9-28-2015

Strong converse rates for quantum communication

Marco Tomamichel
The University of Sydney

Mark M. Wilde
Louisiana State University

Andreas Winter
Universitat Autònoma de Barcelona

Strong Converse Rates for Quantum Communication

Marco Tomamichel*, Mark M. Wilde†, Andreas Winter‡

***Abstract*—We revisit a fundamental open problem in quantum information theory, namely whether it is possible to transmit quantum information at a rate exceeding the channel capacity if we allow for a non-vanishing probability of decoding error. Here we establish that the Rains information of any quantum channel is a strong converse rate for quantum communication: For any sequence of codes with rate exceeding the Rains information of the channel, we show that the fidelity vanishes exponentially fast as the number of channel uses increases. This remains true even if we consider codes that perform classical post-processing on the transmitted quantum data. As an application of this result, for generalized dephasing channels we show that the Rains information is also achievable, and thereby establish the strong converse property for quantum communication over such channels. Thus we conclusively settle the strong converse question for a class of quantum channels that have a non-trivial quantum capacity.**

I. Introduction

The *quantum capacity* of a quantum channel $\mathcal{N}$, denoted $Q(\mathcal{N})$, is defined as the maximum rate (in qubits per channel use) at which it is possible to transmit quantum information over many memoryless uses of the channel with a fidelity that asymptotically converges to one as we increase the number of channel uses. (We will formally introduce capacity in Section II-B.) The question of determining the quantum capacity was set out by Shor in his seminal paper on quantum error correction [1]. Since then, a number of works established a "multi-letter" upper bound on the quantum channel capacity in terms of the coherent information [2]–[4], and the coherent information lower bound on quantum capacity was demonstrated by a sequence of works [5]–[7] which are often said to bear "increasing standards of rigor."[1] In more detail, the work in [5]–[7] showed that the following *coherent information* of the channel $\mathcal{N}$ is an *achievable rate for quantum communication*:

$$I_c(\mathcal{N}) := \sup_{\phi_{RA}} I(R\rangle B)_\rho, \quad \text{where} \quad \rho_{RB} = \mathcal{N}_{A\to B}(\phi_{RA}) \tag{1}$$

and the optimization is over all pure bipartite states ϕ_{RA}. Here, the coherent information of a bipartite state ρ_{RB} is defined as $I(R\rangle B)_\rho := H(\rho_B) - H(\rho_{RB})$, with the von Neumann entropies $H(\rho) := -\mathrm{Tr}\{\rho \log \rho\}$.[2] From the above result, we can also conclude that the rate $I_c(\mathcal{N}^{\otimes \ell})/\ell$ is achievable for any positive integer ℓ, simply by applying the formula in (1) to the "superchannel" $\mathcal{N}^{\otimes \ell}$ and normalizing. By a limiting argument, we find that the regularized coherent information $\lim_{\ell\to\infty} I_c(\mathcal{N}^{\otimes \ell})/\ell$ is also achievable, and Refs. [2]–[4] established that this regularized coherent information is also an upper bound on quantum capacity. This establishes that

$$Q(\mathcal{N}) = \lim_{\ell\to\infty} \frac{I_c(\mathcal{N}^{\otimes \ell})}{\ell}. \tag{2}$$

Clearly, the regularized coherent information is not a tractable characterization of quantum capacity. But the later work of Devetak and Shor proved that the "single-letter" coherent information formula in (1) is equal to the quantum capacity for the class of degradable quantum channels [10]. Degradable channels are such that the receiver of the channel can simulate the channel to the environment by applying a degrading map to the channel output.

All of the above works established an understanding of quantum capacity in the following sense:

1) (Achievability) If the rate of quantum communication is below the quantum capacity, then there exists a scheme for quantum communication such that the fidelity converges to one in the limit of many channel uses.
2) (Weak Converse) If the rate of quantum communication is above the quantum capacity, then there cannot exist an asymptotically error-free quantum communication scheme.

It is crucial to note at this point that practical schemes for quantum communication are restricted to operate on finite block lengths, and it is thus in principle impossible to achieve exactly error-free communication (for most channels). Hence, we are left to wonder whether it is possible to transmit information at a rate larger than the regularized coherent information given in (2) if a non-vanishing error is permissible. Interestingly, it has been known since the early days of classical information theory that the capacity of a classical channel obeys the *strong converse property* [11], [12]: if the rate of communication exceeds the capacity, then the error probability necessarily converges to one in the limit of many channel uses. Furthermore, many works have now confirmed that the strong converse property holds for the classical capacity of several quantum channels [13]–[17] and also for the entanglement-assisted classical capacity of all quantum channels [18]–[20].

Consequently, the present work aims to sharpen the interpretation of quantum capacity in the same spirit.

A. Overview of Results and Outline

In this paper, we extend the Rains relative entropy of a quantum state [21], [22] by defining the *Rains information of*

* School of Physics, The University of Sydney, Sydney, NSW 2006, Australia. (Email: marco.tomamichel@sydney.edu.sg).

† Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA.

‡ ICREA & Física Teórica, Informació i Fenomens Quántics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain.

[1] However, see the later works [8] and [9], which set [5] and [6], respectively, on a firm foundation.

[2] All logarithms in this paper are taken base two.

a quantum channel $\mathcal{N}_{A\to B}$ as follows:[3]

$$R(\mathcal{N}) := \sup_{\rho_{RA}} \inf_{\tau_{RB}\in \mathrm{PPT}'(R:B)} D(\mathcal{N}_{A\to B}(\rho_{RA})\|\tau_{RB}), \quad (3)$$

where we maximize over input states ρ_{RA} with R a reference system, and minimize over a set of subnormalized states (including all PPT states) first considered by Rains, namely the set

$$\mathrm{PPT}'(R\!:\!B) = \{\tau_{RB} : \tau_{RB} \geq 0 \wedge \|T_B(\tau_{RB})\|_1 \leq 1\}. \quad (4)$$

Here, $D(\rho\|\tau) := \mathrm{tr}\{\rho(\log\rho - \log\tau)\}$ is the relative entropy and T_B denotes the partial transpose. The quantity in (3), often called the "Rains bound", was first explored by Rains in the context of entanglement distillation [21] and later refined to the form above in [22].

Our main contribution is that the Rains information of a quantum channel is a *strong converse rate for quantum communication* (Theorem 8), even if we allow for classical pre- and post-processing. (The allowed codes are described in more detail in Section II-B.) That is, if the quantum communication rate of any protocol for a channel $\mathcal{N}_{A\to B}$ exceeds its Rains information, then the fidelity of the scheme decays exponentially fast to zero as the number n of channel uses increases. This result thus provides an operational proof of the inequality $I_c(\mathcal{N}) \leq R(\mathcal{N})$. Note however that Theorem 4 of [23] provides a direct proof of this inequality via operator monotonicity of the logarithm and the notion of the reduction criterion from [24].

We establish this strong converse theorem by exploiting the generalized divergence framework of Sharma and Warsi [25], which generalizes the related framework from the classical context [26] (see also earlier developments from [27]). However, as discussed in Section III, our main departing point from that work is to consider a different class of "useless" channels for quantum data transmission. That is, in [25], the main idea was to exploit a generalized divergence to compare the output of the channel with an operator of the form $\pi_R \otimes \sigma_B$, where π_R denotes the fully mixed state on the reference system R and σ_B is an arbitrary state on the output system. This state can be viewed as the output of a "useless" channel that replaces the input and reference system with the maximally mixed state and an uncorrelated state. The resulting information quantity is a generalization of the coherent information from (1), and one then invokes a data processing inequality to relate this quantity to communication rate and fidelity. Here, we instead compare the output of the channel with an operator in the set $\mathrm{PPT}'(R:B)$, which contains and is closely related to the positive partial transpose states, which in turn are well known to have no distillable entanglement (or equivalently, no quantum data transmission capabilities, so that they constitute another class of "useless" channels for quantum data transmission).

It turns out that a Rényi-like version of the Rains information of a quantum channel appears to be easier to manipulate, so that we can show that it obeys a weak subadditivity property (Theorem 6 in Section IV). From there, some standard limiting arguments conclude the proof of our main result (Theorem 8 in Section V).

The main application of this result is to establish the *strong converse property* for any generalized dephasing channel (Section VI). The action of any channel in this class on an input state ρ is as follows:

$$\mathcal{N}(\rho) = \sum_{x,y=0}^{d-1} \langle x|_A\, \rho\, |y\rangle_A \;\langle\psi_y|\psi_x\rangle\; |x\rangle\langle y|_B, \quad (5)$$

where $\{|x\rangle_A\}$ and $\{|x\rangle_B\}$ are orthonormal bases for the input and output systems, respectively, for some positive integer d, and $\{|\psi_x\rangle_E\}$ is a set of arbitrary pure quantum states. A particular example in this class is the qubit dephasing channel, whose action on a qubit density operator is

$$\mathcal{N}(\rho) = (1-p)\,\rho + pZ\rho Z, \quad (6)$$

where the dephasing parameter $p \in [0,1]$ and Z is the Pauli σ_Z operator. We prove this result by showing that the Rains information of a generalized dephasing channel is equal to its coherent information (Proposition 10). This extends [21, Theorem 6.2] from "maximally correlated states" to generalized dephasing channels.

Finally, in Section VII, we discuss how our technique implies the strong converse property for the classical post-processing assisted quantum capacity of the quantum erasure channel. Note however that this result was previously established in [28].

B. Related Work

A few papers have made some partial progress on or addressed the strong converse for quantum capacity question [18], [19], [25], [28]–[30], with none however establishing that the strong converse holds for any nontrivial class of channels. Refs. [18], [19] prove a strong converse theorem for the entanglement-assisted quantum capacity of a channel, which in turn establishes this as a strong converse rate for the unassisted quantum capacity. Ref. [28] proves that the entanglement cost of a quantum channel is a strong converse rate for the quantum capacity assisted by unlimited forward and backward classical communication, which then demonstrates that this quantity is a strong converse rate for unassisted quantum capacity.

Arguably the most important progress to date for establishing a strong converse for quantum capacity is from [29]. These authors reduced the proof of the strong converse for the quantum capacity of degradable channels to that of establishing it for the simpler class of channels known as the symmetric channels (these are channels symmetric under the exchange of the receiver and the environment of the channel). Along the way, they also demonstrated that a "pretty strong converse" holds for degradable quantum channels, meaning that there is (at least) a jump in the quantum error from zero to $1/2$ as soon as the communication rate exceeds the quantum capacity.

We also mention that our main contribution here sets a previous claim from [31] on a firm foundation. Namely, in the introduction of Ref. [31], the authors claim that the Rains bound for entanglement distillation leads to a weak converse upper bound on the quantum capacity of any channel.

[3]It should be clear from the context whether R refers to "Rains" or to a reference system.

However, Ref. [31] does not appear to sufficiently support this claim. In particular, the authors do not provide an argument for the weak sub-additivity of the Rains information of a quantum channel. (We establish the corresponding property for Rényi Rains information of a quantum channel in Theorem 6 in Section IV.)

Sharma and Warsi established the "generalized divergence" framework for understanding quantum communication and reduced the task of establishing the strong converse to a purely mathematical "additivity" question [25], which hitherto has remained unsolved. The later work in [30] demonstrated that randomly selected codes with a communication rate exceeding the quantum capacity of the quantum erasure channel lead to a fidelity that decreases exponentially fast as the number of channel uses increases. (We stress that a strong converse would imply this behavior for *all* codes whose rate exceeds capacity.)

The methods given in the present paper also imply that the Rains relative entropy of a quantum state is a strong converse rate for entanglement distillation, as established by Hayashi in [32, Section 8.6] using a different method.

