell p_\ell \mathcal{E}_1^\ell(\pi^{\tilde{A}}) \quad \text{and} \quad \varrho_2^{B'^n} = \sum_m q_m \mathcal{E}_2^m(\pi^{\tilde{B}})$$

satisfy

$$\|\varrho_1^{A'^n} - (\rho_1^{A'})^{\otimes n}\|_1 \leq \epsilon \quad \|\varrho_2^{B'^n} - (\rho_2^{B'})^{\otimes n}\|_1 \leq \epsilon \quad (29)$$

and also that there exist isometric extensions  $\mathcal{U}_{\mathcal{D}_1^\ell} : C^n \rightarrow \hat{A}F$  implementing the  $\mathcal{D}_1^\ell$  from Alice's random code, which satisfy

$$F\left(|\Phi_1\rangle^{A\hat{A}}|\lambda\rangle^{FE^n}, \mathcal{U}_{\mathcal{D}_1^\ell} \circ \mathcal{U}_{\mathcal{N}_1}^{\otimes n} \circ \mathcal{E}_1^\ell|\Phi_1\rangle^{A\hat{A}}\right) \geq 1 - \epsilon \quad (30)$$

for every random code index  $\ell$  and the same fixed state  $|\lambda\rangle^{FE^n}$ .

Let the code common randomness between Alice and Charlie be held between the systems  $L_A$  and  $L_C$ , represented by the state

$$\gamma_1^{L_ALC} = \sum_\ell p_\ell |\ell\rangle\langle\ell|^{L_A} \otimes |\ell\rangle\langle\ell|^{L_C}$$

defining a similar state  $\gamma_2^{M_B M_C}$  for the Bob-Charlie common randomness. For convenience, we define a purification

$$|\Gamma_1\rangle^{L_E L_A L_B} = \sum_\ell \sqrt{p_\ell} |\ell\rangle^{L_E} |\ell\rangle^{L_A} |\ell\rangle^{L_C}$$

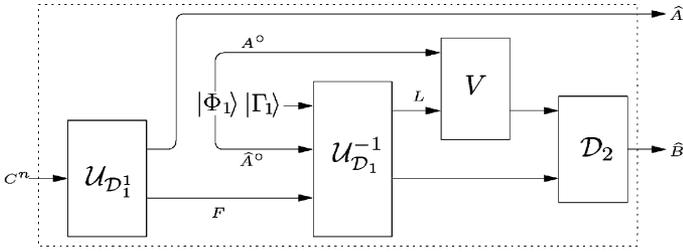


Fig. 3. Structure of decoding  $\mathcal{D}$ .

of  $\gamma_1^{L_A L_C}$ , defining a similar purification  $|\Gamma_2\rangle^{M_E M_B M_C}$  of  $\gamma_2^{M_B M_C}$ . Let us define the controlled encoding isometries  $\mathcal{E}_1: L_A \hat{A} \rightarrow L_A A^n$  and  $\mathcal{E}_2: M_B \hat{B} \rightarrow M_B B^n$  as

$$\mathcal{E}_1 = \sum_{\ell} |\ell\rangle\langle\ell|^{L_A} \otimes \mathcal{E}_1^{\ell} \quad \text{and} \quad \mathcal{E}_2 = \sum_m |m\rangle\langle m|^{M_B} \otimes \mathcal{E}_2^m.$$

Abbreviating  $L = L_E L_A L_C$  and  $M = M_E M_B M_C$ , the states that would arise if Alice and Bob each encoded *coherently* are

$$|\Upsilon_1\rangle^{L A A^n} \equiv \mathcal{E}_1 |\Gamma_1\rangle^L |\Phi_1\rangle^{A \hat{A}} = \sum_{\ell} \sqrt{p_{\ell}} |\ell\rangle^L \mathcal{E}_1^{\ell} |\Phi_1\rangle^{A \hat{A}}$$

$$|\Upsilon_2\rangle^{M B B^n} \equiv \mathcal{E}_2 |\Gamma_2\rangle^M |\Phi_2\rangle^{B \hat{B}} = \sum_m \sqrt{q_m} |m\rangle^M \mathcal{E}_2^m |\Phi_2\rangle^{B \hat{B}}.$$

Because  $|\Upsilon_1\rangle^{L A A^n}$  and  $|\Upsilon_2\rangle^{M B B^n}$  are, respectively, purifications of  $\varrho_1^{A^n}$  and  $\varrho_2^{B^n}$ , together with (29), Uhlmann's theorem tells us that there exist unitaries  $V_1: L A \rightarrow A^n$  and  $V_2: M B \rightarrow B^n$  such that

$$F(V_i |\Upsilon_i\rangle, |\Psi_i\rangle^{\otimes n}) \geq 1 - \epsilon. \quad (31)$$

Further, define a corresponding controlled isometric decoder  $\mathcal{U}_{\mathcal{D}_1}: L_C C^n \rightarrow L_C \hat{A} F$  for Alice's code as

$$\mathcal{U}_{\mathcal{D}_1} = \sum_{\ell} |\ell\rangle\langle\ell|^{L_C} \otimes \mathcal{U}_{\mathcal{D}_1}^{\ell}.$$

Let us now imagine that each of Alice and Bob encodes using the coherent common randomness, resulting in a global pure state  $\mathcal{U}_{\mathcal{N}}^{\otimes n} |\Upsilon_1\rangle^{L A A^n} |\Upsilon_2\rangle^{M B B^n}$  on  $L A M B C^n E^n$ . If Charlie then applies the full controlled decoder from Alice's code, the resulting global pure state would be

$$|\Theta\rangle^{L A \hat{A} M B F E^n} = \mathcal{U}_{\mathcal{D}_1} \circ \mathcal{U}_{\mathcal{N}}^{\otimes n} |\Upsilon_1\rangle^{L A A^n} |\Upsilon_2\rangle^{M B B^n}.$$

For each  $\ell$ , let us define an isometry  $\mathcal{O}^{\ell}: B^n \rightarrow A \hat{A} F E^n$  as

$$\mathcal{O}^{\ell} = \mathcal{U}_{\mathcal{D}_1}^{\ell} \circ \mathcal{U}_{\mathcal{N}}^{\otimes n} \circ \mathcal{E}_1^{\ell} (\Phi_1^{A \hat{A}} \otimes \cdot)$$

which we use to define the pure states

$$|\theta_{\ell}\rangle^{A \hat{A} M B F E^n} = \mathcal{O}^{\ell} |\Upsilon_2\rangle^{M B B^n}.$$

These definitions allow us to express

$$|\Theta\rangle^{L A \hat{A} M B F E^n} = \sum_{\ell} \sqrt{p_{\ell}} |\ell\rangle^L |\theta_{\ell}\rangle^{A \hat{A} M B F E^n}.$$

Further writing  $|\lambda'\rangle^{F M B E^n} \equiv V_2^{-1} |\lambda\rangle^{F B^n E^n}$ , we have

$$\begin{aligned} & F\left(|\Phi_1\rangle^{A \hat{A}} |\lambda'\rangle^{F M B E^n}, |\theta_{\ell}\rangle^{A \hat{A} M B F E^n}\right) \\ &= F\left(|\Phi_1\rangle^{A \hat{A}} |\lambda'\rangle^{F M B E^n}, \mathcal{O}^{\ell} |\Upsilon_2\rangle^{M B B^n}\right) \\ &= F\left(|\Phi_1\rangle^{A \hat{A}} |\lambda\rangle^{F B^n E^n}, V_2 \circ \mathcal{O}^{\ell} |\Upsilon_2\rangle^{M B B^n}\right) \\ &\geq 1 - 2\sqrt{1 - F(V_2 |\Upsilon_2\rangle^{M B B^n}, (|\Psi_2\rangle^{B^n B^n})^{\otimes n})} \\ &\quad - 2\sqrt{1 - F(|\Phi_1\rangle^{A \hat{A}} |\lambda\rangle^{F B^n E^n}, \mathcal{O}^{\ell} (|\Psi_2\rangle^{B^n B^n})^{\otimes n})} \\ &\geq 1 - 2\sqrt{\epsilon} \\ &\quad - 2\sqrt{1 - F(|\Phi_1\rangle^{A \hat{A}} |\lambda\rangle^{F E^n}, \mathcal{U}_{\mathcal{D}_1}^{\ell} \circ \mathcal{U}_{\mathcal{N}_1}^{\otimes n} \circ \mathcal{E}_1^{\ell} |\Phi_1\rangle^{A \hat{A}})} \\ &\geq 1 - 4\sqrt{\epsilon}. \end{aligned}$$

In the above, the first inequality is by the triangle inequality, the fact that fidelity is isometrically invariant, while for the second inequality, we have used (31) for the first term and rewritten the second. The last bound is from (30). Observe that we are still free to specify the global phases of the outputs of the  $\mathcal{U}_{\mathcal{D}_1}^{\ell}$ , further implying that  $\langle\theta_{\ell}||\Phi_1\rangle^{A \hat{A}}|\lambda'\rangle \geq \sqrt{1 - 4\sqrt{\epsilon}}$  for each  $\ell$ . Consequently

$$\begin{aligned} & F(|\Theta\rangle^{L A \hat{A} M B F E^n}, |\Gamma_1\rangle^L |\Phi_1\rangle^{A \hat{A}} |\lambda'\rangle^{M B F E^n}) \\ &= \left| \sum_{\ell \ell'} \sqrt{p_{\ell} p_{\ell'}} \langle\ell||\ell'\rangle \langle\theta_{\ell}||\Phi_1\rangle^{A \hat{A}}|\lambda'\rangle \right|^2 \\ &= \left| \sum_{\ell} p_{\ell} \langle\theta_{\ell}||\Phi_1\rangle^{A \hat{A}}|\lambda'\rangle \right|^2 \\ &\geq 1 - 4\sqrt{\epsilon} \end{aligned}$$

so that essentially, the subsystems  $L$ ,  $A \hat{A}$  and  $M B F E^n$  of  $|\Theta\rangle$  are mutually decoupled.

As mentioned earlier, it will be sufficient for Alice to use *any* deterministic code from the random ensemble to encode. Without loss of generality, we assume that Alice chooses to use the first code ( $\ell = 1$ ) in her ensemble. Bob, on the other hand, will need to use randomness to ensure that Alice's effective channel is close to a product channel. The state on  $A M B C^n$  that results from these encodings is  $\mathcal{N}^{\otimes n} (\mathcal{E}_1^1 (\Phi_1^{A \hat{A}}) \otimes \Upsilon_2^{M B B^n})$ .

