Abstract

This paper establishes the Rains information of a quantum channel as a strong converse bound for quantum communication when using the channel many independent times. We also settle an open question posed by Rains, namely, to show that the Rains bound for entanglement distillation represents a strong converse rate for this task. The main application of our first result is to settle the strong converse question for the quantum capacity of all generalized dephasing channels.

1 Introduction

The quantum capacity of a quantum channel is defined to be the maximum rate at which it is possible to transmit qubits over many independent uses of the channel, such that a receiver can recover the qubits with fidelity approaching one as the number of channel uses becomes large. The question of determining the quantum capacity of a quantum channel was set out by Shor in his seminal paper on quantum error correction [36]. Since then, a number of works established a “multi-letter” upper bound on the quantum channel capacity in terms of the coherent information [34, 3, 4], and the coherent information lower bound on quantum capacity was demonstrated by a sequence of works [24, 37, 13] which are often said to bear “increasing standards of rigor.” In more detail, the work in [24, 37, 13] showed that the following coherent information of channel $N$ is an achievable rate for quantum communication:

$$I_c(N) = \max_{\phi_{RA}} I(R|B)_\rho,$$

(1.1)

where $\rho_{RB} = N_{A\rightarrow B}(\phi_{RA})$, the optimization is over all pure, bipartite states $\phi_{RA}$, and the coherent information of a bipartite state $\rho_{RB}$ is defined as $I(R|B)_\rho = H(B)_\rho - H(RB)_\rho$, with the von Neumann entropies $H(B)_\rho \equiv -\text{Tr}\{\rho_B \log \rho_B\}$ and $H(RB)_\rho \equiv -\text{Tr}\{\rho_{RB} \log \rho_{RB}\}$. From the above result, we can also conclude that the rate $I_c(N \otimes l)/l$ is achievable for any positive integer $l$, simply by applying the formula in (1.1) to the “superchannel” $N \otimes l$ and normalizing. By a limiting

*Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore
†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 Teòrica, Informació i Fenomens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain

1However, see the later works [22] and [17], which set [24] and [37], respectively, on a firm foundation.
2Unless stated otherwise, all logarithms in this paper are taken base two.
argument, we find that the regularized coherent information $\lim_{l \to \infty} I_c(N^\otimes l)/l$ is also achievable, and Refs. [34, 3] established this regularized coherent information as an upper bound on quantum capacity. 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 (1.1) is equal to the quantum capacity for the class of degradable quantum channels [14]. 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 approaches 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.

Although the above understanding of quantum capacity cements its role as a threshold for reliable quantum communication, we are left to wonder whether it is possible to sharpen this interpretation. For example, it has been known since the early days of classical information theory that the classical capacity of a classical channel obeys the strong converse property [52, 1]: if the rate of communication exceeds 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 [28, 51, 23, 50, 48] and also for the entanglement-assisted classical capacity of all quantum channels [6, 8, 16].

A few papers have made some partial progress on or addressed the strong converse for quantum capacity question [6, 8, 7, 25, 35, 49], with none however establishing that the strong converse holds for any nontrivial class of channels. Refs. [6, 8] 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. [7] 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. The most important progress to date for establishing a strong converse for quantum capacity is from [25]. 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 a jump in the quantum error from zero to 1/2 as soon as the communication rate exceeds the quantum capacity threshold. 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 [35], which hitherto has remained unsolved. The later work in [49] demonstrated that the strong converse property holds for randomly selected codes with a communication rate exceeding the quantum capacity of the quantum erasure channel (however, we stress that a true strong converse is one which holds for all codes whose rate exceeds capacity).
2 Overview of results

In this paper, we define the Rains information of a quantum channel \( N_{A \rightarrow B} \) as follows:\(^3\)

\[ R (N) \equiv \max_{\phi_{RA}} R (N_{A \rightarrow B} (\phi_{RA})) , \]

where the Rains relative entropy of a bipartite state \( \rho_{RB} \) is defined as

\[ R (\rho_{RB}) \equiv \min_{\tau_{RB} \in \mathcal{T} (R : B)} D (\rho_{RA} \parallel \tau_{RB}) , \]

with

\[ \mathcal{T} (R : B) \equiv \{ \tau_{RB} : \tau_{RB} \geq 0 \land \| T_B (\tau_{RB}) \|_1 \leq 1 \} , \]

\[ D (\omega \parallel \sigma) \equiv \Tr \{ \omega (\log \omega - \log \sigma) \} , \]

and \( T_B \) the partial transpose (these concepts are explained in more detail in Section 3). The quantity in (2.2), often called the “Rains bound,” was first explored by Rains in the context of entanglement distillation \(^32\) and later refined to the form above in \(^2\).

Our main contribution is that the Rains information of a quantum channel is equal to a strong converse bound on its quantum capacity (Theorem 9). That is, if the quantum communication rate of any protocol for a channel \( N_{A \rightarrow B} \) exceeds its Rains information, then the fidelity of the scheme decays exponentially fast to zero as the number \( n \) of channel uses increases. We establish this theorem by exploiting the generalized divergence framework of Sharma and Warsi \(^35\) (first explored in the classical context in \(^30\)). However, our main departing point from that work is to consider a different class of “useless” channels for quantum data transmission. That is, in \(^35\), the main idea was to exploit a generalized divergence to compare the output of the channel with an operator of the form \( I_R \otimes \sigma_B \), which can be viewed as a “useless” channel that replaces the input and reference system with the operator \( I_R \otimes \sigma_B \). 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 \( \mathcal{T} (R : B) \), 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). From there, some standard arguments from \(^28\) conclude the proof of Theorem 9.

The main application of this result is to settle the strong converse question for any generalized dephasing channel (Conclusion 11). The action of any channel in this class on an input state \( \rho \) is as follows:

\[ N_H (\rho) = \sum_{x,y=0}^{d-1} \langle x | A \rho | y \rangle_A \langle \psi_y | \psi_x \rangle | x \rangle B , \]

\(^3\)It should be clear from the context whether \( R \) refers to “Rains” or to a reference system.