II. Preliminaries

A. States and Channels

States: We denote different physical systems by capital letters (e.g. A, B) and we use these labels as subscripts to indicate with which physical system a mathematical object is associated. Let $\mathcal{H}_A$ denote the Hilbert space corresponding to the system A, where we restrict ourselves to finite-dimensional Hilbert spaces throughout this paper. We define $|A| = \dim\{\mathcal{H}_A\}$. Moreover, let $\mathcal{B}(A)$ and $\mathcal{P}(A)$ denote the algebra of *bounded linear operators* acting on $\mathcal{H}_A$ and the subset of *positive semi-definite operators*, respectively. We also write $X_A \geq 0$ if $X_A \in \mathcal{P}(A)$. An operator ρ_A is in the set $\mathcal{S}(A)$ of *quantum states* if $\rho_A \geq 0$ and $\mathrm{tr}\{\rho_A\} = 1$. We say that a quantum state ρ_A is *pure* if $\mathrm{rank}(\rho_A) = 1$. We denote the *identity operator* in $\mathcal{B}(A)$ by I_A and the *fully mixed* state by $\pi_A = I_A/|A|$.

The fidelity between two density operators ρ and σ is defined as [33]

$$F(\rho, \sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2 = \left[\mathrm{tr}\left\{\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}\right\}\right]^2. \tag{7}$$

A bipartite physical system AB is described by the tensor-product Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ with the sets $\mathcal{B}(AB)$, $\mathcal{P}(AB)$, and $\mathcal{S}(AB)$ defined accordingly. Given a bipartite quantum state $\rho_{AB} \in \mathcal{S}(AB)$, we unambiguously write $\rho_A = \mathrm{tr}_B\{\rho_{AB}\}$ for the reduced state on system A. A state ρ_{AB} is called *maximally entangled* with Schmidt rank M if there exist orthonormal bases $\{|x\rangle_A\}$ and $\{|x\rangle_B\}$ such that $\rho_{AB} = |\Phi\rangle\langle\Phi|_{AB}$, where

$$|\Phi\rangle_{AB} = \frac{1}{\sqrt{M}} \sum_{x=1}^{M} |x\rangle_A \otimes |x\rangle_B. \tag{8}$$

A product state is a state that can be written in the form $\rho_{AB} = \rho_A \otimes \rho_B$.

A state ρ_{AB} is called *separable* if it can be written in the form

$$\rho_{AB} = \sum_{x\in\mathcal{X}} P(x)\rho_A^x \otimes \rho_B^x \tag{9}$$

for a set $\mathcal{X}$, a probability mass function P on $\mathcal{X}$ and $\rho_A^x \in \mathcal{S}(A)$, $\rho_B^x \in \mathcal{S}(B)$ for all $x \in \mathcal{X}$ [34]. We denote the set of all such states by $\mathrm{SEP}(A\!:\!B)$. Furthermore, we say that a bipartite state ρ_{AB} is PPT if it has a positive partial transpose, namely if $T_B(\rho_{AB}) \geq 0$, where T_B indicates the partial transpose operation on system B. This is a necessary condition for a bipartite state to be separable [35].[4] Let $\mathrm{PPT}(A\!:\!B)$ denote the set of all such states.[5]

For our purposes it is useful to further enlarge $\mathrm{PPT}(A\!:\!B)$ to the set $\mathrm{PPT}'(A\!:\!B)$ [22], defined as

$$\mathrm{PPT}'(A:B) := \{\tau_{AB} : \tau_{AB} \geq 0 \wedge \|T_B(\tau_{AB})\|_1 \leq 1\}. \tag{10}$$

The set $\mathrm{PPT}'(A:B)$ includes all PPT states because $\|T_B(\tau_{AB})\|_1 = 1$ if $\tau_{AB} \in \mathrm{PPT}(A:B)$. All operators $\tau_{AB} \in \mathrm{PPT}'(A:B)$ are subnormalized, in the sense that $\mathrm{Tr}\{\tau_{AB}\} \leq 1$, because

$$\mathrm{tr}\{\tau_{AB}\} = \mathrm{tr}\{T_B(\tau_{AB})\} \leq \|T_B(\tau_{AB})\|_1 \leq 1. \tag{11}$$

However, the set PPT' contains strictly sub-normalized states that are not PPT.

Channels: We denote linear maps from the set of bounded linear operators on one system to the bounded linear operators on another system by calligraphic letters. For example, $\mathcal{N}_{A\to B}$ denotes a map from $\mathcal{B}(A)$ to $\mathcal{B}(B)$, and we will drop the subscript $A \to B$ if it is clear from the context. Let id_A denote the *identity map* acting on $\mathcal{B}(A)$. If a linear map is completely positive and trace-preserving (CPTP), we say that it is a *quantum channel*.

We say that a CPTP map $\mathcal{N}_{AB\to A'B'}$ consists of local operations (LO) from $A:B$ to $A':B'$ if it has the form $\mathcal{N}_{AB\to A'B'} = \mathcal{L}_{A\to A'} \otimes \mathcal{M}_{B\to B'}$ for CPTP maps $\mathcal{L}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$. Similarly, we say that $\mathcal{N}$ is LOCC if it consists of local operations and classical communication [37], [38]. Furthermore, LOCC maps are contained in the set of separability preserving maps from $A\!:\!B$ to $A'\!:\!B'$ that take $\mathrm{SEP}(A\!:\!B)$ to $\mathrm{SEP}(A'\!:\!B')$. Finally, $\mathcal{N}$ is a PPT preserving operation from $A\!:\!B$ to $A'\!:\!B'$ if the map $T_{B'} \circ \mathcal{N}_{AB\to A'B'} \circ T_B$ is CPTP [21]. (So we can conclude that any LOCC map is a PPT-preserving map.)

B. Codes, Rates and Capacity

Codes: We consider a general class of classical pre- and post-processing (CPPP) assisted entanglement generation (EG) codes, for which the goal is for the sender (Alice) to use a quantum channel in order to share a state that is close to a maximally entangled state with the receiver (Bob).[6] The CPPP

[4] It is also sufficient for 2×2 and 2×3 systems, but otherwise only necessary [36].

[5] Note that there is no need to have an asymmetric notation here because a state is PPT with respect to transpose on system A if and only if it is PPT with respect to a partial transpose on system B.

[6] This generalizes the standard classical scenario in which we are interested in transmitting a uniformly distributed message.

assisted codes allow classical pre- and post-processing before and after the quantum communication phase. In particular, the parties are allowed to prepare a separable resource state before the quantum communication commences. Note that any converse bounds for these classical communication assisted codes naturally also imply the same bounds for unassisted codes.

Formally, we define a *CPPP assisted EG code* for a channel $\mathcal{N}_{A\to B}$ as a triple

$$\mathcal{C} = (M, \mathcal{E}_{A_0 B_0 \to \tilde{A} A \tilde{B}}, \mathcal{D}_{\tilde{A} B \tilde{B} \to \hat{A} \hat{B}})\,. \tag{12}$$

Here, $A_0 \cong B_0 \cong \mathbb{C}$ are trivial, $\hat{A}$ and $\hat{B}$ are Hilbert spaces of dimension M, and $\tilde{A}$ and $\tilde{B}$ are auxiliary Hilbert spaces of arbitrary dimension. Moreover, $\mathcal{E}_{A_0 B_0 \to \tilde{A} A \tilde{B}}$ is an LOCC quantum channel from $A_0 : B_0$ to $\tilde{A}A : \tilde{B}$, and $\mathcal{D}_{\tilde{A}\tilde{B}B \to \hat{A}\hat{B}}$ is an LOCC quantum channel from $\tilde{A} : \tilde{B}B$ to $\hat{A} : \hat{B}$. We write $|\mathcal{C}| = M$ for the *size* of the EG code. An *unassisted EG code* is defined in the same way, but $\mathcal{E}$ and $\mathcal{D}$ are restricted to be LO instead of LOCC.

The corresponding coding schemes thus begin with Alice and Bob preparing a bipartite state $\rho_{\tilde{A}A\tilde{B}} \in \mathrm{SEP}(\tilde{A}A : \tilde{B})$ using the quantum channel $\mathcal{E}$. (This state is restricted to be a product state for unassisted codes but can be an arbitrary separable state in the CPPP assisted case.) Alice then sends the system A through the channel $\mathcal{N}$, resulting in the state

$$\rho_{\tilde{A}B\tilde{B}} = \mathcal{N}_{A\to B}(\rho_{\tilde{A}A\tilde{B}}). \tag{13}$$

Finally, Alice and Bob perform a decoding $\mathcal{D}$, leading to the state $\omega_{\hat{A}\hat{B}} = \mathcal{D}_{\tilde{A}B\tilde{B}\to\hat{A}\hat{B}}(\rho_{\tilde{A}B\tilde{B}})$.

The *fidelity* of the above code $\mathcal{C}$ on the channel $\mathcal{N}$ is given by

$$F(\mathcal{C}, \mathcal{N}) := \langle \Phi |_{\hat{A}\hat{B}}\ \omega_{\hat{A}\hat{B}}\ | \Phi \rangle_{\hat{A}\hat{B}}\,, \tag{14}$$

where $|\Phi\rangle_{\hat{A}\hat{B}}$ is a (fixed) maximally entangled state on $\hat{A}\hat{B}$ of Schmidt rank M.

Rates And Capacity: The main focus of this paper is on EG codes for many parallel uses of a memoryless quantum channel. That is, we want to investigate product channels $\mathcal{N}^{\otimes n}$ for large n. Note that a CPPP assisted EG code for $\mathcal{N}^{\otimes n}$ (as described above) allows for classical communication before and after the product channel is used, but does not allow for interactive schemes with classical communication between different channel uses.

A rate r is an *achievable rate* for (CPPP assisted) quantum communication over the channel $\mathcal{N}$ if there exists a sequence of codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ where $\mathcal{C}_n$ is a (CPPP assisted) EG code for $\mathcal{N}^{\otimes n}$, such that

$$\liminf_{n\to\infty} \frac{1}{n} \log |\mathcal{C}_n| \geq r, \tag{15}$$

$$\lim_{n\to\infty} F(\mathcal{C}_n, \mathcal{N}^{\otimes n}) = 1. \tag{16}$$

The *quantum capacity* of $\mathcal{N}$, denoted $Q(\mathcal{N})$ is the supremum of all achievable rates. Analogously, the *CPPP assisted quantum capacity* of $\mathcal{N}$, denoted $Q_{\mathrm{pp}}(\mathcal{N})$, is the supremum of all CPPP assisted achievable rates.

On the other hand, r is a *strong converse rate* for (CPPP assisted) quantum communication if for every sequence of codes $\{C_n\}_{n\in\mathbb{N}}$ as above, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log |\mathcal{C}_n| > r \implies \lim_{n\to\infty} F(\mathcal{C}_n, \mathcal{N}^{\otimes n}) = 0. \tag{17}$$

The *strong converse quantum capacity*, denoted $Q^{\dagger}(\mathcal{N})$, is the infimum of all strong converse rates. Analogously, the *CPPP assisted strong converse quantum capacity* of $\mathcal{N}$, denoted $Q^{\dagger}_{\mathrm{pp}}(\mathcal{N})$, is the infimum of all CPPP assisted strong converse rates.

Clearly, the following inequalities hold by definition:

$$Q(\mathcal{N}) \leq Q_{\mathrm{pp}}(\mathcal{N}) \leq Q^{\dagger}_{\mathrm{pp}}(\mathcal{N}), \tag{18}$$

$$Q(\mathcal{N}) \leq Q^{\dagger}(\mathcal{N}) \leq Q^{\dagger}_{\mathrm{pp}}(\mathcal{N})\,. \tag{19}$$

Finally, we say that a channel $\mathcal{N}$ satisfies the *strong converse property for quantum communication* if $Q(\mathcal{N}) = Q^{\dagger}(\mathcal{N})$. Similarly, we say that a channel $\mathcal{N}$ satisfies the *strong converse property for CPPP assisted quantum communication* if $Q_{\mathrm{pp}}(\mathcal{N}) = Q^{\dagger}_{\mathrm{pp}}(\mathcal{N})$.