We will now describe a procedure by which Charlie first decodes Alice's information, then produces a state that is close to  $\Theta$ , making it look like Alice had in fact utilized the coherent coding procedure. This will allow Charlie to apply local unitaries to effectively simulate the channel  $\mathcal{N}_2$  for which Bob's random code was designed, enabling him to decode Bob's information as well. These steps will constitute Charlie's decoding  $\mathcal{D}: M_C C^n \rightarrow M_C \hat{A} \hat{B}$ , which depends on the Bob-Charlie common randomness. The existence of a deterministic decoder will then be inferred.

Charlie first applies the isometric decoder  $\mathcal{U}_{\mathcal{D}_1}^1$ , placing all systems into the state  $|\theta_1\rangle^{A \hat{A} M B F E^n}$ . He then removes his local system  $\hat{A}$  (it is important that he keep  $\hat{A}$  in a safe place, as

it represents the decoder output for Alice's quantum information) and replaces it with the corresponding parts of the locally prepared pure state  $|\Phi_1\rangle^{A^\circ\hat{A}^\circ}$ . Charlie also locally prepares the state  $|\Gamma_1\rangle^L$ . The resulting state

$$\Theta'^{LA\hat{A}MBFE^n} = \Gamma_1^L \otimes \Phi_1^{A^\circ\hat{A}^\circ} \otimes \theta_1^{MBFE^n}$$

satisfies

$$\begin{aligned} F(\Theta', \Theta) &\geq 1 - \left\| \theta_1^{MBFE^n} - \lambda'^{MBFE^n} \right\|_1 \\ &\quad - \left\| \lambda'^{MBFE^n} - \Theta^{MBFE^n} \right\|_1 \\ &\quad - 3 \left( 1 - F(|\Gamma_1\rangle^L |\Phi_1\rangle^{A\hat{A}}, \Theta^{LA\hat{A}}) \right) \\ &\geq 1 - 2\sqrt{4\sqrt{\epsilon}} - 2\sqrt{4\sqrt{\epsilon}} - 12\sqrt{\epsilon} \\ &\geq 1 - 9\epsilon^{1/4} \end{aligned} \quad (32)$$

whenever  $\epsilon \leq 12^{-4}$ . The first inequality combines Lemma 2 and the triangle inequality. The first two estimates in the second inequality are from applying (11) and monotonicity with respect to  $\text{Tr}_{A\hat{A}}$  and  $\text{Tr}_{LA\hat{A}}$  to the previous two estimates. The last estimate in that line is from monotonicity with respect to the map  $\text{Tr}_{MBFE^n}$  applied to the previous estimate. Next, Charlie will apply

$$V_1 \circ \mathcal{U}_{D_1}^{-1}: LA^\circ\hat{A}^\circ F \rightarrow A'^m C^m$$

to  $\Theta'$ , in order to simulate the channel  $\mathcal{N}_2$ . Note that this operation only acts on Charlie's local systems, as pictured in Fig. 3. To see that this will work, define

$$\mathcal{M}: LA\hat{A}FE^n \rightarrow A'^m C^m$$

as  $\mathcal{M} \equiv \text{Tr}_{E^n} V_1 \circ \mathcal{U}_{D_1}^{-1}$  and observe that by monotonicity with respect to  $\mathcal{N}_2^{\otimes n}(\cdot \otimes \Upsilon_2^{M\hat{B}\hat{B}'^n})$  and (31), the states on  $MBA'^m C^m$  satisfy

$$\begin{aligned} &F\left(\mathcal{M}(\Theta), \mathcal{N}_2^{\otimes n}(\Upsilon_2^{M\hat{B}\hat{B}'^n})\right) \\ &= F\left(V_1 \circ \mathcal{N}^{\otimes n}(\Upsilon_1^{LAA'^n} \otimes \Upsilon_2^{M\hat{B}\hat{B}'^n}), \right. \\ &\quad \left. \mathcal{N}^{\otimes n}((\Psi_1^{AA'})^{\otimes n} \otimes \Upsilon_2^{M\hat{B}\hat{B}'^n})\right) \\ &\geq F\left(V_1 |\Upsilon_1\rangle^{LAA'^n}, (|\Psi_1\rangle^{A'A'})^{\otimes n}\right) \\ &\geq 1 - \epsilon. \end{aligned}$$

We may now use the triangle inequality and monotonicity with respect to  $\mathcal{M}$  to combine our last two estimates, yielding

$$\begin{aligned} &F\left(\mathcal{M}(\Theta'), \mathcal{N}_2^{\otimes n}(\Upsilon_2^{M\hat{B}\hat{B}'^n})\right) \\ &\geq 1 - 2\sqrt{1 - F(\mathcal{M}(\Theta'), \mathcal{M}(\Theta))} \\ &\quad - 2\sqrt{1 - F(\mathcal{M}(\Theta), \mathcal{N}_2^{\otimes n}(\Upsilon_2^{M\hat{B}\hat{B}'^n}))} \\ &\geq 1 - 2\sqrt{9\epsilon^{1/4}} - 2\sqrt{\epsilon} \\ &\geq 1 - 7\epsilon^{1/8} \end{aligned} \quad (33)$$

whenever  $\epsilon \leq 2^{-8/3}$ . We have thus far shown that Charlie's decoding procedure succeeds in simulating the channel  $\mathcal{N}_2^{\otimes n}$ , while simultaneously recovering Alice's quantum information. Charlie now uses the controlled decoder

$$\mathcal{D}_2: M_C A'^m C^m \rightarrow M_C \hat{B}$$

defined as

$$\mathcal{D}_2 = \sum_m |m\rangle\langle m|^{M_C} \otimes \mathcal{D}_2^m$$

to decode Bob's quantum information. This entire procedure has defined our decoder

$$\mathcal{D}: M_C C^n \rightarrow M_C \hat{A}\hat{B}$$

which gives rise to a global state  $\Omega^{A\hat{A}B\hat{B}}$  representing the final output state of the protocol, averaged over Bob's common randomness. This state satisfies

$$\begin{aligned} F(|\Phi_1\rangle^{A\hat{A}}, \Omega^{A\hat{A}}) &\geq F(\Theta, \Theta') \\ &\geq 1 - 9\epsilon^{1/4} \end{aligned}$$

because of monotonicity with respect to  $\text{Tr}_{LMBFE^n}$  applied to the bound (32). By using the triangle inequality, the fact that (21) is satisfied for each  $m$ , and monotonicity of the estimate (33) with respect to  $\text{Tr}_M \mathcal{D}_2$ , the global state can further be seen to obey

$$\begin{aligned} &F(|\Phi_2\rangle, \Omega^{B\hat{B}}) \\ &= F(|\Phi_2\rangle^{B\hat{B}}, \text{Tr}_M \mathcal{D}_2 \circ \mathcal{M}(\Theta')) \\ &\geq 1 - 2\sqrt{1 - F(|\Phi_2\rangle^{B\hat{B}}, \text{Tr}_M \mathcal{D}_2 \circ \mathcal{N}_2^{\otimes n}(\Upsilon_2^{M\hat{B}\hat{B}'^n}))} \\ &\quad - 2\sqrt{1 - F(\text{Tr}_M \mathcal{D}_2 \circ \mathcal{N}_2^{\otimes n}(\Upsilon_2^{M\hat{B}\hat{B}'^n}), \text{Tr}_M \mathcal{D}_2 \circ \mathcal{M}(\Theta'))} \\ &\geq 1 - 2\sqrt{\epsilon} - 2\sqrt{7\epsilon^{1/8}} \\ &\geq 1 - 7\epsilon^{1/16} \end{aligned}$$

as long as  $\epsilon \leq 2^{-16/7}$ . Along with (11), a final application of Lemma 2 combines the above two bounds to give

$$\begin{aligned} &F(|\Phi_1\rangle^{A\hat{A}} |\Phi_2\rangle^{B\hat{B}}, \Omega^{A\hat{A}B\hat{B}}) \\ &\geq 1 - \left\| \Phi_1^{A\hat{A}} - \text{Tr}_{B\hat{B}} \Omega^{A\hat{A}B\hat{B}} \right\|_1 \\ &\quad - 3 \left( 1 - F(|\Phi_2\rangle^{B\hat{B}}, \text{Tr}_{A\hat{A}} \Omega^{A\hat{A}B\hat{B}}) \right) \\ &\geq 1 - 2\sqrt{9\epsilon^{1/4}} - 21\epsilon^{1/16} \\ &\geq 1 - 22\epsilon^{1/16} \end{aligned}$$

provided that  $\epsilon \leq 6^{-16}$ . Since this estimate represents an average over Bob's common randomness, there must exist a particular value  $m^*$  of the common randomness so that the corresponding deterministic code is at least as good as the random one, thus concluding the coding theorem.  $\square$

## V. STRONG SUBSPACE TRANSMISSION AND SCENARIO EQUIVALENCES

### A. Scenario III—Strong Subspace Transmission

The criteria of scenarios I and II, both in the cq and qq cases, are directly analogous to the requirement in classical information theory that the average probability of error, averaged over all codewords, be small. Here, we introduce a situation analogous to the stronger classical condition that the *maximal* probability of error be small, or that the probability of error for *each* pair of codewords be small. There are examples of classical multiple-access channels for which, when each encoder is a deterministic function from the set of the messages to the set of input symbols, the maximal error capacity region is *strictly* smaller than the average error region [26]. However, it is known that if stochastic encoders are allowed (see [27, Problem 3.2.4]), the maximal and average error capacity regions are equal.

It is well known that randomization is not necessary for such an equivalence to hold for single-user channels, as Markov's inequality implies that a fraction of the codewords with the worst probability of error can be purged, while incurring a negligible loss of rate. The obstacle to utilizing such an approach for classical multiple-access channels, and hence for quantum ones as well, is that there is no guarantee that a large enough subset of bad pairs of codewords decomposes as the product of subsets of each sender's codewords.

As mentioned in Section II, a particularly attractive feature of the following two scenarios is their *composability*; when combined with other protocols satisfying analogous criteria, the joint protocol will satisfy similar properties.

*III—Classical-Quantum Scenario:* Strong subspace transmission can be considered a more ambitious version of entanglement transmission, whereby rather than requiring Bob to transmit half of a maximally entangled state  $|\Phi\rangle^{B\tilde{B}}$ , it is instead required that he faithfully transmit the  $\tilde{B}$  part, presented to him, of *any* bipartite pure state  $|\Psi\rangle^{R\tilde{B}}$ , where  $|R|$  can be any finite number. The reader should note that this constitutes a generalization of the usual subspace transmission [28], as whenever  $|\Psi\rangle^{R\tilde{B}} = |\psi\rangle^R|\varphi\rangle^{\tilde{B}}$ , this amounts to requiring that  $|\varphi\rangle^{\tilde{B}}$  be transmitted faithfully. We further demand that the maximal error probability for the classical messages be small.

