Entanglement-Assisted Communication of Classical and Quantum Information

Min-Hsiu Hsieh, *Member, IEEE*, and Mark M. Wilde, *Member, IEEE*

***Abstract*—In this paper, we consider the problem of transmitting classical and quantum information reliably over an entanglement-assisted quantum (EAQ) channel. Our main result is a capacity theorem that gives a 3-D achievable rate region. Points in the region are *rate triples*, consisting of the classical communication rate, the quantum communication rate, and the entanglement consumption rate of a particular coding scheme. The crucial protocol in achieving the boundary points of the capacity region is a protocol that we name the *classically enhanced father (CEF) protocol*. The CEF protocol is more general than other protocols in the family tree of quantum Shannon theoretic protocols, in the sense that several previously known quantum protocols are now child protocols of it. The CEF protocol also shows an improvement over a timesharing strategy for the case of a qubit dephasing channel—this result justifies the need for simultaneous coding of classical and quantum information over an EAQ channel. Our capacity theorem is of a multiletter nature (requiring a limit over many uses of the channel), but it reduces to a single-letter characterization for at least three channels: the completely depolarizing channel, the quantum erasure channel, and the qubit dephasing channel.**

***Index Terms*—Classically enhanced father (CEF) protocol, entanglement-assisted classical and quantum (EACQ) coding, entanglement-assisted quantum (EAQ) channel, quantum Shannon theory.**

I. Introduction

THE communication of information over a noisy quantum channel is a fundamental task in quantum communication theory. A sender may wish to transmit classical information, quantum information, or both. The Holevo–Schumacher–Westmoreland (HSW) coding theorem gives an achievable rate at which a sender can transmit *classical* data to a receiver if she transmits the classical information over a noisy quantum channel [1], [2]. The HSW theorem generalizes Shannon's classical channel coding theorem [3] to the quantum setting. The Lloyd–Shor–Devetak (LSD) coding theorem gives an achievable rate at which a sender can transmit *quantum* data to a receiver through a quantum channel [4]–[6]. Devetak and Shor followed up on these results by determining achievable rates at which a sender can simultaneously transmit both classical and quantum information over a quantum channel [7]. The naive scheme is to employ a timesharing strategy, where a sender uses an HSW code for a fraction of the transmitted qubits and an LSD code for the other fraction. The Devetak–Shor (DS) coding strategy outperforms the naive timesharing strategy, at least when the noisy channel is the qubit dephasing channel [8]. This result demonstrates the need to consider nontrivial coding schemes when communicating more than one resource.

A sender can exploit a quantum channel alone, as in the above examples, or she can exploit assisting resources as well. Examples of such assisting resources are a static resource shared with the receiver, as in the case of common randomness, secret key, or entanglement, or a dynamic resource connecting the sender to the receiver, as in the case of a noiseless classical or quantum side channel.

Assisting a quantum channel with noiseless resources sometimes improves communication rates. The simplest and most striking example of this phenomenon occurs when a noiseless ebit assists a noiseless qubit channel. The superdense coding protocol outlines a simple method to transmit two classical bits over a noiseless qubit channel assisted by an ebit [9]. This protocol beats the Holevo bound [8], which limits an unassisted noiseless qubit channel to transmit no more than one classical bit. The superdense coding protocol then led Bennett *et al.* to explore if one could improve the classical capacity of a noisy quantum channel by assisting it with unlimited entanglement [10], [11]. They confirmed their intuition by proving a channel coding theorem that gives an entanglement-assisted classical (EAC) transmission rate higher than that without assistance. Shor then refined this result by determining tradeoffs between the classical communication rate and the entanglement consumption rate [12].

Quantum information theorists have since organized protocols that exploit the different resources of quantum communication, classical communication, and entanglement into a family tree [13]–[16]. One member of the family tree is the *father protocol* [13], [14]. The father protocol is so named because it generates several "child" protocols using the theory of resource inequalities [13], [14]. Devetak *et al.* exploited the father protocol to demonstrate tradeoffs between the quantum communication rate and the entanglement consumption rate over an entanglement-assisted quantum (EAQ) channel [13].

An important natural question, in light of the aforementioned tradeoff solutions for two of the three noiseless resources, is

Manuscript received January 27, 2009; revised November 13, 2009. Date of current version August 18, 2010. The work of M. M. Wilde was supported by the National Research Foundation and Ministry of Education, Singapore and an MDEIE (Québec) PSR-SIIRI International Collaboration Grant.

M.-H. Hsieh is with the ERATO-SORST Quantum Computation and Information Project, Japan Science and Technology Agency, Tokyo 113-0033, Japan (e-mail: minhsiuh@gmail.com).

M. M. Wilde was with the Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore. He is now with the School of Computer Science, McGill University, Montreal, QC H3A 2A7 Canada (e-mail: mark.wilde@mcgill.ca).

Communicated by P. Hayden, Associate Editor for Quantum Information Theory.

Color versions of Figures 1–5 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2053903

then how one might combine all three different resources. Previous work has addressed tradeoffs for the task of remotely preparing quantum states with the aid of classical communication, quantum communication, and entanglement [17], but no one has yet considered the triple tradeoffs for channel coding.

In this paper, we conduct an investigation of the tradeoffs for channel coding both quantum and classical information over a quantum channel assisted by noiseless entanglement. We prove the entanglement-assisted classical and quantum (EACQ) capacity theorem, which gives achievable rates for this task. We extend the family tree of quantum Shannon theory by developing the *classically enhanced father (CEF) protocol*.[1] This protocol is more general than any of the existing protocols in the tree and achieves rates in the 3-D capacity region. We dub this protocol the "CEF protocol" because it is an extension of the father protocol, and it generates five child protocols in the sense of [13], [14]. Two of its child protocols are classically enhanced quantum communication [7] and EAC communication [10]–[12] (we detail the others in Section VI-F). We also demonstrate that isometric encodings are sufficient for achieving our rate formulas, resolving an open problem from [14].

A benefit of the CEF protocol is that it inspires the design of classically enhanced entanglement-assisted quantum error-correcting codes (EAQECCs) [19], [20]. We give evidence in Section VIII-B that it is possible to reach the achievable rates without encoding classical information into the entanglement shared between the sender and the receiver.

We structure this paper as follows. In the next section, we give some definitions and establish notation used in the remainder of the paper. Section III provides a description of a general protocol for communication of classical and quantum information with the assistance of entanglement. We then state the main capacity theorem, Theorem 1, in Section IV and show how the classical capacity theorem [1], [2], the quantum capacity theorem [4]–[6], the classically enhanced quantum capacity region [7], the father capacity region [14], and the EAC capacity region [12] are all special cases of the EACQ capacity region. We prove the converse of Theorem 1 in Section V and prove the direct-coding part of Theorem 1 in Section VI. Section VI-F discusses the child protocols that the CEF protocol generates. We then give three example channels, the completely depolarizing channel, the quantum erasure channel, and the qubit dephasing channel, that admit a single-letter solution for the capacity region (meaning that we have a complete understanding of the capacity region for these channels). We also show that the CEF protocol gives an improvement over a timesharing strategy when the noisy channel is the qubit dephasing channel. We end by summarizing our results and by posing several open questions.

II. Definitions and Notation

The ensemble $\{p(x), \psi_x^{ABE}\}_{x\in\mathcal{X}}$, where each state ψ_x^{ABE} is a pure tripartite state, is essential in the ensuing analysis of this paper. The *coherent information* $I(A\rangle B)_{\psi_x}$ of each state ψ_x^{ABE} in the ensemble is as follows:

$$I(A\rangle B)_{\psi_x} \equiv H(B)_{\psi_x} - H(AB)_{\psi_x}$$

where $H(B)_{\psi_x}$ is the von Neumann entropy of the reduction of the state ψ_x^{ABE} to the system B with a similar definition for $H(AB)_{\psi_x}$. The *quantum mutual information* $I(A;B)_{\psi_x}$ of each state ψ_x^{ABE} is as follows:

$$I(A;B)_{\psi_x} \equiv H(A)_{\psi_x} + I(A\rangle B)_{\psi_x}.$$

We can classically correlate states in some system X with each state ψ_x^{ABE} to produce an augmented ensemble

$$\{p(x), |x\rangle\langle x|^X \otimes \psi_x^{ABE}\}_{x\in\mathcal{X}}$$

where the set $\{|x\rangle\}_{x\in\mathcal{X}}$ is some preferred orthonormal basis for the auxiliary system X. The expected density operator of this augmented ensemble is the following classical-quantum state:

$$\sigma^{XABE} \equiv \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|^X \otimes \psi_x^{ABE}.$$

The *Holevo information* of the classical variable X with the quantum system B is $I(X;B)_\sigma$. For the special case of a classical system X, taking the expectation of the above entropic quantities with respect to the density $p(x)$ gives the respective conditional entropy $H(A\,|\,X)_\sigma$, conditional coherent information $I(A\rangle B\,|\,X)_\sigma$, and conditional mutual information $I(A;B\,|\,X)_\sigma$

$$H(A\,|\,X)_\sigma \equiv \sum_{x\in\mathcal{X}} p\,(x)\,H(A)_{\psi_x}$$
$$I(A\rangle B\,|\,X)_\sigma \equiv \sum_{x\in\mathcal{X}} p(x) I(A\rangle B)_{\psi_x}$$
$$I(A;B\,|\,X)_\sigma \equiv \sum_{x\in\mathcal{X}} p(x) I(A;B)_{\psi_x}.$$

One can easily prove that $I(A\rangle B\,|\,X)_\sigma = I(A\rangle BX)_\sigma$. We use the notation $I(A\rangle BX)_\sigma$ for conditional coherent information in what follows. The above definitions lead to the following useful identities:

$$H(A\,|\,X)_\sigma = \frac{1}{2}I(A;B\,|\,X)_\sigma + \frac{1}{2}I(A;E\,|\,X)_\sigma \quad (1)$$
$$I(A\rangle BX)_\sigma = \frac{1}{2}I(A;B\,|\,X)_\sigma - \frac{1}{2}I(A;E\,|\,X)_\sigma. \quad (2)$$

Proving the above identities is a simple matter of noting that the von Neumann entropy is equal for the reduced systems of a pure bipartite state. Adding the above identities gives the following one:

$$H(A\,|\,X)_\sigma + I(A\rangle BX)_\sigma = I(A;B\,|\,X)_\sigma. \quad (3)$$

The chain rule for quantum mutual information proves to be useful as well

$$I(AX;B)_\sigma = I(A;B\,|\,X)_\sigma + I(X;B)_\sigma. \quad (4)$$

All of the above information quantities possess operational interpretations in the theorems in this paper.

[1]As a side note, we mention that former papers discuss the possibility of this protocol but never fully developed it [18], [14]. In addition, the current authors have both constructed "CEF" error-correcting coding schemes for block codes [19] and for convolutional codes [20].

A noisy quantum channel $\mathcal{N}^{A'\to B}$ acts as a completely positive trace-preserving (CPTP) map. It takes a quantum system A' as an input and produces a noisy output quantum system B.

A *conditional quantum encoder* $\mathcal{E}^{MA\to B}$, or *conditional quantum channel* [21], is a collection $\{\mathcal{E}_m^{A\to B}\}_m$ of CPTP maps. Its inputs are a classical system M and a quantum system A and its output is a quantum system B. A classical-quantum state ρ^{MA}, where

$$\rho^{MA} \equiv \sum_m p(m)|m\rangle\langle m|^M \otimes \rho_m^A$$

can act as an input to the conditional quantum encoder $\mathcal{E}^{MA\to B}$. The action of the conditional quantum encoder $\mathcal{E}^{MA\to B}$ on the classical-quantum state ρ^{MA} is as follows:

$$\mathcal{E}^{MA\to B}(\rho^{MA}) = \mathrm{Tr}_M\left\{\sum_m p(m)|m\rangle\langle m|^M \otimes \mathcal{E}_m^{A\to B}\left(\rho_m^A\right)\right\}.$$

It is actually possible to write *any* quantum channel as a conditional quantum encoder when its input is a classical-quantum state [21]. In this paper, a conditional quantum encoder functions as the sender Alice's encoder of classical and quantum information.

A *quantum instrument* $\mathcal{D}^{A\to BM}$ is a CPTP map whose input is a quantum system A and whose outputs are a quantum system B and a classical system M [14], [21]. A collection $\{\mathcal{D}_m^{A\to B}\}_m$ of completely positive trace-reducing maps specifies the instrument $\mathcal{D}^{A\to BM}$. The action of the instrument $\mathcal{D}^{A\to BM}$ on an arbitrary input state ρ is as follows:

$$\mathcal{D}^{A\to BM}(\rho^A) = \sum_m \mathcal{D}_m^{A\to B}(\rho^A) \otimes |m\rangle\langle m|^M. \quad (5)$$

Tracing out the classical register M gives the induced quantum operation $\mathcal{D}^{A\to B}$ where

$$\mathcal{D}^{A\to B}(\rho^A) \equiv \sum_m \mathcal{D}_m^{A\to B}(\rho^A).$$

This sum map is trace preserving

$$\mathrm{Tr}\left\{\sum_m \mathcal{D}_m^{A\to B}(\rho^A)\right\} = 1.$$

We can think of the following quantity:

$$p(m\,|\,\rho^A) \equiv \mathrm{Tr}\left\{\mathcal{D}_m^{A\to B}(\rho^A)\right\}$$

as a conditional probability $p(m\,|\,\rho^A)$ of receiving the classical message m when the state ρ^A is input. In this paper, a quantum instrument functions as Bob's decoder of classical and quantum information.

We abbreviate a *capacity region* by the noiseless resources involved: classical communication (C), quantum communication (Q), or entanglement (E), but we abbreviate a *protocol* with a different name corresponding either to its inventors or an appropriate acronym. For example, we speak of the C, Q, or CE capacity theorems for classical communication, quantum communication, and EAC communication, respectively, but the corresponding protocols are HSW coding, LSD coding, and EAC coding.

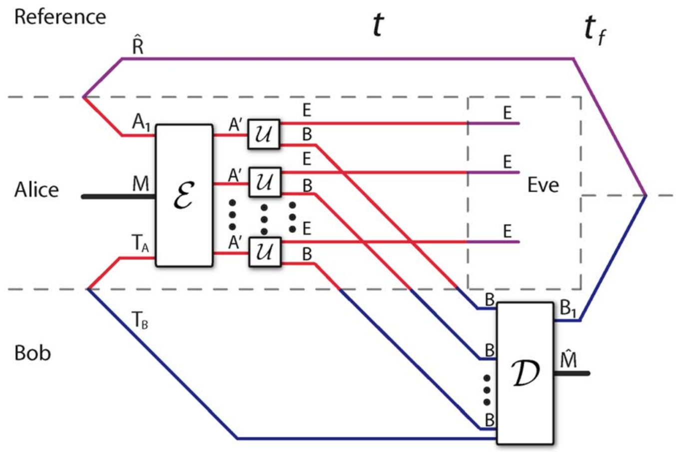

Fig. 1. General EACQ protocol. A sender Alice would like to communicate the quantum information in system A_1 and the classical information in system M. Her system T_A represents shared maximal entanglement with the receiver's system T_B. Alice encodes her information and uses the noisy channel a large number of times. The environment Eve obtains part of the output and the receiver Bob obtains the other part. Bob combines his received systems with his half of the entanglement and performs a decoding operation to recover both the classical and quantum information.

We note some other points before beginning. The trace norm $\|A\|_1$ of an operator A is as follows:

$$\|A\|_1 \equiv \mathrm{Tr}\{\sqrt{A^\dagger A}\}.$$

The maximally entangled state on system T_A and T_B is $\Phi^{T_AT_B}$. The omission of a superscript implies a reduced state, e.g., the state Φ^{T_A} is the reduced state of $\Phi^{T_AT_B}$ on T_A. Yard's thesis [21] provides a good introduction to quantum Shannon theory, and we point the reader there for properties such as strong subadditivity [22] and the quantum data processing inequality [23].

III. A General Protocol for Entanglement-Assisted Communication of Classical and Quantum Information

We begin by defining a general protocol for entanglement-assisted communication of classical and quantum information (EACQ) for a noisy quantum channel connecting a sender Alice to a receiver Bob. Alice would like to communicate two items to Bob:

1) an arbitrary quantum state ρ^{A_1} in a system A_1 with dimension 2^{nQ};
2) one of 2^{nC} classical messages.

Alice and Bob also share entanglement in the form of a maximally entangled state $\Phi^{T_AT_B}$ prior to communication. Alice possesses the system T_A, Bob possesses the system T_B, and the dimension of each system is 2^{nE}. We can think of this state as possessing nE ebits of entanglement because it is equivalent by local isometries to nE "gold standard" ebits in the state $|\Phi^+\rangle^{AB} \equiv (|00\rangle^{AB} + |11\rangle^{AB})/\sqrt{2}$. Alice performs a conditional quantum encoder $\mathcal{E}^{MA_1T_A\to A'^n}$ that encodes both her quantum systems A_1 and T_A and the classical message in system M. The encoding operation $\mathcal{E}^{MA_1T_A\to A'^n}$ prepares a system A'^n for input to a noisy quantum channel $\mathcal{N}^{A'^n\to B^n}$.

The channel $\mathcal{N}^{A'^n\to B^n}$ represents n independent uses of the noisy quantum channel $\mathcal{N}^{A'\to B}$

$$\mathcal{N}^{A'^n\to B^n} \equiv \left(\mathcal{N}^{A'\to B}\right)^{\otimes n}.$$

She then sends her state through the quantum channel $\mathcal{N}^{A'^n\to B^n}$. Bob receives the system B^n and performs a decoding instrument $\mathcal{D}^{B^nT_B\to B_1B_E\hat{M}}$ on the channel output B^n and his half of the entanglement T_B. The instrument $\mathcal{D}^{B^nT_B\to B_1B_E\hat{M}}$ produces a system B_1 with the quantum information that Alice sent, a classical register $\hat{M}$ containing Alice's classical message, and another system B_E that does not contain any useful information. Bob should be able to identify the classical message with high probability and recover the state ρ^{A_1} with high fidelity. Fig. 1 provides a detailed illustration of this protocol.

It is useful to consider the isometric extension $U_{\mathcal{N}}^{A'\to BE}$ of the channel $\mathcal{N}^{A'\to B}$ where Alice controls the channel input system A', Bob has access to the channel output system B, and the environment Eve has access to the system E.[2] For an independent and identically distributed (i.i.d.) channel $\mathcal{N}^{A'^n\to B^n}$ as defined above, we write its isometric extension as $U_{\mathcal{N}}^{A'^n\to B^nE^n}$. Also, it is useful to think of Alice's quantum system ρ^{A_1} as a restriction of some pure state $\varphi^{\hat{R}A_1}$ where Alice does not have access to the purification system $\hat{R}$.