III. Generalized Divergence Framework

A functional $\mathbf{D} : \mathcal{S} \times \mathcal{P} \to \mathbb{R}$ is a generalized divergence if it satisfies the monotonicity inequality

$$\mathbf{D}(\rho\|\sigma) \geq \mathbf{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \tag{20}$$

where $\mathcal{N}$ is a CPTP map. It follows directly from monotonicity that any generalized divergence is invariant under isometries, in the sense that $\mathbf{D}(\rho\|\sigma) = \mathbf{D}(U\rho U^{\dagger}\|U\sigma U^{\dagger})$, where U is an isometry, and that it is invariant under tensoring with another quantum state τ, namely $\mathbf{D}(\rho\|\sigma) = \mathbf{D}(\rho\otimes\tau\|\sigma\otimes\tau)$. Note that to establish isometric invariance from monotonicity, we require a channel that can reverse the action of an isometry (see, e.g., [39, Section 4.6.3] for this standard construction).

A. Rains Relative Entropy and Rains Information

We now define some information measures which play a central role in this paper. They are direct generalizations of the Rains bound on distillable entanglement [21] and the subsequent reformulation of it in [22].

We define the *generalized Rains relative entropy* of a bipartite state ρ_{AB} as follows:

$$R_{\mathbf{D}}(A\!:\!B)_{\rho} := \inf_{\tau_{AB}\in \mathrm{PPT}'(A:B)} \mathbf{D}(\rho_{AB}\|\tau_{AB}), \tag{21}$$

We sometimes abuse notation and write $R_{\mathbf{D}}(A : B)_{\rho} = R_{\mathbf{D}}(\rho_{AB})$ if the bipartition is obvious in context.

One property of $R_{\mathbf{D}}$, critical for our application here, is that it is monotone under PPT preserving operations, in the sense that

$$R_{\mathbf{D}}(A\!:\!B)_{\rho} \geq R_{\mathbf{D}}(A'\!:\!B')_{\omega}, \tag{22}$$

where $\omega_{A'B'} = \mathcal{P}_{AB\to A'B'}(\rho_{AB})$ and $\mathcal{P}_{AB\to A'B'}$ is any PPT-preserving operation from $A : B$ to $A' : B'$. This is because PPT-preserving operations do not take operators τ_{AB} in $\mathrm{PPT}'(A\!:\!B)$ outside of this set, which follows from

$$\begin{aligned}\|T_{B'}(\mathcal{P}_{AB\to A'B'}(\tau_{AB}))\|_1 &= \|T_{B'}(\mathcal{P}_{AB}(T_B(T_B(\tau_{AB}))))\|_1 \\ &\leq \|T_B(\tau_{AB})\|_1 \leq 1.\end{aligned} \tag{23}$$

In the above, the equality follows because the partial transpose T_B is its own inverse, and the first inequality follows because the map $T_{B'} \circ \mathcal{P}_{AB} \circ T_B$ is CPTP (as $\mathcal{P}_{AB}$ is PPT preserving) and the fact that the trace norm is monotone decreasing under CPTP maps. Since the set of PPT preserving operations includes LOCC operations, $R_{\mathbf{D}}(A:B)_\rho$ is also monotone under LOCC operations.

Finally, we define the *generalized Rains information of a quantum channel* as

$$R_{\mathbf{D}}(\mathcal{N}) := \sup_{\rho_{RA}} R_{\mathbf{D}}(R\!:\!B)_\omega, \tag{24}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\rho_{RA})$.

Another critical property of the above quantities has to do with the set $\mathrm{PPT}'(A:B)$ of operators over which we are optimizing. That is, all operators in this set satisfy the property given in Lemma 2 of [40], which we recall now:

Lemma 1 (Lemma 2 of [40]). *Let $\tau_{AB} \in \mathrm{PPT}'(A\!:\!B)$. Then the overlap of τ_{AB} with any maximally entangled state Φ_{AB} of Schmidt rank M is at most $1/M$, i.e.,* $\mathrm{Tr}\{\Phi_{AB}\tau_{AB}\} \le \frac{1}{M}$.

The same is true for $\sigma_{AB} \in \mathrm{PPT}(A:B)$ simply because $\mathrm{PPT}(A\!:\!B) \subseteq \mathrm{PPT}'(A\!:\!B)$.

B. Covariance of Quantum Channels

Covariant quantum channels have symmetries which allow us to simplify the set of states over which we need to optimize their generalized Rains information. Let G be a finite group, and for every $g \in G$, let $g \to U_A(g)$ and $g \to V_B(g)$ be unitary representations acting on the input and output spaces of the channel, respectively. Then a quantum channel $\mathcal{N}_{A\to B}$ is covariant with respect to these representations if the following relation holds for all input density operators $\rho_A \in \mathcal{S}(A)$ and group elements $g \in G$:

$$\mathcal{N}_{A\to B}\Big(U_A(g)\rho_A U_A^\dagger(g)\Big) = V_B(g)\mathcal{N}_{A\to B}(\rho_A)V_B^\dagger(g). \tag{25}$$

We then have the following proposition which allows us to restrict the form of the input states needed to optimize the generalized Rains information of a covariant channel:

Proposition 2. *Let $\mathcal{N}_{A\to B}$ be a covariant channel with group G as above and let $\rho_A \in \mathcal{S}(A)$, ϕ^ρ_{RA} a purification of ρ_A, and $\rho_{RB} = \mathcal{N}_{A\to B}(\phi^\rho_{RA})$. Let $\bar{\rho}_A$ be the group average of ρ_A, i.e.,*

$$\bar{\rho}_A = \frac{1}{|G|}\sum_g U_A(g)\rho_A U_A^\dagger(g), \tag{26}$$

and let $\phi^{\bar{\rho}}_{RA}$ be a purification of $\bar{\rho}_A$ and $\bar{\rho}_{RB} = \mathcal{N}_{A\to B}(\phi^{\bar{\rho}}_{RA})$. Then, $R_{\mathbf{D}}(R\!:\!B)_{\bar{\rho}} \ge R_{\mathbf{D}}(R\!:\!B)_\rho$

Proof: Given the purification ϕ^ρ_{RA}, consider the following state

$$|\psi\rangle_{PRA} := \sum_g \frac{1}{\sqrt{|G|}} |g\rangle_P \left[I_R \otimes U_A(g)\right] |\phi^\rho\rangle_{RA}. \tag{27}$$

Observe that $|\psi\rangle_{PRA}$ is a purification of $\overline{\rho}_A$ with purifying systems P and R. Let τ_{PRB} be an arbitrary operator in $\mathrm{PPT}'(PR:B)$. Then the chain of inequalities in (28)–(32) holds. The first inequality follows from monotonicity of the generalized divergence $\mathbf{D}$ under a dephasing of the P register (where the dephasing operation is given by $\sum_g |g\rangle\langle g|\cdot|g\rangle\langle g|$). The first equality follows from the assumption of channel covariance. The second equality follows from invariance of the generalized divergence under unitaries, with the unitary chosen to be

$$\sum_g |g\rangle\langle g|_P \otimes V_B^\dagger(g). \tag{33}$$

Furthermore, this unitary does not take the state out of the class PPT', i.e.

$$\sum_g p(g)|g\rangle\langle g|_P \otimes V_B^\dagger(g)\tau^g_{RB}V_B(g) \in \mathrm{PPT}'(PR\!:\!B). \tag{34}$$

This is because, in this case, one could also implement this operation as a classically controlled LOCC operation, i.e., a von Neumann measurement $\{|g\rangle\langle g|\}$ of the register P followed by a rotation $V_B^\dagger(g)$ of the B register. One can do so here because both arguments to $\mathbf{D}$ in (29) are classical on P. The second inequality follows because the generalized divergence $\mathbf{D}$ is monotone under the discarding of the register P. The final inequality results from taking a minimization, and the final equality is by definition. Since τ_{PRB} is chosen to be an arbitrary operator in $\mathrm{PPT}'(PR\!:\!B)$, it follows that

$$\inf_{\tau_{PRB}\in \mathrm{PPT}'(PR:B)} \mathbf{D}(\mathcal{N}_{A\to B}(\psi_{PRA})\|\tau_{PRB}) \ge R_{\mathbf{D}}(R\!:\!B)_\rho \tag{35}$$

The conclusion then follows because all purifications are related by an isometry acting on the purifying system and the quantity $R_{\mathbf{D}}$ is invariant under isometries acting on the purifying system. ■

C. Specializing to Rényi Divergence

At this point we specialize our discussion to a particular type of Rényi divergence. For $\rho \in \mathcal{S}$, $\sigma \in \mathcal{P}$ and $\alpha \in (0,1)\cup(1,\infty)$, we define the *sandwiched Rényi relative entropy* of order α as [16], [41]

$$\widetilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log \mathrm{tr}\left\{\left(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha}\right)^\alpha\right\} \tag{36}$$

if $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ or $\alpha \in (0,1)$ and it is equal to $+\infty$ otherwise. The sandwiched Rényi relative entropy is defined for $\alpha \in \{1,\infty\}$ by taking the respective limit. In particular, $\lim_{\alpha\to1}\widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$ [16], [41]. In the following, we restrict our attention to the regime for which $\alpha > 1$. The sandwiched Rényi relative entropy is monotone under CPTP maps $\mathcal{N}$ for such values of α [42]–[44] and thus constitutes a generalized divergence as discussed above. In particular, we note that the trace term

$$(\rho,\sigma) \mapsto \widetilde{Q}_\alpha(\rho\|\sigma) = \mathrm{tr}\left\{\left(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha}\right)^\alpha\right\} \tag{37}$$

is jointly convex in its arguments for $\alpha > 1$ [42].

We denote the corresponding Rains relative entropy by $\widetilde{R}_\alpha(A\!:\!B)_\rho$, or simply use the shorthand $\widetilde{R}_\alpha(\rho_{AB}) = \widetilde{R}_\alpha(A\!:\!B)_\rho$ if the bipartition is evident. The Rényi Rains relative entropy is *subadditive*: namely, for any two bipartite states ρ_{AB} and $\sigma_{A'B'}$ and $\alpha \in (0,1)\cup(1,\infty)$, we have

$$\widetilde{R}_\alpha(AA'\!:\!BB')_{\rho\otimes\sigma} \le \widetilde{R}_\alpha(A\!:\!B)_\rho + \widetilde{R}_\alpha(A'\!:\!B')_\sigma. \tag{38}$$

$$
\begin{aligned}
&\mathbf{D}(\mathcal{N}_{A\to B}(\psi_{PRA})\|\tau_{PRB}) \\
&\geq \mathbf{D}\left(\sum_g \frac{1}{|G|}|g\rangle\langle g|_P \otimes \mathcal{N}_{A\to B}(U_A(g)\phi^\rho_{RA}U_A^\dagger(g))\middle\|\sum_g p(g)|g\rangle\langle g|_P\otimes\tau^g_{RB}\right) &(28)\\
&= \mathbf{D}\left(\sum_g \frac{1}{|G|}|g\rangle\langle g|_P \otimes V_B(g)\mathcal{N}_{A\to B}(\phi^\rho_{RA})V_B^\dagger(g)\middle\|\sum_g p(g)|g\rangle\langle g|_P\otimes\tau^g_{RB}\right) &(29)\\
&= \mathbf{D}\left(\sum_g \frac{1}{|G|}|g\rangle\langle g|_P \otimes \mathcal{N}_{A\to B}(\phi^\rho_{RA})\middle\|\sum_g p(g)|g\rangle\langle g|_P\otimes V_B^\dagger(g)\tau^g_{RB}V_B(g)\right) &(30)\\
&\geq \mathbf{D}\left(\mathcal{N}_{A\to B}(\phi^\rho_{RA})\middle\|\sum_g p(g)V_B^\dagger(g)\tau^g_{RB}V_B(g)\right) &(31)\\
&\geq \inf_{\tau_{RB}\in\mathrm{PPT}'(R:B)} \mathbf{D}(\mathcal{N}_{A\to B}(\phi^\rho_{RA})\|\tau_{RB}) = R_{\mathbf{D}}(R\colon B)_\rho\,. &(32)
\end{aligned}
$$

This follows directly from the additivity of the sandwiched Rényi relative entropy with respect to tensor-product states and from the definition, since we can always restrict the minimization to product states.