As with entanglement transmission, Alice will send classical information at rate  $r$  by preparing one of  $2^{nr}$  pure states  $\{|\phi_m\rangle^{A^{nr}}\}_{m \in 2^{nr}}$ . As previously discussed, our more restrictive information transmission constraints can only be met by allowing Alice to employ a stochastic encoding. We assume that Alice begins by generating some randomness, modeled by the random variable  $X$ . To send message  $M = m$ , she prepares the state  $|f(m)\rangle \equiv |\phi_{f(m)}\rangle$ , where  $f(m) \equiv f_X(m)$  is a random encoding function, depending on the randomness in  $X$ .

Bob will apply an encoding  $\mathcal{E}: \tilde{B} \rightarrow B'^n$ , and Charlie will employ a decoding instrument  $\mathcal{D}: C^n \rightarrow \tilde{M}\tilde{B}$ . These will be constructed by adding an additional layer of processing on top of the entanglement transmission codes that were proved to exist in the previous section. The success probability for the protocol,

conditioned on  $m$  being sent and  $|\Psi\rangle^{R\tilde{B}}$  being presented, can be expressed as

$$P_s^{\text{III}}(m, |\Psi\rangle^{R\tilde{B}}) = F\left(|f(m)\rangle^{\tilde{M}}|\Psi\rangle^{R\tilde{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\phi_{f(m)}^{A^{nr}} \otimes \mathcal{E}(\Psi^{R\tilde{B}}))\right).$$

We will say that  $(f, X, \{|\phi_m\rangle\}_{m \in 2^{nr}}, \mathcal{E}, \mathcal{D})$  is a  $(2^{nr}, 2^{nS}, n, \epsilon)$  cq strong subspace transmission code for the channel  $\mathcal{N}$  if, for every  $m \in 2^{nr}$  and every  $|\Psi\rangle^{R\tilde{B}}$

$$\mathbb{E}_X P_s^{\text{III}}(m, |\Psi\rangle^{R\tilde{B}}) \geq 1 - \epsilon. \quad (34)$$

The rate pair  $(r, S)$  is an achievable cq rate pair for strong subspace transmission if there is a sequence of  $(2^{nr}, 2^{nS}, n, \epsilon_n)$  cq random strong subspace transmission codes with  $\epsilon_n \rightarrow 0$ , and the capacity region  $\mathcal{C}\mathcal{Q}_{\text{III}}(\mathcal{N})$  is closure of the collection of all such achievable rates.

*III—Quantum-Quantum Scenario:* This scenario is the obvious combination of the relevant concepts from the previous scenario and the qq entanglement transmission scenario. Alice and Bob are, respectively, presented with the  $\tilde{A}$  and  $\tilde{B}$  parts of some pure bipartite states  $|\Psi_1\rangle^{Q\tilde{A}}$  and  $|\Psi_2\rangle^{R\tilde{B}}$ . As before, we place no restriction on  $|Q|$  and  $|R|$ , other than that they are finite. They employ their respective encodings  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , while Charlie decodes with  $\mathcal{D}$ . As in the above cq case, the structure of these maps will be more complicated than in the previous two scenarios.  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$  is then a  $(2^{nR}, 2^{nS}, n, \epsilon)$  qq strong subspace transmission code if

$$F\left(|\Psi_1\rangle^{Q\tilde{A}}|\Psi_2\rangle^{R\tilde{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n} \circ (\mathcal{E}_1 \otimes \mathcal{E}_2)(\Psi_1^{Q\tilde{A}} \otimes \Psi_2^{R\tilde{B}})\right) \geq 1 - \epsilon \quad (35)$$

for every pair of pure bipartite states  $|\Psi_1\rangle^{Q\tilde{A}}$  and  $|\Psi_2\rangle^{R\tilde{B}}$ . Achievable rates and the capacity region  $\mathcal{Q}_{\text{III}}(\mathcal{N})$  are defined as in the cq case.

### B. Entanglement Transmission Implies Entanglement Generation

1) *Proof That  $\mathcal{C}\mathcal{Q}_{\text{II}} \subseteq \mathcal{C}\mathcal{Q}_{\text{I}}$ :* Suppose there exists a  $(2^{nr}, 2^{nS}, n, \epsilon)$  cq entanglement transmission code consisting of classical message states  $\{|\phi_m\rangle^{A^{nr}}\}_{m \in 2^{nr}}$ , an encoding map  $\mathcal{E}: \tilde{B} \rightarrow B'^n$ , and a decoding instrument  $\mathcal{D}: C^n \rightarrow \tilde{M}\tilde{B}$ . Write any pure state decomposition of the encoded state

$$(1^B \otimes \mathcal{E})(\Phi^{B\tilde{B}}) = \sum_i p_i |\Upsilon_i\rangle\langle \Upsilon_i|^{BB'^n}.$$

Then, the success condition (3) for a cq entanglement transmission code can be rewritten as

$$\begin{aligned} 1 - \epsilon &\leq 2^{-nr} \sum_{m=1}^{2^{nr}} P_s^{\text{II}}(m) \\ &= 2^{-nr} \sum_{m=1}^{2^{nr}} P_s^{\text{I}}\left(m, \sum_i p_i \Upsilon_i^{BB'^n}\right) \\ &= \sum_i p_i \left(2^{-nr} \sum_{m=1}^{2^{nr}} P_s^{\text{I}}(m, |\Upsilon_i\rangle^{BB'^n})\right) \end{aligned}$$

so that there is a particular value  $i^*$  of  $i$  for which

$$2^{-nr} \sum_{m=1}^{2^{nr}} P_s^I(m, |\Upsilon_{i^*}\rangle^{BB'^n}) \geq 1 - \epsilon.$$

Hence,  $(\{|\phi_m\rangle^{A^n}\}_{m \in 2^{nr}}, |\Upsilon_{i^*}\rangle^{BB'^n}, \mathcal{D})$  comprises an  $(2^{nr}, 2^{nS}, n, \epsilon)$  cq entanglement generation code.  $\square$

2) *Proof That  $\mathcal{Q}_{\text{II}} \subseteq \mathcal{Q}_{\text{I}}$ :* Suppose there exists a  $(2^{nR}, 2^{nS}, n, \epsilon)$  entanglement transmission code  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$  that transmits the maximally entangled states  $|\Phi_1\rangle^{AA'}, |\Phi_2\rangle^{BB'}$ . As in the cq case, the encoded states can be decomposed as

$$(\mathbf{1}^A \otimes \mathcal{E}_1)(\Phi_1^{AA'}) = \sum_i p_i \Upsilon_{1i}^{AA'}$$

and

$$(\mathbf{1}^B \otimes \mathcal{E}_2)(\Phi_2^{BB'}) = \sum_j q_j \Upsilon_{2j}^{BB'}$$

The reliability condition (6) can then be rewritten as

$$\sum_{ij} p_i q_j F(|\Phi_1\rangle^{AA'} |\Phi_2\rangle^{BB'}, \mathcal{D} \otimes \mathcal{N}^{\otimes n}(\Upsilon_{1i}^{AA'} \otimes \Upsilon_{2j}^{BB'})) \geq 1 - \epsilon$$

which implies the existence of a particular pair  $(i^*, j^*)$  of values of  $(i, j)$  such that

$$F(|\Phi_1\rangle^{AA'} |\Phi_2\rangle^{BB'}, \mathcal{D} \otimes \mathcal{N}^{\otimes n}(\Upsilon_{1i^*}^{AA'} \otimes \Upsilon_{2j^*}^{BB'})) \geq 1 - \epsilon.$$

Hence,  $(|\Upsilon_{1i^*}\rangle^{AA'}, |\Upsilon_{2j^*}\rangle^{BB'}, \mathcal{D})$  comprises a  $(2^{nR}, 2^{nS}, n, \epsilon)$  qq entanglement generation code.  $\square$

### C. Entanglement Transmission Implies Strong Subspace Transmission

1) *Proof That  $\mathcal{CQ}_{\text{II}} \subseteq \mathcal{CQ}_{\text{III}}$ :* Suppose there exists a  $(2^{nr}, 2^{nS}, n, \epsilon^2/2)$  entanglement transmission codes with classical message states  $\{|\phi_m\rangle^{A^n}\}_{m \in 2^{nr}}$ , quantum encoding  $\mathcal{E}: \tilde{B} \rightarrow \hat{B}$ , and decoding instrument  $\mathcal{D}: C^n \rightarrow \hat{M}\hat{B}$  with trace-reducing components  $\{\mathcal{D}_m: C^n \rightarrow \hat{B}\}$ .

We will initially prove the equivalence by constructing a code that requires two independent sources of shared common randomness  $X$  and  $Y$ .  $X$  is assumed to be available to Alice and to Charlie, while  $Y$  is available to Bob and to Charlie. Then, we will argue that it is possible to eliminate the dependence on the shared randomness, by using the channel to send a negligibly small “random seed,” which can be recycled to construct a code that asymptotically achieves the same performance as the randomized one.

We begin by demonstrating how shared common randomness between Alice and Charlie allows Alice to send any message with low probability of error. Setting  $\mu = 2^{nr}$ , let the random variable  $X$  be uniformly distributed on the set  $\{1, \dots, \mu\}$ . To send message  $M = m$ , Alice computes  $m' = m + X$  modulo  $\mu$ . She then prepares the state  $|\phi_{m'}\rangle^{A^n}$  for transmission through the channel. Bob encodes the  $\tilde{B}$  part of  $|\Phi\rangle^{B\tilde{B}}$  with  $\mathcal{E}$ , and each sends appropriately through the channel. Charlie decodes as usual with the instrument  $\mathcal{D}$ . Denoting the classical output as

$\hat{M}'$ , his declaration of Alice’s message is then  $\hat{M} = \hat{M}' - X$  modulo  $\mu$ . Defining the trace-reducing maps  $\mathcal{M}_m: \tilde{B} \rightarrow \hat{B}$  by

$$\mathcal{M}_m: \tau_{\tilde{B}} \mapsto \mathcal{D}_m \circ \mathcal{N}^{\otimes n}(\phi_m^{A^n} \otimes \mathcal{E}(\tau_{\tilde{B}}))$$

and the trace-reducing average map as

$$\mathcal{M}: \tau_{\tilde{B}} \rightarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \mathcal{M}_m(\tau_{\tilde{B}})$$

we can rewrite the success criterion (3) for entanglement transmission as

$$F(|\Phi\rangle^{B\hat{B}}, \mathcal{M}(\Phi^{B\tilde{B}})) \geq 1 - \epsilon^2/2$$

which, together with (11), implies that for the identity map  $\text{id}: \tilde{B} \rightarrow \hat{B}$

$$\|(\mathcal{M} - \text{id})(\Phi^{B\tilde{B}})\|_1 \leq \epsilon. \quad (36)$$

The above randomization of the classical part of the protocol can be mathematically expressed by replacing the  $\mathcal{M}_m$  with  $\mathcal{M}_{m+X}$ . As tracing over the common randomness  $X$  is equivalent to computing the expectation with respect to  $X$ , we see that  $\mathbb{E}_X \mathcal{M}_{m+X} = \mathcal{M}$ , or rather

$$\mathbb{E}_X F(|\Phi\rangle^{B\hat{B}}, \mathcal{M}_{m+X}(\Phi^{B\tilde{B}})) = F(|\Phi\rangle^{B\hat{B}}, \mathcal{M}(\Phi^{B\tilde{B}})).$$

It is thus clear that the maximal error criterion for the randomized protocol is equal to the average criterion for the original one.