We formalize the EACQ quantum information processing task as follows. Define an (n, C, Q, E, ϵ) EACQ code by the following.

- Alice's conditional quantum encoder $\mathcal{E}^{MA_1T_A\to A'^n}$ with encoding maps $\{\mathcal{E}_m^{A_1T_A\to A'^n}\}_{m\in[2^{nC}]}$. This encoder encodes both her quantum information and classical information. Define the following states for each classical message m

$$\omega_m^{\hat{R}A'^nT_B} \equiv \mathcal{E}_m^{A_1T_A\to A'^n}(\varphi^{\hat{R}A_1}\otimes\Phi^{T_AT_B}) \tag{6}$$

  where the dimension of system A_1 is 2^{nQ} and the dimension of system T_A is 2^{nE}. The density operator that includes the classical register M and averages over all classical messages is as follows:

$$\omega^{M\hat{R}A'^nT_B} \equiv \frac{1}{|M|}\sum_m |m\rangle\langle m|^M \otimes \omega_m^{\hat{R}A'^nT_B} \tag{7}$$

  where $|M|$ is the size of the classical register M. The output of the channel given that Alice sent classical message m is then as follows:

$$\omega_m^{\hat{R}B^nE^nT_B} \equiv U_{\mathcal{N}}^{A'^n\to B^nE^n}\left(\omega_m^{\hat{R}A'^nT_B}\right).$$

  The average output of the channel is as follows:

$$\omega^{M\hat{R}B^nE^nT_B} \equiv U_{\mathcal{N}}^{A'^n\to B^nE^n}\left(\omega^{M\hat{R}A'^nT_B}\right). \tag{8}$$

- Bob's decoding instrument $\mathcal{D}^{B^nT_B\to B_1B_E\hat{M}}$, whose action is defined in (5), is a collection of completely positive trace-reducing maps $\{\mathcal{D}_m^{B^nT_B\to B_1B_E}\}_{m\in[2^{nC}]}$. The decoding instrument decodes both the quantum information and classical information that Alice sends. The density operator corresponding to Bob's output state is as follows:

$$\omega^{M\hat{R}B_1B_E\hat{M}E^n} \equiv \mathcal{D}^{B^nT_B\to B_1B_E\hat{M}}\left(\omega^{M\hat{R}B^nE^nT_B}\right).$$

The classical probability of successful transmission of message m is as follows:

$$\Pr\{\hat{M}=m\,|\,M=m\} = \mathrm{Tr}\left\{\left(\mathcal{D}_m^{B^nT_B\to B_1B_E}\right)\left(\omega_m^{\hat{R}B^nE^nT_B}\right)\right\}$$

where $\hat{M}$ denotes the random variable corresponding to Bob's received classical message. The final state on the reference system $\hat{R}$ and Bob's quantum system B_1 is $\Upsilon^{\hat{R}B_1}$ where

$$\Upsilon^{\hat{R}B_1} \equiv \mathrm{Tr}_{\hat{M}B_EE^n}\left\{\mathcal{D}^{B^nT_B\to B_1B_E\hat{M}}\left(\omega_m^{\hat{R}B^nE^nT_B}\right)\right\}.$$

For the (n, C, Q, E, ϵ) EACQ code to be "ϵ-good," the following two conditions should hold for all classical messages $m \in [2^{nC}]$.

1) Bob decodes any of the classical messages m with high probability

$$\Pr\{\hat{M}=m\,|\,M=m\} \geq 1-\epsilon. \tag{9}$$

2) The state $\Upsilon^{\hat{R}B_1}$ should be ϵ-close to the ideal state $\varphi^{\hat{R}B_1} \equiv \mathrm{id}^{A_1\to B_1}(\varphi^{\hat{R}A_1})$

$$\|\Upsilon^{\hat{R}B_1} - \varphi^{\hat{R}B_1}\|_1 \leq \epsilon \tag{10}$$

so that Bob recovers the quantum information in system A_1 with high fidelity.

A rate triple (C, Q, E) is *achievable* if there exists an $(n, C-\delta, Q-\delta, E+\delta, \epsilon)$ EACQ code for any $\epsilon, \delta > 0$ and sufficiently large n. The capacity region $\mathcal{C}(\mathcal{N})$ is a 3-D region containing all achievable rate triples (C, Q, E).

IV. The EACQ Capacity Theorem

We now state our main theorem: the EACQ capacity (CQE) theorem that involves all three noiseless resources.

Theorem 1: The capacity region $\mathcal{C}_{\mathrm{CQE}}(\mathcal{N})$ of an EAQ channel $\mathcal{N}$ for simultaneously transmitting both quantum information and classical information is equal to the following expression:

$$\mathcal{C}_{\mathrm{CQE}}(\mathcal{N}) = \overline{\bigcup_{k=1}^{\infty}\frac{1}{k}\mathcal{C}_{\mathrm{CQE}}^{(1)}(\mathcal{N}^{\otimes k})} \tag{11}$$

where the overbar indicates the closure of a set. The "one-shot" region $\mathcal{C}_{\mathrm{CQE}}^{(1)}(\mathcal{N})$ is the union of the regions $\mathcal{C}_{\mathrm{CQE},\sigma}^{(1)}(\mathcal{N})$

$$\mathcal{C}_{\mathrm{CQE}}^{(1)}(\mathcal{N}) \equiv \bigcup_{\sigma}\mathcal{C}_{\mathrm{CQE},\sigma}^{(1)}(\mathcal{N})$$

where $\mathcal{C}_{\mathrm{CQE},\sigma}^{(1)}(\mathcal{N})$ is the set of all $C, Q, E \geq 0$, such that

$$C + 2Q \leq I(AX;B)_\sigma \tag{12}$$
$$Q \leq I(A\rangle BX)_\sigma + E \tag{13}$$
$$C + Q \leq I(X;B)_\sigma + I(A\rangle BX)_\sigma + E. \tag{14}$$

[2]It should be clear from context when E refers to Eve's system or when it refers to the entanglement consumption rate.

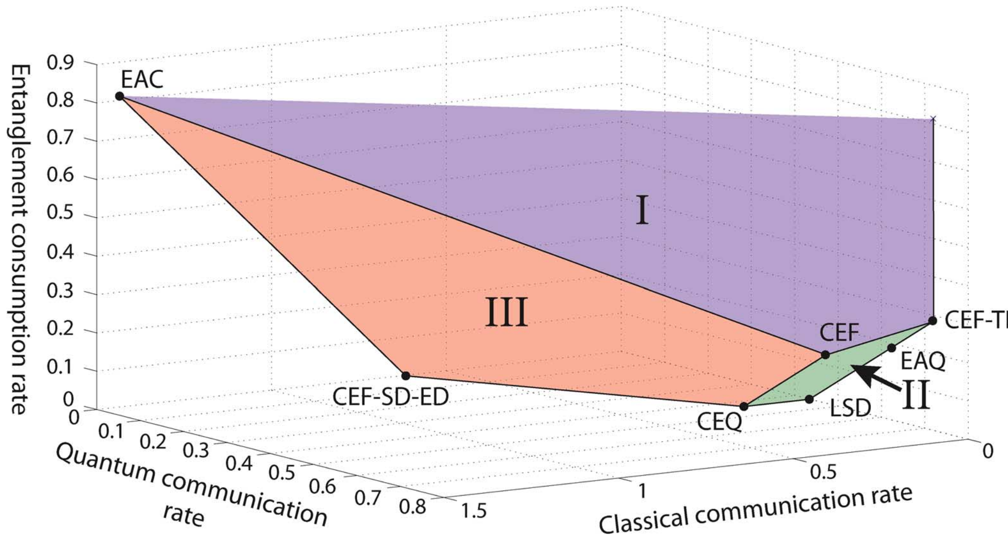

Fig. 2. Example of the one-shot, one-state achievable region $\mathcal{C}_{\text{CQE},\sigma}^{(1)}(\mathcal{N})$ corresponding to a state σ^{XABE} that arises from a qubit dephasing channel with dephasing parameter $p = 0.2$. The state input to the channel $\mathcal{N}$ is $\sigma^{XAA'}$, defined in (16). The plot features seven achievable corner points of the one-shot, one-state region. We can achieve the convex hull of these eight points by timesharing any two different coding strategies. We can also achieve any point above an achievable point by consuming more entanglement than necessary. The seven achievable points correspond to the father protocol (EAQ) [13], [14], the DS protocol for classically enhanced quantum communication (CEQ) [7], Shor's protocol for EAC communication with limited entanglement [12], quantum communication (LSD) [4]–[6], combining CEF with entanglement distribution and superdense coding (CEF-SD-ED) as detailed in Section VI-F, the CEF protocol outlined in Section VI, and combining the CEF protocol with teleportation [24] (CEF-TP). Observe that we can obtain EAC by combining CEF with superdense coding as detailed in Section VI-F, so that the points CEQ, CEF, EAC, and CEF-SD-ED all lie in plane III. Observe that we can obtain CEQ from CEF by entanglement distribution and we can obtain LSD from EAQ and EAQ from CEF-TP, both by entanglement distribution. Thus, the points CEF, CEQ, LSD, EAQ, and CEF-TP all lie in plane II. Finally, observe that we can obtain all corner points by combining CEF with the unit protocols in (61)–(63). This one-shot, one-state achievable region for the state σ^{XABE} is tight. The bounds in (12)–(14) uniquely specify the respective planes I–III. We obtain the full achievable region by taking the union over all states σ of the one-shot, one-state regions $\mathcal{C}_{\sigma}^{(1)}(\mathcal{N})$ and taking the regularization, as outlined in Theorem 1. The above region is a translation of the unit resource capacity region to the CEF protocol.

The above entropic quantities are with respect to a "one-shot" quantum state σ^{XABE} where

$$\sigma^{XABE} \equiv \sum_x p(x)|x\rangle\langle x|^X \otimes U_{\mathcal{N}}^{A'\to BE}\left(\phi_x^{AA'}\right) \qquad (15)$$

the states $\phi_x^{AA'}$ are pure, and it is sufficient to consider $|\mathcal{X}| \leq \min\{|A'|, |B|\}^2 + 1$ by the method in [25].

The capacity region in Theorem 1 is a union of general polyhedra, each specified by (12)–(14), where the union is over all possible states of the form (15) and a potentially infinite number of uses of the channel. Fig. 2 illustrates an example of the general polyhedron specified by (12)–(14), where the channel is the qubit dephasing channel[3] with dephasing parameter $p = 0.2$, and the input state is

$$\sigma^{XAA'} \equiv \frac{1}{2}\left(|0\rangle\langle 0|^X \otimes \phi_0^{AA'} + |1\rangle\langle 1|^X \otimes \phi_1^{AA'}\right) \qquad (16)$$

where

$$|\phi_0\rangle^{AA'} \equiv \sqrt{1/4}|00\rangle^{AA'} + \sqrt{3/4}|11\rangle^{AA'}$$
$$|\phi_1\rangle^{AA'} \equiv \sqrt{3/4}|00\rangle^{AA'} + \sqrt{1/4}|11\rangle^{AA'}.$$

The state σ^{XABE} resulting from the channel is $U_{\mathcal{N}}^{A'\to BE}(\sigma^{XAA'})$ where $U_{\mathcal{N}}$ is an isometric extension of the qubit dephasing channel. The figure caption provides a detailed explanation of the one-shot, one-state region $\mathcal{C}_{\text{CQE},\sigma}^{(1)}$ (note that Fig. 2 displays the one-shot, one-state region and does not display the full capacity region).

The above capacity region has a simple interpretation. In [26], we determined a unit resource capacity region. This unit resource region outlines what is achievable if one does not possess a noisy channel, but only possesses the three noiseless resources of classical communication, quantum communication, and entanglement. There, we found that the optimal strategy is to combine teleportation, superdense coding, and entanglement distribution. Interestingly, the above set of inequalities demonstrates that the one-shot, one-state region is a translation of the unit resource capacity region to the CEF protocol. Indeed, eliminating the entropic quantities from (12)–(14) reveals that the inequalities are the same as those that specify the unit resource capacity region.

Proving that Theorem 1 holds consists of proving it in two steps, traditionally called the *direct coding theorem* and the *converse*. For our case, the *direct coding theorem* proves that the region corresponding to the right-hand side of (11) is an achievable rate region. It constructs an EACQ protocol whose rates are in the region of the right-hand side of (11) and shows that its fidelity of quantum communication is high and its probability of error of classical communication is small. The *converse* assumes that a good code with high fidelity and low probability of error exists and shows that the region on the right-hand side of (11) bounds the achievable rate region. We prove the converse in Section V and the direct coding theorem in Section VI.

[3]The action of the qubit dephasing channel with dephasing parameter p on a density operator ρ is $\rho \to (1-p)\rho + pZ\rho Z$.

A. *Special Cases of the Capacity Theorem*

We first consider five special cases of the above capacity theorem that arise when Q and E both vanish, C and E both vanish, or one of C, Q, or E vanishes. The first two cases correspond respectively to the HSW coding theorem and the LSD coding theorem. Each of the other special cases traces out a 2-D achievable rate region in the 3-D capacity region. The five coding scenarios are as follows.

1) Classical communication (C) when there is no entanglement assistance or quantum communication [1], [2]. The achievable rate region lies on the $(C,0,0)$ ray extending from the origin.
2) Quantum communication (Q) when there is no entanglement assistance or classical communication [4]–[6]. The achievable rate region lies on the $(0,Q,0)$ ray extending from the origin.
3) EAQ communication (QE) when there is no classical communication [13], [14]. The achievable rate region lies in the $(0,Q,E)$ quarter plane of the 3-D region in (11).
4) Classically enhanced quantum communication (CQ) when there is no entanglement assistance [7]. The achievable rate region lies in the $(C,Q,0)$ quarter plane of the 3-D region in (11).
5) EAC communication (CE) when there is no quantum communication [12]. The achievable rate region lies in the $(C,0,E)$ quarter plane of the 3-D region in (11).

1) Classical Capacity: The following theorem gives the 1-D capacity region $\mathcal{C}_{\mathrm{C}}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ for classical communication [1], [2].

Theorem 2: The classical capacity region $\mathcal{C}_{\mathrm{C}}(\mathcal{N})$ is given by

$$\mathcal{C}_{\mathrm{C}}(\mathcal{N}) = \overline{\bigcup_{k=1}^{\infty} \frac{1}{k}\mathcal{C}_{\mathrm{C}}^{(1)}(\mathcal{N}^{\otimes k})}. \quad (17)$$

The "one-shot" region $\mathcal{C}_{\mathrm{C}}^{(1)}(\mathcal{N})$ is the union of the regions $\mathcal{C}_{\mathrm{C},\sigma}^{(1)}(\mathcal{N})$, where $\mathcal{C}_{\mathrm{C},\sigma}^{(1)}(\mathcal{N})$ is the set of all $C \geq 0$, such that

$$C \leq I(X;B)_\sigma + I(A\rangle BX)_\sigma. \quad (18)$$

The entropic quantity is with respect to the state σ^{XABE} in (15).

The bound in (18) is a special case of the bound in (14) with $Q = 0$ and $E = 0$. The above characterization of the classical capacity region may seem slightly different from the original HSW characterization, until we make a few observations. First, we rewrite the coherent information $I(A\rangle BX)_\sigma$ as $H(B\,|\,X)_\sigma - H(E\,|\,X)_\sigma$. Then, $I(X;B)_\sigma + I(A\rangle BX)_\sigma = H(B)_\sigma - H(E\,|\,X)_\sigma$. Next, pure states of the form $|\varphi\rangle_x^{A'}$ are sufficient to attain the classical capacity of a quantum channel [12]. We briefly recall this argument. An ensemble of the following form realizes the classical capacity of a quantum channel:

$$\rho^{XA'} \equiv \sum_x p_X(x)|x\rangle\langle x|^X \otimes \rho_x^{A'}.$$

This ensemble itself is a restriction of the ensemble in (15) to the systems X and A'. Each mixed state $\rho_x^{A'}$ admits a spectral decomposition of the form $\rho_x^{A'} = \sum_y p_{Y\,|\,X}(y\,|\,x)\psi_{x,y}^{A'}$ where $\psi_{x,y}^{A'}$ is a pure state. We can define an augmented classical-quantum state $\theta^{XYA'}$ as follows:

$$\theta^{XYA'} \equiv \sum_{x,y} p_{Y\,|\,X}(y\,|\,x)p_X(x)|x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y \otimes \psi_{x,y}^{A'}$$

so that $\mathrm{Tr}_Y\{\theta^{XYA'}\} = \rho^{XA'}$. Sending the A' system of the states $\rho^{XA'}$ and $\theta^{XYA'}$ leads to the respective states ρ^{XB} and θ^{XYB}. Then, the following equality and inequality hold:

$$\begin{aligned} I(X;B)_\rho &= I(X;B)_\theta \\ &\leq I(XY;B)_\theta \end{aligned}$$

where the equality holds because $\mathrm{Tr}_Y\{\theta^{XYA'}\} = \rho^{XA'}$ and the inequality follows from quantum data processing. Redefining the classical variable as the joint random variable X, Y reveals that it is sufficient to consider pure state ensembles for the classical capacity. Returning to our main argument, then $H(E\,|\,X)_\sigma = H(B\,|\,X)_\sigma$ so that $I(X;B)_\sigma + I(A\rangle BX)_\sigma = H(B)_\sigma - H(B\,|\,X)_\sigma = I(X;B)_\sigma$ for states of this form. Thus, the expression in (18) can never exceed the classical capacity and finds its maximum exactly at the Holevo information.

2) Quantum Capacity: The following theorem gives the 1-D quantum capacity region $\mathcal{C}_{\mathrm{Q}}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ [4]–[6].

Theorem 3: The quantum capacity region $\mathcal{C}_{\mathrm{Q}}(\mathcal{N})$ is given by

$$\mathcal{C}_{\mathrm{Q}}(\mathcal{N}) = \overline{\bigcup_{k=1}^{\infty} \frac{1}{k}\mathcal{C}_{\mathrm{Q}}^{(1)}(\mathcal{N}^{\otimes k})}. \quad (19)$$

The "one-shot" region $\mathcal{C}_{\mathrm{Q}}^{(1)}(\mathcal{N})$ is the union of the regions $\mathcal{C}_{\mathrm{Q},\sigma}^{(1)}(\mathcal{N})$, where $\mathcal{C}_{\mathrm{Q},\sigma}^{(1)}(\mathcal{N})$ is the set of all $Q \geq 0$, such that

$$Q \leq I(A\rangle BX)_\sigma. \quad (20)$$

The entropic quantity is with respect to the state σ^{XABE} in (15) with the restriction that the density $p(x)$ is degenerate.

The bound in (20) is a special case of the bound in (13) with $E = 0$. The other bounds in Theorem 1 are looser than the bound in (13) when $C, E = 0$.