\(^4\)We also mention that our main contribution here sets some previous claims from \(^38\), \(^40\) on a firm foundation. Namely, in Refs. \(^38\), \(^40\), the authors claim that the Rains bound for entanglement distillation leads to a weak converse upper bound on the quantum capacity of any channel. However, these papers do not support this claim in any way. Our strong converse result in the present paper thus provides a proof of these claims.
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

\[
N_H (\rho) = (1 - p) \rho + pZ \rho Z,
\]

(2.6)

where the dephasing parameter \( p \in [0, 1] \) and \( Z \) is the Pauli \( \sigma_Z \) operator. We prove this result by showing that the Rains information of a generalized dephasing channel is equal to its coherent information (Proposition 10), and invoking the well known result that the coherent information of this channel is equal to its quantum capacity [14].

Finally, we settle an open question posed by Rains in [32]. Namely, we show that the Rains bound on the rate of entanglement distillation of an i.i.d. state can be interpreted as a strong converse bound.

The rest of the paper proceeds as follows. First, we establish some notation and definitions that are used throughout the paper. Section 4 reviews the definition of an entanglement generation code, and Section 5 introduces the generalized divergence framework for our setting here and some lemmas that are useful for later sections. Proposition 5 then establishes a “one-shot” bound for quantum communication, in terms of a Rényi-like Rains information of a quantum channel. The rest of the paper from there proceeds along the lines mentioned above, and we conclude in Section 11 with a summary and some open questions.

3 Notation and definitions

Norms, states, channels, and measurements. Let \( \mathcal{B}(\mathcal{H}) \) denote the algebra of bounded linear operators acting on a Hilbert space \( \mathcal{H} \). We restrict ourselves to finite-dimensional Hilbert spaces throughout this paper. For \( \alpha \geq 1 \), we define the \( \alpha \)-norm of an operator \( X \) as

\[
\|X\|_\alpha \equiv \text{Tr}\{ (\sqrt{X^\dagger X})^\alpha \}^{1/\alpha}.
\]

(3.1)

The trace norm of an operator \( A \) is equal to \( \|A\|_1 \) and has the following alternate characterization:

\[
\|A\|_1 = \max_U \text{Tr}\{UA\},
\]

(3.2)

where the optimization is over all unitaries. Let \( \mathcal{P}(\mathcal{H}) \) denote the subset of positive semi-definite operators (we often simply say that an operator is “positive” if it is positive semi-definite). We also write \( X \geq 0 \) if \( X \in \mathcal{P}(\mathcal{H}) \). An operator \( \rho \) is in the set \( \mathcal{S}(\mathcal{H}) \) of density operators if \( \rho \in \mathcal{P}(\mathcal{H}) \) and \( \text{Tr}\{\rho\} = 1 \). The fidelity between two density operators \( \rho \) and \( \sigma \) is defined as

\[
F(\rho, \sigma) \equiv \|\sqrt{\rho} \sqrt{\sigma}\|_1^2.
\]

(3.3)

The tensor product of two Hilbert spaces \( \mathcal{H}_A \) and \( \mathcal{H}_B \) is denoted by \( \mathcal{H}_A \otimes \mathcal{H}_B \). Given a multiparticle density operator \( \rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B \), we unambiguously write \( \rho_A = \text{Tr}_B\{\rho_{AB}\} \) for the reduced density operator on system \( A \). A linear map \( \mathcal{N}_{A\rightarrow B} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B) \) is positive if \( \mathcal{N}_{A\rightarrow B}(\sigma_A) \in \mathcal{P}(\mathcal{H}_B) \) whenever \( \sigma_A \in \mathcal{P}(\mathcal{H}_A) \). Let \( \text{id}_A \) denote the identity map acting on a system \( A \). A linear map \( \mathcal{N}_{A\rightarrow B} \) is completely positive if the map \( \text{id}_R \otimes \mathcal{N}_{A\rightarrow B} \) is positive for a reference system \( R \) of arbitrary size. A linear map \( \mathcal{N}_{A\rightarrow B} \) is trace-preserving if \( \text{Tr}\{\mathcal{N}_{A\rightarrow B}(\tau_A)\} = \text{Tr}\{\tau_A\} \) for all input operators \( \tau_A \in \mathcal{B}(\mathcal{H}_A) \). If a linear map is completely positive and trace-preserving (CPTP), we say that it
The sandwiched Rényi relative entropy is monotone under a CPTP map.

Lemma 1
Suppose From the assumption that
Proof. If supp (ρ) ⊆ supp (σ) or α ∈ (0, 1) then
\( D_\alpha (\rho\|\sigma) \leq D_\alpha (\rho\|\sigma) \).

Entropies. We define the sandwiched Rényi quasi-relative entropy of order α, for ρ ∈ \( \mathcal{S}(\mathcal{H}) \) and σ ∈ \( \mathcal{P}(\mathcal{H}) \), as follows [27, 50]:
\[
\tilde{Q}_\alpha (\rho\|\sigma) \equiv \begin{cases} 
\text{Tr} \left\{ \left( \sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha} \right)^\alpha \right\} & \text{if supp (ρ) \subseteq supp (σ) or } \alpha \in (0, 1) \\
\infty & \text{else}
\end{cases}.
\]

We define the sandwiched Rényi relative entropy (also known as the “quantum Rényi divergence”) as [27, 50]
\[
\tilde{D}_\alpha (\rho\|\sigma) \equiv \frac{1}{\alpha - 1} \log \tilde{Q}_\alpha (\rho\|\sigma).
\]