We continue by randomizing the quantum part of the classically randomized protocol. Setting  $d = 2^{nS} = |\tilde{B}|$ , let  $\{U_y\}_{y \in d^2}$  be the collection of Weyl unitaries, or generalized Pauli operators, on the  $d$ -dimensional input space. Observe that for any  $\rho$ , acting with a uniformly random choice of Weyl unitary has a completely randomizing effect, in the sense that

$$\frac{1}{d^2} \sum_{y=1}^{d^2} U_y \rho U_y^{-1} = \pi_d.$$

Let the random variable  $Y$  be uniformly distributed on  $\{1, \dots, d^2\}$ . It will be convenient to define the common randomness state

$$\Upsilon^{Y_B Y_C} = \frac{1}{d^2} \sum_{y=1}^{d^2} |y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C}$$

where the system  $Y_B$  is in the possession of Bob, while  $Y_C$  is possessed by Charlie. Define now the controlled unitaries  $\mathcal{U}_B: Y_B \tilde{B} \rightarrow Y_B \hat{B}$  and  $\mathcal{U}_C: Y_C \tilde{B} \rightarrow Y_C \hat{B}$  by

$$\mathcal{U}_B = \sum_{y=1}^{d^2} |y\rangle\langle y|^{Y_B} \otimes U_y$$

and

$$\mathcal{U}_C = \sum_{y=1}^{d^2} |y\rangle\langle y|^{Y_C} \otimes U_y^{-1}.$$

Suppose Bob is given the  $\tilde{B}$  part of an arbitrary pure state  $|\Psi\rangle^{R\tilde{B}}$ , where  $|R| < \infty$ , and Alice sends the classical message  $M = m$ . For encoding, Bob will apply  $\mathcal{E} \circ \mathcal{U}_B$  to the combined system  $\Upsilon^Y \otimes \Psi^{B\tilde{B}}$ , where we abbreviate  $Y \equiv Y_B Y_C$ . Charlie decodes with  $\mathcal{U}_C \circ \mathcal{D}$ . If  $\mathcal{M}$  were equal to the perfect quantum channel id:  $\tilde{B} \rightarrow \hat{B}$ , this procedure would result in the state

$$\frac{1}{d^2} \sum_{y=1}^{d^2} |y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C} \otimes \Psi^{R\hat{B}}.$$

Note that the common randomness is still available for reuse. Abbreviating  $|y\rangle\langle y|^Y = |y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C}$ , and also  $|\Psi_y\rangle^{R\tilde{B}} = (\mathbb{1}^R \otimes U_y)|\Psi\rangle^{R\tilde{B}}$ , we write

$$\sigma^{YR\tilde{B}} = \mathcal{U}_B(\Upsilon^Y \otimes \Psi^{R\tilde{B}}) \quad (37)$$

$$= \frac{1}{d^2} \sum_{y=1}^{d^2} |y\rangle\langle y|^Y \otimes \Psi_y^{R\tilde{B}}. \quad (38)$$

Observe that  $\sigma^{YR\tilde{B}}$  is an extension of the maximally mixed state  $\pi^{\tilde{B}}$  and can be seen to arise by storing in  $Y$  the result of a von Neumann measurement along the basis  $\{|y\rangle^{R'}\}_{y \in d^2}$  on the  $R'$  part of the pure state

$$|\Gamma\rangle^{R'R\tilde{B}} = \frac{1}{d} \sum_{y=1}^{d^2} |y\rangle^{R'} |\Psi_y\rangle^{R\tilde{B}}.$$

Since

$$\text{Tr}_{R'R} \Gamma^{R'R\tilde{B}} = \text{Tr}_{YR} \sigma^{YR\tilde{B}} = \pi^{\tilde{B}}$$

it follows that  $|\Gamma\rangle^{R'R\tilde{B}}$  is maximally entangled between  $R'R$  and  $\tilde{B}$ ; so there exists an isometry  $V: B \rightarrow R'R$  such that

$$(V \otimes \mathbb{1}^{\tilde{B}})|\Phi\rangle^{B\tilde{B}} = |\Gamma\rangle^{R'R\tilde{B}}.$$

This implies that there is a quantum operation  $\mathcal{O}: B \rightarrow YR$  satisfying  $(\mathcal{O} \otimes \mathbb{1}^{\tilde{B}})(\Phi^{B\tilde{B}}) = \sigma^{YR\tilde{B}}$ . Define the trace-reducing map  $\mathcal{T}: \tilde{B} \rightarrow \hat{B}$ , which represents the coded channel with common randomness accounted for, by

$$\mathcal{T}: \tau^{\tilde{B}} \mapsto \text{Tr}_Y \mathcal{U}_C \circ \mathcal{M} \circ \mathcal{U}_B(\Upsilon^Y \otimes \tau^{\tilde{B}}).$$

Recalling our denotation of the noiseless quantum channel id:  $\tilde{B} \rightarrow \hat{B}$ , as well as our convention that id acts as the identity on any system that is not  $\tilde{B}$ , we now bound

$$\begin{aligned} 1 - F(|\Psi\rangle^{R\tilde{B}}, \mathcal{T}(\Psi^{R\tilde{B}})) &\leq \|(\mathcal{T} - \text{id})(\Psi^{R\tilde{B}})\|_1 \\ &\leq \|(\mathcal{U}_C \circ \mathcal{M} \circ \mathcal{U}_B - \text{id})(\Upsilon^Y \otimes \Psi^{R\tilde{B}})\|_1 \\ &= \|(\mathcal{M} - \text{id}) \circ \mathcal{U}_B(\Upsilon^Y \otimes \Psi^{R\tilde{B}})\|_1 \\ &= \|(\mathcal{M} - \text{id})(\sigma^{YR\tilde{B}})\|_1 \\ &\leq \|(\mathcal{M} - \text{id})(\Phi^{B\tilde{B}})\|_1 \\ &\leq \epsilon \end{aligned}$$

where the first line is by (9) and the second by monotonicity with respect to  $\text{Tr}_Y$ . The third follows from unitary invariance of the trace. The second to last inequality is a consequence of monotonicity with respect to  $\mathcal{O}$ , while the last is by (36). Note

that by monotonicity, this implies that any density matrix  $\Omega^{R\tilde{B}}$  satisfies

$$\|\mathcal{T}(\Omega^{R\tilde{B}}) - \Omega^{R\hat{B}}\|_1 \leq \epsilon. \quad (39)$$

We have thus shown that if Alice and Charlie have access to a common randomness source of rate  $r$ , while Bob and Charlie can access one of rate  $2S$ , the conditions for strong subspace transmission can be satisfied. Next we will illustrate that by modifying our protocol, it is possible to reduce the amount of shared randomness required. Using the previous blocklength- $n$  construction, we will concatenate  $N$  such codes, where each utilizes the same shared randomness, to construct a new code with blocklength  $nN$ . For an arbitrary  $|\Psi\rangle^{R\tilde{B}^N}$ , define the commuting operations  $\{\mathcal{T}_i\}_{i \in N}$ , where  $\mathcal{T}_i: \tilde{B}_i \rightarrow \hat{B}_i$  is  $\mathcal{T}$  acting on the  $i$ th tensor factor of  $\Psi^{(N)}$ . Setting  $\xi_0^{R\tilde{B}^N} \equiv \Psi^{R\tilde{B}^N}$ , we then recursively define the density operators

$$\xi_i^{R\hat{B}_i \tilde{B}^{N-i}} = \mathcal{T}_i(\xi_{i-1}^{R\tilde{B}_i \tilde{B}^{N-i+1}}).$$

Note that

$$\xi_N^{R\hat{B}^N} = \mathcal{T}^{\otimes N}(\Psi^{R\tilde{B}^N}).$$

Identifying the  $\tilde{B}_i$  with the  $\hat{B}_i$ , it follows from (39) that

$$\|\xi_{i+1} - \xi_i\|_1 = \|\mathcal{T}_{i+1}(\xi_i) - \xi_i\|_1 \leq \epsilon.$$

Therefore, we can use the triangle inequality to estimate

$$\begin{aligned} \|\mathcal{T}^{\otimes N}(\Psi^{R\tilde{B}^N}) - \Psi^{R\hat{B}^N}\|_1 &= \|\xi_N - \xi_0\|_1 \\ &\leq \sum_{i=1}^N \|\xi_i - \xi_{i-1}\|_1 \\ &\leq N\epsilon. \end{aligned}$$

By choosing  $N = \frac{1}{\sqrt{\epsilon}}$ , it is clear that we have reduced Alice's and Bob's shared randomness rates, respectively, to  $\sqrt{\epsilon}r$  and  $2\sqrt{\epsilon}S$ , while the error on the  $N$ -blocked protocol is now  $\sqrt{\epsilon}$ . Next, we argue that by using two more blocks of length  $n$ , it is possible to simulate the shared randomness by having Alice send  $nr$  random bits  $X$  using the first block, while Bob locally prepares two copies of  $|\Phi\rangle^{B\tilde{B}}$ , written  $|\Phi\rangle^{B_1\tilde{B}_1} |\Phi\rangle^{B_2\tilde{B}_2}$ , and transmits the  $\tilde{B}_1\tilde{B}_2$  parts over the channel using both blocks. Charlie decodes each block separately, obtaining a random variable  $\hat{X}$  and the  $\hat{B}_1$  and  $\hat{B}_2$  parts of the postdecoded states  $\Omega_1^{B_1\hat{B}_1}$  and  $\Omega_2^{B_2\hat{B}_2}$ . Bob and Charlie then measure their respective parts of  $\Omega_1^{B_1\hat{B}_1} \otimes \Omega_2^{B_2\hat{B}_2}$  in some previously agreed upon orthogonal bases to obtain a simulation  $\hat{Y}^Y$  of the perfect shared randomness state which, by monotonicity and telescoping, satisfies

$$\begin{aligned} \|\Upsilon^Y - \hat{Y}^Y\|_1 &\leq \|\Phi^{B_1\hat{B}_1} \otimes \Phi^{B_2\hat{B}_2} - \Omega_1^{B_1\hat{B}_1} \otimes \Omega_2^{B_2\hat{B}_2}\|_1 \\ &\leq \epsilon^2. \end{aligned}$$

Writing  $\text{dist}(Z, Z')$  for the joint probability distribution of any pair of random variables  $(Z, Z')$ , we also have that

$$\begin{aligned} \|\text{dist}(X, X) - \text{dist}(X, \hat{X})\|_1 &= 2\Pr\{X = \hat{X}\} \\ &\leq \epsilon^2. \end{aligned}$$

By monotonicity of trace distance and the triangle inequality, using the noisy common randomness state  $\hat{\Upsilon}^Y$  increases the estimate for each block by  $2\epsilon^2$ . For identical reasons, the same increase is incurred by using the noisy common randomness  $(X, \hat{X})$ . Thus, accounting for both sources of noisy common randomness, the estimate (39) is changed to  $2\epsilon$ , provided that  $\epsilon \leq \frac{1}{4}$ . The noisy common randomness thus increases the bound on the error of the  $N$ -blocked protocol to  $2\sqrt{\epsilon}$ , while costing each of Alice and Bob a negligible rate overhead of  $\frac{2}{N+2}$  in order to seed the protocol.