As a consequence of the joint convexity of the expression in (37) for $\alpha > 1$ [42, Proposition 3], its point-wise minimum over σ is still a convex function of ρ. Moreover, taking into account the logarithm and prefactor in the definition of the Rényi divergence, we conclude that the Rényi Rains relative entropy is *quasi-convex*: i.e., if ρ_{AB} decomposes as $\rho_{AB} = \int \mathrm{d}\mu(x)\rho^x_{AB}$ then

$$\widetilde{R}_\alpha(\rho_{AB}) \leq \sup_x \widetilde{R}_\alpha(\rho^x_{AB})\,. \tag{39}$$

The *Rényi Rains information of a quantum channel* $\mathcal{N}_{A\to B}$ is defined as

$$\widetilde{R}_\alpha(\mathcal{N}) := \sup_{\rho_{RA}} \inf_{\tau_{RB}\in\mathrm{PPT}'(R:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\tau_{RB}). \tag{40}$$

It suffices to perform the maximization in $\widetilde{R}_\alpha(\mathcal{N})$ over pure bipartite states ρ_{RA}, due to the quasi-convexity of $\widetilde{R}_\alpha$ whenever $\alpha > 1$ (as discussed above). As a result, it suffices for the dimension of the reference system R to be no larger than the dimension of the channel input A, due to the well known Schmidt decomposition.

The Rényi Rains information converges to $R(\mathcal{N})$ in the limit as α approaches one from above. This is shown in the following lemma, whose proof is provided in Appendix A.

Lemma 3. *For any quantum channel $\mathcal{N}$ and $\alpha > \beta > 1$, we have*

$$\widetilde{R}_\alpha(\mathcal{N}) \geq \widetilde{R}_\beta(\mathcal{N}) \geq R(\mathcal{N}), \tag{41}$$

$$\lim_{\alpha\to 1^+} \widetilde{R}_\alpha(\mathcal{N}) = R(\mathcal{N}). \tag{42}$$

D. Relating Fidelity of an Entanglement Generation Code to the Rényi Rains Information of a Channel

The power of the generalized divergence framework is that it allows us to relate rate and fidelity to an information quantity. The usual approach is to compare the states resulting from any code to a set of states resulting from a "useless channel." For the transmission of classical information, the only set of useless channels are those which trace out the input to the channel and replace it with an arbitrary density operator, effectively "cutting the communication line." However, for the transmission of quantum information, there are more interesting classes of "useless channels" [45]. For example, it is well known that a PPT entanglement binding channel has zero quantum capacity [46]. More generally, the bound in Lemma 1 establishes that if both the input to the channel and the reference system are replaced with an operator $\tau_{RB} \in \mathrm{PPT}'(R:B)$, then the fidelity with a maximally entangled state can never be larger than $1/M$. Since for a memoryless channel we are taking $M = 2^{nQ}$, this overlap will be exponentially small with the number of channel uses, so that "channels" that replace with τ_{RB} cannot send any quantum information reliably.

The following proposition gives a "one-shot" bound on the fidelity of any CPPP assisted EG code:

Proposition 4. *Let $\mathcal{N}$ be a quantum channel. Any CPPP assisted EG code $\mathcal{C}$ on $\mathcal{N}$ obeys the following bound. For all $\alpha > 1$,*

$$F(\mathcal{C},\mathcal{N}) \leq 2^{-\left(\frac{\alpha-1}{\alpha}\right)\left(\log|\mathcal{C}| - \widetilde{R}_\alpha(\mathcal{N})\right)}\,. \tag{43}$$

Proof: Let $\mathcal{C} = (M, \mathcal{E}_{A_0B_0\to\tilde{A}A\tilde{B}}, \mathcal{D}_{\tilde{A}B\tilde{B}\to\hat{A}\hat{B}})$ as in (12) and recall the state $\rho_{\tilde{A}B\tilde{B}}$ in (13) prior to decoding and the state $\omega_{\hat{A}\hat{B}} = \mathcal{D}_{\tilde{A}B\tilde{B}\to\hat{A}\hat{B}}(\rho_{\tilde{A}B\tilde{B}})$. The binary test channel $\mathcal{B}_{\hat{A}\hat{B}\to Z}$ outputs a flag indicating if the state is maximally entangled in the state $\Phi_{\hat{A}\hat{B}}$ or orthogonal to it:

$$
\begin{aligned}
\mathcal{B}_{\hat{A}\hat{B}\to Z}(\cdot) := \mathrm{tr}\left\{\Phi_{\hat{A}\hat{B}}(\cdot)\right\}|1\rangle\langle 1| \\
+ \mathrm{tr}\left\{\left(I_{\hat{A}\hat{B}} - \Phi_{\hat{A}\hat{B}}\right)(\cdot)\right\}|0\rangle\langle 0|.
\end{aligned} \tag{44}
$$

Furthermore, consider an arbitrary subnormalized state $\tau_{\tilde{A}B\tilde{B}} \in \mathrm{PPT}'(\tilde{A} : B\tilde{B})$ and observe that $\mathcal{D}_{\tilde{A}B\tilde{B}\to\hat{A}\hat{B}}(\tau_{\tilde{A}B\tilde{B}}) \in \mathrm{PPT}'(\hat{A} : \hat{B})$ because the decoding operator is restricted to be LOCC. For ease of presentation, we set

$$p = \mathrm{tr}\left\{\Phi_{\hat{A}\hat{B}}\,\mathcal{D}_{\tilde{A}B\tilde{B}\to\hat{A}\hat{B}}(\tau_{\tilde{A}B\tilde{B}})\right\}, \tag{45}$$

$$F = F(\mathcal{C},\mathcal{N}) = \mathrm{tr}\left\{\Phi_{\hat{A}\hat{B}}\,\omega_{\hat{A}\hat{B}}\right\}, \tag{46}$$

and assume without loss of generality that the operator $\tau_{\tilde{A}B\tilde{B}}$ is chosen such that $p \in (0, \operatorname{tr}\{\tau_{\tilde{A}B\tilde{B}}\})$ (otherwise, $\tau_{\tilde{A}B\tilde{B}}$ would not be a good choice because we would have $\widetilde{D}_\alpha(\rho_{\tilde{A}B\tilde{B}} \| \tau_{\tilde{A}B\tilde{B}}) = +\infty$). Applying monotonicity of the divergence under the decoding map $\mathcal{D}$ and the test $\mathcal{B}$, we find

$$\begin{aligned}
&\widetilde{D}_\alpha(\rho_{\tilde{A}B\tilde{B}} \| \tau_{\tilde{A}B\tilde{B}}) \\
&\geq \widetilde{D}_\alpha(\mathcal{B}_{\hat{A}\hat{B}\to Z}(\omega_{\hat{A}\hat{B}}) \| \mathcal{B}_{\hat{A}\hat{B}\to Z}(\mathcal{D}_{\tilde{A}B\tilde{B}\to\hat{A}\hat{B}}(\tau_{\tilde{A}B\tilde{B}}))) \quad (47)\\
&= \frac{1}{\alpha-1}\log\left[F^\alpha p^{1-\alpha} + (1-F)^\alpha \left[\operatorname{tr}\{\tau_{\tilde{A}B\tilde{B}}\} - p\right]^{1-\alpha}\right] \quad (48)\\
&\geq \frac{1}{\alpha-1}\log\left[F^\alpha p^{1-\alpha}\right] \quad (49)\\
&\geq \frac{1}{\alpha-1}\log\left[F^\alpha (1/M)^{1-\alpha}\right] \quad (50)\\
&= \frac{\alpha}{\alpha-1}\log F + \log M. \quad (51)
\end{aligned}$$

The second inequality follows by discarding the second term $(1-F)^\alpha \left[\operatorname{tr}\{\tau_{RB}\} - p\right]^{1-\alpha}$ (recall that we are considering $\alpha > 1$). The third inequality follows from (45) and Lemma 1.

Now, recall that the state $\rho_{\tilde{A}A\tilde{B}}$ is in $\operatorname{SEP}(\tilde{A}A : \tilde{B})$ and can thus be decomposed as a convex combination of tensor products of pure states. Using the quasi-convexity of $\widetilde{D}_\alpha$ in the first argument, we find that there exist pure states $\sigma_{\tilde{A}A}$ and $\sigma_{\tilde{B}}$ such that for every $\tau_{\tilde{A}B} \in \operatorname{PPT}'(\tilde{A} : B)$, we have

$$\begin{aligned}
&\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\sigma_{\tilde{A}A}) \| \tau_{\tilde{A}B}) \\
&= \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\sigma_{\tilde{A}A}) \otimes \sigma_{\tilde{B}} \| \tau_{\tilde{A}B} \otimes \sigma_{\tilde{B}}) \quad (52)\\
&\geq \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{\tilde{A}A\tilde{B}}) \| \tau_{\tilde{A}B} \otimes \sigma_{\tilde{B}}) \quad (53)\\
&\geq \frac{\alpha}{\alpha-1}\log F + \log M, \quad (54)
\end{aligned}$$

where in the last line we apply the development in (47)-(51) given that $\tau_{\tilde{A}B} \otimes \sigma_{\tilde{B}} \in \operatorname{PPT}'(\tilde{A} : B\tilde{B})$. Since the above bound holds for all $\tau_{\tilde{A}B} \in \operatorname{PPT}'(\tilde{A} : B)$, we can conclude that

$$\inf_{\tau_{\tilde{A}B}\in\operatorname{PPT}'(\tilde{A}:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\sigma_{\tilde{A}A}) \| \tau_{\tilde{A}B}) \geq \frac{\alpha}{\alpha-1}\log F + \log M. \quad (55)$$

We can finally remove the dependence on any particular code by optimizing over all inputs to the channel. Identifying $\tilde{A}$ with R to simplify notation, we find

$$\begin{aligned}
\widetilde{R}_\alpha(\mathcal{N}) &= \sup_{\rho_{RA}} \inf_{\tau_{RB}\in\operatorname{PPT}'(R:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA}) \| \tau_{RB}) \quad (56)\\
&\geq \frac{\alpha}{\alpha-1}\log F + \log M. \quad (57)
\end{aligned}$$

This bound is then equivalent to (43). ∎

IV. Weak Subadditivity of the α-Rains Information for Memoryless Channels

In this section, we prove an important theorem, which is critical for concluding that the Rains information of a channel is a strong converse rate for quantum communication. Before we commence, we need the following technical property.

Lemma 5. *Let $\alpha > 1$, $\rho, \rho' \in \mathcal{S}$ and $\sigma \in \mathcal{P}$. If $\rho \leq \gamma\rho'$ for some $\gamma \geq 1$, then*

$$\widetilde{D}_\alpha(\rho \| \sigma) \leq \frac{\alpha}{\alpha-1}\log\gamma + \widetilde{D}_\alpha(\rho' \| \sigma). \quad (58)$$

Proof: From the assumption that $\rho \leq \gamma\rho'$, we get $\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha} \leq \gamma\sigma^{(1-\alpha)/2\alpha}\rho'\sigma^{(1-\alpha)/2\alpha}$. Then we have that

$$\begin{aligned}
&\operatorname{tr}\{(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha})^\alpha\} \\
&\qquad \leq \gamma^\alpha \operatorname{tr}\{(\sigma^{(1-\alpha)/2\alpha}\rho'\sigma^{(1-\alpha)/2\alpha})^\alpha\} \quad (59)
\end{aligned}$$

because $\operatorname{tr}\{f(P)\} \leq \operatorname{tr}\{f(Q)\}$ for $P \leq Q$ and f a monotone increasing function (see, e.g., [44, Lemma III.6]). Taking logarithms and dividing by $\alpha - 1$ gives the statement of the lemma. ∎

We are ready to prove that the Rains information of the channel obeys a weak subadditivity property.