3) EAQ Capacity: The following theorem gives the 2-D EAQ capacity region $\mathcal{C}_{\mathrm{QE}}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ [13], [14].

Theorem 4: The EAQ capacity region $\mathcal{C}_{\mathrm{QE}}(\mathcal{N})$ is given by

$$\mathcal{C}_{\mathrm{QE}}(\mathcal{N}) = \overline{\bigcup_{k=1}^{\infty} \frac{1}{k}\mathcal{C}_{\mathrm{QE}}^{(1)}(\mathcal{N}^{\otimes k})}. \quad (21)$$

The "one-shot" region $\mathcal{C}_{\mathrm{QE}}^{(1)}(\mathcal{N})$ is the union of the regions $\mathcal{C}_{\mathrm{QE},\sigma}^{(1)}(\mathcal{N})$, where $\mathcal{C}_{\mathrm{QE},\sigma}^{(1)}(\mathcal{N})$ is the set of all $Q, E \geq 0$, such that

$$Q \leq \frac{1}{2}I(AX;B)_\sigma \quad (22)$$

$$Q \leq E + I(A\rangle BX)_\sigma. \quad (23)$$

The entropic quantities are with respect to the state σ^{XABE} in (15) with the restriction that the density $p(x)$ is degenerate.

The bounds in (22) and (23) are a special case of the respective bounds in (12) and (13) with $C = 0$. The other bounds in Theorem 1 are looser than the bounds in (12) and (13) when $C = 0$. Observe that the region is a union of general pentagons (see the QE-plane in Fig. 2 for an example of one of these general pentagons in the union).

4) Classically Enhanced Quantum Capacity: The following theorem gives the 2-D capacity region $\mathcal{C}_{\mathrm{CQ}}(\mathcal{N})$ for classically enhanced quantum communication through a quantum channel $\mathcal{N}$ [7].

Theorem 5: The classically enhanced quantum capacity region $\mathcal{C}_{\mathrm{CQ}}(\mathcal{N})$ is given by

$$\mathcal{C}_{\mathrm{CQ}}(\mathcal{N}) = \overline{\bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{C}_{\mathrm{CQ}}^{(1)}(\mathcal{N}^{\otimes k})}. \tag{24}$$

The "one-shot" region $\mathcal{C}_{\mathrm{CQ}}^{(1)}(\mathcal{N})$ is the union of the regions $\mathcal{C}_{\mathrm{CQ},\sigma}^{(1)}(\mathcal{N})$, where $\mathcal{C}_{\mathrm{CQ},\sigma}^{(1)}(\mathcal{N})$ is the set of all $C, Q \geq 0$, such that

$$C + Q \leq I(X;B)_\sigma + I(A\rangle BX)_\sigma \tag{25}$$

$$Q \leq I(A\rangle BX)_\sigma. \tag{26}$$

The entropic quantities are with respect to the state σ^{XABE} in (15).

The bounds in (25) and (26) are a special case of the respective bounds in (13) and (14) with $E = 0$. Observe that the region is a union of trapezoids (see the CQ-plane in Fig. 2 for an example of one of these trapezoids in the union).

The above characterization is a slightly improved characterization of the DS region from [7]. Indeed, the one-shot, one-state region there was a union of rectangles given by the following set of inequalities:

$$C \leq I(X;B)_\sigma \tag{27}$$

$$Q \leq I(A\rangle BX)_\sigma. \tag{28}$$

These rectangles are inside the trapezoids above, though, our characterization in (25)–(26) is the same as theirs when we consider the union over all the one-shot, one-state regions.

5) EAC Capacity With Limited Entanglement:

Theorem 6: The EAC capacity region $\mathcal{C}_{\mathrm{CE}}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is

$$\mathcal{C}_{\mathrm{CE}}(\mathcal{N}) = \overline{\bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{C}_{\mathrm{CE}}^{(1)}(\mathcal{N}^{\otimes k})}. \tag{29}$$

The "one-shot" region $\mathcal{C}_{\mathrm{CE}}^{(1)}(\mathcal{N})$ is the union of the regions $\mathcal{C}_{\mathrm{CE},\sigma}^{(1)}(\mathcal{N})$, where $\mathcal{C}_{\mathrm{CE},\sigma}^{(1)}(\mathcal{N})$ is the set of all $C, E \geq 0$, such that

$$C \leq I(AX;B)_\sigma \tag{30}$$

$$C \leq I(X;B)_\sigma + I(A\rangle BX)_\sigma + E. \tag{31}$$

where the entropic quantities are with respect to the state σ^{XABE} in (15).

The bounds in (30) and (31) are a special case of the respective bounds in (12) and (14) with $Q = 0$. Observe that the region is a union of general polyhedra (see the CE-plane in Fig. 2 for an example of one of these general polyhedra in the union).

The above characterization of the CE achievable region is again an improvement over the characterization in [11], [12], and [14]. It specifies a union of general trapezoids. The region in [11], [12], and [14] was a union of general rectangles of the form

$$C \leq I(AX;B)_\sigma \tag{32}$$

$$E \geq H(A\,|\,X)_\sigma. \tag{33}$$

These general rectangles are inside the above general trapezoids [note that the bounds in (30)–(31) intersect at $E = H(A\,|\,X)_\sigma$], but the regions coincide when we take the union over all the one-shot, one-state regions.

V. The Converse Proof

Our method for proving the converse of Theorem 1 is to apply standard entropic bounds that are available in [8]. We first recall the Fannes inequality for continuity of entropy, the Alicki–Fannes inequality for continuity of coherent information, and another inequality of the Fannes class for continuity of quantum mutual information.

Theorem 7 (Fannes Inequality [27]): Suppose two states ρ^A and σ^A are close

$$\|\rho^A - \sigma^A\|_1 \leq \epsilon.$$

Then, their respective entropies are close

$$|H(A)_\rho - H(A)_\sigma| \leq \epsilon \log |A| + H_2(\epsilon). \tag{34}$$

$|A|$ is the dimension of the system A and $H_2(\epsilon)$ is the binary entropy function that has the property $\lim_{\epsilon\to 0} H_2(\epsilon) = 0$.

Theorem 8 (Alicki–Fannes Inequality [28]): Suppose two states ρ^{AB} and σ^{AB} are close

$$\|\rho^{AB} - \sigma^{AB}\|_1 \leq \epsilon.$$

Then, their respective coherent informations are close

$$|I(A\rangle B)_\rho - I(A\rangle B)_\sigma| \leq 4\epsilon \log |A| + 2H_2(\epsilon). \tag{35}$$

Corollary 1: Suppose two states ρ^{AB} and σ^{AB} are close

$$\|\rho^{AB} - \sigma^{AB}\|_1 \leq \epsilon.$$

Then, their respective quantum mutual informations are close

$$|I(A;B)_\rho - I(A;B)_\sigma| \leq 5\epsilon \log |A| + 3H_2(\epsilon). \tag{36}$$

Proof: The proof follows in two steps by applying Theorems 7 and 8. First, monotonicity of the trace distance under the discarding of subsystems implies that $\|\rho^A - \sigma^A\|_1 \leq \epsilon$. Theorem 7 then applies. The corollary then follows from the equality $I(A;B) = H(A) + I(A\rangle B)$ and the triangle inequality. □

Converse: Section III describes the most general EACQ protocol and this most general case is the one we consider in proving the converse. Suppose Alice shares the maximally entangled state $\Phi^{\hat{R}A_1}$ with the reference system $\hat{R}$ (the protocol should be able to transmit the entanglement in state $\Phi^{\hat{R}A_1}$ with ϵ-accuracy if it can approximately transmit the entanglement with system $\hat{R}$ for any pure state on $\hat{R}$ and A_1). Alice also shares the maximally entangled state $\Phi^{T_A T_B}$ with Bob. Alice combines her system A_1 of the quantum state $\Phi^{\hat{R}A_1}$ with her system T_A of the state $\Phi^{T_A T_B}$ and the classical register M that contains her classical information. The most general encoding operation that she can perform on her three registers M, A_1, and T_A is a conditional quantum encoder $\mathcal{E}^{MA_1T_A\to A'^n}$ consisting of a collection $\{\mathcal{E}_m^{A_1T_A\to A'^n}\}_m$ of CPTP maps. For now, we assume this general form of the encoder but later show in Appendix V that it is only necessary to consider a collection of isometries. Each element $\mathcal{E}_m^{A_1T_A\to A'^n}$ of the conditional quantum encoder produces the following state:

$$\omega_m^{\hat{R}A'^nE'T_B} \equiv U_{\mathcal{E}_m}^{A_1T_A\to A'^nE'}(\Phi^{\hat{R}A_1}\otimes\Phi^{T_AT_B})$$

where we consider the isometric extension $U_{\mathcal{E}_m}^{A_1T_A\to A'^nE'}$ of each element $\mathcal{E}_m^{A_1T_A\to A'^n}$. The average density operator over all classical messages is then as follows:

$$\frac{1}{|M|}\sum_m |m\rangle\langle m|^M\otimes\omega_m^{\hat{R}A'^nE'T_B}.$$

Alice sends the A'^n system through the noisy channel $U_{\mathcal{N}}^{A'^n\to B^nE^n}$, producing the following state:

$$\omega^{M\hat{R}B^nE^nE'T_B} \equiv \frac{1}{|M|}\sum_m |m\rangle\langle m|^M\otimes U_{\mathcal{N}}^{A'^n\to B^nE^n}\left(\omega_m^{\hat{R}A'^nE'T_B}\right). \quad (37)$$

Define $A\equiv\hat{R}T_B$ so that the state in (37) is a particular nth extension of the state in (15). The above state is the state at time t in Fig. 1. Bob receives the above state and performs a decoding instrument $\mathcal{D}^{B^nT_B\to B_1B_E\hat{M}}$. The protocol ends at time t_f. Let $(\omega')^{M\hat{R}B_1B_E\hat{M}E^nE'}$ be the state at time t_f after Bob processes $\omega^{M\hat{R}B^nE^nE'T_B}$ with the decoding instrument $\mathcal{D}^{B^nT_B\to B_1B_E\hat{M}}$. Suppose that an $(n, C-\delta, Q-\delta, E+\delta,\epsilon)$ EACQ protocol as given above exists. We prove that the following bounds apply to the elements of its rate triple $(C-\delta, Q-\delta, E+\delta)$

$$C+2Q-\delta\le\frac{1}{n}I(AM;B^n)_\omega \quad (38)$$

$$Q-\delta\le\frac{1}{n}I(A\rangle B^nM)_\omega+E \quad (39)$$

$$C+Q-\delta\le\frac{1}{n}(I(M;B^n)_\omega+I(A\rangle B^nM)_\omega)+E \quad (40)$$

for any $\epsilon,\delta>0$ and all sufficiently large n. In the ideal case, the identity quantum channel acts on system A_1 to produce the maximally entangled state $\Phi^{\hat{R}B_1}$. So for our case, the following inequality:

$$\|(\omega')^{\hat{R}B_1}-\Phi^{\hat{R}B_1}\|_1\le\epsilon \quad (41)$$

holds because the protocol is ϵ-good for quantum communication according to the criterion in (10). Also, in the ideal case, the identity classical channel acts on system M to produce the maximally correlated state $\bar{\Phi}^{M\hat{M}}$ where

$$\bar{\Phi}^{M\hat{M}}\equiv\frac{1}{|M|}\sum_m|m\rangle\langle m|^M\otimes|m\rangle\langle m|^{\hat{M}}. \quad (42)$$

So for our case, the following inequality:

$$\|(\omega')^{M\hat{M}}-\bar{\Phi}^{M\hat{M}}\|_1\le\epsilon \quad (43)$$

holds because the protocol is ϵ-good for classical communication according to the criterion in (9). We first prove the upper bound in (38) on the classical and quantum rates. Shor's version [12] of the EAC capacity theorem [10], [11] states that the rate $I(AM;B^n)/n$ is achievable and serves as a multiletter upper bound. This bound implies that the unlimited EAQ capacity is $I(AM;B^n)/2n$. If it were not so, then one could convert all of the quantum communication to classical communication by superdense coding and beat the rate $I(AM;B^n)/n$. But this result contradicts the optimality of the unlimited EAC capacity. These two results imply the bounds $C\le I(AM;B^n)/n$ and $2Q\le I(AM;B^n)/n$. But we can go further and prove that the sum rate is bounded as well. Suppose there exists a protocol that beats the sum rate in (38). With more entanglement, one could convert all of the quantum communication to classical communication by superdense coding. But this result again contradicts the optimality of the unlimited EAC capacity. So the bound $C+2Q-\delta\le I(AM;B^n)_\omega/n$ holds. We next prove the upper bound in (39) on the quantum communication rate

$$\begin{aligned}
&n(Q-\delta)\\
&\quad=I(\hat{R}\rangle B_1)_{\Phi^{\hat{R}B_1}}\\
&\quad\le I(\hat{R}\rangle B_1)_{\omega'}+4nQ\epsilon+H_2(\epsilon)\\
&\quad\le I(\hat{R}\rangle B_1M)_{\omega'}+4nQ\epsilon+H_2(\epsilon)\\
&\quad\le I(\hat{R}\rangle B^nT_BM)_\omega+4nQ\epsilon+H_2(\epsilon)\\
&\quad\le I(\hat{R}T_B\rangle B^nM)_\omega+H(T_B|M)_\omega+4nQ\epsilon+H_2(\epsilon)\\
&\quad\le I(A\rangle B^nM)_\omega+nE+4nQ\epsilon+H_2(\epsilon).
\end{aligned} \quad (44)$$

The first equality follows by evaluating the coherent information for the state $\Phi^{\hat{R}B_1}$. The first inequality follows from (41) and the Alicki–Fannes inequality in Theorem 8. The second inequality is from strong subadditivity, and the third inequality is quantum data processing. The fourth inequality follows because $H(T_B|B^nM)\le H(T_B|M)$ (conditioning reduces entropy). The last inequality follows from the definition $A\equiv\hat{R}T_B$ and the fact that $H(T_B\,|\,M)_\omega\le nE$. The inequality in (39) follows by redefining δ as $\delta'\equiv\delta+4Q\epsilon+\frac{H_2(\epsilon)}{n}$. We prove the upper bound in (40) on the classical and quantum rates

$$\begin{aligned}
&n(C+Q-\delta)\\
&\quad=I(M;\hat{M})_{\bar{\Phi}^{M\hat{M}}}+I(\hat{R}\rangle B_1)_{\Phi^{\hat{R}B_1}}\\
&\quad\le I(M;\hat{M})_{\omega'}+I(\hat{R}\rangle B_1)_{\omega'}+5nC\epsilon+4nQ\epsilon+5H_2(\epsilon)\\
&\quad\le I(M;B^nT_B)_{\omega'}+I(\hat{R}\rangle B^nT_BM)_{\omega'}+n\delta'\\
&\quad=I(M;B^n)_{\omega'}+I(M;T_B\,|\,B^n)_{\omega'}+H(B^nT_B\,|\,M)_{\omega'}\\
&\qquad-H(\hat{R}B^nT_B\,|\,M)_{\omega'}+n\delta'
\end{aligned}$$

$$
\begin{aligned}
&= I(M;B^n)_{\omega'} + H(T_B\,|\,B^n)_{\omega'} \\
&\quad + H(B^n\,|\,M)_{\omega'} - H(\hat{R}B^nT_B\,|\,M)_{\omega'} + n\delta' \\
&\le I(M;B^n)_{\omega'} + H(T_B)_{\omega'} + I(\hat{R}T_B\rangle B^nM)_{\omega'} + n\delta' \\
&= I(M;B^n)_{\omega'} + I(A\rangle B^nM)_{\omega'} + nE + n\delta'.
\end{aligned}
$$

The first equality follows because the mutual information $I(M;\hat{M})$ of the maximally correlated state $\bar{\Phi}^{M\hat{M}}$ is equal to nC. The first inequality follows by applying (43) and Corollary 1 to the mutual information $I(M;\hat{M})$, and (41) and the Alicki–Fannes' inequality to the coherent information $I(\hat{R}\rangle B_1)$. The second inequality follows by applying the quantum data processing inequality and strong subadditivity as we did in the proof of the previous bound and by defining $\delta' \equiv 5C\epsilon + 4Q\epsilon + 5H_2(\epsilon)/n$. The second and third equalities follow by manipulating entropies. The third inequality follows from the definition of coherent information and because conditioning does not increase entropy. The last inequality follows from the definition $A \equiv \hat{R}T_B$ and because nE is the maximal value that $H(T_B)$ can take. □

VI. The Direct Coding Theorem

In this section, we prove the direct coding theorem for entanglement-assisted communication of classical and quantum information by giving a combination of strategies that can achieve the rates in Theorem 1. The most important development is the introduction of the CEF protocol and its corresponding proof in the next section. This protocol yields a corner point in the achievable region (see, for example, the point labeled CEF in Fig. 2). Section VI-F shows that combining this protocol with teleportation, superdense coding, and entanglement distribution allows us to obtain all other corner points of the achievable rate region. Thus, this protocol is the most general one available for the channel coding scenario.

A. The CEF Protocol

We can phrase the CEF protocol as a *resource inequality* (see [14] for the theory of resource inequalities)

$$
\begin{aligned}
&\langle \mathcal{N}^{A'\to B}\rangle + \frac{1}{2}I(A;E\,|\,X)_\sigma[qq] \\
&\qquad \ge \frac{1}{2}I(A;B\,|\,X)_\sigma[q\to q] + I(X;B)_\sigma[c\to c]. \quad (45)
\end{aligned}
$$

The precise statement of the CEF resource inequality is a statement of achievability. For any $\epsilon,\delta > 0$ and sufficiently large n, there exists a protocol that consumes n uses of the noisy channel $\mathcal{N}^{A'\to B}$ and consumes $\approx nI(A;E\,|\,X)_\sigma/2$ ebits. In doing so, the protocol communicates $\approx nI(A;B\,|\,X)_\sigma/2$ qubits with $1-\epsilon$ fidelity and $\approx nI(X;B)_\sigma$ classical bits with ϵ probability of error. The entropic quantities are with respect to the state σ^{XABE} in (15).

The proof of the achievability of the CEF protocol proceeds in several steps. We first establish some definitions relevant to an EAQ code, or *father* code for short, and recall the direct coding theorem for EAQ communication [13]–[15]. We then define a *random* father code, give a few relevant definitions and properties, and prove a version of the EAQ coding theorem that applies to random father codes. In particular, we show random father codes exist whose expected channel input is close to a product state (similar to result of the random quantum coding theorem in [6, App. D]). We follow this development by showing how to "paste" random father codes together so that the expected channel input of the pasted random code is close to a product state containing a classical message. A *random CEF code* is then a collection of "pasted" father codes. The closeness of each expected channel input to a product state allows us to apply the HSW coding theorem [1], [2] so that Bob can decode the classical message while causing almost no disturbance to the encoded quantum information. Based on the classical message, Bob determines which random father code he should be decoding for. This method of efficiently coding classical and quantum information is the "piggybacking" technique introduced in [7] and applied again in [25] and [29]. The final arguments consist of a series of Shannon-theoretic arguments of derandomization and expurgation. The result is a *deterministic* CEF code that performs well and achieves the rates in the capacity region in Theorem 1.