The above protocol can be considered as defining an encoding map  $\mathcal{E}': \tilde{B}^N \rightarrow B^{(N+2)n}$  and decoding instrument  $\mathcal{D}: C^{(N+2)n} \rightarrow \tilde{B}^N \tilde{M}^N$ . Thus, the protocol takes a  $(2^{nr}, 2^{nS}, n, \epsilon_n)$  cq entanglement transmission code and constructs a  $(2^{n'r'}, 2^{n'S'}, n', \epsilon'_n)$  strong subspace transmission code with cq rate pair  $(r', S') = \left(\frac{r}{1+\epsilon'_n}, \frac{S}{1+\epsilon'_n}\right)$ , where  $n' = \left(2 + \frac{1}{\sqrt{\epsilon_n}}\right)n$ , and  $\epsilon'_n = 2\sqrt{\epsilon_n}$ . Now, if the rates  $(r, S)$  are achievable cq rates for entanglement transmission, there must exist a sequence of  $(2^{nr}, 2^{nS}, n, 2\epsilon_n^2)$  entanglement transmission codes with  $\epsilon_n \rightarrow 0$ . Since this means that  $\frac{1}{1+2\sqrt{\epsilon_n}}$  increases to unity, we have shown that for any  $\delta > 0$ , every rate pair  $(r - \delta, S - \delta)$  is an achievable cq rate pair for strong subspace transmission. Since the capacity regions for each scenario are defined as the closure of the achievable rates, this completes the proof.  $\square$

2) *Proof That  $\mathcal{Q}_{\text{II}} \subseteq \mathcal{Q}_{\text{III}}$ :* We will employ similar techniques as were used in the previous proof to obtain this implication. Suppose there exists a  $(2^{nR}, 2^{nS}, n, \frac{1}{2}\epsilon^2)$  qq entanglement transmission code  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$ , with  $\mathcal{E}_1: A \rightarrow A^n$ ,  $\mathcal{E}_2: \tilde{B} \rightarrow B^n$ , and  $\mathcal{D}: C^n \rightarrow \tilde{A}\tilde{B}$ . Setting  $a = |\tilde{A}| = 2^{nR}$  and  $b = |\tilde{B}| = 2^{nS}$ , define the common randomness states

$$\Upsilon_1^{X_A X_C} = \frac{1}{a^2} \sum_{x=1}^{a^2} |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_C}$$

and

$$\Upsilon_2^{Y_B Y_C} = \frac{1}{b^2} \sum_{y=1}^{b^2} |y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C}.$$

These states will be used as partial inputs to the controlled unitaries

$$\begin{aligned} \mathcal{U}_A &= \sum_{x=1}^{a^2} |x\rangle\langle x|^{X_A} \otimes U_x, \\ \mathcal{U}_C &= \sum_{x=1}^{a^2} |x\rangle\langle x|^{X_C} \otimes U_x^{-1}, \\ \mathcal{V}_B &= \sum_{y=1}^{b^2} |y\rangle\langle y|^{Y_B} \otimes V_x, \\ \mathcal{V}_C &= \sum_{y=1}^{b^2} |y\rangle\langle y|^{Y_C} \otimes V_x^{-1} \end{aligned}$$

where, as before, we have utilized the Weyl unitaries  $\{U_x\}_{x \in a^2}$  and  $\{V_y\}_{y \in b^2}$ , which, respectively, completely randomize any states on  $a$ -dimensional and  $b$ -dimensional spaces. Suppose Alice and Bob are, respectively, presented with the  $\tilde{A}$  and  $\tilde{B}$  parts of the arbitrary pure states  $|\Psi_1\rangle^{Q\tilde{A}}$  and  $|\Psi_2\rangle^{R\tilde{B}}$ . Defining the channel  $\mathcal{M}: \tilde{A}\tilde{B} \rightarrow \tilde{A}\tilde{B}$  by

$$\mathcal{M} = \mathcal{D} \circ \mathcal{N}^{\otimes n} \circ (\mathcal{E}_1 \otimes \mathcal{E}_2)$$

and defining the map  $\mathcal{T}: \tilde{A}\tilde{B} \rightarrow \tilde{A}\tilde{B}$  by

$$\mathcal{T}: \tau^{\tilde{A}\tilde{B}} \mapsto (\mathcal{U}_C \otimes \mathcal{V}_C) \circ \mathcal{M} \circ (\mathcal{U}_A \otimes \mathcal{V}_B)(\tau^{\tilde{A}\tilde{B}} \otimes \Upsilon_1^X \otimes \Upsilon_2^Y)$$

the overall joint state of the randomized protocol is given by  $\mathcal{T}(\Psi_1^{Q\tilde{A}} \otimes \Psi_2^{R\tilde{B}})$ . Abbreviating

$$|xy\rangle\langle xy|^{XY} = |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_C} \otimes |y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C}$$

and defining the pure states

$$|\Psi_x\rangle^{Q\tilde{A}} = (\mathbb{1}^Q \otimes U_x)|\Psi_1\rangle^{Q\tilde{A}} \quad |\Psi_y\rangle^{R\tilde{B}} = (\mathbb{1}^R \otimes V_y)|\Psi_2\rangle^{R\tilde{B}}$$

we write

$$\sigma^{XYQR\tilde{A}\tilde{B}} = \frac{1}{a^2 b^2} \sum_{xy} |xy\rangle\langle xy|^{XY} \otimes \Psi_x^{Q\tilde{A}} \otimes \Psi_y^{R\tilde{B}}.$$

By similar arguments as in the cq case, there exists a map  $\mathcal{O}: AB \rightarrow XYQR$  so that

$$(\mathcal{O} \otimes \mathbb{1}^{\tilde{A}\tilde{B}})(\Phi_1^{A\tilde{A}} \otimes \Phi_2^{B\tilde{B}}) = \sigma^{XYQR\tilde{A}\tilde{B}}.$$

Again, for the same reasons as in the cq case, we have

$$\begin{aligned} \|(\mathcal{T} - \text{id})(\Psi_1^{A\tilde{A}} \otimes \Psi_2^{B\tilde{B}})\|_1 &\leq \|(\mathcal{M} - \text{id})(\sigma^{XYQR\tilde{A}\tilde{B}})\|_1 \\ &\leq \|(\mathcal{M} - \text{id})(\Phi_1^{A\tilde{A}} \otimes \Phi_2^{B\tilde{B}})\|_1 \\ &\leq \epsilon. \end{aligned}$$

The rest of the proof is nearly identical to that from the previous section and is thus omitted.  $\square$

#### D. Strong Subspace Transmission Implies Entanglement Transmission

1) *Proof That  $\mathcal{C}\mathcal{Q}_{\text{III}} \subseteq \mathcal{C}\mathcal{Q}_{\text{II}}$ :* Given a strong subspace transmission code, if Alice uses any deterministic value  $x$  for her locally generated randomness  $X$ , the average classical error will be equal to the expected maximal classical error of the randomized code. Since the ability to transmit any state includes the maximally entangled case, this completes the claim.

2) *Proof That  $\mathcal{Q}_{\text{III}} \subseteq \mathcal{Q}_{\text{II}}$ :* This implication is immediate. As any states can be transmitted, this certainly includes the case of a pair of maximally entangled states.

## VI. SINGLE-LETTER EXAMPLES

Due to the regularized form of our Theorems 1 and 2, the possibility of actually computing the capacity regions seems out of reach for the time being. Here, we give some examples of channels whose capacity region does in fact admit a single-letter characterization, in the sense that no regularization is necessary. We begin by reviewing a class of single-user channels for which

a single-letter formula for the quantum capacity is known. We also prove that the coherent information over these channels is concave in the input density operator.

### A. Degradable Quantum Channels

Recall that given a channel  $\mathcal{N}: A' \rightarrow B$ , one may define a complementary channel  $\mathcal{N}_c: A' \rightarrow E$  as  $\text{Tr}_B \mathcal{U}$ , where  $\mathcal{U}: A' \rightarrow BE$  is any isometric extension of  $\mathcal{N}$ . The channel  $\mathcal{N}$  is then said to be *degradable* [29] if there exists a channel  $\mathcal{D}: B \rightarrow E$  which degrades  $\mathcal{N}$  to  $\mathcal{N}_c$ , in the sense that  $\mathcal{N}_c = \mathcal{D} \circ \mathcal{N}$ . It was shown there that if  $\mathcal{N}_1$  and  $\mathcal{N}_2$  are degradable, then so is  $\mathcal{N}_1 \otimes \mathcal{N}_2$ , while

$$I_c(\rho^{A_1 A_2}, \mathcal{N}_1 \otimes \mathcal{N}_2) \leq I_c(\rho^{A_1}, \mathcal{N}_1) + I_c(\rho^{A_2}, \mathcal{N}_2). \quad (40)$$

By induction, this implies that a single-letter description of the quantum capacity of any degradable channel is  $\max_{\rho^{A'}} I_c(\rho^{A'}, \mathcal{N})$ . We remark that an analogous property of mutual information over classical channels is responsible for Shannon's famous single-letter expression for the capacity. Another useful feature of degradable channels that parallels the classical case is the following.

*Lemma 5:* If  $\mathcal{N}$  is degradable, then  $I_c(\rho^{A'}, \mathcal{N})$  is a concave function of  $\rho^{A'}$ .