Theorem 6. *Let $\mathcal{N}_{A\to B}$ be a quantum channel. For all $\alpha > 1$ and $n \in \mathbb{N}$, we have*

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}) \leq n\widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha|A|^2}{\alpha-1}\log n. \quad (60)$$

Proof: To begin with, we observe that a tensor-power channel is covariant with respect to permutations of the input and output systems, in the sense that

$$\begin{aligned}
&\forall\pi \in S_n : W^\pi_{B^n}\mathcal{N}^{\otimes n}(\rho_{A^n})\,(W^\pi_{B^n})^\dagger \\
&\qquad = \mathcal{N}^{\otimes n}\left(W^\pi_{A^n}\rho_{A^n}\,(W^\pi_{A^n})^\dagger\right), \quad (61)
\end{aligned}$$

where $W^\pi_{A^n}$ and $W^\pi_{B^n}$ are unitary representations of the permutation π, acting on the input space A^n and the output space B^n, respectively. So, letting $|\phi^\rho\rangle_{RA^n}$ denote a purification of ρ_{A^n}, we can apply Proposition 2 to find that

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\phi^\rho_{RA^n})) \leq \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\phi^{\overline{\rho}}_{RA^n})), \quad (62)$$

where $\phi^{\overline{\rho}}_{RA^n}$ is a purification of the permutation invariant state $\overline{\rho}_{A^n}$. Now, this purification $\phi^{\overline{\rho}}_{RA^n}$ is related by a unitary on the reference system R to a state $|\psi\rangle_{\hat{A}^nA^n} \in \operatorname{Sym}((\hat{A}\otimes A)^{\otimes n})$, where $\hat{A} \simeq A$ [47, Lemma 4.3.1]. So it follows that

$$\begin{aligned}
\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\phi^\rho_{RA^n})) &\leq \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\phi^{\overline{\rho}}_{RA^n})) \\
&= \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\psi_{\hat{A}^nA^n})). \quad (63)
\end{aligned}$$

For such a state in $\operatorname{Sym}((\hat{A}\otimes A)^{\otimes n})$, we observe (see, e.g., [48]) that

$$\psi_{\hat{A}^nA^n} \leq n^{|A|^2}\omega^{(n)}_{\hat{A}^nA^n}, \quad \text{where} \quad \omega^{(n)}_{\hat{A}^nA^n} := \int d\mu(\varphi)\,\varphi^{\otimes n}_{\hat{A}A}, \quad (64)$$

with $\mu(\varphi)$ denoting the uniform probability measure on the unit sphere consisting of pure bipartite states $\varphi_{\hat{A}A}$.

Employing Lemma 5, we find that

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\psi_{\hat{A}^nA^n})) \leq \frac{\alpha|A|^2}{\alpha-1}\log n + \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\omega_{\hat{A}^nA^n})). \quad (65)$$

and since the right hand side does not depend on the state ρ_{A^n} anymore, this yields

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}) \leq \frac{\alpha\,|A|^2}{\alpha-1}\log n + \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\omega_{\hat{A}^n A^n})) \qquad (66)$$

$$\leq \frac{\alpha\,|A|^2}{\alpha-1}\log n + \sup_{\phi_{\hat{A}A}} \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}(\phi_{\hat{A}A}^{\otimes n})) \qquad (67)$$

where we used the quasi-convexity of $\widetilde{R}_\alpha$ in the state (cf. (39)) and the definition of $\omega^{(n)}_{\hat{A}^n A^n}$ in (64) to establish the second inequality. Finally, $\widetilde{R}_\alpha$ is subadditive for product states as seen in (38), and thus

$$\widetilde{R}_\alpha\big(\mathcal{N}^{\otimes n}(\phi_{\hat{A}A}^{\otimes n})\big) \leq n\widetilde{R}_\alpha(\mathcal{N}(\phi_{\hat{A}A})). \qquad (68)$$

Combining this with (67) concludes the proof. ■

A consequence of Theorem 6 is that the Rains information of a channel is weakly subadditive. This corollary is required in order to set some of the claims in [31], [49] on a firm foundation. We provide its proof in Appendix B.

Corollary 7. *The Rains information of a quantum channel is weakly subadditive, in the sense that*

$$\limsup_{n\to\infty} \frac{1}{n} R(\mathcal{N}^{\otimes n}) \leq R(\mathcal{N}). \qquad (69)$$

V. The Rains Information is a Strong Converse Rate for Quantum Communication

We are ready to state our main result, which is that the Rains information of the channel is an upper bound on the strong converse CPPP assisted quantum capacity.

Theorem 8. *For any quantum channel $\mathcal{N}$, we have $Q^\dagger_{\mathrm{pp}}(\mathcal{N}) \leq \inf_{\ell\in\mathbb{N}} \frac{1}{\ell} R(\mathcal{N}^{\otimes \ell})$.*

Before we prove this result, we first state a technical proposition which implies that the fidelity decreases exponentially fast as the number of channel uses increases, with the exponent bounded in (70) below. (However, we do not know if the exponent in (70) is optimal.)

Proposition 9. *Let $\mathcal{N}$ be a quantum channel. Consider any sequence of codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where $\mathcal{C}_n$ is a CPPP assisted EG code for $\mathcal{N}^{\otimes n}$. Then the rate of this sequence, $r = \liminf_{n\to\infty} \frac{1}{n}\log|\mathcal{C}_n|$, satisfies*

$$\liminf_{n\to\infty}\left\{-\frac{1}{n}\log F(\mathcal{C}_n,\mathcal{N}^{\otimes n})\right\} \geq \sup_{\alpha>1}\left\{\frac{\alpha-1}{\alpha}\big(r-\widetilde{R}_\alpha(\mathcal{N})\big)\right\}. \qquad (70)$$

Proof: First, by definition of r, for any $\delta>0$, there exists an $N_0\in\mathbb{N}$ such that $\frac{1}{n}\log|\mathcal{C}_n| \geq r-\delta$ for all $n\geq N_0$. For such n and any $\alpha>1$, we employ Proposition 4 to find the following bound:

$$F\left(\mathcal{C}_n,\mathcal{N}^{\otimes n}\right) \leq 2^{-\left(\frac{\alpha-1}{\alpha}\right)\left(\log|\mathcal{C}_n| - \widetilde{R}_\alpha(\mathcal{N}^{\otimes n})\right)} \qquad (71)$$

$$\leq 2^{-\left(\frac{\alpha-1}{\alpha}\right)\left(n(r-\delta) - \widetilde{R}_\alpha(\mathcal{N}^{\otimes n})\right)} \qquad (72)$$

$$\leq 2^{-\left(\frac{\alpha-1}{\alpha}\right)\left(n(r-\delta) - n\widetilde{R}_\alpha(\mathcal{N}) - \frac{\alpha|A|^2}{\alpha-1}\log n\right)} \qquad (73)$$

$$= n^{|A|^2}2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(r-\delta-\widetilde{R}_\alpha(\mathcal{N})\right)}. \qquad (74)$$

Here, we used the weak subadditivity result from Theorem 6 to establish (73). Hence, we find

$$\liminf_{n\to\infty}\left\{-\frac{1}{n}\log F(\mathcal{C}_n,\mathcal{N}^{\otimes n})\right\} \geq \frac{\alpha-1}{\alpha}\big(r-\delta-\widetilde{R}_\alpha(\mathcal{N})\big) \qquad (75)$$

and the statement of the theorem follows since we can choose δ and α arbitrarily. ■

We briefly sketch a proof of the main theorem here for the case in which we want to show that $Q^\dagger_{\mathrm{pp}}(\mathcal{N}) \leq R(\mathcal{N})$ and provide the full proof below. Consider any code with r defined as in Proposition 9. If $r > R(\mathcal{N})$, then by continuity of $\widetilde{R}_\alpha(\mathcal{N})$ as $\alpha\to 1^+$, there always exists an $\alpha>1$ such that $r > \widetilde{R}_\alpha(\mathcal{N})$ as well. Thus, the right hand side of (70) is strictly positive and the fidelity thus vanishes.

Proof of Theorem 8: Consider any sequence of codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where $\mathcal{C}_n$ is a CPPP assisted EG code for $\mathcal{N}^{\otimes n}$ such that $r = \liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{C}_n| > \inf_{\ell\in\mathbb{N}}\frac{1}{\ell}R(\mathcal{N}^{\otimes\ell})$. Then, by definition of the infimum there exists a value $\ell\in\mathbb{N}$ such that $\ell r > R(\mathcal{N}^{\otimes\ell})$. Furthermore, by continuity of $\widetilde{R}_\alpha(\mathcal{N}^{\otimes\ell})$ as $\alpha\to 1^+$ (cf. Lemma 3), there exists an $\alpha>1$ such that $\ell r > \widetilde{R}_\alpha(\mathcal{N}^{\otimes\ell})$.

Next, for any $j\in\{1,2,\ldots,\ell\}$, consider the subsequence of codes $\{\mathcal{C}_{k\ell+j}\}_{k\in\mathbb{N}}$ and their embeddings as codes for the channels $\mathcal{N}^{\otimes(k+1)\ell} = \left(\mathcal{N}^{\otimes\ell}\right)^{\otimes(k+1)}$. (These codes are defined for $k\ell+j$ channels, and their embeddings will simply ignore the last $\ell-j$ channels.) Proposition 9 applied to the channel $\mathcal{N}^{\otimes\ell}$ yields

$$\liminf_{k\to\infty}\left\{-\frac{1}{k+1}\log F(\mathcal{C}_{k\ell+j},\mathcal{N}^{\otimes(k+1)\ell})\right\}$$
$$\geq \frac{\alpha-1}{\alpha}\big(\ell r - \widetilde{R}_\alpha(\mathcal{N}^{\otimes\ell})\big) =: c > 0. \qquad (76)$$

Here we used that $\frac{1}{\ell}\liminf_{k\to\infty}\frac{1}{k+1}\log|\mathcal{C}_{k\ell+j}| = \liminf_{k\to\infty}\frac{1}{k\ell+j}\log|\mathcal{C}_{k\ell+j}| - \frac{\ell-j}{(k\ell+l)(k\ell+j)}\log|\mathcal{C}_{k\ell+j}| \geq r$ since the $\liminf$ of the subsequence $\{\frac{1}{k\ell+j}\log|\mathcal{C}_{k\ell+j}|\}_k$ is lower bounded by the $\liminf$ of the sequence $\{\frac{1}{n}\log|\mathcal{C}_n|\}_n$ and the second term vanishes. Moreover, (76) yields

$$\liminf_{k\to\infty}\left\{-\frac{1}{k\ell+j}\log F(\mathcal{C}_{k\ell+j},\mathcal{N}^{\otimes(k\ell+j)})\right\}$$
$$= \liminf_{k\to\infty}\left\{-\frac{1}{k\ell+j}\log F(\mathcal{C}_{k\ell+j},\mathcal{N}^{\otimes(k+1)\ell})\right\} \qquad (77)$$
$$\geq \liminf_{k\to\infty}\left\{-\frac{1}{k\ell+\ell}\log F(\mathcal{C}_{k\ell+j},\mathcal{N}^{\otimes(k+1)\ell})\right\} \qquad (78)$$
$$\geq \frac{c}{\ell}. \qquad (79)$$