The sandwiched Rényi relative entropy is defined for all \( \alpha \in (0, 1) \cup (1, \infty) \), with it being defined for \( \alpha \in \{0, 1, \infty\} \) in the limit as \( \alpha \) approaches 0, 1, and \( \infty \), respectively. Our focus in this paper will be on the regime \( \alpha > 1 \). The quantity \( \tilde{Q}_\alpha \) is clearly multiplicative under tensor-product operators:
\[
\tilde{Q}_\alpha (\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) = \tilde{Q}_\alpha (\rho_1\|\sigma_1) \cdot \tilde{Q}_\alpha (\rho_2\|\sigma_2),
\]
so that \( \tilde{D}_\alpha \) is additive under tensor-product operators:
\[
\tilde{D}_\alpha (\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) = \tilde{D}_\alpha (\rho_1\|\sigma_1) + \tilde{D}_\alpha (\rho_2\|\sigma_2).
\]

The sandwiched Rényi relative entropy is monotone under a CPTP map \( \mathcal{N} \) for \( \alpha \in [1/2, \infty] \) [15, 5, 26]:
\[
\tilde{D}_\alpha (\rho\|\sigma) \geq \tilde{D}_\alpha (\mathcal{N}(\rho)\|\mathcal{N}(\sigma)).
\]

Lemma 1 Suppose \( \alpha > 0 \). If ρ ≤ ρ', then
\[
\tilde{Q}_\alpha (\rho\|\sigma) \leq \tilde{Q}_\alpha (\rho'\|\sigma).
\]

Proof. From the assumption that ρ ≤ ρ', we get \( \sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha} \leq \sigma^{(1-\alpha)/2\alpha} \rho' \sigma^{(1-\alpha)/2\alpha} \). The statement of the lemma then follows because \( \text{Tr} \{ f (P) \} \leq \text{Tr} \{ f (Q) \} \) for \( P \leq Q \) and \( f \) a monotone increasing function (see, e.g., [26, Lemma III.6]).

Positive partial transpose. A bipartite state \( \rho_{AB} \) is “PPT” if it has a positive partial transpose:
\[
T_B (\rho_{AB}) \geq 0,
\]
where \( T_B \) indicates the partial transpose operation on system \( B \) [29]. This is a necessary condition for a bipartite state to be separable—it is also sufficient for \( 2 \times 2 \) and \( 2 \times 3 \) systems, but otherwise only necessary [18]. Let PPT(A : B) denote the set of all such states (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). Let \( |\Phi\rangle_{AB} \) denote the maximally entangled state:
\[
|\Phi\rangle_{AB} \equiv \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle_A |i\rangle_B,
\]
and let $\Phi_{AB} = |\Phi\rangle_{AB}$. Let $F_{AB}$ denote the swap operator:

$$F_{AB} \equiv \sum_{i,j=0}^{d-1} |i\rangle_A \otimes |j\rangle_B \langle j| \otimes \langle i|.$$  (3.12)

Then from these definitions, it is clear that

$$T_B(\Phi_{AB}) = \frac{1}{d} F_{AB}. \quad (3.13)$$

A CPTP map $N_{AB\rightarrow A'B'}$ is a PPT preserving operation if the map $T_{B'} \circ N_{AB\rightarrow A'B'} \circ T_B$ is CPTP [32]. A CPTP map $N_{AB\rightarrow A'B'}$ is a separable operation (or separability preserving operation) if it can be written in the following form:

$$N_{AB\rightarrow A'B'}(\cdot) = \sum_x (C_x^{A \rightarrow A'} \otimes D_x^{B \rightarrow B'}) (\cdot) (C_x^{A \rightarrow A'} \otimes D_x^{B \rightarrow B'})^\dagger, \quad \text{(3.14)}$$

for operators $\{C_x^{A \rightarrow A'}, D_x^{B \rightarrow B'}\}$ such that

$$\sum_x (C_x^{A \rightarrow A'} \otimes D_x^{B \rightarrow B'})^\dagger (C_x^{A \rightarrow A'} \otimes D_x^{B \rightarrow B'}) = I_{AB}. \quad \text{(3.15)}$$

4  Entanglement generation codes

In this paper, we focus on entanglement generation codes, for which the goal is for the sender Alice to use a channel in order to share a state with the receiver Bob, such that this state is indistinguishable from a maximally entangled state. We focus on this task because the entanglement generation capacity of a quantum channel serves as an upper bound on its quantum capacity (this in turn is because a protocol for noiseless quantum communication can always be used to generate entanglement between sender and receiver). Thus, if one establishes an upper bound on the entanglement generation capacity, then this bound serves as an upper bound on the quantum capacity.

More formally, we now define an $(M, \varepsilon, \phi, D)$ entanglement generation code for a channel $\mathcal{N}$. Such a protocol begins with Alice preparing a bipartite state, she sends one share of the state through the channel, and then Bob decodes. That is, such a code begins with Alice preparing a state $|\phi\rangle_{RA}$. The reduced state on system $R$ has its rank equal to $M$. Alice then transmits the system $A$ through the channel, leading to the state

$$\rho_{RB} \equiv \mathcal{N}_{A \rightarrow B}(\phi_{RA}). \quad \text{(4.1)}$$

Finally, Bob performs a decoding $D_{B \rightarrow \hat{B}}$, leading to the state

$$\omega_{R\hat{B}} \equiv D_{B \rightarrow \hat{B}} (\mathcal{N}_{A \rightarrow B}(\phi_{RA})). \quad \text{(4.2)}$$

The fidelity of the code is given by

$$F \equiv \langle \Phi |_{R\hat{B}} \omega_{R\hat{B}} | \Phi \rangle_{R\hat{B}}, \quad \text{(4.3)}$$
where $|\Phi\rangle_{RB}$ is the maximally entangled state
\[
|\Phi\rangle_{RB} \equiv \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} |i\rangle_R |i\rangle_B.
\] (4.4)

An $(M, \varepsilon, \phi, D)$ code uses the state $\phi$, the decoder $D$, the channel $N_{A \rightarrow B}$, and is such that the fidelity $F \geq 1 - \varepsilon$. Note that without loss of generality, we can restrict our consideration to pure-state entanglement generation codes. For if the initial state is a mixed state $\rho_{RA}$ and the following condition holds
\[
\langle \Phi | \hat{D}_{B \rightarrow B} (N_{A \rightarrow B} (\rho_{RA})) |\Phi\rangle_{RB} \geq 1 - \varepsilon,
\] (4.5)
then there always exists at least one pure state in the spectral decomposition of $\rho_{RA}$ which meets the same fidelity constraint given above.