*Proof:* Fixing density matrices  $\rho_0^{A'}$  and  $\rho_1^{A'}$ , we write  $\rho^{U A'} = \lambda_0 |0\rangle\langle 0|^U \otimes \rho_0 + \lambda_1 |1\rangle\langle 1|^U \otimes \rho_1$ , where  $0 \leq \lambda_0, \lambda_1$  satisfy  $\lambda_0 + \lambda_1 = 1$ . Defining states  $\sigma^{U B E} = \mathcal{U}(\rho^{U A'})$  and  $\sigma^{U E' F} = \mathcal{V}(\text{Tr}_E \sigma)$ , where  $\mathcal{V}: B \rightarrow E' F$  isometrically extends the associated degrading map  $\mathcal{D}$ , we write

$$\begin{aligned} \lambda_0 I_c(\rho_0, \mathcal{N}) + \lambda_1 I_c(\rho_1, \mathcal{N}) &= H(B|U)_\sigma - H(E|U)_\sigma \\ &= H(FE'|U)_{\sigma'} - H(E'|U)_{\sigma'} \\ &= H(F|E'U)_{\sigma'} \\ &\leq H(F|E')_{\sigma'} \\ &= H(B)_\sigma - H(E)_\sigma \\ &= I_c(\lambda_0 \rho_0 + \lambda_1 \rho_1, \mathcal{N}). \end{aligned}$$

Here, we used the fact that  $\mathcal{V}$  preserves entropies and the form of strong subadditivity which states that conditioning cannot increase entropy. This proves the claim.  $\square$

A particularly simple subclass of degradable channels consists of the generalized dephasing channels [29]. These are channels  $\mathcal{N}: A' \rightarrow B$  with  $|A| = |B|$  and an isometric extension  $\mathcal{U}_\mathcal{N}: A' \rightarrow BE$  given by

$$\mathcal{U}_\mathcal{N} = \sum_i |i\rangle^B |\phi_i\rangle^E \langle i|^{A'}$$

where the  $|i\rangle^{A'}$  and  $|i\rangle^B$  are orthonormal bases and the  $|\phi_i\rangle^E$  are arbitrary. Equivalently, these are precisely those channels whose Kraus matrices can be simultaneously diagonalized. These channels act on density matrices as

$$\rho^{A'} \mapsto \sum_{i,j} \langle i|\rho^{A'}|j\rangle \cdot \langle \phi_j|\phi_i\rangle^E \cdot |i\rangle\langle j|^B.$$

Because their action on operators is diagonalized in the basis  $|i\rangle\langle j|^{A'}$ , they have also been called *diagonal* or *Hadamard diagonal* channels. A useful fact regarding generalized dephasing channels is that  $\max_{\rho} I_c(\rho, \mathcal{N})$  is achieved on a  $\rho$ , which is diagonal in the dephasing basis  $\{|i\rangle^{A'}\}$ .

### B. A Class of Channels for Which $\mathcal{CQ} = \mathcal{CQ}^{(1)}$

Here, we define a class of channels  $\mathcal{N}: A' B' \rightarrow C$  for which  $\mathcal{CQ}(\mathcal{N})$  admits a single-letter description.

*Theorem 3:* Let  $C = K C_B$  and  $|A'| = |K|$ . Suppose that  $K$  has the orthonormal basis  $\{|k\rangle^K\}$  and define a corresponding basis  $\{|k\rangle^{A'}\}$  for  $A'$ . If  $\{\mathcal{N}_k: B' \rightarrow C_B\}$  is any collection of degradable quantum channels, then the channel defined by linearly extending

$$\mathcal{N}(\rho^{A'} \otimes \sigma^{B'}) = \sum_k \langle k|\rho^{A'}|k\rangle \cdot |k\rangle\langle k|^K \otimes \mathcal{N}_k(\sigma^{B'}) \quad (41)$$

to all states satisfies  $\mathcal{CQ}(\mathcal{N}) = \mathcal{CQ}^{(1)}(\mathcal{N})$ .

*Proof:* For any given  $n$ , fix an arbitrary ensemble  $\{p(x), |\psi_x\rangle^{A^n}\}$  and an arbitrary state  $|\Psi\rangle^{B B'^n}$ . These give rise to the global state

$$\omega^{X B C^n} = \sum_x p(x) |x\rangle\langle x|^X \otimes \omega_x^{B C^n}$$

where

$$\omega_x^{B C^n} = \sum_{k^n} p(k^n | x) |k^n\rangle\langle k^n|^K \otimes (\mathcal{N}_{k_1} \otimes \cdots \otimes \mathcal{N}_{k_n})(\Psi^{B B'^n})$$

and  $p(k^n | x) = |\langle k^n | \psi_x \rangle|^2$ . Because

$$\begin{aligned} I(X; C^n) &\leq I(K^n; C^n) = H(K^n) \\ I_c(B) C^n X &= I_c(B) C^n K^n \end{aligned}$$

it suffices for Alice to signal with the pure states  $|k^n\rangle^{A^n}$  with probabilities  $p(k^n) = \sum_x p(x) p(k^n | x)$ . For each  $1 \leq j \leq n$ , we write  $\Psi_j^{B'}$  for the partial trace of the restriction  $\Psi^{B_j B'_j}$  of  $\Psi^{B B'^n}$  to  $j$ th system acting on some fixed system  $B'$ , and fix purifications  $|\Psi_j\rangle^{B'' B'}$  of these states. Recursively, using (40), we may now write

$$\begin{aligned} I_c(B) C^n K^n &= \sum_{k^n} p(k^n) I_c(\Psi^{B'^n}, \mathcal{N}_{k_1} \otimes \cdots \otimes \mathcal{N}_{k_n}) \\ &\leq \sum_{k^n} p(k^n) \sum_{j=1}^n I_c(\Psi_j^{B'}, \mathcal{N}_{k_j}) \\ &= \sum_{j=1}^n \sum_{k_j} p(k_j) I_c(\Psi_j^{B'}, \mathcal{N}_{k_j}) \\ &= \sum_{j=1}^n I_c(B'' C K) \omega_j \end{aligned}$$

where

$$\omega_j^{B'' C K} = \sum_{k_j} p(k_j) |k_j\rangle\langle k_j|^K \otimes \mathcal{N}_{k_j}(\Psi_j^{B'' B'}).$$

Similarly, concavity of entropy implies that

$$I(K^n; C^n)_\omega = H(K^n)_\omega \leq \sum_{j=1}^n H(K_j)_\omega = \sum_{j=1}^n I(K; C)_{\omega_j}.$$

We may, therefore, conclude that every cq rate pair  $(r, S) \in \mathcal{Q}(\mathcal{N}^{\otimes n})$  is contained in the convex hull of  $n\mathcal{Q}^{(1)}(\mathcal{N})$ . This concludes the proof.  $\square$

As a concrete application of this theorem, consider the erasure multiple-access channel described in Section III. Up to a local isometry on  $C$ , this is equivalent to one of the form (41) where  $\mathcal{N}_1$  is an identity qubit channel and  $\mathcal{N}_2$  is a 50% erasure channel  $\mathcal{N}_2(\sigma) = \frac{1}{2}\sigma + \frac{1}{2}|e\rangle\langle e|^C$ . Here,  $|e\rangle^C$  is a state orthogonal to the support of  $\sigma$ . The convex hull of  $\mathcal{CQ}^{(1)}$  for this channel can be readily evaluated by either using the main result of [30], or by otherwise noting that for whatever preparation probabilities used by Alice, the coherent information from Bob to Charlie is maximized by a maximally mixed input. This follows because that channel is always degradable, so by convexity and symmetry, its coherent information is maximized accordingly. The region is given in Section III and plotted there in Fig. 1 for  $d = 2$ .

### C. Single-Letter Description of $\mathcal{Q}$ for a Phase Flip Channel

While the description of the capacity region  $\mathcal{Q}$  in Theorem 2 generally requires taking a many-letter limit, we give here an example of a channel  $\mathcal{N}_p: A'B' \rightarrow C$  for which that description can be single-letterized. The channel  $\mathcal{N}_p$  takes as input two qubits, one from Alice and the other from Bob. With probability  $p$ , the channel causes each qubit to undergo a phase flip, by rotating the state of each by  $180^\circ$  about its  $z$ -axis on the Bloch sphere, before it is received by the receiver Charlie. The action of  $\mathcal{N}_p$  on an input density operator  $\rho^{A'B'}$  is

$$\mathcal{N}_p(\rho^{A'B'}) = (1-p)\rho^{A'B'} + p(\sigma_z \otimes \sigma_z)\rho^{A'B'}(\sigma_z \otimes \sigma_z)$$

where  $\sigma_z$  is the Pauli phase flip matrix. We will demonstrate that  $\mathcal{Q}(\mathcal{N}_p)$  is equal to the collection of all pairs of nonnegative rates  $(R, S)$ , which satisfy

$$\begin{aligned} R &\leq 1 \\ S &\leq 1 \\ R + S &\leq 2 - H(p). \end{aligned}$$

*Proof:* We first observe that for any input state of the form (8),  $\frac{1}{k}I_c(AB)C^k$  is upper bounded by the quantum capacity of  $\mathcal{N}_p$  when both senders may act together. Since  $\mathcal{N}_p$  is a generalized dephasing channel, its quantum capacity can be calculated by a single-letter optimization of the coherent information over input density operators that are diagonal in the dephasing basis which, incidentally, is just the computational basis. A short calculation reveals that it suffices to check inputs of the form

$$\rho_\alpha = \frac{1}{2}(\alpha(|00\rangle\langle 00| + |11\rangle\langle 11|) + (1-\alpha)(|10\rangle\langle 10| + |01\rangle\langle 01|))$$

when computing the quantum capacity of  $\mathcal{N}_p$ . Furthermore, since  $\mathcal{N}_p$  is also degradable, Lemma 5 implies that  $I_c(\rho, \mathcal{N}_p)$  is concave as a function of  $\rho$ . Since  $I_c(\rho_\alpha, \mathcal{N}_p) = I_c(\rho_{1-\alpha}, \mathcal{N}_p)$ ,

one then concludes by symmetry that the maximum is achieved for  $\alpha = \frac{1}{2}$ , in which case  $\rho_{\frac{1}{2}}$  is maximally mixed, yielding

$$\max_{\rho} I_c(\rho, \mathcal{N}_p) = I_c(\pi^{A'B'}, \mathcal{N}_p) = 2 - H(p).$$

Note that this maximizing input is a product state, since  $\pi^{A'B'} = \pi^{A'} \otimes \pi^{B'}$ . Define the Bell states

$$|\psi_{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle).$$

As  $|\psi_{+}\rangle$  purifies the maximally mixed state  $\pi_2$ , let us write the global state

$$\omega^{ABC} = \mathcal{N}(\psi_{+}^{AA'} \otimes \psi_{+}^{BB'}).$$

Identifying  $C = \hat{A}\hat{B}$  in the obvious way, we reexpress

$$\omega^{A\hat{A}B\hat{B}} = (1-p)\psi_{+}^{A\hat{A}} \otimes \psi_{+}^{B\hat{B}} + p\psi_{-}^{A\hat{A}} \otimes \psi_{-}^{B\hat{B}}.$$