Since this holds for all j, we conclude

$$\liminf_{n\to\infty}\left\{-\frac{1}{n}\log F(\mathcal{C}_n,\mathcal{N}^{\otimes n})\right\}$$
$$= \min_{j\in\{0,1,\ldots\ell-1\}}\liminf_{k\to\infty}\left\{-\frac{1}{k\ell+j}\log F(\mathcal{C}_{k\ell+j},\mathcal{N}^{\otimes k\ell+j})\right\}$$
$$\geq \frac{c}{\ell}. \qquad (80)$$

Hence, we conclude that the fidelity vanishes (exponentially fast in n) and that $\inf_{\ell\in\mathbb{N}}\frac{1}{\ell}R(\mathcal{N}^{\otimes\ell})$ is a strong converse rate. ■

Theorem 8 establishes the Rains information of a quantum channel as an upper bound on the strong converse capacity $Q^{\dagger}_{\text{pp}}(\mathcal{N})$ for CPPP assisted quantum communication over a channel $\mathcal{N}$. Thus, in summary, the following inequalities hold for all quantum channels:

$$I_{\text{c}}(\mathcal{N}) \leq \lim_{\ell\to\infty} \frac{1}{\ell} I_{\text{c}}(\mathcal{N}^{\otimes \ell}) = Q(\mathcal{N}) \leq Q^{\dagger}(\mathcal{N}) \\ \leq Q^{\dagger}_{\text{pp}}(\mathcal{N}) \leq \inf_{\ell \geq 1} \frac{1}{\ell} R(\mathcal{N}^{\otimes \ell}) \leq R(\mathcal{N}). \quad (81)$$

VI. Strong Converse Property for Quantum Communication over Dephasing Channels

In this section, we show that the Rains information of a generalized dephasing channel [10], [50]–[52] (also known as "Hadamard diagonal" channels [53] and "Schur multiplier" channels [29]) is equal to the coherent information of this channel. This extends [21, Theorem 6.2] from "maximally correlated states" to generalized dephasing channels. Consequently, the hierarchy in (81) collapses for generalized dephasing channels. In particular, we establish that the quantum capacity of this class of channels obeys the strong converse property and also that classical pre- and post-processing does not increase the capacity for these channels.

A generalized dephasing channel is any channel with an isometric extension of the form

$$U^{\mathcal{N}}_{A\to BE} := \sum_{x=0}^{d-1} |x\rangle_B \langle x|_A \otimes |\psi_x\rangle_E \,, \quad (82)$$

where the states $|\psi_x\rangle$ are arbitrary (not necessarily orthonormal). Note that these channels are degradable [10].

Proposition 10. *Let $\mathcal{N}$ be a generalized dephasing channel of the form* (82)*. Then $I_c(\mathcal{N}) = R(\mathcal{N})$.*

Proof: We have already seen in (81) that $I_c(\mathcal{N}) \leq R(\mathcal{N})$ holds for all channels (recall also that this follows from [23, Theorem 4]). We now establish that the opposite inequality holds for a generalized dephasing channel $\mathcal{N}$. Consider that any generalized dephasing channel $\mathcal{N}$ obeys the following covariance property:[7]

$$\mathcal{N}\left(Z_A(z)\rho Z_A^{\dagger}(z)\right) = Z_B(z)\mathcal{N}(\rho)Z_B^{\dagger}(z), \quad (83)$$

for $z \in \{0, \cdots, d-1\}$, where

$$Z_A(z)|x\rangle_A = \exp\{2\pi i x z/d\}\,|x\rangle_A, \quad (84)$$
$$Z_B(z)|x\rangle_B = \exp\{2\pi i x z/d\}\,|x\rangle_B. \quad (85)$$

Furthermore, a uniform mixing of these operators is equivalent to a "completely dephasing" channel:

$$\frac{1}{d}\sum_{z=0}^{d-1} Z_A(z)(\cdot)Z_A^{\dagger}(z) = \sum_{x=0}^{d-1} |x\rangle\langle x|_A(\cdot)|x\rangle\langle x|_A, \quad (86)$$

with the same true for the operators $\{Z_B(z)\}_{z\in\{0,\ldots,d-1\}}$. Then we can apply Proposition 2 to conclude that the Rains information of a generalized dephasing channel is maximized by a state with a Schmidt decomposition of the following form:

$$|\varphi^p\rangle_{RA} := \sum_x \sqrt{p_X(x)}|x\rangle_R|x\rangle_A, \quad (87)$$

for some probability distribution $p_X(x)$ and some orthonormal basis $\{|x\rangle_R\}$ for the reference system R (with the key result being that the basis $\{|x\rangle_A\}$ is "aligned with" the basis of the channel). That is,

$$R(\mathcal{N}) = \sup_{p_X} \inf_{\tau_{RB}\in \text{PPT}'(R:B)} D(\mathcal{N}(\varphi^p_{RA})\|\tau_{RB}). \quad (88)$$

Let Δ_P be a CPTP map constructed as follows:

$$\Delta_P(\cdot) = P(\cdot)P + (I-P)\,(\cdot)\,(I-P)\,, \quad (89)$$
$$\text{with} \qquad P = \sum_x |x\rangle\langle x|_R \otimes |x\rangle\langle x|_B\,. \quad (90)$$

Then the following chain of inequalities holds:

$$I_c(\mathcal{N}) = \sup_{\varphi_{RA}} \inf_{\sigma_B} D(\mathcal{N}(\varphi_{RA})\|I_R\otimes\sigma_B) \quad (91)$$
$$\geq \sup_{p_X} \inf_{\sigma_B} D(\mathcal{N}(\varphi^p_{RA})\|I_R\otimes\sigma_B) \quad (92)$$
$$\geq \sup_{p_X} \inf_{\sigma_B} D(\Delta_P(\mathcal{N}(\varphi^p_{RA}))\|\Delta_P(I_R\otimes\sigma_B)) \quad (93)$$
$$= \sup_{p_X} \inf_{\sigma_B} D(P(\mathcal{N}(\varphi^p_{RA}))P\|P(I_R\otimes\sigma_B)P) \quad (94)$$
$$= \sup_{p_X} \inf_{q} D(\mathcal{N}(\varphi^p_{RA})\|\sum_x q(x)|x\rangle\langle x|_R\otimes|x\rangle\langle x|_B) \quad (95)$$
$$\geq \sup_{p_X} \inf_{\tau_{RB}\in\text{PPT}'(R:B)} D(\mathcal{N}(\varphi^p_{RA})\|\tau_{RB}) = R(\mathcal{N}). \quad (96)$$

The first inequality follows by restricting the maximization to be over pure bipartite vectors of the form in (87). The second inequality follows from monotonicity of the relative entropy under the CPTP map Δ_P. The second equality follows because

$$D(\Delta_P[(\text{id}_R\otimes\mathcal{N})(\varphi^p_{RA})]\|\Delta_P(I_R\otimes\sigma_B)) \\ = D(P[(\text{id}_R\otimes\mathcal{N})(\varphi^p_{RA})]P\|P(I_R\otimes\sigma_B)P) \\ +D((I-P)[(\text{id}_R\otimes\mathcal{N})(\varphi^p_{RA})](I-P)\|(I-P)(I_R\otimes\sigma_B)(I-P)) \quad (97)$$

since $P\perp I-P$, and because the state $(\text{id}_R\otimes\mathcal{N})\,(\varphi^p_{RA})$ has no support in the subspace onto which $I-P$ projects, so that the last term above is equal to zero. The third equality follows because

$$P\left[(\text{id}_R\otimes\mathcal{N})\,(\varphi^p_{RA})\right]P = (\text{id}_R\otimes\mathcal{N})\,(\varphi^p_{RA}), \quad (98)$$
$$P\,(I_R\otimes\sigma_B)\,P = \sum_x q(x)|x\rangle\langle x|_R\otimes|x\rangle\,\langle x|_B, \quad (99)$$

for some distribution $q(x) = \langle x|_B\sigma_B|x\rangle_B$. The last inequality follows because the state $\sum_x q(x)|x\rangle\langle x|_R\otimes|x\rangle\langle x|_B \in \text{PPT}'(R\!:\!B)$, and the final equality follows from (88). ■

[7] The covariance in (83) in fact holds for any operators of the form $\sum_{x=0}^{d-1}\exp\{i\varphi_x\}\,|x\rangle\langle x|_A$ and $\sum_{x=0}^{d-1}\exp\{i\varphi_x\}\,|x\rangle\langle x|_B$ with $\varphi_x\in\mathbb{R}$, but it suffices to consider only the operators in (85) for our proof here.

VII. Strong Converse for Classical Communication Assisted Capacity of Erasure Channels

In this section, we consider a class of erasure channels of the form

$$\mathcal{N}_{A\to B} : \rho_A \mapsto (1-p)\rho_B + p|e\rangle\langle e|_B, \tag{100}$$

where $p \in [0,1]$ is the erasure probability, ρ_B is an isometric embedding of ρ_A into B, and $|e\rangle$ is a quantum state orthogonal to ρ_B.

Using the notion of entanglement cost of a channel, it was shown in [28] that the strong converse two-way assisted quantum capacity of the erasure channel is equal to $(1-p)\log|A|$. An alternate way of proving this (implicit in prior work) is as follows:

1) Use the fact that the erasure channel can be realized by the action of teleportation on the state $\mathcal{N}_{A\to B}(\Phi_{RA})$. As a consequence, any two-way assisted protocol for quantum communication can be simulated by an entanglement distillation protocol for the state $\mathcal{N}_{A\to B}(\Phi_{RA})$. The authors of [37, Section V] devised this teleportation simulation argument for discrete-variable channels, and it was subsequently generalized to more channels, including continuous-variable channels, in [54].
2) Apply the strong converse theorem from [32] for distillable entanglement of a state and evaluate the Rains relative entropy.[8]

This latter approach is similar in spirit to what we do below, which however utilizes the methods given in this paper.

Proposition 11. *Let $\mathcal{N}$ be an erasure channel of the form (100). Then, $Q_{\rm pp}(\mathcal{N}) = Q^\dagger_{\rm pp}(\mathcal{N}) = (1-p)\log|A|$.*

Proof: First note that erasure channels are covariant under the discrete Heisenberg–Weyl unitary group acting on A, and thus Proposition 2 implies that

$$\begin{aligned} R(\mathcal{N}) &= R(\mathcal{N}(\psi_{RA})) && (101)\\ &\leq D(\mathcal{N}(\psi_{RA})\|\tau_{RB}) && (102)\\ &= H(\tau_{RB}) - H(\rho_{RB}) && (103)\end{aligned}$$

where ψ_{RA} is a maximally entangled state, $\rho_{RB} = \mathcal{N}(\psi_{RA})$, and τ_{RB} is chosen as follows:

$$|\psi\rangle_{RA} = \sum_x \frac{1}{\sqrt{|A|}}|x\rangle_R \otimes |x\rangle_A, \tag{104}$$

$$\tau_{RB} = \sum_x \frac{1}{|A|}|x\rangle\langle x|_R \otimes \mathcal{N}\big(|x\rangle\langle x|_A\big). \tag{105}$$

[8] Note here that one could also apply the relative entropy of entanglement to obtain an upper bound as done in [54], but this bound would be generally weaker than that given by the Rains relative entropy.