B. Father Codes

The unencoded state of a father code is as follows:

$$|\varphi\rangle^{\hat{R}A_1} \otimes |\Phi\rangle^{T_AT_B} \quad (46)$$

where

$$|\varphi\rangle^{\hat{R}A_1} \equiv \sum_{k=1}^{2^{nQ}} \alpha_k |k\rangle^{\hat{R}}|k\rangle^{A_1}$$

$$|\Phi\rangle^{T_AT_B} \equiv \frac{1}{\sqrt{2^{nE}}}\sum_{m=1}^{2^{nE}} |m\rangle^{T_A}|m\rangle^{T_B}.$$

The isometric encoder $\mathcal{E}^{A_1T_A\to A'^n}$ of the father code maps kets on the systems A_1 and T_A as follows:

$$|\phi_{k,m}\rangle^{A'^n} \equiv \mathcal{E}^{A_1T_A\to A'^n}(|k\rangle^{A_1}|m\rangle^{T_A})$$

where the states $|\phi_{k,m}\rangle^{A'^n}$ are mutually orthogonal. Therefore, the encoder $\mathcal{E}^{A_1T_A\to A'^n}$ maps the unencoded state in (46) to the following encoded state:

$$\mathcal{E}^{A_1T_A\to A'^n}(|\varphi\rangle^{\hat{R}A_1}\otimes|\Phi\rangle^{T_AT_B}) = \sum_{k=1}^{2^{nQ}} \alpha_k|k\rangle^{\hat{R}}|\phi_k\rangle^{A'^nT_B}$$

where we define the states $|\phi_k\rangle^{A'^nT_B}$ in the following definition.

Definition 1: The set $\mathcal{C} \equiv \{|\phi_k\rangle^{A'^nT_B}\}_k$ is a representation of the father code. The EAQ *codewords* are as follows:

$$|\phi_k\rangle^{A'^nT_B} \equiv \frac{1}{\sqrt{2^{nE}}}\sum_{m=1}^{2^{nE}} |\phi_{k,m}\rangle^{A'^n}|m\rangle^{T_B}. \quad (47)$$

The *EAQ code density operator* $\rho^{A'^nT_B}(\mathcal{C})$ is a uniform mixture of the EAQ codewords

$$\rho^{A'^nT_B}(\mathcal{C}) \equiv \frac{1}{2^{nQ}}\sum_{k=1}^{2^{nQ}} |\phi_k\rangle\langle\phi_k|^{A'^nT_B}.$$

The *channel input density operator* $\rho^{A'^n}(\mathcal{C})$ is the part of the code density operator $\rho^{A'^nT_B}(\mathcal{C})$ that is input to the channel

$$\rho^{A'^n}(\mathcal{C}) \equiv \mathrm{Tr}_{T_B}\{\rho^{A'^nT_B}(\mathcal{C})\}.$$

The above definitions imply the following two results:

$$\rho^{A'^n T_B}(\mathcal{C}) = \mathcal{E}^{A_1 T_A \to A'^n}(\pi^{A_1} \otimes \Phi^{T_A T_B})$$
$$\rho^{A'^n}(\mathcal{C}) = \frac{1}{2^{n(Q+E)}} \sum_{k=1}^{2^{nQ}} \sum_{m=1}^{2^{nE}} |\phi_{k,m}\rangle\langle\phi_{k,m}|^{A'^n}.$$

The direct coding theorem for EAQ communication gives a method for achieving the multiletter quantum communication rate and entanglement consumption rate.

Proposition 1 (EAQ Coding Theorem): Consider a quantum channel $\mathcal{N}^{A'\to B}$ and its isometric extension $U_{\mathcal{N}}^{A'\to BE}$. For any $\epsilon, \delta > 0$ and all sufficiently large n, there exists an (n, ϵ) EAQ code defined by isometries $(\mathcal{E}, \mathcal{D})$, such that the trace distance between the actual output

$$\left(\mathcal{D}^{B^n T_B \to B_1 B_E} \circ U_{\mathcal{N}}^{A'^n \to B^n E^n} \circ \mathcal{E}^{A_1 T_A \to A'^n}\right)\left(\varphi^{\hat{R}A_1} \otimes \Phi^{T_A T_B}\right)$$

and the ideal decoupled output

$$\varphi^{\hat{R}B_1} \otimes \rho^{E^n B_E} \tag{48}$$

is no larger than ϵ, for any state $\varphi^{\hat{R}A_1}$ with dimension 2^{nQ} in the system A_1 and any maximally entangled $\Phi^{T_A T_B}$ equivalent to nE ebits. The rate of quantum communication is $Q - \delta = \frac{1}{2}I(A;B)_\phi$ provided that the rate of entanglement consumption is at least $E + \delta = \frac{1}{2}I(A;E)_\phi$. The entropic quantities are with respect to the following state:

$$|\phi\rangle^{ABE} \equiv U_{\mathcal{N}}^{A'\to BE}|\psi\rangle^{AA'} \tag{49}$$

where $|\psi\rangle^{AA'}$ is the purification of some state $\rho^{A'}$.

Proof: See [15]. □

C. Random Father Codes

We cannot say much about the channel input density operator $\rho^{A'^n}(\mathcal{C})$ for a particular EAQ code $\mathcal{C}$. But we can say something about the expected channel input density operator of a *random EAQ code* $\mathbb{C}$ (where $\mathcal{C}$ itself becomes a random variable).

Definition 2: A *random EAQ code* is an ensemble $\{p_{\mathcal{C}}, \mathcal{C}\}$ of codes where each code $\mathcal{C}$ occurs with probability $p_{\mathcal{C}}$. The *expected code density operator* $\bar{\rho}^{A'^n T_B}$ is as follows:

$$\bar{\rho}^{A'^n T_B} \equiv \mathbb{E}_{\mathcal{C}}\left\{\rho^{A'^n T_B}(\mathcal{C})\right\}.$$

The *expected channel input density operator* $\bar{\rho}^{A'^n}$ is as follows:

$$\bar{\rho}^{A'^n} \equiv \mathbb{E}_{\mathcal{C}}\left\{\rho^{A'^n}(\mathcal{C})\right\}.$$

A random EAQ code is "ρ-like" if the expected channel input density operator is close to a tensor power of some state ρ:

$$\left\|\bar{\rho}^{A'^n} - \rho^{\otimes n}\right\|_1 \le \epsilon. \tag{50}$$

It follows from the above definition that

$$\bar{\rho}^{A'^n T_B} = \sum_{\mathcal{C}} p_{\mathcal{C}} \rho^{A'^n T_B}(\mathcal{C})$$
$$\bar{\rho}^{A'^n} = \mathrm{Tr}_{T_B}\left\{\bar{\rho}^{A'^n T_B}\right\}.$$

We now state a version of the direct coding theorem that applies to random father codes. The proof shows that we can produce a random father code with an expected channel input density operator close to a tensor power state.

Proposition 2: For any $\epsilon, \delta > 0$ and all sufficiently large n, there exists a random $\rho^{A'}$-like EAQ code for a channel $\mathcal{N}^{A'\to B}$. In particular, the random EAQ code has quantum rate $\frac{1}{2}I(A;B)_\phi - \delta$ and entanglement consumption rate $\frac{1}{2}I(A;E)_\phi + \delta$. The entropic quantities are with respect to the state in (49) and the state $\rho^{A'}$ is that state's restriction to the system A'.

Proof: The proof is in Appendix I. □

D. Associating a Random Father Code With a Classical String

Suppose that we have an ensemble $\{p(x), \rho_x\}_{x\in\mathcal{X}}$ of quantum states. Let $x^n \equiv x_1 \cdots x_n$ denote a classical string generated according to the probability distribution $p(x)$ where each symbol $x_i \in \mathcal{X}$. Then, there is a density operator ρ_{x^n} corresponding to the string x^n where

$$\rho_{x^n} \equiv \bigotimes_{i=1}^{n} \rho_{x_i}.$$

Suppose that we label a random father code by the string x^n and let $\bar{\rho}_{x^n}^{A'^n}$ denote its expected channel input density operator.

Definition 3: A random father code is (ρ_{x^n})-like if the expected channel input density operator $\bar{\rho}_{x^n}^{A'^n}$ is close to the state ρ_{x^n}

$$\left\|\bar{\rho}_{x^n}^{A'^n} - \rho_{x^n}\right\|_1 \le \epsilon.$$

Proposition 3: Suppose we have an ensemble as above. Consider a quantum channel $\mathcal{N}^{A'\to B}$ with its isometric extension $U_{\mathcal{N}}^{A'\to BE}$. Then, there exists a random (ρ_{x^n})-like EAQ code for the channel $\mathcal{N}^{A'\to B}$ for any $\epsilon, \delta > 0$, for all sufficiently large n, and for any classical string x^n in the typical set $T_\delta^{X^n}$ [30]. Its quantum communication rate is $I(A;B\,|\,X)/2 - c'\delta$ and its entanglement consumption rate is $I(A;E\,|\,X)/2 + c''\delta$ for some constants c', c'' where the entropic quantities are with respect to the state in (15) with a trivial system E'. The state ρ_{x^n} is generated from the restriction of the ensemble $\{p(x), \phi_x^{AA'}\}_{x\in\mathcal{X}}$ to the A' system. The states $\phi_x^{AA'}$ in the ensemble correspond to the states $\phi_x^{AA'}$ in (15).

Proof: The method of proof involves "pasting" random father codes together. The proof is in Appendix II. □

E. Construction of a CEF Code

The HSW coding theorem gives an achievable method for sending classical information over a noisy quantum channel. The crucial property that we exploit is that it uses a product-state input for sending classical information. This tensor-product structure is what allows us to "piggyback" classical information onto father codes.

Proposition 4 (HSW Coding Theorem [1], [2]): Consider an input ensemble $\{p(x), \rho_x^{A'}\}$ that gives rise to a classical-quantum state σ^{XB} where

$$\sigma^{XB} \equiv \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|^X \otimes \mathcal{N}^{A'\to B}\left(\rho_x^{A'}\right).$$

Let $C = I(X;B)_\sigma - c'\delta$ for any $\delta > 0$ and for some constant c'. Then, for all $\epsilon > 0$ and for all sufficiently large n, there exists a classical encoding map

$$f : [2^{nC}] \to T_\delta^{X^n}$$

and a decoding positive operator-valued measure (POVM)

$$\Lambda^{B^n} \equiv \left(\Lambda_m^{B^n}\right)_{m\in[2^{nC}]}$$

that allows Bob to decode any classical message $m \in [2^{nC}]$ with high probability

$$\mathrm{Tr}\{\tau_m^{B^n}\Lambda_m^{B^n}\} \geq 1-\epsilon.$$

The density operators $\tau_m^{B^n}$ are the channel outputs

$$\tau_m^{B^n} \equiv \mathcal{N}^{A'^n\to B^n}\left(\rho_{f(m)}^{A'^n}\right) \tag{51}$$

and the channel input states $\rho_{x^n}^{A'^n}$ are a tensor product of states in the ensemble

$$\rho_{x^n}^{A'^n} \equiv \bigotimes_{i=1}^{n} \rho_{x_i}^{A'}.$$

We are now in a position to prove the direct coding part of the CEF capacity theorem. The proof is similar to that in [7].

Direct Coding Theorem: Define the classical message set $[2^{nC}]$, the classical encoding map f, the channel output states $\tau_m^{B^n}$, and the decoding POVM Λ^{B^n} as in Proposition 4. Invoking Proposition 3, we know that for each $m \in [2^{nC}]$, there exists a random $(\rho_{f(m)}^{A'^n})$-like father code $\mathcal{C}_m$ whose probability density is $p_{\mathcal{C}_m}$. The random father code $\mathcal{C}_m$ has encoding–decoding isometry pairs $(\mathcal{E}_{\mathcal{C}_m}^{A_1T_A\to A'^n}, \mathcal{D}_{\mathcal{C}_m}^{B_nT_B\to B_1B_E})$ for each of its realizations. It transmits $n[I(A;B\,|\,X)/2 - c'\delta]$ qubits provided Alice and Bob share at least $n[I(A;E\,|\,X)/2 + c''\delta]$ ebits. Let $\mathcal{C}$ denote the *random CEF code* that is the collection of random father codes $\{\mathcal{C}_m\}_{m\in[2^nC]}$. We first prove that the expectation of the classical error probability for message m is small. The expectation is with respect the random father code $\mathcal{C}_m$. Let $\tau_{\mathcal{C}_m}^{B^n}$ denote the *channel output density operator* corresponding to the father code $\mathcal{C}_m$

$$\tau_{\mathcal{C}_m}^{B^n} \equiv \mathrm{Tr}_{T_B}\left\{\mathcal{N}^{A'^n\to B^n}\left(\mathcal{E}_{\mathcal{C}_m}^{A_1T_A\to A'^n}(\pi^{A_1}\otimes\Phi^{T_AT_B})\right)\right\}.$$

Let $\bar{\tau}_m^{B^n}$ denote the *expected channel output density operator* of the random father code $\mathcal{C}_m$

$$\bar{\tau}_m^{B^n} \equiv \mathbb{E}_{\mathcal{C}_m}\left\{\tau_{\mathcal{C}_m}^{B^n}\right\} = \sum_{\mathcal{C}_m} p_{\mathcal{C}_m}\tau_{\mathcal{C}_m}^{B^n}.$$

The following inequality holds:

$$\left\|\bar{\rho}_{f(m)}^{A'^n} - \rho_{f(m)}^{A'^n}\right\|_1 \leq |\mathcal{X}|\epsilon$$

because the random father code $\mathcal{C}_m$ is $(\rho_{f(m)}^{A'^n})$-like. Then, the expected channel output density operator $\bar{\tau}_m^{B^n}$ is close to the tensor product state $\tau_m^{B^n}$ in (51)

$$\left\|\bar{\tau}_m^{B^n} - \tau_m^{B^n}\right\|_1 \leq |\mathcal{X}|\epsilon \tag{52}$$

because the trace distance is monotone under the quantum operation $\mathcal{N}^{A'^n\to B^n}$. It then follows that the POVM element $\Lambda_m^{B^n}$ has a high probability of detecting the expected channel output density operator $\bar{\tau}_m^{B^n}$

$$\begin{aligned}\mathrm{Tr}\left\{\Lambda_m^{B^n}\bar{\tau}_m^{B^n}\right\} &\geq \mathrm{Tr}\left\{\Lambda_m^{B^n}\tau_m^{B^n}\right\} - \left\|\bar{\tau}_m^{B^n} - \tau_m^{B^n}\right\|_1 \\ &\geq 1-\epsilon-|\mathcal{X}|\epsilon.\end{aligned} \tag{53}$$

The first inequality follows from the following lemma [21] that holds for any two quantum states ρ and σ and a positive operator Π where $0 \leq \Pi \leq I$

$$\mathrm{Tr}\{\Pi\rho\} \geq \mathrm{Tr}\{\Pi\sigma\} - \|\rho-\sigma\|_1.$$

The second inequality follows from Proposition 4 and (52). We define Bob's decoding instrument $\mathcal{D}_{\mathcal{C}}^{B^nT_B\to B_1B_E\hat{M}}$ for the random CEF code $\mathcal{C}$ as follows:

$$\mathcal{D}_{\mathcal{C}}^{B^nT_B\to B_1B_E\hat{M}}\left(\rho^{B^nT_B}\right) \equiv \sum_m \mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}\left(\sqrt{\Lambda_m^{B^n}}\rho^{B^nT_B}\sqrt{\Lambda_m^{B^n}}\right)\otimes|m\rangle\langle m|^{\hat{M}}$$

where $\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}$ is the decoding isometry for the father code $\mathcal{C}_m$ and each map $\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}(\sqrt{\Lambda_m^{B^n}}\rho\sqrt{\Lambda_m^{B^n}})$ is trace reducing. The induced quantum operation corresponding to this instrument is as follows:

$$\mathcal{D}_{\mathcal{C}}^{B^nT_B\to B_1B_E}(\rho) = \mathrm{Tr}_{\hat{M}}\left\{\mathcal{D}_{\mathcal{C}}^{B^nT_B\to B_1B_E\hat{M}}(\rho)\right\}.$$

Let $p_e(\mathcal{C}_m)$ denote the classical error probability for each classical message m of the CEF code $\mathcal{C}$

$$\begin{aligned}p_e(\mathcal{C}_m) &\equiv 1 - \Pr\{M'=m\,|\,M=m\} \\ &= 1 - \mathrm{Tr}\left\{\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}\left(\sqrt{\Lambda_m^{B^n}}\tau_{\mathcal{C}_m}^{B^n}\sqrt{\Lambda_m^{B^n}}\right)\right\}.\end{aligned}$$

Then, by the above definition, (53), and the fact that the trace does not change under the isometry $\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}$, it holds that the expectation of the classical error probability $p_e(\mathcal{C}_m)$ with respect to the random father code $\mathcal{C}_m$ is low

$$\mathbb{E}_{\mathcal{C}_m}\{p_e(\mathcal{C}_m)\} \leq (1+|\mathcal{X}|)\epsilon. \tag{54}$$

We now prove that the expectation of the quantum error is small (the expectation is with respect to the random father code $\mathcal{C}_m$). Input the state $\Phi^{\hat{R}A_1}\otimes\Phi^{T_AT_B}$ to the encoder $\mathcal{E}_{\mathcal{C}_m}^{A_1T_A\to A'^n}$ followed by the channel $\mathcal{N}^{A'^n\to B^n}$. The resulting state is an extension $\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}$ of $\tau_{\mathcal{C}_m}^{B^n}$

$$\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n} \equiv \mathcal{N}^{A'^n\to B^n}\left(\mathcal{E}_{\mathcal{C}_m}^{A_1T_A\to A'^n}(\Phi^{\hat{R}A_1}\otimes\Phi^{T_AT_B})\right).$$

Let $\bar{\Omega}_m^{\hat{R}T_BB^n}$ denote the expectation of $\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}$ with respect to the random code $\mathcal{C}_m$

$$\bar{\Omega}_m^{\hat{R}T_BB^n} \equiv \mathbb{E}_{\mathcal{C}_m}\left\{\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\right\}.$$