One readily calculates  $H(ABC) = H(p)$ ,  $H(C) = 2$ , and  $H(AC) = H(BC) = H(p) + 1$ . Combining these gives the relevant coherent informations

$$\begin{aligned} I_c(AB)C &= 2 - H(p) \\ I_c(A)BC &= I_c(B)AC = 1. \end{aligned}$$

Finally, because any state arising from  $\mathcal{N}_p$  must satisfy

$$I_c(A)BC \leq \log |A'| = 1 \text{ and } I_c(B)AC \leq \log |B'| = 1$$

the individual rate bounds are saturated and the claim follows.  $\square$

## VII. DISCUSSION

There have been a number of results analyzing multiterminal coding problems in quantum Shannon theory. For an i.i.d. cq source  $XB$ , Devetak and Winter [31] have proved a Slepian–Wolf-like coding theorem achieving the cq rate pair  $(H(X|B), H(B))$  for classical data compression with quantum side information. Such codes extract classical side information from  $B^n$  to aid in compressing  $X^n$ . The extraction of side information is done in such a way as to cause a negligible disturbance to  $B^n$ . Our Theorem 1 is somewhat of this flavor. There, the quantum state of  $C^n$  is measured to extract Alice's classical message which, in turn, is used as side information for decoding Bob's quantum information. Analogous results to ours were obtained by Winter in his analysis of a multiple-access channel with classical inputs and a quantum output, whereby the classical decoded message of one sender can be used as side information to increase the classical capacity of another sender. After a preprint of this manuscript was made available, analogous results have been found for the entanglement-assisted capacity regions of quantum multiple-access channels [32].

We further mention the obvious connection between our coding theorems and the subject of channel codes with side information available to the receiver. A more difficult problem is that of determining classical and quantum capacities when side information is available at the *encoder*, constituting quantum generalizations of results obtained by Gelfand and Pinsker [33]

for classical channels with side information. Some time after this preprint became available, the corresponding problems for data compression were solved in a “fully quantum” setting for cases where either the sender or receiver has quantum side information (the so-called fully quantum reverse Shannon and fully quantum Slepian–Wolf theorems) [34], [35] and more recently were generalized to the case where both parties have quantum side information, resulting in the so-called state redistribution protocol [36], [37].

In an earlier draft of this paper, we characterized  $\mathcal{Q}(\mathcal{N})$  as the closure of a regularized union of rectangles

$$\begin{aligned} 0 \leq R &\leq \frac{1}{k} I_c(A) C^k \\ 0 \leq S &\leq \frac{1}{k} I_c(B) C^k. \end{aligned}$$

This solution had been conjectured on the basis of a duality between classical Slepian–Wolf distributed source coding and classical multiple-access channels [27], [23], as well as on a purported no-go theorem for distributed data compression of the so-called irreducible pure state ensembles that appeared in an early version of [38]. After the earlier preprint was made available, Winter announced [39] recent progress with Oppenheim and Horodecki [40] on the quantum Slepian–Wolf problem, offering a characterization identical in functional form to the classical one, while also supplying an interpretation of negative rates and apparently evading the no-go theorem. Motivated by the earlier mentioned duality, he informed us that the qq capacity region could also be characterized in direct analogy to the classical case. Subsequently, we found that we could modify our previous coding theorem to achieve the new region, provided that the rates are nonnegative. After those events unfolded, the authors of [38] found an error in the proof of their no-go theorem, leading to a revised version consistent with the newer developments. Our earlier characterization of  $\mathcal{Q}(\mathcal{N})$ , while correct, is contained in the rate region of Theorem 2 for any finite  $k$ , frequently strictly so. The newer theorem, therefore, gives a more accurate approximation to the rate region for finite  $k$ . In fact, for any state arising from the channel that does not saturate the strong subadditivity inequality [41], the corresponding pentagon and rectangle regions are distinct. Another beneficial feature of this characterization is that it is possible to show that the maximum sum rate bound  $R + S \leq \max I_c(AB)C$  is additive, where the maximization is over all states of the form (8), for any degradable channel.

The same technique used to prove the new characterization of  $\mathcal{Q}(\mathcal{N})$  implies a new cq coding theorem, and thus a new characterization of  $\mathcal{CQ}(\mathcal{N})$ . By techniques nearly identical to those employed in the coding theorem for Theorem 2, it is possible to achieve the cq rate pair

$$(r, S) = (I(X; BC), I_c(B)C)$$

corresponding to Bob’s quantum information being used as side information for decoding Alice’s classical message. This is accomplished by having Charlie isometrically decode Bob’s quantum information, then coherently decode to produce an effective channel  $\mathcal{N}_1: A' \rightarrow BC$  so that Alice can transmit

classically at a higher rate. The new characterization is then a regularized union of pentagons, consisting of pairs of nonnegative rates  $(r, S)$  satisfying

$$\begin{aligned} r &\leq I(X; BC) \\ S &\leq I_c(B)CX \\ r + S &\leq I(X; C) + I_c(B)CX = I(X; BC) + I_c(B)C. \end{aligned}$$

Surprisingly, it is thus possible to characterize each of  $\mathcal{CQ}(\mathcal{N})$  and  $\mathcal{Q}(\mathcal{N})$  in terms of pentagons, in analogy to the original classical result. This situation makes apparent the dangers of being satisfied with regularized expressions for capacity regions. Without being able to prove single-letterization steps in the converses, it is hard to differentiate which characterization is the “right” one. While it is intuitively satisfying to see analogous formulas appear in both the classical and quantum theories, the regularized nature of the quantum results blurs the similarity. Indeed, the problems with single-letterization for single-user channels appear to be amplified when analyzing quantum networks (see, e.g., [42]). Perhaps this indicates that the necessity of understanding the capacities of single-user channels at a level beyond regularized optimizations is even more pressing than previously thought. We should mention that for the channels analyzed in Section VI-B, the newer description of  $\mathcal{CQ}(\mathcal{N})$  is not an issue, as the new corner point is contained in the old rectangle for any state arising from any number of parallel instances of the erasure channel. On the other hand, we demonstrate in [43] that for the collective phase flip channel, both characterizations single-letterize, yielding a cq region, which is identical in form to that obtained for  $\mathcal{Q}(\mathcal{N})$  in Section VI-C, replacing the quantum rate with a classical one.

Consider the full simultaneous cq region  $\mathcal{S}(\mathcal{N})$  for two senders, where each sends classical and quantum information at the same time. A formal operational definition of  $\mathcal{S}(\mathcal{N})$  is found in [43] and [44]. This region can be characterized in a way that generalizes Theorems 1 and 2 as the regularization of the region  $\mathcal{S}^{(1)}(\mathcal{N})$ , defined as the vectors of nonnegative rates  $(r, s, R, S)$  satisfying

$$\begin{aligned} r &\leq I(X; C|Y) \\ s &\leq I(Y; C|X) \\ r + s &\leq I(XY; C) \\ R &\leq I_c(A)BCXY \\ S &\leq I_c(B)ACXY \\ R + S &\leq I_c(AB)CXY \end{aligned}$$

for some state on  $XYABC$  of the form

$$\sum_{x,y} p(x)p(y)|x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y \otimes \mathcal{N}(\psi_x^{AA'} \otimes \phi_y^{BB'})$$

arising from the action of  $\mathcal{N}$  on the  $A'$  and  $B'$  parts of some pure state ensembles  $\{p(x), |\psi_x\rangle^{AA'}\}$  and  $\{p(y), |\phi_y\rangle^{BB'}\}$ . Briefly, achievability of this region is obtained as follows. Using techniques introduced in [29], each sender “shapes” their quantum information into HSW codewords. Decoding is accomplished by first decoding all of the classical information, then using that information as side information for a quantum decoder. The

main result of [29], the regularized optimization of the cq result from [5] over pairs of input ensembles, and our Theorems 1 and 2 follow as corollaries of the corresponding capacity theorem. Indeed, the six two-dimensional “shadows” of the above region, obtained by setting pairs of rates equal to zero, reproduce those aforementioned results. This characterization, however, only utilizes the rectangle description of  $\mathcal{CQ}(\mathcal{N})$ . It is indeed possible to write a more accurate regularized description of  $\mathcal{S}(\mathcal{N})$ , which generalizes the pentagon characterizations of  $\mathcal{CQ}(\mathcal{N})$  and  $\mathcal{Q}(\mathcal{N})$ , although we will not pursue that at this time.

## APPENDIX

### A. Proof of Convexity of $\mathcal{CQ}$ and $\mathcal{Q}$

Let  $\mathcal{N} : A'B' \rightarrow C$  be a quantum multiple-access channel. We will prove that  $\mathcal{Q}(\mathcal{N})$  is convex, as the proof for  $\mathcal{CQ}$  is identical.

*Proof:* Let  $k_0$  and  $k_1$  be positive integers, and fix any two states of the form (8),  $\sigma_0^{A_0B_0C^{k_0}}$ , and  $\sigma_1^{A_1B_1C^{k_1}}$ . Then,  $(R_0, S_0), (R_1, S_1) \in \mathcal{Q}(\mathcal{N})$ , where for  $i \in \{0, 1\}$

$$R_i = \frac{1}{k_i} I_c(A_i) C^{k_i} \sigma_i$$

$$S_i = \frac{1}{k_i} I_c(B_i) C^{k_i} \sigma_i.$$

We will now show that for any rational  $0 \leq \lambda \leq 1$ ,  $\lambda(R_0, S_0) + (1 - \lambda)(R_1, S_1) \in \mathcal{Q}(\mathcal{N})$ . We first write  $\lambda = \frac{\alpha}{\beta}$ , for integers satisfying  $\beta > 0$ ,  $\beta \geq \alpha \geq 0$ . Setting  $p_0 = \alpha k_1$ ,  $p_1 = (\beta - \alpha)k_0$ , and  $k = p_0 k_0 + p_1 k_1$ , define the composite systems  $A = A_0^{p_0} A_1^{p_1}$  and  $B = B_0^{p_0} B_1^{p_1}$ , as well as the density matrix  $\sigma^{ABC^k} = \sigma_0^{\otimes p_0} \otimes \sigma_1^{\otimes p_1}$ , which is also of the form (8). Additivity of coherent information across product states and some simple algebra gives

$$\begin{aligned} \frac{1}{k} I_c(A) C^k \sigma &= \frac{p_0}{k} I_c(A_0) C^{k_0} \sigma_0 + \frac{p_1}{k} I_c(A_1) C^{k_1} \sigma_1 \\ &= \frac{p_0 k_0 R_0 + p_1 k_1 R_1}{p_0 k_0 + p_1 k_1} \\ &= \lambda R_0 + (1 - \lambda) R_1. \end{aligned}$$