Evaluating this, we find

$$\begin{aligned} H(RB)_\tau &= H(R)_\tau + H(B|R)_\tau && (106)\\ &= \log|A| + \sum_x \frac{1}{|A|} H(B)_{\mathcal{N}(|x\rangle\langle x|_A)} && (107)\\ &= \log|A| + \sum_x \frac{1}{|A|}[H((1-p)|x\rangle\langle x|_A + p|e\rangle\langle e|)] && (108)\\ &= \log|A| + h_2(p), && (109)\\ H(RB)_\rho &= H((1-p)\psi_{RA} + p\pi_R \otimes |e\rangle\langle e|) && (110)\\ &= h_2(p) + (1-p)H(\psi_{RA}) + pH(\pi_R \otimes |e\rangle\langle e|) && (111)\\ &= h_2(p) + p\log|A|. && (112)\end{aligned}$$

In the above, $H(B|R) := H(BR) - H(R)$ is the conditional entropy. Therefore, we conclude that $Q^\dagger_{\rm pp}(\mathcal{N}) \leq R(\mathcal{N}) \leq (1-p)\log|A|$.

It is well known that any rate $r < (1-p)\log|A|$ can be achieved with CPPP assistance [55]. We review this argument briefly. Let Alice prepare and send n maximally entangled states through $\mathcal{N}^{\otimes n}$. Bob then records which instances of the channel led to an erasure, and communicates this to Alice. They will then use the correctly transmitted states to distill a maximally entangled state of dimension rn. This will succeed whenever $rn/\log|A|$ is smaller than the number of correctly transmitted states, which will happen with probability 1 as $n\to\infty$ as long as $r < (1-p)\log|A|$. This establishes that $Q_{\rm pp}(\mathcal{N}) \geq (1-p)\log|A|$ and concludes the proof. ∎

Finally, recall that the erasure channel is degradable for $p \leq 1/2$ [10], and we can thus calculate its unassisted quantum capacity to be $Q(\mathcal{N}) = \max\{(1-2p)\log|A|, 0\}$ [55]. Thus, we have an example of a channel where $Q(\mathcal{N}) < R(\mathcal{N})$, which means that our upper bound in terms of the Rains information of the channel is not sufficient to establish the strong converse property for general degradable channels.

VIII. Conclusion

This paper has established that the Rains information of a quantum channel is a strong converse rate for quantum communication. The main application of the first result is to establish the strong converse property for the quantum capacity of all generalized dephasing channels. Going forward from here, there are several questions to consider. First, are there any other channels besides the generalized dephasing ones for which the Rains information is equal to the coherent information? If true, the theorems established here would establish the strong converse property for these channels. For example, can we prove a strong converse theorem for the quantum capacity of general Hadamard channels?

Is it possible to show weak subadditivity of a Rényi coherent information quantity for some class of channels, addressing the original question posed in [25]? To this end, the developments in [56] might be helpful. Next, is it possible to show that the Rains information of a general quantum channel represents a strong converse rate for quantum communication assisted by interactive forward and backward classical communication? Similarly, can one show that the squashed entanglement of a channel [57] is a strong converse rate for this task? Recent work has proved that the squashed entanglement of a quantum channel is an upper bound on the quantum capacity with interactive forward and backward classical communication, and it could be that the quantities defined in [58] would be helpful for settling this question.

Finally, now that the strong converse holds for the classical capacity, the quantum capacity, and the entanglement-assisted capacity of all generalized dephasing channels, can we establish that the strong converse holds for trade-off capacities of these channels, in the sense of [52], [59]? Can we establish second-order characterizations for this class of channels, in the sense of [60]–[63]? See [64] for recent progress on this last question.

Acknowledgements. We acknowledge many discussions with our colleagues regarding the strong converse for quantum capacity. This includes Mario Berta, Frederic Dupuis, Will Matthews, Ciara Morgan, Naresh Sharma, Stephanie Wehner, and Dong Yang. We are grateful to the anonymous referees for many helpful comments about our paper. MT is funded by the Ministry of Education (MOE) and National Research Foundation Singapore, as well as MOE Tier 3 Grant "Random numbers from quantum processes" (MOE2012-T3-1-009). MMW acknowledges startup funds from the Department of Physics and Astronomy at LSU, support from the NSF through Award No. CCF-1350397, and support from the DARPA Quiness Program through US Army Research Office award W31P4Q-12-1-0019. MMW is also grateful to the quantum information theory group at the Universitat Autònoma de Barcelona and to Stephanie Wehner's quantum information group at the Centre for Quantum Technologies, National University of Singapore, for hosting him for research visits during May-June 2014. AW acknowledges financial support by the Spanish MINECO, project FIS2008-01236 with the support of FEDER funds, the EC STREP "RAQUEL", the ERC Advanced Grant "IRQUAT", and the Philip Leverhulme Trust.

Appendix A
Proof of Lemma 3

Proof of Lemma 3: The first statement follows because the underlying sandwiched relative entropy, $\widetilde{D}_\alpha(\rho\|\sigma)$, is monotonically increasing in α [41, Theorem 7] for all $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, i.e.

$$\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\beta(\rho\|\sigma) \tag{113}$$

for all $\alpha \geq \beta \geq 0$. This already establishes that the limit exists and satisfies

$$\lim_{\alpha\to 1^+} \widetilde{R}_\alpha(\mathcal{N}) \geq R(\mathcal{N}). \tag{114}$$

We would like to show the opposite inequality. Consider the following bound [65, Lemma 8] (see also [16, Eq. (21)])

$$\widetilde{D}_{1+\delta}(\rho\|\sigma) \leq D(\rho\|\sigma) + 4\delta\left[\log\nu\left(\rho,\sigma\right)\right]^2, \tag{115}$$

which holds for $\delta \in (0, \log 3/\left(4\log\nu\left(\rho,\sigma\right)\right))$ where

$$\nu(\rho,\sigma) := \operatorname{tr}(\rho^{3/2}\sigma^{-1/2}) + \operatorname{tr}(\rho^{1/2}\sigma^{1/2}) + 1. \tag{116}$$

Let us for the moment assume that $\sigma > 0$ with $\operatorname{tr}(\sigma) \leq 1$, and its smallest eigenvalue is denoted as λ. In this case, we can bound $\nu(\rho,\sigma) \leq 2 + 1/\sqrt{\lambda}$.

For any state ρ_{RB} and $\delta > 0$ sufficiently small, we can apply the bound above to arrive at

$$\begin{aligned}
&\inf_{\sigma_{RB}\in\mathrm{PPT}'(R:B)} \widetilde{D}_{1+\delta}(\rho_{RB}\|\sigma_{RB}) && (117)\\
&\leq \inf_{\tau_{RB}\in\mathrm{PPT}'(R:B)} \widetilde{D}_{1+\delta}(\rho_{RB}\|(1-\delta)\tau_{RB}+\delta\pi_{RB}) && (118)\\
&\leq \inf_{\tau_{RB}\in\mathrm{PPT}'(R:B)} D(\rho_{RB}\|(1-\delta)\tau_{RB}+\delta\pi_{RB}) \\
&\qquad + 4\delta\left(\log\left(2+\frac{\sqrt{|A||B|}}{\sqrt{\delta}}\right)\right)^2 && (119)\\
&\leq \inf_{\tau_{RB}\in\mathrm{PPT}'(R:B)} D(\rho_{RB}\|\tau_{RB}) + \log\frac{1}{1-\delta} \\
&\qquad + 4\delta\left(\log\left(2+\frac{\sqrt{|A||B|}}{\sqrt{\delta}}\right)\right)^2 && (120)
\end{aligned}$$

The first inequality follows by picking σ_{RB} to be of the form $\sigma_{RB} = (1-\delta)\,\tau_{RB} + \delta\pi_{RB}$ where π_{RB} is the fully mixed state on RB. Also, note that $(1-\delta)\,\tau_{RB} + \delta\pi_{RB} \in \mathrm{PPT}'(R:B)$. To verify the second inequality, note that the minimum eigenvalue of $(1-\delta)\,\tau_{RB} + \delta\pi_{RB}$ is always larger than $\delta/(|A||B|)$. (The system R can be chosen to be of size $|A|$ without loss of generality, where $|A|$ is the dimension of the channel input.) Finally, recall that $D(\rho\|\sigma) \leq D\left(\rho\|\sigma'\right)$ whenever $\sigma' \leq \sigma$ to verify the last inequality.

Since, crucially, the upper bound in (120) is uniform in ρ_{RB}, we can immediately conclude that

$$\begin{aligned}
\widetilde{R}_{1+\delta}(\mathcal{N}) &\leq R(\mathcal{N}) + \log\frac{1}{1-\delta} \\
&\qquad + 4\delta\left(\log\left(2+\frac{\sqrt{|A||B|}}{\sqrt{\delta}}\right)\right)^2 \qquad (121)
\end{aligned}$$

by maximizing over channel output states as in the definition of $\widetilde{R}_{1+\delta}$ and R. Thus, $\lim_{\alpha\to1^+}\widetilde{R}_\alpha(\mathcal{N}) \leq R(\mathcal{N})$, concluding the proof. ■

Appendix B
Proof of Corollary 7

Proof: We can focus only on operators τ_{RB} such that $\operatorname{supp}(\rho_{RB}) \subseteq \operatorname{supp}(\tau_{RB})$, where $\rho_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})$. This is because the quantity of interest contains a minimization over all τ_{RB}. For any sufficiently small $\delta > 0$

$$R(\mathcal{N}^{\otimes n}) \leq \widetilde{R}_{1+\delta}(\mathcal{N}^{\otimes n}) \leq \frac{1+\delta}{\delta}\,|A|^2\log n + n\widetilde{R}_{1+\delta}(\mathcal{N}). \tag{122}$$

The first inequality is from the monotonicity of the sandwiched Rényi relative entropy in the Rényi parameter [41, Theorem 7]. The second inequality follows from Theorem 6.

This invites an application of (121), which gives

$$\frac{1}{n} R(\mathcal{N}^{\otimes n}) \leq R(\mathcal{N}) + \frac{1+\delta}{\delta} |A|^2 \frac{\log n}{n} + \log \frac{1}{1-\delta} + 4\delta \left(\log \left(2 + \frac{\sqrt{|A||B|}}{\sqrt{\delta}} \right) \right)^2 \quad (123)$$

Choosing $\delta = 1/\sqrt{n}$ and taking the limit $n \to \infty$ then immediately yields the desired result. ■

References

[1] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory," *Physical Review A*, vol. 52, pp. R2493–R2496, October 1995.

[2] B. Schumacher and M. A. Nielsen, "Quantum data processing and error correction," *Physical Review A*, vol. 54, pp. 2629–2635, October 1996. arXiv:quant-ph/9604022.

[3] H. Barnum, E. Knill, and M. A. Nielsen, "On quantum fidelities and channel capacities," *IEEE Transactions on Information Theory*, vol. 46, pp. 1317–1329, July 2000. arXiv:quant-ph/9809010.

[4] H. Barnum, M. A. Nielsen, and B. Schumacher, "Information transmission through a noisy quantum channel," *Physical Review A*, vol. 57, pp. 4153–4175, June 1998. arXiv:quant-ph/9702049.

[5] S. Lloyd, "Capacity of the noisy quantum channel," *Physical Review A*, vol. 55, pp. 1613–1622, March 1997. arXiv:quant-ph/9604015.

[6] P. W. Shor, "The quantum channel capacity and coherent information," in *Lecture Notes, MSRI Workshop on Quantum Computation*, 2002.

[7] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 51, pp. 44–55, January 2005. arXiv:quant-ph/0304127.

[8] R. Klesse, "A random coding based proof for the quantum coding theorem," *Open Systems & Information Dynamics*, vol. 15, pp. 21–45, March 2008. arXiv:0712.2558.

[9] P. Hayden, P. W. Shor, and A. Winter, "Random quantum codes from Gaussian ensembles and an uncertainty relation," *Open Systems & Information Dynamics*, vol. 15, pp. 71–89, March 2008. arXiv:0712.0975.

[10] I. Devetak and P. W. Shor, "The capacity of a quantum channel for simultaneous transmission of classical and quantum information," *Communications in Mathematical Physics*, vol. 256, pp. 287–303, June 2005. arXiv:quant-ph/0311131.