It follows that $\bar{\Omega}_m^{\hat{R}T_BB^n}$ is an extension of $\bar{\tau}_m^{B^n}$. The following inequality follows from (53):

$$\mathrm{Tr}\left\{\bar{\Omega}_m^{\hat{R}T_BB^n}\Lambda_m^{B^n}\right\} \geq 1-(1+|\mathcal{X}|)\epsilon. \tag{55}$$

The above inequality is then sufficient for us to apply a modified version of the gentle measurement lemma (Lemma 1 in Appendix III) so that the following inequality holds:

$$\mathbb{E}_{\mathcal{C}_m}\left\{\left\|\sqrt{\Lambda_m^{B^n}}\,\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\sqrt{\Lambda_m^{B^n}}-\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\right\|_1\right\}\leq\sqrt{8(1+|\mathcal{X}|)\epsilon}. \quad (56)$$

Monotonicity of the trace distance gives an inequality for the trace-reducing maps of the quantum decoding instrument

$$\mathbb{E}_{\mathcal{C}_m}\left\{\left\|\begin{array}{c}\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}\left(\sqrt{\Lambda_m^{B^n}}\,\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\sqrt{\Lambda_m^{B^n}}\right)\\ -\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}\left(\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\right)\end{array}\right\|_1\right\} \leq\sqrt{8(1+|\mathcal{X}|)\epsilon}. \quad (57)$$

The following inequality also holds:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_m}\left\{\left\|\begin{array}{c}\mathcal{D}_{\mathcal{C}}^{B^nT_B\to B_1B_E}\left(\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\right)\\ -\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}\left(\sqrt{\Lambda_m^{B^n}}\,\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\sqrt{\Lambda_m^{B^n}}\right)\end{array}\right\|_1\right\}\\
&\leq\mathbb{E}_{\mathcal{C}_m}\left\{\sum_{m'\neq m}\left\|\mathcal{D}_{\mathcal{C}_{m'}}^{B^nT_B\to B_1B_E}\left(\sqrt{\Lambda_{m'}^{B^n}}\,\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\sqrt{\Lambda_{m'}^{B^n}}\right)\right\|_1\right\}\\
&=\mathbb{E}_{\mathcal{C}_m}\left\{\sum_{m'\neq m}\left\|\sqrt{\Lambda_{m'}^{B^n}}\,\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\sqrt{\Lambda_{m'}^{B^n}}\right\|_1\right\}\\
&=\mathbb{E}_{\mathcal{C}_m}\left\{\sum_{m'\neq m}\mathrm{Tr}\left\{\Lambda_{m'}^{B^n}\,\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\right\}\right\}\\
&=1-\mathrm{Tr}\left\{\Lambda_m^{B^n}\,\bar{\Omega}_m^{\hat{R}T_BB^n}\right\}\\
&\leq(1+|\mathcal{X}|)\epsilon.
\end{aligned} \quad (58)$$

The first inequality follows from definitions and the triangle inequality. The first equality follows because the trace distance is invariant under isometry. The second equality follows because the operator $\Lambda_m^{B^n}\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}$ is positive. The third equality follows from some algebra, and the second inequality follows from (53). The fidelity of quantum communication for all classical messages m and codes $\mathcal{C}_m$ is high

$$F\left(\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}\left(\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\right),\Phi^{\hat{R}B_1}\right)\geq1-\epsilon$$

because each code $\mathcal{C}_m$ in the random father code is good for quantum communication. It then follows that

$$\mathbb{E}_{\mathcal{C}_m}\left\{\left\|\mathcal{D}_{\mathcal{C}_m}^{B^nT_B\to B_1B_E}\left(\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\right)-\Phi^{\hat{R}B_1}\right\|_1\right\}\leq2\sqrt{\epsilon} \quad (59)$$

because of the relation between the trace distance and fidelity [21]. Application of the triangle inequality to (59), (58), and (57) gives the following bound on the expected quantum error:

$$\mathbb{E}_{\mathcal{C}_m}\{q_e(\mathcal{C}_m)\}\leq\epsilon' \quad (60)$$

where

$$\epsilon'\equiv(1+|\mathcal{X}|)\epsilon+\sqrt{8(1+|\mathcal{X}|)\epsilon}+2\sqrt{\epsilon}$$

and where we define the quantum error $q_e(\mathcal{C}_m)$ of the code $\mathcal{C}_m$ as follows:

$$q_e\left(\mathcal{C}_m\right)\equiv\left\|\mathcal{D}_{\mathcal{C}}^{B^nT_B\to B_1B_E}\left(\Omega_{\mathcal{C}_m}^{\hat{R}T_BB^n}\right)-\Phi^{\hat{R}B_1}\right\|_1.$$

The above random CEF code relies on Alice and Bob having access to a source of common randomness. We now show that they can eliminate the need for common randomness and select a good CEF code $\mathcal{C}$ that has a low quantum error $q_e(\mathcal{C}_m)$ and low classical error $p_e(\mathcal{C}_m)$ for all classical messages m in a large subset of $[2^{nC}]$. By the bounds in (54) and (60), the following bound holds for the expectation of the averaged summed error probabilities:

$$\mathbb{E}_{\mathcal{C}_m}\left\{\frac{1}{2^{nC}}\sum_m p_e\left(\mathcal{C}_m\right)+q_e\left(\mathcal{C}_m\right)\right\}\leq\epsilon'+(1+|\mathcal{X}|)\epsilon.$$

If the above bound holds for the expectation over all random codes, it follows that there exists a particular CEF code $\mathcal{C}=\{\mathcal{C}_m\}_{m\in[2^{nC}]}$ with the following bound on its averaged summed error probabilities:

$$\frac{1}{2^{nC}}\sum_m p_e\left(\mathcal{C}_m\right)+q_e\left(\mathcal{C}_m\right)\leq\epsilon'+(1+|\mathcal{X}|)\epsilon.$$

We fix the code $\mathcal{C}$ and expurgate the worst half of the father codes—those father codes with classical messages m that have the highest value of $p_e(\mathcal{C}_m)+q_e(\mathcal{C}_m)$. This derandomization and expurgation yields a CEF code that has each classical error $p_e(\mathcal{C}_m)$ and each quantum error $q_e(\mathcal{C}_m)$ upper bounded by $2(\epsilon'+(1+|\mathcal{X}|)\epsilon)$ for the remaining classical messages m. This expurgation decreases the classical rate by a negligible factor of $\frac{1}{n}$. □

Note that the above proof is a scheme for entanglement transmission. This task is equivalent to the task of subspace transmission (quantum communication) by the methods in [31].

F. Child Protocols

We detail five protocols that are children of the CEF protocol in the sense of [14]. Recall the CEF resource inequality in (45). Recall the three respective unit resource inequalities for teleportation, super-dense coding, and entanglement distribution

$$2[c\to c]+[qq]\geq[q\to q] \quad (61)$$

$$[qq]+[q\to q]\geq2[c\to c] \quad (62)$$

$$[q\to q]\geq[qq]. \quad (63)$$

We can first append entanglement distribution to the CEF resource inequality. This appending gives rise to the classically enhanced quantum communication protocol in [7]. The development proceeds as follows:

$$\begin{aligned}
&\langle\mathcal{N}^{A'\to B}\rangle+\frac{1}{2}I(A;E\,|\,X)[qq]\\
&\geq\frac{1}{2}I(A;B\,|\,X)[q\to q]+I(X;B)[c\to c]\\
&=\left(\frac{1}{2}I(A;E\,|\,X)+I(A\rangle BX)\right)[q\to q]+I(X;B)[c\to c]\\
&\geq\frac{1}{2}I(A;E\,|\,X)[qq]+I(A\rangle BX)[q\to q]+I(X;B)[c\to c]
\end{aligned}$$

where the first inequality is the CEF resource inequality, the first equality exploits the identity in (2), and the last inequality follows from entanglement distribution. By the cancellation lemma [14, Lemma 4.6], the following resource inequality holds:

$$\langle \mathcal{N}^{A'\to B}\rangle + o[qq] \geq I(A\rangle BX)[q\to q] + I(X;B)[c\to c] \tag{64}$$

where $o[qq]$ represents a sublinear amount of entanglement. The above resource inequality is equivalent to the classically enhanced quantum communication protocol in [7] (modulo the sublinear entanglement). Combining the above resource inequality further with entanglement distribution gives the classically enhanced entanglement generation protocol from [7]

$$\langle \mathcal{N}^{A'\to B}\rangle + o[qq] \geq I(A\rangle BX)[qq] + I(X;B)[c\to c].$$

We can combine the CEF protocol with superdense coding and entanglement distribution. Let CEF-SD-ED denote the resulting protocol. The development proceeds by first using qubits at a rate $\frac{1}{2}H(A\,|\,X)$ for entanglement distribution

$$\begin{aligned}
&\langle \mathcal{N}^{A'\to B}\rangle + \frac{1}{2}I(A;E\,|\,X)[qq]\\
&\quad\geq \frac{1}{2}I(A;B\,|\,X)[q\to q] + I(X;B)[c\to c]\\
&\quad= \left(\frac{1}{2}H(A\,|\,X)+\frac{1}{2}I(A\rangle BX)\right)[q\to q]+I(X;B)[c\to c]\\
&\quad\geq \frac{1}{2}H(A\,|\,X)[qq]+\frac{1}{2}I(A\rangle BX)[q\to q]+I(X;B)[c\to c].
\end{aligned}$$

After this step, the above protocol is equivalent to the following one:

$$\langle \mathcal{N}\rangle + o[qq] \geq \frac{1}{2}I(A\rangle BX)([qq]+[q\to q]) + I(X;B)[c\to c]$$

so that it has generated entanglement at a net rate of $\frac{1}{2}I(A\rangle BX)$ ebits. We can then further combine with superdense coding to achieve the protocol CEF-SD-ED

$$\langle \mathcal{N}^{A'\to B}\rangle + o[qq] \geq I(A\rangle BX)[c\to c] + I(X;B)[c\to c].$$

We can combine the CEF protocol with superdense coding to get Shor's EAC communication protocol [12]

$$\begin{aligned}
&\langle \mathcal{N}^{A'\to B}\rangle + H(A\,|\,X)[qq]\\
&\quad= \langle \mathcal{N}^{A'\to B}\rangle + \frac{1}{2}I(A;E\,|\,X)[qq] + \frac{1}{2}I(A;B\,|\,X)[qq]\\
&\quad\geq \frac{1}{2}I(A;B\,|\,X)[q\to q]+I(X;B)[c\to c]+\frac{1}{2}I(A;B\,|\,X)[qq]\\
&\quad\geq I(X;B)[c\to c] + I(A;B\,|\,X)[c\to c]\\
&\quad= I(AX;B)[c\to c].
\end{aligned}\tag{65}$$

The first equality uses the identity in (1). The first inequality uses the CEF resource inequality. The second inequality uses superdense coding, and the last equality uses the chain-rule identity in (4). The above rates are the same as those in [12] and [14].

Teleportation is the last unit resource inequality with which we can combine the CEF protocol. Let CEF-TP (classically enhanced father combined with teleportation) denote the resulting protocol. Consider that the CEF protocol generates classical communication at a rate $I(X;B)$. Alice and Bob can teleport quantum information if they have an extra $I(X;B)/2$ ebits of entanglement. The development proceeds as follows:

$$\begin{aligned}
&\langle \mathcal{N}^{A'\to B}\rangle + \frac{1}{2}I(A;E\,|\,X)[qq] + \frac{1}{2}I(X;B)[qq]\\
&\quad\geq \frac{1}{2}I(A;B\,|\,X)[q\to q]+I(X;B)[c\to c]+\frac{1}{2}I(X;B)[qq]\\
&\quad\geq \frac{1}{2}I(A;B\,|\,X)[q\to q] + \frac{1}{2}I(X;B)[q\to q]\\
&\quad= \frac{1}{2}I(AX;B)[q\to q].
\end{aligned}$$

We apply teleportation to get the second inequality and the chain rule in (4) to get the last equality. We can rewrite the above protocol as follows:

$$\langle \mathcal{N}^{A'\to B}\rangle + \frac{1}{2}(I(A;E\,|\,X) + I(X;B))[qq] \geq \frac{1}{2}I(AX;B)[q\to q].$$

This protocol is the same as the father protocol if random variable X has a degenerate distribution.

VII. Single-Letter Examples

Theorem 1 is a general theorem that determines the capacity region of any entanglement-assisted channel for classical and quantum communication. Unfortunately, the theorem is of a multiletter nature, implying that it is an intractable problem to compute the capacity region corresponding to an arbitrary channel.

In the forthcoming subsections, we provide several examples of channels for which we can exactly compute their corresponding capacity regions. The first example is the trivial completely depolarizing channel (the channel that replaces the input state with the maximally mixed state). We find this example interesting despite its triviality because it coincides with our results in [26]. The second example is the quantum erasure channel [32]. The advantage of the quantum erasure channel is that we can apply simple reasoning to determine the outer bound of its corresponding capacity region. We then show that the inner bound corresponding to the achievable region of this channel matches the outer bound. Thus, we know the full capacity region for the quantum erasure channel. The final channel that we single-letterize is the qubit dephasing channel. Perhaps surprisingly, we are able to do so by arguing that the DS CQ region and the Shor CE region each single-letterize.

A. The Completely Depolarizing Channel

The first single-letter example that we consider is the completely depolarizing channel. This channel simply replaces the input state with the maximally mixed state. Therefore, no classical or quantum information can traverse it, even with the help of entanglement.

Corollary 2: The following set of inequalities specifies the entanglement-assisted capacity of the completely depolarizing channel:

$$\begin{aligned}
C + 2Q &\leq 0\\
Q &\leq E\\
C + Q &\leq E.
\end{aligned}$$

Proof: The proof follows by considering that the mutual information $I(AX;B)$ and the Holevo information $I(X;B)$ in Theorem 1 vanish for any k-qudit state transmitted through the completely depolarizing channel and the coherent information is either negative or zero for any input state. Then, the inequalities (12)–(14) there become the respective inequalities above. □

One should observe that the region is actually trivial (it is empty) because $C+2Q\leq 0$. Nevertheless, we still find the inequalities in Corollary 2 interesting because they coincide with those that we found in [26] for the "unit resource capacity" region[4] (modulo a different sign convention with the entanglement rate E). The proof techniques in [26] involve *reductio ad absurdum* arguments that show how points outside the region conflict with physical law, whereas the arguments in the converse proof of Theorem 1 are information theoretic. One should expect that the set of inequalities in Corollary 2 coincide with those for the unit resource capacity region because having access to the completely depolarizing channel is equivalent to having no quantum channel at all—Bob can actually simulate this resource locally merely by preparing the maximally mixed state in his laboratory.

B. The Quantum Erasure Channel

The quantum erasure channel is perhaps one of the simplest noisy quantum channels [32], because it has a simple specification and its known transmission capacities admit simple formulas [33]. A quantum erasure channel passes the input state along to the environment and gives Bob an erasure state $|e\rangle$ with probability ϵ. It passes the input state along to Bob and gives the environment an erasure state $|e\rangle$ with probability $1-\epsilon$. It induces the following map on a density operator $\rho^{A'}$

$$\rho^{A'}\rightarrow(1-\epsilon)\rho^{B}+\epsilon|e\rangle\langle e|^{B}$$

and its isometric extension acts as follows:

$$|\psi\rangle^{AA'}\rightarrow\sqrt{1-\epsilon}|\psi\rangle^{AB}|e\rangle^{E}+\sqrt{\epsilon}|\psi\rangle^{AE}|e\rangle^{B}$$

where $|\psi\rangle^{AA'}$ is some purification of $\rho^{A'}$.

Table I lists the known optimal transmission capacities for the quantum erasure channel. Bennett *et al.* determined the classical capacity of the quantum erasure channel with an intuitive argument (the outer bound exploits the Holevo bound [8] and the inner bound uses an encoding with orthogonal states), and they determined its quantum capacity with a different intuitive argument (the well-known no-cloning argument combined with linear interpolation for the outer bound and one-way random hashing for the inner bound [33]). The optimality of the classical rate $2(1-\epsilon)$ and the quantum rate $1-\epsilon$ of an EAQ erasure channel follows from the arguments in [10]. The optimality of the respective entanglement consumption rates follows from our forthcoming arguments. Finally, note that we can obtain the quantum capacity result by combining the father protocol (EAQ communication) with entanglement distribution at a rate ϵ, and we can obtain EAC communication from EAQ communication by consuming all of its quantum communication at rate $1-\epsilon$ with superdense coding.

[4]The unit resource capacity region is the set of rates that are achievable without the aid of a noisy resource.

TABLE I
THE LEFT COLUMN GIVES A PARTICULAR TYPE OF CAPACITY FOR THE QUANTUM ERASURE CHANNEL, AND THE RIGHT COLUMN GIVES THE CORRESPONDING OPTIMAL RATE TRIPLE

Capacity	Rate Triple (C,Q,E)
Entanglement-assisted classical capacity (EAC)	$(2(1-\epsilon),0,1)$
Quantum capacity (LSD)	$(0,1-2\epsilon,0)$
Classical capacity (HSW)	$(1-\epsilon,0,0)$
Entanglement-assisted quantum capacity (EAQ)	$(0,1-\epsilon,\epsilon)$

Corollary 3 shows that the CQE capacity region of a quantum erasure channel admits a simple characterization in terms of three inequalities. We prove the converse by intuitive reasoning that one would perhaps expect to be able to apply to the quantum erasure channel, given earlier intuitive reasoning that authors have applied to this channel. We prove the direct coding theorem by giving an explicit ensemble that reaches all of the bounds in the inequalities in Corollary 3. The result is that timesharing[5] between the four protocols in Table I is the optimal coding strategy.

Corollary 3: Suppose a quantum erasure channel has an erasure probability ϵ. The following set of inequalities specifies the capacity region of this entanglement-assisted channel for transmitting classical and quantum information:

$$C+2Q\leq 2(1-\epsilon) \tag{66}$$

$$\frac{1-2\epsilon}{1-\epsilon}C+Q\leq E+(1-2\epsilon) \tag{67}$$

$$C+(1+\epsilon)Q\leq(1-\epsilon)(1+E). \tag{68}$$

Proof: We first prove the converse. The first bound in (66) holds because the sum rate $C+2Q$ can never exceed $2(1-\epsilon)$. Otherwise, one could beat the EAC capacity by dense coding or one could beat the EAQ capacity by teleportation. We next consider the second bound in (67). We first prove that timesharing between the HSW point and the LSD point is an optimal strategy[6] and then show that this result implies the bound in (67). Consider a scheme of quantum error correction for an erasure channel with erasure parameter ϵ. If Alice transmits n qubits, then Bob receives $n(1-\epsilon)$ of these and the environment receives $n\epsilon$ of them (for the case of large n). From these $n(1-\epsilon)$ physical qubits, Bob can perform a decoding to obtain $n(1-2\epsilon)$ logical qubits, by the quantum capacity result for the erasure channel. This implies an optimal "decoding ratio" of $n(1-2\epsilon)$ decoded qubits for the $n(1-\epsilon)$ received qubits: $(1-2\epsilon)/(1-\epsilon)$. Now let us consider a DS-like code for the erasure channel. Suppose that Alice can achieve the rate triple $(\lambda(1-\epsilon),(1-\lambda)(1-2\epsilon)+\delta,0)$

[5]Timesharing is a simple method of combining coding strategies [30]. As an example, consider the case of timesharing a channel between an (n,Q_1,ϵ) quantum code and another (n,Q_2,ϵ) quantum code. For any λ where $0<\lambda<1$, the sender uses the first code for a fraction λ of the channel uses and uses the other code for a fraction $(1-\lambda)$ of the channel uses. This timesharing strategy produces a quantum code with rate $\lambda Q_1+(1-\lambda)Q_2$ and error at most 2ϵ. Timesharing immediately gives that the convex hull of any set of achievable points is an achievable region.