The main focus of this paper is on entanglement generation codes for a memoryless quantum channel $N \otimes^n$. In such a case, we define the rate of entanglement generation as
\[
Q \equiv \frac{1}{n} \log M.
\]
A rate $Q$ of entanglement generation is achievable if for all $\varepsilon > 0$ and sufficiently large $n$, there exists a $(2^{nQ}, \varepsilon, \phi, D)$ entanglement generation code. The entanglement generation capacity of a channel $N$ is defined as the supremum of all achievable rates.

5 Generalized divergence framework

We say that a function $D : S(\mathcal{H}) \times \mathcal{P}(\mathcal{H}) \rightarrow \mathbb{R}$ is a generalized divergence if it satisfies the following monotonicity inequality:
\[
D(\rho||\sigma) \geq D(N(\rho)||N(\sigma)),
\] (5.1)
where $N$ is a completely positive trace preserving map. It follows directly from monotonicity that any generalized divergence is invariant under unitaries, in the sense that
\[
D(\rho||\sigma) = D(U\rho U^\dagger ||U\sigma U^\dagger),
\] (5.2)
where $U$ is a unitary operator, and that it is invariant under tensoring with another quantum state $\tau$:
\[
D(\rho||\sigma) = D(\rho \otimes \tau || \sigma \otimes \tau).
\] (5.3)

5.1 Rains relative entropy and Rains information

We now define some information measures which play a central role in this paper. They are inspired by the Rains bound on distillable entanglement [32] and the subsequent reformulation of it in [2].

We define the generalized Rains relative entropy of a bipartite state $\rho_{RB}$ as follows:
\[
R_D (\rho_{RB}) \equiv \min_{\tau_{RB} \in \mathcal{T}(R:B)} D(\rho_{RB}||\tau_{RB}),
\] (5.4)
where $\mathcal{T}(R:B)$ is the following set of bipartite operators:
\[
\mathcal{T}(R:B) \equiv \{ \tau_{RB} : \tau_{RB} \geq 0 \land \| T_B (\tau_{RB}) \|_1 \leq 1 \}.
\] (5.5)
The set $\mathcal{T}(R:B)$ includes all PPT states because $\| T_B (\tau_{RB}) \|_1 = 1$ if $\tau_{RB} \in \text{PPT}(R:B)$. Furthermore, all operators $\tau_{RB} \in \mathcal{T}(R:B)$ are subnormalized, in the sense that $\text{Tr} \{ \tau_{RB} \} \leq 1$, because
\[
\text{Tr} \{ \tau_{RB} \} = \text{Tr} \{ T_B (\tau_{RB}) \} \leq \| T_B (\tau_{RB}) \|_1 \leq 1.
\] (5.6)
One property of $R_D(\rho_{RB})$, critical for our application here, is that it is monotone under PPT preserving operations, in the sense that:

$$R_D(\rho_{RB}) \geq R_D(\mathcal{P}_{RB}(\rho_{RB})),$$

where $\mathcal{P}_{RB}$ is a PPT-preserving operation. This is because PPT-preserving operations do not take operators $\tau_{RB}$ in $\mathcal{T}(R:B)$ outside of this set, which follows from

$$\|T_B(\mathcal{P}_{RB}(\tau_{RB}))\|_1 = \|T_B(\mathcal{P}_{RB}(T_B(\tau_{RB})))\|_1 \leq \|T_B(\tau_{RB})\|_1 \leq 1.$$ (5.8)

In the above, the first equality follows because the partial transpose $T_B$ is its own inverse, and the first inequality follows because the map $T_B \circ \mathcal{P}_{RB} \circ T_B$ is CPTP (as it 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 ones (local operations and classical communication), $R_D(\rho_{RB})$ is also monotone under LOCC operations. We then define the generalized Rains information of a quantum channel as follows:

$$R_D(N) \equiv \max_{\rho_{RA}} R_D(N_{A \rightarrow B}(\rho_{RA})).$$ (5.9)

Our focus in this paper will be on the case when the generalized divergence is the von Neumann or Umegaki relative entropy, defined in (2.4), or the sandwiched Rényi relative entropy defined in (3.5).

Another critical property of the above quantities has to do with the set $\mathcal{T}(R:B)$ of operators over which we are optimizing. That is, all operators in this set satisfy the property given in Lemma 2 of [31]. We repeat the proof of this lemma for the reader’s convenience.

**Lemma 2 (Lemma 2 of [31])** Let $\tau_{RB} \in \mathcal{T}(R:B)$. Then the overlap of $\tau_{RB}$ with a maximally entangled state of Schmidt rank $M$ is no larger than $1/M$:

$$\text{Tr}\{\Phi_{RB}\tau_{RB}\} \leq \frac{1}{M}.$$ (5.10)

The same is true for $\sigma_{RB} \in \text{PPT}(R:B)$ simply because $\text{PPT}(R:B) \subseteq \mathcal{T}(R:B)$.