An identical calculation shows that

$$\frac{1}{k} I_c(B) C^k \sigma = \lambda S_0 + (1 - \lambda) S_1.$$

As  $\mathcal{Q}(\mathcal{N})$  was defined as the topological closure of rate pairs corresponding to states which appropriately arise from the channel, the result follows because the set of previously considered  $\lambda$ 's comprises a dense subset of the unit interval.  $\square$

### B. Proof of Cardinality Bound on $\mathcal{X}$

Begin by fixing a finite set  $\mathcal{X}$ , a labeled collection of pure states  $\{|\phi_x\rangle^{A'}\}_{x \in \mathcal{X}}$ , and a pure bipartite state  $|\Psi\rangle^{BB'}$ . For each  $x$ , these define the states  $\sigma_x^{BC} = \mathcal{N}(\phi_x \otimes \Psi)$  and  $\omega_x^C = \text{Tr}_B \sigma_x$ . Assume for now that  $|A'| \geq |C|$ . Define a mapping  $f: \mathcal{X} \rightarrow \mathbb{R}^{|C|^2+1}$ , via

$$f: x \mapsto f_x \equiv (\omega_x, H(\omega_x), I_c(B)C)_{\sigma_x}$$

where we identify  $\omega_x$  with its  $|C|^2 - 1$  dimensional parameterization. By linearity, this extends to a map from probability mass functions on  $\mathcal{X}$  to  $\mathbb{R}^{|C|^2+1}$ , where

$$f: p(x) \mapsto \sum_x p(x) f_x \equiv (\omega_p, H(C|X)_p, I_c(B)C X)_p.$$

Our use of the subscript  $p$  should be clear from the context. The use of Caratheodory's theorem for bounding the support sizes of auxilliary random variables in information theory (see [27]) is well known. Perhaps less familiar is the observation [45], [46] that a better bound can often be obtained by use of a related theorem by Fenchel and Eggleston [47], which states that if  $S \subseteq \mathbb{R}^n$  is the union of at most  $n$  connected subsets, and if  $y$  is contained in the convex hull of  $S$ , then  $y$  is also contained in the convex hull of at most  $n$  points in  $S$ . As the map  $f$  is linear, it maps the simplex of distributions on  $\mathcal{X}$  into a single connected subset of  $\mathbb{R}^{|C|^2+1}$ . Thus, for any distribution  $p(x)$ , there is another distribution  $p'(x)$ , which puts positive probability on at most  $|C|^2 + 1$  states, while satisfying  $f(p) = f(p')$ . If it is instead the case that  $|A'| < |C|$ , this bound can be reduced to  $|A'|^2 + 1$  by replacing the first components of the map  $f$  with a parameterization of  $\phi_x^{A'}$ , as specification of a density matrix on  $A'$  is enough to completely describe the resulting state on  $C$ . It is, therefore, sufficient to consider  $|X| \leq \min\{|A'|, |C|\}^2 + 1$  in computing  $\mathcal{CQ}^{(1)}(\mathcal{N})$ .  $\square$

## ACKNOWLEDGMENT

J. Yard would like to thank T. Cover for much useful input and feedback at many stages during the process of writing of this manuscript, and Y. H. Kim for useful discussions regarding classical multiple-access channels.

## REFERENCES

- [1] R. Ahlswede, “Multi-way communication channels,” in *Proc. 2nd Int. Symp. Inf. Theory*, 1973, pp. 23–52.
- [2] Liao, “Multiple access channels,” Ph.D. dissertation, Dept. Electr. Eng., Univ. Hawaii, , Manoa, HI, 1972.
- [3] T. Cover, A. El Gamal, and M. Salehi, “Multiple access channels with arbitrarily correlated sources,” *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.
- [4] P. W. Shor and J. A. Smolin, “Quantum error-correcting codes need not completely reveal the error syndrome,” quant-ph/9604006.
- [5] A. Winter, “The capacity of the quantum multiple access channel,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3059–3065, Nov. 2001.
- [6] G. Klimovitch, “On the classical capacity of a quantum multiple access channel,” in *Proc. IEEE Int. Symp. Inf. Theory*, Washington, DC, Jun. 2001, p. 278.
- [7] G. Klimovitch and A. Winter, “Classical capacity of quantum binary adder channels,” quant-ph/0502055.
- [8] M. Nielsen and I. Chuang, *Quantum Information and Quantum Computation*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
- [9] A. Uhlmann, “The ‘transition probability’ in the state space of a \*-algebra,” *Rep. Math. Phys.*, vol. 9, pp. 273–279, 1976.
- [10] I. Devetak and A. Winter, “Distilling common randomness from bipartite quantum states,” *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3183–3196, Dec. 2004.
- [11] A. S. Holevo, “Bounds for the quantity of information transmitted by a quantum channel,” *Probl. Inf. Transm.*, vol. 9, pp. 177–183, 1973.
- [12] B. Schumacher and M. A. Nielsen, “Quantum data processing and error correction,” *Phys. Rev. A, Gen. Phys.*, vol. 54, no. 4, pp. 2629–2629, 1996.
- [13] E. B. Davies and J. T. Lewis, “An operational approach to quantum probability,” *Comm. Math. Phys.*, vol. 17, pp. 239–260, 1970.

- [14] H. Barnum, C. M. Caves, C. A. Fuchs, R. Josza, and B. Schumacher, "Noncommuting mixed states cannot be broadcast," *Phys. Rev. Lett.*, vol. 76, no. 15, pp. 2818–2821, 1996.
- [15] M. B. Ruskai, "Beyond strong subadditivity: Improved bounds on the contraction of generalized relative entropy," *Rev. Math. Phys.*, vol. 6, no. 5A, pp. 1147–1161, 1994.
- [16] R. Alicki and M. Fannes, "Continuity of quantum conditional information," *J. Phys. A, Math. Gen.*, vol. 37, pp. L55–L57, Jan. 2004.
- [17] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2481–2485, Nov. 1999.
- [18] S. Lloyd, "Capacity of the noisy quantum channel," *Phys. Rev. A, Gen. Phys.*, vol. 55, no. 3, pp. 1613–1622, 1997.
- [19] P. Shor, "Quantum error correction," in *Proc. MSRI Workshop Quantum Inf. Cryptogr.*, Nov. 2002 [Online]. Available: <http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/>
- [20] I. Devetak, "The private classical information capacity and quantum information capacity of a quantum channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 44–55, Jan. 2005.
- [21] A. S. Holevo, "The capacity of the quantum channel with general input states," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 269–273, Jan. 1998.
- [22] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A, Gen. Phys.*, vol. 56, no. 1, pp. 131–138, 1997.
- [23] T. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [24] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *J. Math. Phys.*, vol. 14, pp. 1938–1941, Dec. 1973.
- [25] M. Nielsen and D. Petz, "A simple proof the strong subadditivity inequality," quant-ph/0408130.
- [26] G. Dueck, "Maximal error regions are strictly smaller than average error regions for multi-user channels," *Probl. Control Inf. Theory*, vol. 7, pp. 11–19, 1978.
- [27] I. Cs szar and J. K rner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akad miai Kiad , 1982.
- [28] H. Barnum, E. Knill, and M. Nielsen, "On quantum fidelities and quantum capacities," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1317–1329, Jul. 2000.
- [29] I. Devetak and P. Shor, "The capacity of a quantum channel for simultaneous transmission of classical and quantum information," *Commun. Math. Phys.*, vol. 256, no. 2, pp. 287–303, Mar. 2005.
- [30] C. Bennett, D. DiVincenzo, and J. Smolin, "Capacities of quantum erasure channels," *Phys. Rev. Lett.*, vol. 78, pp. 3217–3220, Apr. 1997.
- [31] I. Devetak and A. Winter, "Classical data compression with quantum side information," *Phys. Rev. A, Gen. Phys.*, vol. 68, pp. 042301–042306, Oct. 2003.
- [32] M. Hsieh, I. Devetak, and A. Winter, "Entanglement assisted capacity of quantum multiple access channels," quant-ph/0511228.
- [33] S. I. Gelfand and M. S. Pinsker, "Coding for a channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [34] I. Devetak, "A triangle of dualities: Reversibly decomposable quantum channels, source channel duality, and time reversal," *Phys. Rev. Lett.*, vol. 97, p. 140503, Oct. 2006.
- [35] A. Abeyesinge, I. Devetak, P. Hayden, and A. Winter, "The mother of all protocols: Restructuring quantum information's family tree," *IEEE Trans. Inf. Theory*, quant-ph/0606225, submitted for publication.
- [36] I. Devetak and J. Yard, "An operational interpretation of quantum conditional information," *Phys. Rev. Lett.*, quant-ph/0612059, to be published.
- [37] J. Yard and I. Devetak, "Quantum source coding with quantum side information at the encoder and decoder," arXiv:0706.2907.
- [38] C. Ahn, P. Doherty, P. Hayden, and A. Winter, "On the distributed compression of quantum information," *IEEE Trans. Inf. Theory*, quant-ph/0403042, to be published.
- [39] A. Winter, F. Verstraete, J. Smolin, J. Oppenheim, and M. Horodecki, "Entanglement of assistance and applications to multiuser quantum information theory, or transmitting partial (quantum) information – Which can be negative!," presented at the 8th Workshop Quantum Inf. Process., Cambridge, MA, Jan. 2005.
- [40] M. Horodecki, J. Oppenheim, and A. Winter, "Partial quantum information," *Nature*, vol. 436, no. 4, pp. 673–676.
- [41] P. Hayden, R. Josza, D. Petz, and A. Winter, "Structure of states which satisfy strong subadditivity of quantum entropy with equality," *Commun. Math. Phys.*, vol. 246, no. 2, pp. 359–374, Feb. 2004.
- [42] W. D r, J. I. Cirac, and P. Horodecki, "Non-additivity of quantum capacity for multiparty communication channels," *Phys. Rev. Lett.*, vol. 93, pp. 020503–02506, Jul. 2004, quant-ph/0403068.
- [43] J. Yard, I. Devetak, and P. Hayden, "Capacity theorems for quantum multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sep. 2005, pp. 884–888, cs.IT/0508031.
- [44] J. Yard, "Simultaneous classical-quantum capacities of quantum multiple access channels," Ph.D. dissertation, Electr. Eng. Dept., Stanford Univ., Stanford, CA, 2005.
- [45] A. W. J. Ziv, "The rate-distortion function for source decoding with side-information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
- [46] M. Salehi, "Cardinality bounds on auxiliary random variables in multiple-user theory via the method of Ahlswede and K rner," Dept. Statist., Stanford Univ., Stanford, CA, Tech. Rep. No. 33, Aug. 1978.
- [47] H. G. Eggleston, *Convexity*. New York: Cambridge Univ. Press, 1963.