[6]Devetak and Shor stated (but did not explicitly prove) that timesharing between HSW and LSD is optimal for the erasure channel [7].

where δ is some small positive number (so that this rate triple represents any point that beats the timesharing limit). Now if Alice transmits n qubits, Bob receives $n(1-\epsilon)$ of them and the environment again receives $n\epsilon$ of them. But this time, Bob performs measurements on $n\lambda(1-\epsilon)$ of them in order to obtain the classical information. Thus, these qubits are no longer available for decoding quantum information because the measurements completely dephase them. This leaves $n(1-\epsilon)-n\lambda(1-\epsilon)=n(1-\lambda)(1-\epsilon)$ qubits available for decoding the quantum information. If Bob could decode $n((1-\lambda)(1-2\epsilon)+\delta)$ logical qubits, this would contradict the optimality of the above "decoding ratio" because $n((1-\lambda)(1-2\epsilon)+\delta)/(n(1-\lambda)(1-\epsilon))=(1-2\epsilon)/(1-\epsilon)+\delta/(1-\lambda)(1-\epsilon)$ is greater than the optimal decoding ratio $(1-2\epsilon)/(1-\epsilon)$. Therefore, he must only be able to decode $n(1-\lambda)(1-2\epsilon)$ logical qubits. This proves that timesharing between HSW and LSD is an optimal strategy for the quantum erasure channel. Now, the capacity region excludes any point lying above the CQ-plane with which we can combine with entanglement distribution to reach a point on the CQ-plane outside the DS timesharing bound (otherwise, we would be able to beat the timesharing bound between HSW and LSD by combining this point with entanglement distribution). In particular, this means that achievable points cannot be outside the plane containing the vector connecting LSD to HSW and the vector of entanglement distribution. It is straightforward to calculate the equation for this plane. The vector connecting LSD to HSW is

$$(0,1-2\epsilon,0)-(1-\epsilon,0,0)=(-(1-\epsilon),1-2\epsilon,0).$$

The vector of entanglement distribution is $(0,-1,-1)$. A normal vector for the plane containing the two vectors is

$$\left(-\frac{1-2\epsilon}{1-\epsilon},-1,1\right).$$

Then, the equation for the plane is

$$-\frac{1-2\epsilon}{1-\epsilon}(C-(1-\epsilon))-Q+E=0$$

implying that achievable points must obey the bound in (67) because they cannot lie outside this plane. The above argument also shows that the EAQ rate triple $(0,1-\epsilon,\epsilon)$ is optimal (in particular, that the entanglement consumption rate is optimal) because it lies at the intersection of the two bounds in (66) and (67). We now prove the last bound in (68) in three steps. We first prove that the entanglement consumption rate of the EAC protocol is optimal. We then prove that timesharing between EAC and HSW is optimal, and finally rule out all points outside a plane containing the vector connecting EAC to HSW and the vector of superdense coding. Consider the EAC rate triple $(2(1-\epsilon),0,1)$. The entanglement consumption rate of one ebit per channel use is optimal, i.e., one cannot achieve the classical rate of $2(1-\epsilon)$ with less than one ebit per channel use. The state that achieves capacity is the maximally entangled state $|\Phi^+\rangle$. The minimum amount of entanglement that this capacity-achieving state requires is $H(A)=1$ ebit (we give a more detailed proof in Appendix IV). Thus, no lower amount of entanglement could suffice for achieving the maximal classical rate. We now prove that timesharing between EAC and HSW is optimal by an argument similar to the argument for our other timesharing bound. Consider a scheme of EAC communication for an erasure channel with erasure parameter ϵ. If Alice transmits n qubits (that could potentially be entangled with n qubits of Bob's), then Bob receives $n(1-\epsilon)$ of these and the environment receives $n\epsilon$ of them (for the case of large n). From these $n(1-\epsilon)$ physical qubits (and his halves of the ebits), Bob can perform a decoding to obtain $n2(1-\epsilon)$ classical bits, by the EAC capacity result for the erasure channel. This implies an optimal "EA decoding ratio" of $n2(1-\epsilon)$ decoded bits for the $n(1-\epsilon)$ received qubits: $2(1-\epsilon)/(1-\epsilon)$. Now let us consider a Shor-like code[7] for the erasure channel. Suppose that Alice can achieve the rate triple $(\lambda(1-\epsilon)+(1-\lambda)2(1-\epsilon)+\delta,0,1-\lambda)$ where δ is some small positive number (so that this rate triple represents any point that beats the timesharing limit). Now if Alice transmits n qubits, then Bob receives $n(1-\epsilon)$ of them and the environment again receives $n\epsilon$ of them. But this time, Bob performs some measurement on $n\lambda(1-\epsilon)$ of them in order to obtain some of the classical information. Thus, these qubits are no longer available for decoding any more classical information because they have already been decoded. This leaves $n(1-\epsilon)-n\lambda(1-\epsilon)=n(1-\lambda)(1-\epsilon)$ qubits available for decoding the extra classical information. If Bob could decode $n((1-\lambda)2(1-\epsilon)+\delta)$ extra classical bits, this would contradict the optimality of the above "EA decoding ratio" because $n((1-\lambda)2(1-\epsilon)+\delta)/(n(1-\lambda)(1-\epsilon))=2(1-\epsilon)/(1-\epsilon)+\delta/(1-\lambda)(1-\epsilon)$ is greater than the optimal decoding ratio $2(1-\epsilon)/(1-\epsilon)$. Therefore, he must only be able to decode $n(1-\lambda)2(1-\epsilon)$ classical bits. This proves that timesharing between HSW and EAC is an optimal strategy for the quantum erasure channel. Now, the capacity region excludes any point lying to the right of the CE-plane with which we can combine with superdense coding to reach a point on the CE-plane outside the timesharing bound (otherwise, we would be able to beat the timesharing bound between HSW and EAC by combining this point with superdense coding). In particular, this means that achievable points cannot be outside the plane containing the vector connecting HSW to EAC and the vector of superdense coding. It is straightforward to calculate the equation for this plane. Consider that the vector between EAC and HSW is

$$(2(1-\epsilon),0,1)-(1-\epsilon,0,0)=(1-\epsilon,0,1).$$

The vector of dense coding is $(2,-1,1)$. A normal vector for this plane is

$$(-1,-(1+\epsilon),1-\epsilon).$$

The equation for the plane is

$$-(C-(1-\epsilon))-(1+\epsilon)Q+(1-\epsilon)E=0$$

implying that achievable points must obey the bound in (68) because they cannot lie outside this plane. We have now completed the proof of the outer bound. We prove the direct coding theorem. The simple way to prove it follows simply by timesharing between the four protocols HSW, LSD, EAQ, and EAC, but it is interesting to explore a particular ensemble of states of the

[7] "Shor-like" in the sense of [12].

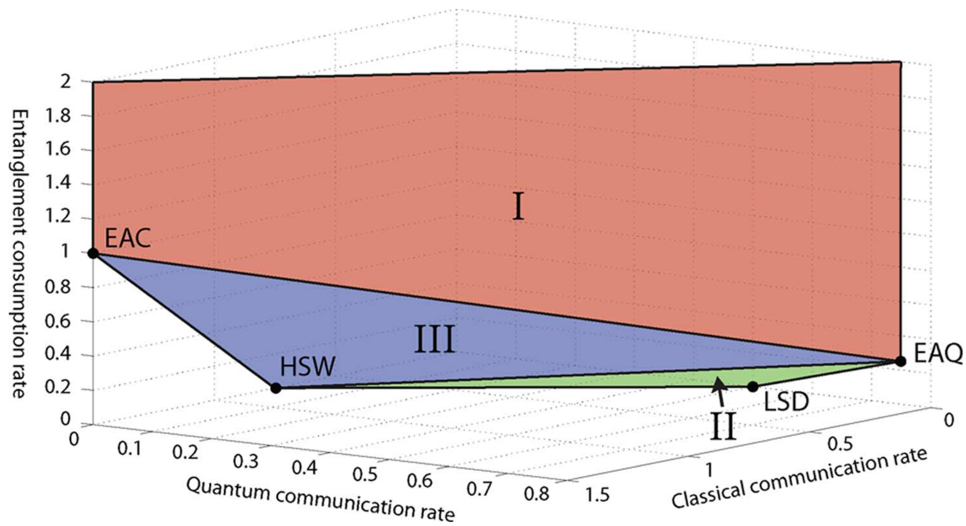

Fig. 3. Capacity region of the quantum erasure channel with erasure parameter $\epsilon = 1/4$. Planes I–III correspond to the respective bounds in (66)–(68). The optimal strategy is to timeshare between classical coding (HSW), quantum coding (LSD), EAQ coding, and EAC coding. The CEF protocol does not give any improvement over timesharing for a quantum erasure channel.

form (15) in Theorem 1 that achieves the capacity. We consider transmitting the A' system of the following classical-quantum state through the channel:

$$\sigma^{XAA'} \equiv \frac{1}{2}\left(|0\rangle\langle 0|^X \otimes \psi_0^{AA'} + |1\rangle\langle 1|^X \otimes \psi_1^{AA'}\right) \quad (69)$$

where

$$|\psi_0\rangle^{AA'} \equiv \sqrt{\mu}|00\rangle^{AA'} + \sqrt{1-\mu}|11\rangle^{AA'}$$
$$|\psi_1\rangle^{AA'} \equiv \sqrt{1-\mu}|00\rangle^{AA'} + \sqrt{\mu}|11\rangle^{AA'}$$

and $\mu \in [0, \frac{1}{2}]$. This classical-quantum state is a purified version of the ensemble considered in [7]. We can evaluate various relevant entropic quantities for this state

$$H(B)_\sigma = 1 - \epsilon + H_2(\epsilon)$$
$$H(A)_\sigma = H_2(\mu)$$
$$H(A\,|\,X)_\sigma = H_2(\mu)$$
$$H(B\,|\,X)_\sigma = (1-\epsilon)H_2(\mu) + H_2(\epsilon)$$
$$H(E\,|\,X)_\sigma = \epsilon H_2(\mu) + H_2(\epsilon)$$

where the state σ is the state resulting from sending the A' system through the erasure channel. It then follows that

$$I(X;B)_\sigma = (1-\epsilon)(1 - H_2(\mu))$$
$$I(A\rangle BX)_\sigma = (1-2\epsilon)H_2(\mu)$$
$$\frac{1}{2}I(A;B\,|\,X)_\sigma = (1-\epsilon)H_2(\mu)$$
$$\frac{1}{2}I(A;E\,|\,X)_\sigma = (1-\epsilon)H_2(\mu)$$
$$I(AX;B) = (1 + H_2(\mu))(1-\epsilon).$$

A quick glance over the above information quantities reveals that exploiting coding strategies such as the CEF protocol gives no improvement over timesharing because $H_2(\mu)$ varies between zero and one as μ varies between zero and $1/2$ (the CEF protocol gives exactly the same performance as timesharing, as does the classically enhanced quantum communication strategy of Devetak and Shor [7]). Thus, the region obtained as the union of the one-shot, one-state regions is indeed equivalent to the outer bound given above. Fig. 3 plots this region for a quantum erasure channel with erasure parameter $\epsilon = 1/4$, demonstrating that this region is equivalent to the outer bound. □

The following corollary applies to the noiseless qubit channel by simply plugging in $\epsilon = 0$.

Corollary 4: The following set of inequalities specifies the entanglement-assisted capacity of the noiseless qubit channel for transmitting classical and quantum information:

$$C + 2Q \leq 2$$
$$C + Q \leq E + 1.$$

C. *The Qubit Dephasing Channel*

In this section, we show that we can compute the full capacity region of a qubit dephasing channel and plot it in Fig. 4 for a channel with dephasing parameter $p = 0.2$. We show also that the CEF protocol can beat timesharing for a qubit dephasing channel (the example is an extension of the argument in [7]).

1) Single-Letterization: We first show that the CEF tradeoff curve is optimal in the sense that it lies along the boundary of the capacity region for the qubit dephasing channel. A surprisingly simple argument proves this result by resorting to the result of Devetak and Shor in [7]. There, they showed that the following tradeoff curve in the CQ-plane is optimal:

$$\{(C_{\rm CQ}(\mu), Q_{\rm CQ}(\mu), 0) : 0 \leq \mu \leq 1/2\} \quad (70)$$

where

$$C_{\rm CQ}(\mu) \equiv 1 - H_2(\mu)$$
$$Q_{\rm CQ}(\mu) \equiv H_2(\mu) - H_2(g(p,\mu))$$
$$g(p,\mu) \equiv \frac{1}{2} + \frac{1}{2}\sqrt{1 - 16 \cdot \frac{p}{2}\left(1 - \frac{p}{2}\right)\mu(1-\mu)}.$$

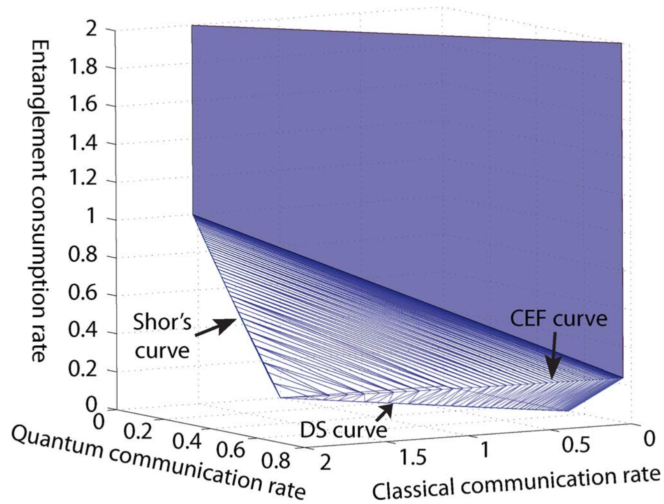

Fig. 4. Full capacity region for the qubit dephasing channel with dephasing parameter $p = 0.2$. It outlines Shor's tradeoff curve, the DS tradeoff curve, and the CEF tradeoff curve. The surface between Shor's curve and the CEF curve is that in (73). The surface between the CEF curve and the DS curve is that in (71). Finally, (74) specifies the solid plane. This region is a union of regions formed by translating the unit resource capacity region from [26] along the CEF tradeoff curve. This point is perhaps more clear in [26] where we plot the full triple tradeoff.

Now, consider the surface formed by the following set of points:

$$\{(C_{\rm CQ}(\mu), Q_{\rm CQ}(\mu)+E, E) : 0 \leq \mu \leq 1/2, E \geq 0\}. \quad (71)$$

This surface is an outer bound for the capacity region [if it were not so, one could combine points outside this surface with entanglement distribution and beat the optimal bound in (70) for the DS case].

Now consider sending the μ-parametrized ensemble in (69), where $\mu \in [0, 1/2]$, through the qubit dephasing channel with dephasing parameter p. It is straightforward to show that the various entropic quantities in the CEF protocol are as follows for the μ-parametrized ensemble:

$$\begin{aligned} C_{\rm CEF}(\mu) &\equiv I(X;B)_\sigma = 1 - H_2(\mu) \\ Q_{\rm CEF}(\mu) &\equiv \frac{1}{2}I(A;B\,|\,X) = H_2(\mu) - \frac{1}{2}H_2(g(p,\mu)) \\ E_{\rm CEF}(\mu) &\equiv \frac{1}{2}I(A;E\,|\,X) = \frac{1}{2}H_2(g(p,\mu)). \end{aligned}$$

Thus, the following set of points contains all points along the CEF tradeoff curve:

$$\{(C_{\rm CEF}(\mu), Q_{\rm CEF}(\mu), E_{\rm CEF}(\mu)) : 0 \leq \mu \leq 1/2\}.$$

All points along the CEF tradeoff curve lie along the boundary because they are of the form in (71) with $E = H_2(g(p,\mu))/2$. This proves that the points along the CEF tradeoff curve are optimal. One can also achieve any point along the surface in (71) with entanglement consumption below the CEF by combining the CEF with entanglement distribution.

We now outline the proof that Shor's tradeoff curve for EAC communication single-letterizes for the qubit dephasing channel (full details appear in [34]—the argument complements the argument in [7, App. B]). Any point along Shor's tradeoff curve achieves a classical communication rate of $I(AX;B^n)$ at an entanglement consumption rate of $H(A\,|\,X)$ [14]. Therefore, to determine a point along the tradeoff curve, we would like to maximize the classical communication rate while minimizing the entanglement consumption rate. To do so, we can define the following function:

$$f_\lambda(\mathcal{N}^{\otimes n}) \equiv \max_\sigma (I(AX;B^n) - \lambda H(A\,|\,X))$$

where $\lambda > 0$ and the maximization is over all states of the form (15), with the exception that the E' system is not necessary for Shor's tradeoff curve [14]. By a sequence of arguments similar to those in [7, App. B], we can show that

$$f_\lambda(\mathcal{N}^{\otimes n}) \leq n h_\lambda(\mathcal{N})$$

where

$$h_\lambda(\mathcal{N}) \equiv \max_{\sigma_\mu}(H(Y) + (1-\lambda)H(A\,|\,X) - H(E\,|\,X))$$

Y is the completely dephased version of B, and σ_μ is a state that arises after sending the A' system of a state of the form in (69) through a *single use* of the qubit dephasing channel. This then shows that the region single-letterizes and that states of the form in (69) give rise to optimal points that lie along Shor's tradeoff curve. Shor's tradeoff curve in the CE-plane has the following form:

$$\{(C_{\rm CE}(\mu), 0, E_{\rm CE}(\mu)) : 0 \leq \mu \leq 1/2\} \quad (72)$$

where

$$\begin{aligned} C_{\rm CE}(\mu) &\equiv 1 + H_2(\mu) - H_2(g(p,\mu)) \\ E_{\rm CE}(\mu) &\equiv H_2(\mu). \end{aligned}$$

We can now exploit Shor's tradeoff curve to outline a bounding surface in the CQE space (just as we did before with the DS curve and entanglement distribution). Consider the surface formed by the following set of points:

$$\{(C_{\rm CE}(\mu) - 2E, E, E_{\rm CE}(\mu) - E) : 0 \leq \mu \leq 1/2,\ E \geq 0\}. \quad (73)$$

This surface is an outer bound for the capacity region [if it were not so, one could combine points outside this surface with superdense coding and beat the optimal bound in (72)]. Interestingly, this surface intersects the surface in (71) at exactly the CEF tradeoff curve.