**Proof.** Consider that the partial transpose is its own adjoint and inverse, so that

$$\text{Tr}\{\Phi_{RB}\tau_{RB}\} = \text{Tr}\{T_B(\Phi_{RB})T_B(\tau_{RB})\}$$

$$= \frac{1}{M}\text{Tr}\{F_{RB}T_B(\tau_{RB})\}$$

$$\leq \frac{1}{M}\|T_B(\tau_{RB})\|_1$$

$$\leq \frac{1}{M}.$$ (5.14)

The second equality follows because the partial transpose acting on the maximally entangled state $\Phi_{RB}$ is equivalent to the swap operator $F_{RB}$ normalized by the Schmidt rank of $\Phi_{RB}$. The first inequality follows from (3.2), and the second from the assumption that $\tau_{RB} \in \mathcal{T}(R:B)$. ■

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 \( g \rightarrow U_A (g) \) and \( g \rightarrow V_B (g) \) be unitary representations acting on the input and output spaces of the channel, respectively. Then a quantum channel \( \mathcal{N}_{A \rightarrow B} \) is covariant with respect to these representations if the following relation holds for all input density operators \( \rho \) and group elements \( g \in G \):

\[
\mathcal{N}_{A \rightarrow B} \left( U_A (g) \rho U_A^\dagger (g) \right) = V_B (g) \mathcal{N}_{A \rightarrow B} (\rho) V_B^\dagger (g) .
\]  

(5.15)

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 3** Let \( \rho_A \) be a density operator which can serve as an input to a covariant channel \( \mathcal{N}_{A \rightarrow B} \) (as defined above) and let \( |\phi^0\rangle_{RA} \) be a purification of it. Let \( \overline{\rho}_A \) be the group average of \( \rho_A \), i.e.,

\[
\overline{\rho}_A \equiv \frac{1}{|G|} \sum_g U_A (g) \rho U_A^\dagger (g) ,
\]  

(5.16)

and let \( |\phi^{\overline{\rho}}\rangle_{RA} \) be a purification of it. Then

\[
R_D (\mathcal{N}_{A \rightarrow B} (\phi^{\overline{\rho}}_{RA})) \geq R_D (\mathcal{N}_{A \rightarrow B} (\phi^0_{RA})) .
\]  

(5.17)

**Proof.** Given the purification \( |\phi^0\rangle_{RA} \), consider the following state

\[
|\psi\rangle_{PRA} \equiv \sum_g \frac{1}{\sqrt{|G|}} |g\rangle_P \left[ I_R \otimes U_A (g) \right] |\phi^0\rangle_{RA} .
\]  

(5.18)

Observe that \( |\psi\rangle_{PRA} \) is a purification of \( \overline{\rho}_A \) with purifying systems \( P \) and \( R \). Let \( \tau_{PAB} \) be an arbitrary operator in \( \mathcal{T} (PR : B) \). Then the following chain of inequalities holds

\[
D (\mathcal{N}_{A \rightarrow B} (|\psi\rangle_{PRA}) \| \tau_{PRB}) \\
\geq D \left( \sum_g \frac{1}{|G|} |g\rangle_P \otimes \mathcal{N}_{A \rightarrow B} (U_A (g) \phi^0_{RA} U_A^\dagger (g)) \| \sum_g p (g) |g\rangle_P \otimes \tau^0_{RB} \right) \\
= D \left( \sum_g \frac{1}{|G|} |g\rangle_P \otimes V_B (g) \mathcal{N}_{A \rightarrow B} (\phi^0_{RA}) V_B^\dagger (g) \| \sum_g p (g) |g\rangle_P \otimes \tau^0_{RB} \right) \\
= D \left( \sum_g \frac{1}{|G|} |g\rangle_P \otimes \mathcal{N}_{A \rightarrow B} (\phi^0_{RA}) \| \sum_g p (g) V_B^\dagger (g) \tau^0_{RB} V_B (g) \right) \\
\geq D \left( \mathcal{N}_{A \rightarrow B} (\phi^0_{RA}) \| \sum_g p (g) V_B^\dagger (g) \tau^0_{RB} V_B (g) \right) \\
\geq \min_{\tau_{RB} \in \mathcal{T} (RB)} D (\mathcal{N}_{A \rightarrow B} (\phi^0_{RA}) \| \tau_{RB}) \\
= R_D (\mathcal{N}_{A \rightarrow B} (\phi^0_{RA})) .
\]  

(5.19) 

(5.20) 

(5.21) 

(5.22) 

(5.23) 

(5.24)

The first inequality follows from monotonicity of the generalized divergence \( D \) 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

$$
\sum_g |g\rangle \langle g|_P \otimes V_B^\dagger (g).
$$

(5.25)

Furthermore, this unitary does not take the state out of the class $\mathcal{T}$, i.e.

$$
\sum_g p(g) |g\rangle \langle g|_P \otimes V_B^\dagger (g) \tau_{RB}^{P} V_B (g) \in \mathcal{T} (PR : B).
$$

(5.26)

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 $D$ in (5.20) are classical on $P$. The second inequality follows because the generalized divergence $D$ 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 $\tau_{PRB}$ is chosen to be an arbitrary operator in $\mathcal{T} (PR : B)$, it follows that

$$
R_D (\mathcal{N}_{A\to B} (\psi_{PRA})) \geq R_D (\mathcal{N}_{A\to B} (\phi_{PRA})).
$$

(5.27)

The conclusion then follows because all purifications are related by a unitary on the purifying system and the quantity $R_D$ is invariant under unitaries on the purifying system.

### 5.2 Relating rate and fidelity of an entanglement generation code to the generalized 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” [39]. For example, it is well known that a PPT entanglement binding channel has zero quantum capacity [19]. More generally, the bound in Lemma [2] establishes that if both the input to the channel and the reference system are replaced with an operator $\tau_{RB} \in \mathcal{T} (R : B)$, 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 $\tau_{RB}$ cannot send any quantum information reliably.

So our starting point is to consider the generalized divergence between the state $\rho_{RB}$ in (4.11) and an operator $\tau_{RB} \in \mathcal{T} (R : B)$:

$$
D (\rho_{RB} \| \tau_{RB}).
$$

(5.28)

By monotonicity under the application of the decoder $D_{B\to \hat{B}}$, the following inequality holds

$$
D (\rho_{RB} \| \tau_{RB}) \geq D (\omega_{RB} \| D_{B\to \hat{B}} (\tau_{RB})).
$$

(5.29)

Observe that $D_{B\to \hat{B}} (\tau_{RB}) \in \mathcal{T} (R : \hat{B})$ because a local operation does not take an operator out of the class $\mathcal{T}$. 