[11] J. Wolfowitz, *Coding Theorems of Information Theory*, vol. 31. Springer, 1964.

[12] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 19, pp. 357–359, May 1973.

[13] T. Ogawa and H. Nagaoka, "Strong converse to the quantum channel coding theorem," *IEEE Transactions on Information Theory*, vol. 45, pp. 2486–2489, November 1999. arXiv:quant-ph/9808063.

[14] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, pp. 2481–2485, November 1999.

[15] R. Koenig and S. Wehner, "A strong converse for classical channel coding using entangled inputs," *Physical Review Letters*, vol. 103, p. 070504, August 2009. arXiv:0903.2838.

[16] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy," *Communications in Mathematical Physics*, vol. 331, pp. 593–622, October 2014. arXiv:1306.1586.

[17] M. M. Wilde and A. Winter, "Strong converse for the classical capacity of the pure-loss bosonic channel," *Problems of Information Transmission*, vol. 50, pp. 117–132, April 2014. arXiv:1308.6732.

[18] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Transactions on Information Theory*, vol. 60, pp. 2926–2959, May 2014. arXiv:0912.5537.

[19] M. Berta, M. Christandl, and R. Renner, "The quantum reverse Shannon theorem based on one-shot information theory," *Communications in Mathematical Physics*, vol. 306, pp. 579–615, August 2011. arXiv:0912.3805.

[20] M. Gupta and M. M. Wilde, "Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity," *Communications in Mathematical Physics*, vol. 334, pp. 867–887, March 2015. arXiv:1310.7028.

[21] E. M. Rains, "A semidefinite program for distillable entanglement," *IEEE Transactions on Information Theory*, vol. 47, pp. 2921–2933, November 2001. arXiv:quant-ph/0008047.

[22] K. Audenaert, B. De Moor, K. G. H. Vollbrecht, and R. F. Werner, "Asymptotic relative entropy of entanglement for orthogonally invariant states," *Physical Review A*, vol. 66, p. 032310, September 2002. arXiv:quant-ph/0204143.

[23] M. B. Plenio, S. Virmani, and P. Papadopoulos, "Operator monotones, the reduction criterion and the relative entropy," *Journal of Physics A: Mathematical and General*, vol. 33, p. L193, June 2000. arXiv:quant-ph/0002075.

[24] M. Horodecki and P. Horodecki, "Reduction criterion of separability and limits for a class of distillation protocols," *Physical Review A*, vol. 59, pp. 4206–4216, June 1999.

[25] N. Sharma and N. A. Warsi, "On the strong converses for the quantum channel capacity theorems," June 2012. arXiv:1205.1712.

[26] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computation*, pp. 1327–1333, September 2010.

[27] H. Nagaoka, "Strong converse theorems in quantum information theory," *Proceedings of ERATO Workshop on Quantum Information Science*, p. 33, 2001. Also appeared in Asymptotic Theory of Quantum Statistical Inference, ed. M. Hayashi, World Scientific, 2005.

[28] M. Berta, F. G. S. L. Brandao, M. Christandl, and S. Wehner, "Entanglement cost of quantum channels," *IEEE Transactions on Information Theory*, vol. 59, pp. 6779–6795, October 2013. arXiv:1108.5357.

[29] C. Morgan and A. Winter, ""Pretty strong" converse for the quantum capacity of degradable channels," *IEEE Transactions on Information Theory*, vol. 60, pp. 317–333, January 2014. arXiv:1301.4927.

[30] M. M. Wilde and A. Winter, "Strong converse for the quantum capacity of the erasure channel for almost all codes," *Proceedings of the 9th Conference on the Theory of Quantum Computation, Communication and Cryptography*, May 2014. arXiv:1402.3626.

[31] G. Smith, J. A. Smolin, and A. Winter, "The quantum capacity with symmetric side channels," *IEEE Transactions on Information Theory*, vol. 54, pp. 4208–4217, September 2008. arXiv:quant-ph/0607039.

[32] M. Hayashi, *Quantum Information: An Introduction*. Springer, 2006.

[33] A. Uhlmann, "The "transition probability" in the state space of a *-algebra," *Reports on Mathematical Physics*, vol. 9, no. 2, pp. 273–279, 1976.

[34] R. F. Werner, "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model," *Physical Review A*, vol. 40, pp. 4277–4281, October 1989.

[35] A. Peres, "Separability criterion for density matrices," *Physical Review Letters*, vol. 77, pp. 1413–1415, August 1996. arXiv:quant-ph/9604005.

[36] M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of mixed states: necessary and sufficient conditions," *Physics Letters A*, vol. 223, pp. 1–8, November 1996. arXiv:quant-ph/9605038.

[37] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-state entanglement and quantum error correction," *Physical Review A*, vol. 54, pp. 3824–3851, November 1996. arXiv:quant-ph/9604024.

[38] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter, "Everything you always wanted to know about LOCC (but were afraid to ask)," *Communications in Mathematical Physics*, vol. 328, pp. 303–326, May 2014. arXiv:1210.4583.

[39] M. M. Wilde, *From Classical to Quantum Shannon Theory*. December 2015. arXiv:1106.1445v6.

[40] E. M. Rains, "Bound on distillable entanglement," *Physical Review A*, vol. 60, pp. 179–184, July 1999. arXiv:quant-ph/9809082.

[41] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: a new definition and some properties," *Journal of Mathematical Physics*, vol. 54, p. 122203, December 2013. arXiv:1306.3142.

[42] R. L. Frank and E. H. Lieb, "Monotonicity of a relative Rényi entropy," *Journal of Mathematical Physics*, vol. 54, p. 122201, December 2013. arXiv:1306.5358.

[43] S. Beigi, "Sandwiched Rényi divergence satisfies data processing inequality," *Journal of Mathematical Physics*, vol. 54, p. 122202, December 2013. arXiv:1306.5920.

[44] M. Mosonyi and T. Ogawa, "Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies," September 2013. arXiv:1309.3228.

[45] G. Smith and J. A. Smolin, "Detecting incapacity of a quantum channel," *Physical Review Letters*, vol. 108, p. 230507, June 2012. arXiv:1108.1807.

[46] P. Horodecki, M. Horodecki, and R. Horodecki, "Binding entanglement channels," *Journal of Modern Optics*, vol. 47, pp. 347–354, July 2000. arXiv:quant-ph/9904092.

[47] R. Renner, *Security of Quantum Key Distribution*. PhD thesis, ETH Zürich, December 2005. arXiv:quant-ph/0512258.

[48] M. Christandl, R. König, and R. Renner, "Postselection technique for quantum channels with applications to quantum cryptography," *Physical Review Letters*, vol. 102, p. 020504, January 2009. arXiv:0809.3019.

[49] G. Smith and J. A. Smolin, "Additive extensions of a quantum channel," *Proceedings of the IEEE Information Theory Workshop*, pp. 368–372, May 2008. arXiv:0712.2471.

[50] J. Yard, P. Hayden, and I. Devetak, "Capacity theorems for quantum multiple-access channels: Classical-quantum and quantum-quantum capacity regions," *IEEE Transactions on Information Theory*, vol. 54, pp. 3091–3113, July 2008. arXiv:quant-ph/0501045.

[51] M.-H. Hsieh, I. Devetak, and A. Winter, "Entanglement-assisted capacity of quantum multiple-access channels," *IEEE Transactions on Information Theory*, vol. 54, pp. 3078–3090, July 2008. arXiv:quant-ph/0511228.

[52] K. Brádler, P. Hayden, D. Touchette, and M. M. Wilde, "Trade-off capacities of the quantum Hadamard channels," *Physical Review A*, vol. 81, p. 062312, June 2010. arXiv:1001.1732.

[53] C. King, K. Matsumoto, M. Nathanson, and M. B. Ruskai, "Properties of conjugate channels with applications to additivity and multiplicativity," *Markov Processes and Related Fields*, vol. 13, no. 2, pp. 391–423, 2007. J. T. Lewis memorial issue, arXiv:quant-ph/0509126.

[54] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, "Fundamental limits of repeaterless quantum communications," 2015. arXiv:1510.08863.

[55] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, "Capacities of quantum erasure channels," *Physical Review Letters*, vol. 78, pp. 3217–3220, April 1997. arXiv:quant-ph/9701015.

[56] M. Tomamichel, M. Berta, and M. Hayashi, "Relating different quantum generalizations of the conditional Rényi entropy," *Journal of Mathematical Physics*, vol. 55, p. 082206, August 2014. arXiv:1311.3887.

[57] M. Takeoka, S. Guha, and M. M. Wilde, "The squashed entanglement of a quantum channel," *IEEE Transactions on Information Theory*, vol. 60, pp. 4987–4998, August 2014. arXiv:1310.0129.

[58] M. Berta, K. Seshadreesan, and M. M. Wilde, "Rényi generalizations of the conditional quantum mutual information," *Journal of Mathematical Physics*, vol. 56, p. 022205, February 2015. arXiv:1403.6102.

[59] M. M. Wilde and M.-H. Hsieh, "The quantum dynamic capacity formula of a quantum channel," *Quantum Information Processing*, vol. 11, pp. 1431–1463, December 2012. arXiv:1004.0458.

[60] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," *IEEE Transactions on Information Theory*, vol. 59, pp. 7693–7710, November 2013. arXiv:1208.1478.

[61] M. Tomamichel and V. Y. F. Tan, "Second-order asymptotics for the classical capacity of image-additive quantum channels," *Communication in Mathematical Physics*, vol. 338, pp. 103–137, August 2015. arXiv:1308.6503.

[62] N. Datta and F. Leditzky, "Second-order asymptotics for source coding, dense coding and pure-state entanglement conversions," *IEEE Transactions on Information Theory*, vol. 61, pp. 582–608, March 2016. arXiv:1403.2543.

[63] N. Datta, M. Tomamichel, and M. M. Wilde, "On the second-order asymptotics for entanglement-assisted communication," *Quantum Information Processing*, vol. 15, pp. 2569–2591, May 2016. arXiv:1405.1797.

[64] M. Tomamichel, M. Berta, and J. M. Renes, "Quantum coding with finite resources," *Nature Communications*, vol. 7, p. 11419, April 2016. arXiv:1504.04617.

[65] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Transactions on Information Theory*, vol. 55, pp. 5840–5847, December 2009. arXiv:0811.1221.

Marco Tomamichel (M'13) received the M.Sc. in Electrical Engineering and Information Technology degree from ETH Zurich (Switzerland) in 2007. He then graduated with a Ph.D. in Physics at the Institute of Theoretical Physics at ETH Zurich in 2012. From 2012 to 2014 he was a Research Fellow and then Senior Research Fellow at the Centre for Quantum Technologies at the National University of Singapore. Currently he is a Lecturer in the School of Physics at the University of Sydney. His research interests include classical and quantum information theory with finite resources as well as applications to cryptography.

Mark M. Wilde (M'99-SM'13) was born in Metairie, Louisiana, USA. He received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, California, in 2008. He is an Assistant Professor in the Department of Physics and Astronomy and the Center for Computation and Technology at Louisiana State University. His current research interests are in quantum Shannon theory, quantum optical communication, quantum computational complexity theory, and quantum error correction.

Andreas Winter received a Diploma degree in Mathematics from the Freie Universität Berlin, Berlin, Germany, in 1997, and a Ph.D. degree from the Fakultät für Mathematik, Universität Bielefeld, Bielefeld, Germany, in 1999. He was Research Associate at the University of Bielefeld until 2001, and then with the Department of Computer Science at the University of Bristol, Bristol, UK. In 2003, still with the University of Bristol, he was appointed Lecturer in Mathematics, and in 2006 Professor of Physics of Information. Since 2012 he has been ICREA Research Professor with the Universitat Autònoma de Barcelona, Barcelona, Spain.