We can finally outline the full capacity region by combining the two surfaces in (71) and (73) with the bound

$$C + 2Q \leq 2 - H_2(g(p,1/2)). \quad (74)$$

The above bound is the largest that the EAC capacity can be and therefore bounds the sum rate $C + 2Q$ as we have argued previously. The intersection of these three surfaces forms a single-letter bound for the capacity region, and all points on the boundary are achievable by combining the CEF tradeoff curve with entanglement distribution, superdense coding, or the wasting of entanglement. Fig. 4 plots the full capacity region.

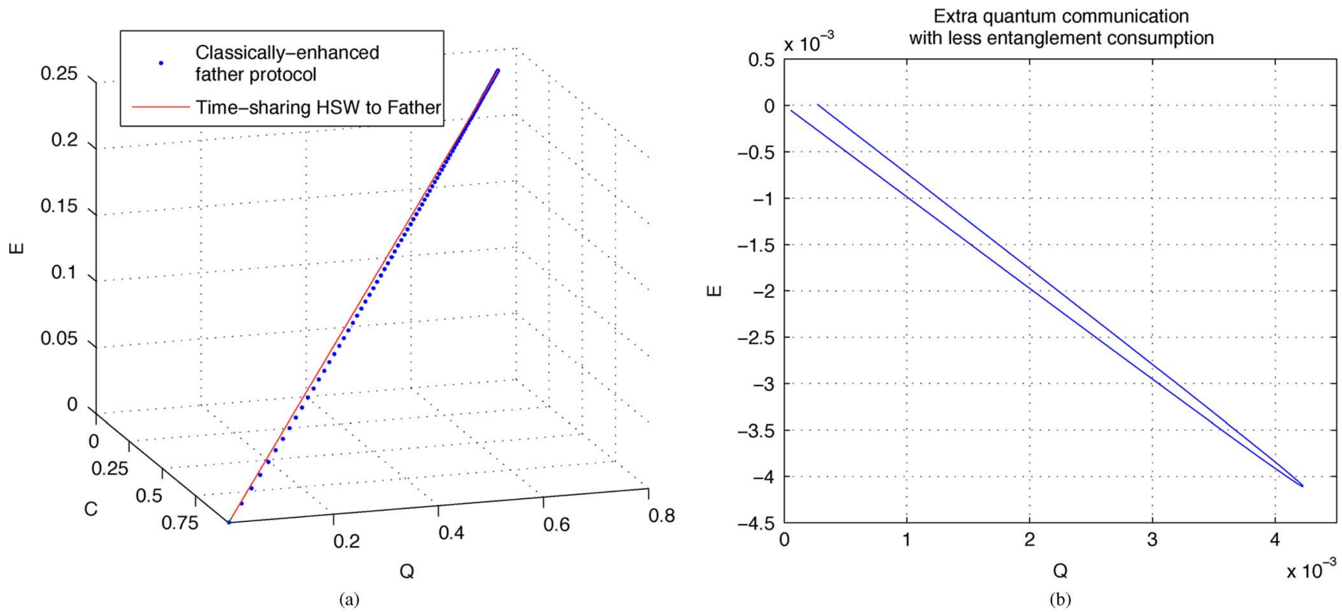

Fig. 5. (a) Points achievable by timesharing between EAQ coding and classical coding on the solid red line, and points achievable with the CEF protocol on the dotted blue line. The channel for which we are coding is the qubit dephasing channel with dephasing parameter $p = 0.2$. The figure demonstrates that one can achieve more quantum communication with less entanglement consumption, while having the same rate of classical communication, by employing the CEF protocol instead of a timesharing strategy. (b) This figure makes the previous statement precise, by showing the difference between quantum communication and entanglement consumption for achievable points on the CEF tradeoff curve that attain the same rate of classical communication as a timesharing strategy.

2) The CEF Protocol Can Beat Timesharing: An important question for EAC quantum coding is whether a timesharing strategy is optimal for all channels or if the CEF protocol can give an improvement over timesharing. There are three timesharing strategies that one could employ in EACQ coding. In all three strategies, we suppose that the sender and receiver share some finite amount of entanglement E. The three strategies are as follows.

1) Use an EAQ code with rate triple $(0, Q_1, E_1)$ and an HSW code with rate triple $(C_2, 0, 0)$. If $E = \lambda E_1$, then timesharing produces an EACQ code with rate triple $((1 - \lambda)C_2, \lambda Q_1, E)$.
2) Use an EAC code with rate triple $(C_1, 0, E_1)$ and a quantum channel code with rate triple $(0, Q_2, 0)$. If $E = \lambda E_1$, then timesharing produces an EACQ code with rate triple $(\lambda C_1, (1 - \lambda)Q_2, E)$.
3) Use an EAQ code with rate triple $(0, Q_1, E_1)$ and an EAC code with rate triple $(C_2, 0, E_2)$. If $E = \lambda E_1 + (1-\lambda)E_2$, then timesharing produces an EACQ code with rate triple $((1 - \lambda)C_2, \lambda Q_1, E)$.

We should compare the CEF protocol to the first timesharing strategy because the two points EAQ and HSW are special cases of it. For the second timesharing strategy, it is clear that this strategy is not optimal because the line connecting EAC to LSD is strictly inside the capacity region. For the third timesharing strategy, timesharing is the optimal strategy. If it were not (in the sense that one could achieve a higher quantum or classical rate than a point along the timesharing bound), then one could beat the bound in (12) by combining this protocol with either teleportation or superdense coding.

We now consider the first case for the qubit dephasing channel and show that the CEF protocol can beat a timesharing strategy. Consider the qubit dephasing channel with dephasing parameter $p = 0.2$. The classical capacity of this channel is one bit per channel use, and the EAQ capacity is about 0.7655 qubits per channel use while using about 0.2345 ebits per channel use. The solid red line in Fig. 5(a) corresponds to the timesharing line between these two optimal points. The blue dotted line in Fig. 5(a) corresponds to the various points along the CEF protocol. In comparing the timesharing line to the CEF tradeoff curve, we see that the CEF protocol achieves more quantum communication for less entanglement consumption for any point along the timesharing line that achieves the same amount of classical communication. Fig. 5(b) makes this statement precise by comparing the difference in quantum communication and entanglement consumption for all points along the tradeoff curve that achieve the same amount of classical communication as a timesharing point.

VIII. Conclusion

We have proven the EACQ capacity theorem. This theorem determines the ultimate rates at which a noisy quantum channel can communicate both classical and quantum information reliably, while consuming entanglement to do so. The coding strategy exploits a new EACQ coding strategy, the *CEF protocol*, and the unit protocols of teleportation, superdense coding, and entanglement distribution. Several protocols in the family tree of quantum Shannon theory are now child protocols of the CEF. We also have provided example channels whose corresponding CQE capacity regions single-letterize, so that we can actually determine the region for these channels, and we have shown that CEF protocol beats a timesharing strategy for the case of a qubit dephasing channel. We discuss follow-up work and several open problems in what follows.

A. The Full Triple Tradeoff

This paper addresses only one octant of the channel coding scenario—the octant where we consume entanglement and generate classical and quantum communication. We characterize the full triple tradeoff region in [26] and [35], where we show that the CEF protocol combined with the unit resource protocols in (61)–(63) achieves the full capacity region for all octants.

B. The Structure of CEF Codes

In [20], one of the authors constructed a CEF code that uses only ancilla qubits for encoding classical information. In [19], the other author constructed a CEF code that uses both ancilla qubits and ebits for encoding classical information. One might think that using ebits in addition to ancilla qubits for encoding classical information could improve performance and it was unclear which coding structure might perform better.

The structure of our CEF protocol actually gives a hint for constructing CEF codes that achieve the rates in Theorem 1. Consider the protocol in the proof of the direct coding part of Theorem 1. Bob decodes the classical information by measuring the channel outputs only. He does not need to measure his half of the entanglement to decode the classical information. This decoding implies that he is not using the entanglement for sending classical information—if he were, he would need to measure his half of the entanglement as well. This observation lends credence to the conjecture that it is sufficient to encode classical information into ancilla qubits when attempting to construct codes that achieve the tradeoff rate triple in Theorem 1.

C. Other Issues

Another issue remains with the "pasting" proof technique. It relies on the assumption that the channel is i.i.d. and thus does not apply in a straightforward way to channels with memory. Many proof techniques in quantum Shannon theory rely on a "one-shot" lemma applied to the i.i.d. case. The usefulness of this method of proof is that the one-shot result can apply to more general scenarios such as channels that have memory. So it may be useful to develop a one-shot result for the code pasting technique.

Appendix I
Proof of Proposition 2

The proof of Proposition 2 is an extension of the development in [6, App. D].

Proposition 2: Consider an arbitrary density operator $\rho^{A'}$ whose spectral decomposition is as follows:

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

The nth extension of the above state as a tensor power state is as follows:

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

We define the pruned distribution p'^n as follows:

$$p'^n(x^n) \equiv \begin{cases} p^n(x^n)/\sum_{x^n\in T_\delta^{X^n}} p^n(x^n): & x^n \in T_\delta^{X^n} \\ 0: & \text{else} \end{cases}$$

where $T_\delta^{X^n}$ denotes the δ-typical set of sequences with length n. Let $\widetilde{\rho}^{A'^n}$ denote the following "pruned state":

$$\widetilde{\rho}^{A'^n} \equiv \sum_{x^n\in T_\delta^{X^n}} p'^n(x^n)|x^n\rangle\langle x^n|^{A'^n}. \tag{75}$$

For any $\epsilon > 0$ and sufficiently large n, the state $\rho^{A'^n}$ is close to $\widetilde{\rho}^{A'^n}$ by the gentle measurement lemma [36] and the typical subspace theorem [8]

$$\left\|\rho^{A'^n} - \widetilde{\rho}^{A'^n}\right\|_1 \le 2\epsilon.$$

For any density operator $\rho^{A'}$, it is possible to construct an EAQ code that achieves the quantum communication rate and entanglement consumption rate in Proposition 2. Abeyesinghe *et al.* [15] provide group-theoretical and other clever arguments to show how to achieve the rates in Proposition 2. Another method for achieving the rates in Proposition 2 is to exploit the connection between quantum privacy and quantum coherence in constructing quantum codes [6], [37]. Indeed, in [38], one of the current authors showed how to construct secret-key-assisted private classical codes for a quantum channel. Using the methods of [6] and [37], it is possible to make "coherent" versions, i.e., EAQ codes, of these secret-key-assisted private classical codes. Let $[k]$ denote a set of size $\sim 2^{nQ}$ and let $[m]$ denote a set of size $\sim 2^{nE}$. Let $U_{k,m}$ denote $\sim 2^{n(Q+E)}$ random variables that we choose according to the pruned distribution $p'^n(x^n)$. The realizations $u_{k,m}$ of the random variables $U_{k,m}$ are sequences in $\mathcal{X}^n$ and are the basis for constructing an EAQ code $\mathcal{C}$ whose codewords are as follows:

$$\mathcal{C} = \{|\phi_k\rangle^{A^nT_B}\}_k.$$

The EAQ codewords $|\phi_k\rangle^{A^nT_B}$ in $\mathcal{C}$ are as follows:

$$|\phi_k\rangle^{A^nT_B} \equiv \frac{1}{\sqrt{2^{nE}}}\sum_{m=1}^{2^{nE}} |\phi_{u_{k,m}}\rangle^{A'^n}|m\rangle^{T_B}$$

where

$$|\phi_{u_{k,m}}\rangle^{A'^n} \equiv |u_{k,m}\rangle^{A'^n}.$$

We then expurgate this code to improve its performance and this expurgation has a minimal impact on the rate of the code. After expurgation, the code forms a good EAQ code, resulting in failure with probability $\epsilon + 10\sqrt[4]{\epsilon}$ by the arguments in [6] and [37]. Suppose that we choose a particular EAQ code $\mathcal{C}$ according to the above prescription. Its code density operator is

$$\rho^{A'^nT_B}(\mathcal{C}) = \frac{1}{2^{nQ}}\sum_{k=1}^{2^{nQ}} |\phi_k\rangle\langle\phi_k|^{A'^nT_B}$$

and its input code density operator is

$$\begin{aligned}\rho^{A'^n}(\mathcal{C}) &= \mathrm{Tr}_{T_B}\left\{\rho^{A'^nT_B}(\mathcal{C})\right\} \\ &= \frac{1}{2^{n(Q+E)}}\sum_{m=1}^{2^{nE}}\sum_{k=1}^{2^{nQ}} |\phi_{u_{k,m}}\rangle\langle\phi_{u_{k,m}}|^{A'^n}.\end{aligned}$$

Suppose we now consider the entanglement-assisted code chosen according to the above prescription as a *random* code $\mathcal{C}$ (where $\mathcal{C}$ is now a random variable). Let $\rho'^{A'^n}(\mathcal{C})$ be the channel input density operator for the random code before expurgation and $\rho^{A'^n}(\mathcal{C})$ its channel input density operator after expurgation

$$\rho'^{A'^n}(\mathcal{C}) \equiv \frac{1}{2^{n(Q'+E')}} \sum_{k=1}^{2^{nQ'}} \sum_{m=1}^{2^{nE'}} |\phi_{U_{k,m}}\rangle\langle\phi_{U_{k,m}}|^{A'^n}$$

$$\rho^{A'^n}(\mathcal{C}) \equiv \frac{1}{2^{n(Q+E)}} \sum_{k=1}^{2^{nQ}} \sum_{m=1}^{2^{nE}} |\phi_{U_{k,m}}\rangle\langle\phi_{U_{k,m}}|^{A'^n}$$

where the primed rates are the rates before expurgation and the unprimed rates are those after expurgation (they are slightly different but identical for large n). Let $\bar{\rho}'^{A'^n}$ and $\bar{\rho}^{A'^n}$ denote the expectation of the above channel input density operators

$$\bar{\rho}'^{A'^n} \equiv \mathbb{E}_{\mathcal{C}}\left\{\rho'^{A'^n}(\mathcal{C})\right\}$$
$$\bar{\rho}^{A'^n} \equiv \mathbb{E}_{\mathcal{C}}\left\{\rho^{A'^n}(\mathcal{C})\right\}.$$

Choosing our code in the particular way that we did leads to an interesting consequence. The expectation of the density operator corresponding to Alice's restricted codeword $|\phi_{U_{k,m}}\rangle^{A'^n}$ is equal to the pruned state in (75)

$$\mathbb{E}_{\mathcal{C}}\left\{|\phi_{U_{k,m}}\rangle\langle\phi_{U_{k,m}}|^{A'^n}\right\} = \sum_{x^n} p'^n(x^n)|\phi_{x^n}\rangle\langle\phi_{x^n}|^{A'^n}$$

because we choose the codewords $|\phi_{U_{k,m}}\rangle$ randomly according to the pruned distribution $p'^n(x^n)$. Then, the expected channel input density operator $\bar{\rho}'^{A'^n}$ is as follows:

$$\bar{\rho}'^{A'^n} = \mathbb{E}_{\mathcal{C}}\left\{\rho'^{A'^n}(\mathcal{C})\right\} \tag{76}$$
$$= \frac{1}{2^{n(Q'+E')}} \sum_{k=1}^{2^{nQ'}} \sum_{m=1}^{2^{nE'}} \mathbb{E}_{\mathcal{C}}\left\{|\phi_{U_{k,m}}\rangle\langle\phi_{U_{k,m}}|^{A'^n}\right\} \tag{77}$$
$$= \sum_{x^n} p'^n|\phi_{x^n}\rangle\langle\phi_{x^n}|^{A'^n}. \tag{78}$$

Then, we know that the following inequality holds for $\bar{\rho}'^{A'^n}$ and the tensor power state $\rho^{A'^n}$:

$$\left\|\bar{\rho}'^{A'^n} - \rho^{A'^n}\right\|_1 \leq 2\epsilon \tag{79}$$

by the typical subspace theorem and the gentle measurement lemma. The expurgation of any entanglement-assisted code $\mathcal{C}$ has a minimal effect on the resulting channel input density operator [6]

$$\left\|\rho'^{A'^n}(\mathcal{C}) - \rho^{A'^n}(\mathcal{C})\right\|_1 \leq 4\sqrt[4]{\epsilon}.$$

The above inequality implies that the following one holds for the expected channel input density operators $\bar{\rho}'^{A'^n}$ and $\bar{\rho}^{A'^n}$:

$$\left\|\bar{\rho}'^{A'^n} - \bar{\rho}^{A'^n}\right\|_1 \leq 4\sqrt[4]{\epsilon} \tag{80}$$

because the trace distance is convex. The following inequality holds:

$$\left\|\bar{\rho}^{A'^n} - \rho^{A'^n}\right\|_1 \leq 2\epsilon + 4\sqrt[4]{\epsilon} \tag{81}$$

by applying the triangle inequality to (79) and (80). Therefore, the random EAQ code is ρ-like. □

Appendix II
Proof of Proposition 3

We now prove Proposition 3 that applies to a random father code that has an associated classical string.