10
Next, consider the following binary test $B_{R\hat{B}\rightarrow Z}$ (a completely positive trace-preserving map), which outputs a flag indicating whether a state is maximally entangled or not:

$$B_{R\hat{B}\rightarrow Z}(\cdot) \equiv \text{Tr} \{ \Phi_{\hat{R}\hat{B}}(\cdot) \} |1\rangle \langle 1| + \text{Tr} \{ \{(I_{RB} - \Phi_{RB}) (\cdot)\} |0\rangle \langle 0| .$$  \hspace{1cm} (5.30)

Intuitively, this test is simply asking, “Is the entanglement decoded or not?” Applying monotonicity of the generalized divergence under this test, we find that the following inequality holds

$$D(\omega_{RB} \| D_{B\rightarrow \hat{B}}(\tau_{RB})) \geq D(B_{R\hat{B}\rightarrow Z}(\omega_{RB}) \| B_{R\hat{B}\rightarrow Z}(D_{B\rightarrow \hat{B}}(\tau_{RB}))).$$  \hspace{1cm} (5.31)

The next section continues this development by selecting the generalized divergence to be the sandwiched Rényi relative entropy.

### 6 Specializing to Rényi relative entropies

We now define the Rényi-like Rains information of a quantum channel as

$$\tilde{R}_{\alpha}(N) \equiv \max_{\phi_{RA}} \min_{\tau_{RB} \in \mathcal{T}(R:B)} D_{\alpha}(N_{A\rightarrow B}(\phi_{RA})\|\tau_{RB}).$$  \hspace{1cm} (6.1)

In this paper, we focus on the regime for which $\alpha > 1$. In this case, it suffices to perform the maximization in $\tilde{R}_{\alpha}(N)$ over pure bipartite states $\phi_{RA}$, due to the joint convexity of $Q_{\alpha}$ whenever $\alpha > 1$ [15, Proposition 3]. 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 theorem.

Since it should be clear from the context, we can simply refer to the quantity $\tilde{R}_{\alpha}(N)$ as the Rains information of a quantum channel. The critical property of this quantity that we require is that it converges to $R(N)$ in the limit as $\alpha$ approaches one from above. This is shown in the following Lemma, whose proof is provided in Appendix A.

**Lemma 4** For any quantum channel $N$ and $\alpha > \beta > 1$, we have

$$\tilde{R}_{\alpha}(N) \geq \tilde{R}_{\beta}(N) \geq R(N) \quad \text{and} \quad \lim_{\alpha \rightarrow 1^+} \tilde{R}_{\alpha}(N) = R(N).$$  \hspace{1cm} (6.2)

The following proposition gives a “one-shot” bound on the fidelity of any entanglement generation code:

**Proposition 5** The fidelity of any $(M, \varepsilon, \phi, D)$ entanglement generation code obeys the following bound for all $\alpha > 1$:

$$F \leq 2^{-\frac{M}{2}(\log M - \tilde{R}_{\alpha}(N))}.$$  \hspace{1cm} (6.3)

**Proof.** We simply set the generalized divergence to be the sandwiched Rényi relative entropy, defined in (6.5). In this case, defining

$$p \equiv \text{Tr} \{ \Phi_{RB} D_{B\rightarrow \hat{B}}(\tau_{RB}) \},$$  \hspace{1cm} (6.4)
the bound from (5.31) becomes
\[ \tilde{D}_\alpha (\rho_{RB} \parallel \tau_{RB}) \geq \tilde{D}_\alpha \left( B_{RB} \parallel \left( \frac{1}{M} \right)^{1/\alpha} \right) \]
\[ = \frac{1}{\alpha - 1} \log \left[ F^{\alpha (1 - F)} \right] \] (6.6)
\[ \geq \frac{1}{\alpha - 1} \log \left[ F^{\alpha (1/\alpha)} \right] \] (6.7)
\[ \geq \frac{1}{\alpha - 1} \log \left[ F^{\alpha (1/\alpha)} \right] \] (6.8)
\[ = \frac{\alpha}{\alpha - 1} \log F + \log M. \] (6.9)

The second inequality follows by discarding the second term \((1 - F)^\alpha [\text{Tr} \{ \tau_{RB} \} - p]^{1-\alpha}\) (recall that we are considering \(\alpha > 1\)). The third inequality follows from (6.4) and Lemma 2. Since the bound holds for all \(\tau_{RB} \in \mathcal{T} (R : B)\), we can take a minimization over all \(\tau_{RB} \in \mathcal{T} (R : B)\) to establish the following bound:
\[ \min_{\tau_{RB} \in \mathcal{T} (R : B)} \tilde{D}_\alpha (\rho_{RB} \parallel \tau_{RB}) \geq \frac{\alpha}{\alpha - 1} \log F + \log M. \] (6.10)

We can finally remove the dependence on any particular code by optimizing over all inputs to the channel:
\[ \tilde{R}_\alpha (N) = \max_{\phi_{RA} \in \mathcal{T} (R : B)} \min_{\tau_{RB} \in \mathcal{T} (R : B)} \tilde{D}_\alpha (N_{A \rightarrow B} (\phi_{RA}) \parallel \tau_{RB}) \geq \frac{\alpha}{\alpha - 1} \log F + \log M. \] (6.11)
This bound is then equivalent to (6.3).

7 Weak subadditivity of the \(\alpha\)-Rains information for memoryless channels

In this section, we prove the following theorem, which is critical for concluding that the Rains information of a channel is a strong converse bound for quantum communication.

**Theorem 6** For all \(\alpha > 1\), the Rains information \(\tilde{R}_\alpha (N)\) of a quantum channel \(N\) obeys a weak subadditivity property, in the sense that
\[ \tilde{R}_\alpha (N^{\otimes n}) \leq n \tilde{R}_\alpha (N) + \frac{\alpha |A|^2}{\alpha - 1} \log n, \] (7.1)
where \(|A|\) is the dimension of the input to the channel \(N\).