Proposition 3: The proof is similar to the proof of Proposition 5 in [7]. Suppose that we have an ensemble $\{p_x, \rho_x^{A'}\}$ where each density operator $\rho_x^{A'}$ has a purification $\psi_x^{AA'}$ and state $\phi_x^{ABE} = U_{\mathcal{N}}^{A'\to BE}(\psi_x^{AA'})$ arising from the channel $\mathcal{N}^{A'\to B}$. By Proposition 2, for sufficiently large n and for all $x \in \mathcal{X}$, there exists a random $\rho_x^{A'}$-like entanglement-assisted $(n[p_x - \delta], \epsilon)$ code of quantum rate $Q_x = I(A;B)_{\phi_x}/2 - \delta$ and entanglement consumption rate $E_x = I(A;E)_{\phi_x}/2 + \delta$. Its expected channel input density operator $\bar{\rho}_x^{A'^{n[p_x-\delta]}}$ is close to a tensor power of the state $\rho_x^{A'}$

$$\left\|\bar{\rho}_x^{A'^{n[p_x-\delta]}} - \rho_x^{\otimes n[p_x-\delta]}\right\|_1 \leq \epsilon.$$

The code's quantum rate is $Q_x = \frac{1}{2}I(A;B)_{\phi_x} - \delta$ because it transmits $n[p_x - \delta]Q_x$ qubits for $n[p_x - \delta]$ uses of the channel. The code's entanglement consumption rate is $E_x = \frac{1}{2}I(A;E)_{\phi_x} + \delta$ because it consumes at least $n[p_x - \delta]E_x$ ebits for $n[p_x - \delta]$ uses of the channel. We produce an $(n - |\mathcal{X}|\delta, |\mathcal{X}|\epsilon)$ entanglement-assisted code with expected channel input density operator

$$\bar{\rho}^{A'^{n(1-|\mathcal{X}|\delta)}} = \bigotimes_x \bar{\rho}_x^{A'^{n[p_x-\delta]}}$$

by "pasting" $|\mathcal{X}|$ of these codes together (one for each x). Applying the triangle inequality $|\mathcal{X}|$ times, the expected channel input density operator $\bar{\rho}^{A'^{n(1-|\mathcal{X}|\delta)}}$ of the pasted code is close to a pasting of the tensor power states $\{\rho_x^{\otimes n[p_x-\delta]}\}_x$

$$\left\|\bar{\rho}^{A'^{n(1-|\mathcal{X}|\delta)}} - \bigotimes_x \rho_x^{\otimes n[p_x-\delta]}\right\|_1 \leq |\mathcal{X}|\epsilon. \tag{82}$$

Consider the classical sequence x^n. Let random variable X have the probability distribution p and define the typical set

$$T_\delta^{X^n} = \{x^n : \forall x |n_x - np_x| \leq \delta n\}$$

where $n_x \equiv N(x|x^n)$ is the number of occurrences of the symbol x in x^n. If x^n lies in the typical set $T_\delta^{X^n}$, then we can construct a conditional permutation operation that permutes the elements of the input sequence as follows [39]:

$$x^n \to \underbrace{x_1\cdots x_1}_{n[p_{x_1}-\delta]}\underbrace{x_2\cdots x_2}_{n[p_{x_2}-\delta]}\cdots\underbrace{x_{|\mathcal{X}|}\cdots x_{|\mathcal{X}|}}_{n[p_{x_{|\mathcal{X}|}}-\delta]}x_g$$

where x_g (for "x garbage") denotes the remaining $n|\mathcal{X}|\delta$ symbols in x^n. The density operator ρ_{x^n} corresponds to the input sequence x^n. We can construct a conditional permutation unitary that acts on the density operator ρ_{x^n} and changes the ordering of the state ρ_{x^n} as follows:

$$\rho_{x^n} \to \bigotimes_x \rho_x^{n[p_x-\delta]} \otimes \rho_{x_g}$$

where $\dim(\rho_{x_g}) \leq n|\mathcal{X}|\delta \log d_{A'}$. We modify the random entanglement-assisted code slightly by inserting $|\mathcal{X}|\delta$ "garbage states" with density operator ρ_{x_g} and define the expected channel input density operator $\bar{\rho}^{A'^n}$ for the full code as follows:

$$\bar{\rho}^{A'^n} \equiv \bar{\rho}^{A'^{n(1-|\mathcal{X}|\delta)}} \otimes \rho_{x_g}.$$

Then, the expected channel input density operator $\bar{\rho}^{A'^n}$ is close to the permuted version of ρ_{x^n}

$$\left\| \bar{\rho}^{A'^n} - \bigotimes_x \rho_x^{n[p_x-\delta]} \otimes \rho_{x_g} \right\|_1 \leq |\mathcal{X}|\epsilon.$$

The quantum rate Q for the random "pasted" father code is as follows:

$$\begin{aligned} Q &= \frac{\sum_x nQ_x\,[p_x-\delta]}{n} \\ &= \sum_x Q_x\,[p_x-\delta] \\ &= \sum_x p_x \left(\frac{I(A;B)_{\phi_x}}{2} - \delta \right) - \delta Q_x \\ &= \frac{I(A;B\,|\,X)}{2} - c'\delta \end{aligned}$$

where

$$c' \equiv 1 + \sum_x Q_x.$$

The entanglement consumption rate E is as follows:

$$\begin{aligned} E &= \frac{\sum_x nE_x[p_x-\delta]}{n} \\ &= \sum_x E_x[p_x-\delta] \\ &= \sum_x p_x \left(\frac{I(A;E)_{\phi_x}}{2} - \delta \right) - \delta E_x \\ &= \frac{I(A;E\,|\,X)}{2} - c''\delta \end{aligned}$$

where

$$c'' \equiv 1 + \sum_x E_x.$$

A permutation relates the states ρ_{x^n} and $\bigotimes_x \rho_x^{n[p_x-\delta]} \otimes \rho_{x_g}$. Therefore, there exists an $(n, |\mathcal{X}|\epsilon)$ random entanglement-assisted code of the same quantum communication rate and entanglement consumption rate with an expected channel input density operator $\bar{\rho}'^{A'^n}$ that is close to the tensor power state ρ_{x^n}

$$\left\| \bar{\rho}'^{A'^n} - \rho_{x^n} \right\|_1 \leq |\mathcal{X}|\epsilon$$

because the action of the i.i.d. channel $\mathcal{N}^{\otimes n}$ is invariant under permutations of the input Hilbert spaces. □

Appendix III
Gentle Measurement for Ensembles

Lemma 1 (Gentle Measurement for Ensembles): Let $\{p_x, \rho_x\}$ be an ensemble with average $\bar{\rho} \equiv \sum_x p_x \rho_x$. Given a positive operator X with $X \leq I$ and $\text{Tr}\{\bar{\rho}X\} \geq 1-\epsilon$ where $\epsilon \leq 1$, then

$$\sum_x p_x \|\rho_x - \sqrt{X}\rho_x\sqrt{X}\|_1 \leq \sqrt{8\epsilon}.$$

Proof: We can apply the same steps in the proof of the gentle measurement lemma [40] to get the following inequality:

$$\|\rho_x - \sqrt{X}\rho_x\sqrt{X}\|_1^2 \leq 8(1 - \text{Tr}\{\rho_x X\}).$$

Summing over both sides produces the following inequality:

$$\begin{aligned} \sum_x p_x \|\rho_x - \sqrt{X}\rho_x\sqrt{X}\|_1^2 &\leq 8\left(1 - \text{Tr}\left\{\rho X\right\}\right) \\ &\leq 8\epsilon. \end{aligned}$$

Taking the square root of the above inequality gives the following one:

$$\sqrt{\sum_x p_x \left\|\rho_x - \sqrt{X}\rho_x\sqrt{X}\right\|_1^2} \leq \sqrt{8\epsilon}.$$

Concavity of the square root then implies the result

$$\sum_x p_x \sqrt{\left\|\rho_x - \sqrt{X}\rho_x\sqrt{X}\right\|_1^2} \leq \sqrt{8\epsilon}. \qquad \square$$

Appendix IV
Entanglement Consumption Rate of the EAC Capacity

We prove that the entanglement consumption rate corresponding to the maximal EAC rate is one ebit. Consider a general qubit density operator $\rho^{A'}$ that Alice can input to the erasure channel. Let $\psi^{AA'}$ denote the purification of $\rho^{A'}$. Suppose that ρ has the spectral decomposition $\rho = p|\phi_0\rangle\langle\phi_0| + (1-p)|\phi_1\rangle\langle\phi_1|$ for some orthonormal states $|\phi_0\rangle, |\phi_1\rangle$. After Alice transmits this density operator through an erasure channel with erasure parameter ϵ, Bob has the following state:

$$\sigma^B \equiv (1-\epsilon)\rho + \epsilon|e\rangle\langle e|$$

and Eve has

$$\sigma^E \equiv \epsilon\rho + (1-\epsilon)|e\rangle\langle e|$$

where $|e\rangle$ is an erasure state. The entropies $H(A)$, $H(B)$, and $H(E)$ are as follows:

$$\begin{aligned} H(A) &= H_2(p) \\ H(B) &= (1-\epsilon)H_2(p) + H_2(\epsilon) \\ H(E) &= \epsilon H_2(p) + H_2(\epsilon) \end{aligned}$$

and the mutual information $I(A;B)$ is as follows:

$$I(A;B) = H(A) + H(B) - H(E) = 2(1-\epsilon)H_2(p).$$

This quantity is maximized only when $p = \frac{1}{2}$, implying that the entanglement consumed for this state is exactly one ebit because

$H(A) = H_2(p)$. Thus, Alice and Bob cannot consume entanglement at a lower rate than this amount in order to achieve the EAC capacity.

APPENDIX V
ISOMETRIC ENCODINGS SUFFICE IN THE CQE THEOREM

We prove that it is only necessary to consider isometric encodings for achieving points in the EACQ capacity region. Our argument follows the technique of [14], by showing that a protocol can only improve upon measuring the environment of a nonisometric encoder.

Suppose that we exploit the following state that results from a nonisometric encoder, rather than the state in (15):

$$\widetilde{\sigma}^{XABEE'} \equiv \sum_x p(x)|x\rangle\langle x|^X \otimes U_{\mathcal{N}}^{A'\to BE}\left(\phi_x^{AA'E'}\right). \quad (83)$$

The inequalities in (12)–(14) for the EACQ capacity region involve the mutual information $I(AX;B)_{\widetilde{\sigma}}$, the Holevo information $I(X;B)_{\widetilde{\sigma}}$, and the coherent information $I(A\rangle BX)_{\widetilde{\sigma}}$. As we show below, each of these entropic quantities can only improve if Alice measures the system E'. This improvement then implies that it is only necessary to consider isometric encodings in the EACQ capacity theorem.

Suppose that Alice sends the system E' through a completely dephasing channel $\Delta^{E'\to Y}$ to obtain a classical variable Y (this simulates a measurement). Let $\bar{\sigma}^{XYABE}$ denote this later state, a state of the form

$$\bar{\sigma}^{XYABE} \equiv \sum_x p(x,y)|x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y \otimes U_{\mathcal{N}}^{A'\to BE}\left(\psi_{x,y}^{AA'}\right). \quad (84)$$

This state is actually a state of the form in (15) if we subsume the classical variables X and Y into one classical variable.

The following three inequalities each follow from an application of the quantum data processing inequality (or, equivalently, strong subadditivity):

$$I(X;B)_{\widetilde{\sigma}} = I(X;B)_{\bar{\sigma}} \leq I(XY;B)_{\bar{\sigma}} \quad (85)$$
$$I(AX;B)_{\widetilde{\sigma}} = I(AX;B)_{\bar{\sigma}} \leq I(AXY;B)_{\bar{\sigma}} \quad (86)$$
$$I(A\rangle BX)_{\widetilde{\sigma}} = I(A\rangle BX)_{\bar{\sigma}} \leq I(A\rangle BXY)_{\bar{\sigma}}. \quad (87)$$

Each of these inequalities proves the desired result for the respective Holevo information, mutual information, and coherent information.

ACKNOWLEDGMENT

The authors would like to thank K. Brádler, I. Devetak, P. Hayden, D. Touchette, and A. Winter for useful discussions. They would also like to thank C.-H. Chou and the National Center for Theoretical Science (South) for hosting M.-H. Hsieh as a visitor and M. Rötteler and NEC Laboratories America for hosting M. M. Wilde as a visitor.

REFERENCES

[1] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 269–273, Jan. 1998.
[2] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, pp. 131–138, July 1997.
[3] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
[4] S. Lloyd, "Capacity of the noisy quantum channel," *Phys. Rev. A*, vol. 55, no. 3, pp. 1613–1622, Mar. 1997.
[5] P. W. Shor, "The quantum channel capacity and coherent information," in *Lecture Notes, MSRI Workshop on Quantum Computation*, 2002.
[6] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 44–55, Jan. 2005.
[7] I. Devetak and P. W. Shor, "The capacity of a quantum channel for simultaneous transmission of classical and quantum information," *Commun. Math. Phys.*, vol. 256, pp. 287–303, 2005.
[8] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[9] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states," *Phys. Rev. Lett.*, vol. 69, no. 20, pp. 2881–2884, Nov. 1992.
[10] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted classical capacity of noisy quantum channels," *Phys. Rev. Lett.*, vol. 83, no. 15, pp. 3081–3084, Oct. 1999.
[11] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002.
[12] P. W. Shor, "The classical capacity achievable by a quantum channel assisted by limited entanglement," in *Quantum Information, Statistics, Probability (Dedicated to A. S. Holevo on the Occasion of His 60th Birthday)*. Paramus, NJ: Rinton, 2004, pp. 144–152.
[13] I. Devetak, A. W. Harrow, and A. Winter, "A family of quantum protocols," *Phys. Rev. Lett.*, vol. 93, no. 23, Dec. 2004, 230504.
[14] I. Devetak, A. W. Harrow, and A. Winter, "A resource framework for quantum Shannon theory," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4587–4618, Oct. 2008.
[15] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, "The mother of all protocols: Restructuring quantum information's family tree," *Proc. R. Soc. A*, vol. 465, no. 2108, pp. 2537–2563, 2009.
[16] J. Oppenheim, "State redistribution as merging: Introducing the coherent relay," 2008 [Online]. Available: arXiv:0805.1065
[17] A. Abeyesinghe and P. Hayden, "Generalized remote state preparation: Trading cbits, qubits, and ebits in quantum communication," *Phys. Rev. A*, vol. 68, no. 6, Dec. 2003, 062319.
[18] M.-H. Hsieh, I. Devetak, and A. Winter, "Entanglement-assisted capacity of quantum multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3078–3090, Jul. 2008.
[19] I. Kremsky, M.-H. Hsieh, and T. A. Brun, "Classical enhancement of quantum-error-correcting codes," *Phys. Rev. A*, vol. 78, no. 1, 2008, 012341.
[20] M. M. Wilde and T. A. Brun, "Unified quantum convolutional coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 359–363.
[21] J. Yard, "Simultaneous classical-quantum capacities of quantum multiple access channels," Ph.D. dissertation, Dept. Electr. Eng., Stanford Univ., Stanford, CA, Jun. 2005 [Online]. Available: arXiv:quant-ph/0506050
[22] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *J. Math. Phys.*, vol. 14, pp. 1938–1941, 1973.
[23] B. Schumacher and M. A. Nielsen, "Quantum data processing and error correction," *Phys. Rev. A*, vol. 54, no. 4, pp. 2629–2635, Oct. 1996.
[24] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Phys. Rev. Lett.*, vol. 70, no. 13, pp. 1895–1899, Mar. 1993.
[25] J. Yard, P. Hayden, and I. Devetak, "Capacity theorems for quantum multiple-access channels: Classical-quantum and quantum-quantum capacity regions," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3091–3113, Jul. 2008.
[26] M.-H. Hsieh and M. M. Wilde, "Trading classical communication, quantum communication, and entanglement in quantum Shannon theory," *IEEE Trans Inf. Theory*, 2010 [Online]. Available: arXiv:0901.3038, to be published
[27] M. Fannes, "A continuity property of entropy density for spin lattice systems," *Commun. Math. Phys.*, vol. 31, pp. 291–294, 1973.
[28] R. Alicki and M. Fannes, "Continuity of quantum conditional information," *J. Phys. A, Math. General*, vol. 37, no. 5, pp. L55–L57, 2004.

[29] J. Yard and I. Devetak, “Optimal quantum source coding with quantum information at the encoder and decoder,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5339–5351, Nov. 2009.
[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.
[31] H. Barnum, E. Knill, and M. A. Nielsen, “On quantum fidelities and channel capacities,” *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1317–1329, Jul. 2000.
[32] M. Grassl, T. Beth, and T. Pellizzari, “Codes for the quantum erasure channel,” *Phys. Rev. A*, vol. 56, no. 1, pp. 33–38, Jul. 1997.
[33] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, “Capacities of quantum erasure channels,” *Phys. Rev. Lett.*, vol. 78, no. 16, pp. 3217–3220, Apr. 1997.
[34] K. Brádler, P. Hayden, D. Touchette, and M. M. Wilde, “Trade-off capacities of the quantum Hadamard channels,” *Phys. Rev. A*, vol. 81, no. 6, Jun. 2010, 062312.
[35] M. M. Wilde and M.-H. Hsieh, “The quantum dynamic capacity formula of a quantum channel,” Apr. 2010 [Online]. Available: arXiv:1004.0458
[36] A. Winter, “Coding theorem and strong converse for quantum channels,” *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2481–2485, Nov. 1999.
[37] I. Devetak and A. Winter, “Relating quantum privacy and quantum coherence: An operational approach,” *Phys. Rev. Lett.*, vol. 93, no. 8, Aug. 2004, 080501.
[38] M.-H. Hsieh, Z. Luo, and T. Brun, “Secret-key-assisted private classical communication capacity over quantum channels,” *Phys. Rev. A*, vol. 78, no. 4, 2008, 042306.
[39] H. Krovi and I. Devetak, “Local purity distillation with bounded classical communication,” *Phys. Rev. A*, vol. 76, no. 1, 2007, 012321.
[40] A. Winter, “Coding theorems of quantum information theory,” Ph.D. dissertation, Dept. Math., Universität Bielefeld, Bielefeld, Germany, 1999 [Online]. Available: arXiv:quant-ph/9907077

Min-Hsiu Hsieh (M’09) was born in Tainan, Taiwan. He studied electrical engineering with an emphasis in wireless communications at the National Taiwan University, Taipei, Taiwan, where he received the B.S. and M.S. degrees in 1999 and 2001, respectively. He received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 2008.

Currently, he is a Researcher at the ERATO-SORST Quantum Computation and Information Project, Japan Science and Technology Agency, Tokyo, Japan, between 2008 and 2010.

Dr. Hsieh received the Best Student Paper Award in 2007, the Best Student Paper Award with honorable mention in 2008 from the Ming Hsieh Department of Electrical Engineering, University of Southern California, and the Academic Achievement Award from The Office of International Service, University of Southern California in 2008.

Mark M. Wilde (M’99) was born in Metairie, LA. He received the B.S. degree in computer engineering from Texas A&M University, College Station, in 2002, the M.S. degree in electrical engineering from Tulane University, New Orleans, LA, in 2004, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 2008.

He was a Visiting Researcher at NEC Laboratories America, Princeton, NJ, and at the Centre for Quantum Technologies, Singapore, during the latter half of 2008. He then was employed as a Quantum Information Scientist at Science Applications International Corporation during 2009. Currently, he is a Postdoctoral Fellow at the School of Computer Science, McGill University, Montreal, QC, Canada. He has published over 30 articles and preprints in the area of quantum information processing. His current research interests are in quantum Shannon theory, quantum error correction, and the study of quantum effects in biological systems.

Dr. Wilde is a member of the American Physical Society, a frequent writer for *The Quantum Times*, and has been a reviewer for the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE International Symposium on Information Theory.