**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
\[ \forall \pi \in S_n : W_{B^n}^\pi (N^{\otimes n} (\rho_{A^n})) = N^{\otimes n} (W_{A^n}^\pi \rho_{A^n} (W_{A^n}^\pi)^\dagger), \] (7.2)
where \(W_{A^n}^\pi\) and \(W_{B^n}^\pi\) are unitary representations of the permutation \(\pi\), 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 \(\rho_{A^n}\), we can apply Proposition 3 to conclude that
\[ \tilde{R}_\alpha (N_{A^n \rightarrow B^n} (\phi^\rho_{RA^n})) \leq \tilde{R}_\alpha (N_{A^n \rightarrow B^n} (\phi^\rho_{RA^n})), \] (7.3)
where \( \phi_{\mathcal{R}A^n} \) is a purification of the permutation invariant state \( \rho_{A^n} \), defined as

\[
\rho_{A^n} \equiv \frac{1}{n!} \sum_{\pi \in S_n} W_{A^n}^\pi \rho_{A^n} (W_{A^n}^\pi)^\dagger.
\] (7.4)

Now, this purification \( \phi_{\mathcal{R}A^n} \) is related by a unitary on the reference system \( \mathcal{R} \) to a state

\[
|\psi\rangle_{\hat{A}^n A^n} \in \text{Sym}((\hat{A} \otimes A)^{\otimes n}),
\] (7.5)

where \( \hat{A} \simeq A \) [33, Lemma 4.3.1]. So it follows that

\[
\bar{R}_\alpha (\mathcal{N}_{A^n \to B^n} (\phi_{\mathcal{R}A^n}^\rho)) \leq \bar{R}_\alpha (\mathcal{N}_{A^n \to B^n} (\phi_{\mathcal{R}A^n}^\rho)) = \bar{R}_\alpha (\mathcal{N}_{A^n \to B^n} (\psi_{\hat{A}^n A^n})).
\] (7.6)

For such a state in \( \text{Sym}((\hat{A} \otimes A)^{\otimes n}) \), we can apply the postselection lemma from [11] (see also [25, Lemma 12]), to conclude the following operator inequality:

\[
\psi_{\hat{A}^n A^n} \leq n! |A|^2 \omega^{(n)},
\] (7.7)

where \( \omega^{(n)} \) is the universal de Finetti state:

\[
\omega^{(n)} \equiv \int d\mu (\varphi) \varphi^{\otimes n}_{\hat{A}A},
\] (7.8)

with \( \mu (\varphi) \) denoting the uniform probability measure on the unit sphere \( S_1 (\mathcal{H}_{\hat{A}} \otimes \mathcal{H}_A) \), consisting of pure bipartite states \( \varphi_{\hat{A}A} \). Employing the definition of \( \bar{R}_\alpha \), we find that

\[
\bar{R}_\alpha (\mathcal{N}_{A^n \to B^n} (\psi_{\hat{A}^n A^n})) = \min_{\tau_{\hat{A}^n B^n} \in \mathcal{T}(\hat{A}^n : B^n)} D_\alpha (\mathcal{N}_{A^n \to B^n} (\psi_{\hat{A}^n A^n}) \| \tau_{\hat{A}^n B^n}) \leq \frac{1}{\alpha - 1} \log \bar{Q}_\alpha (\mathcal{N}_{A^n \to B^n} (\psi_{\hat{A}^n A^n}) \| \tau_{\hat{A}^n B^n}) \leq \frac{1}{\alpha - 1} \log \min_{\tau_{\hat{A}^n B^n} \in \mathcal{T}(\hat{A}^n : B^n)} \bar{Q}_\alpha (\mathcal{N}_{A^n \to B^n} (\psi_{\hat{A}^n A^n}) \| \tau_{\hat{A}^n B^n}) \leq \frac{1}{\alpha - 1} \log \min_{\tau_{\hat{A}^n B^n} \in \mathcal{T}(\hat{A}^n : B^n)} \bar{Q}_\alpha (\mathcal{N}_{A^n \to B^n} (n! |A|^2 \omega^{(n)}) \| \tau_{\hat{A}^n B^n}) \leq \frac{\alpha |A|^2}{\alpha - 1} \log n + \frac{1}{\alpha - 1} \log \min_{\tau_{\hat{A}^n B^n} \in \mathcal{T}(\hat{A}^n : B^n)} \bar{Q}_\alpha (\mathcal{N}_{A^n \to B^n} (\omega^{(n)}) \| \tau_{\hat{A}^n B^n}) \leq \frac{\alpha |A|^2}{\alpha - 1} \log n + \frac{1}{\alpha - 1} \log \min_{\tau_{\hat{A}^n B^n} \in \mathcal{T}(\hat{A}^n : B^n)} \bar{Q}_\alpha (\mathcal{N}_{A^n \to B^n} (\int d\mu (\varphi) \varphi^{\otimes n}_{\hat{A}A}) \| \tau_{\hat{A}^n B^n})
\] (7.9) (7.10) (7.11) (7.12) (7.13) (7.14)

The first two equalities follow from definitions, and the third follows from our assumption that \( \alpha > 1 \). The inequality is a consequence of Lemma [11] and our assumption that \( \alpha > 1 \). The fourth equality follows easily from the definition of \( \bar{Q}_\alpha \). The last equality is from (7.8). We now focus on
bounding the quantity inside the second logarithm on the last line above:

\[
\min_{\tau_{\hat{A}^nB^n} \in T(\hat{A}^n:B^n)} \tilde{Q}_\alpha \left( N_{A^n \rightarrow B^n} \left( \int d\mu (\varphi) \varphi^{\otimes n}_{\hat{A}A} \right) \| \tau_{\hat{A}^nB^n} \right)
\leq \tilde{Q}_\alpha \left( N_{A^n \rightarrow B^n} \left( \int d\mu (\varphi) \varphi^{\otimes n}_{\hat{A}A} \right) \left\| \tau_{\hat{A}B}^{\varphi} \right\|^n \right)
\leq \int d\mu (\varphi) \tilde{Q}_\alpha \left( N_{A^n \rightarrow B^n} \left( \varphi^{\otimes n}_{\hat{A}A} \right) \left\| \tau_{\hat{A}B}^{\varphi} \right\|^n \right)
= \int d\mu (\varphi) \left[ \tilde{Q}_\alpha \left( N_{A \rightarrow B} (\varphi^{\hat{A}A}) \left\| \tau_{\hat{A}B}^{\varphi} \right\|^n \right) \right] (7.15)
\]

\[
\text{The first inequality follows because we have a minimization over all states } \tau_{\hat{A}^nB^n} \in T(\hat{A}^n:B^n) \text{ and we choose the particular state } \int d\mu (\varphi) (\tau_{\hat{A}B}^{\varphi})^{\otimes n} \text{ where each } \tau_{\hat{A}B}^{\varphi} \text{ is in } T(\hat{A} : B) \text{ and the notation indicates that each } \tau_{\hat{A}B}^{\varphi} \text{ is “tied” to a state } \varphi^{\hat{A}A} \text{ in } \int d\mu (\varphi) \varphi^{\otimes n}_{\hat{A}A}. \text{ It follows that } \int d\mu (\varphi) (\tau_{\hat{A}B}^{\varphi})^{\otimes n} \in T(\hat{A}^n : B^n) \text{ because}
\]

\[
\left\| T_{B^n} \left( \int d\mu (\varphi) (\tau_{\hat{A}B}^{\varphi})^{\otimes n} \right) \right\|_1 \leq \int d\mu (\varphi) \left\| T_{B^n} \left( (\tau_{\hat{A}B}^{\varphi})^{\otimes n} \right) \right\|_1 (7.18)
\]

\[
\leq \int d\mu (\varphi) \left\| T_{B^n} \left( (\tau_{\hat{A}B}^{\varphi})^{\otimes n} \right) \right\|_1 \leq 1, (7.19)
\]

\[
\text{with the first inequality from the convexity of } \| \cdot \|_1 \text{ and the last inequality from the fact that } \tau_{\hat{A}B}^{\varphi} \in T(\hat{A} : B). \text{ The inequality in (7.16) follows from joint convexity of } \tilde{Q}_\alpha \text{ for } \alpha > 1 \text{ [15 Proposition 3]. Now, since each operator } \tau_{\hat{A}B}^{\varphi} \in T(\hat{A} : B) \text{ can be chosen arbitrarily, it follows from the development in (7.15)-(7.17) that}
\]

\[
\min_{\tau_{\hat{A}^nB^n} \in T(\hat{A}^n:B^n)} \tilde{Q}_\alpha \left( N_{A^n \rightarrow B^n} \left( \int d\mu (\varphi) \varphi^{\otimes n}_{\hat{A}A} \right) \left\| \tau_{\hat{A}^nB^n} \right\| \right)
\leq \int d\mu (\varphi) \left[ \min_{\tau_{\hat{A}B} \in T(\hat{A} : B)} \tilde{Q}_\alpha \left( N_{A \rightarrow B} (\varphi^{\hat{A}A}) \left\| \tau_{\hat{A}B} \right\|^n \right) \right] (7.21)
\]

\[
\leq \max_{\varphi^{\hat{A}A}} \left[ \min_{\tau_{\hat{A}B} \in T(\hat{A} : B)} \tilde{Q}_\alpha \left( N_{A \rightarrow B} (\varphi^{\hat{A}A}) \left\| \tau_{\hat{A}B} \right\|^n \right) \right]^n (7.22)
\]

\[
= \max_{\tau_{\hat{A}B} \in T(\hat{A} : B)} \left[ \min_{\varphi^{\hat{A}A}} \tilde{Q}_\alpha \left( N_{A \rightarrow B} (\varphi^{\hat{A}A}) \left\| \tau_{\hat{A}B} \right\|^n \right) \right]^n (7.23)
\]
Combining the development in (7.3)-(7.6), (7.9)-(7.14), and (7.15)-(7.23), we finally recover the bound
\[
\tilde{R}_\alpha (\mathcal{N}_{A^n \rightarrow B^n} (\phi_{RA^n})) \leq \frac{\alpha |A|^2}{\alpha - 1} \log n + \frac{1}{\alpha - 1} \log \left[ \max_{\varphi_{AA}} \min_{\tau_{AB} \in T(A:B)} \tilde{Q}_\alpha (\mathcal{N}_{A \rightarrow B} (\varphi_{A}) \parallel \tau_{AB}) \right]^n
\]
(7.24)
and
\[
\tilde{R}_\alpha (\mathcal{N}_{A^n} (\varphi^\rho_{RA^n})) \leq \frac{\alpha |A|^2}{\alpha - 1} \log n + n \max_{\varphi_{AA}} \min_{\tau_{AB} \in T(A:B)} \frac{1}{\alpha - 1} \log \tilde{Q}_\alpha (\mathcal{N}_{A \rightarrow B} (\varphi_{A}) \parallel \tau_{AB})
\]
(7.25)
which we can write as
\[
\tilde{R}_\alpha (\mathcal{N}_{A^n} (\varphi^\rho_{RA^n})) \leq \frac{\alpha |A|^2}{\alpha - 1} \log n + n \tilde{R}_\alpha (\mathcal{N}).
\]
(7.26)
Since the bound holds for any choice of initial state \(\phi_{RA^n}\), we recover the inequality in (7.1).

**Corollary 7** For all \(\alpha > 1\) and an \(n\) for which \(n = l \cdot k\), the Rains information \(\tilde{R}_\alpha (\mathcal{N})\) of a quantum channel \(\mathcal{N}\) obeys the following weak subadditivity property, in the sense that
\[
\tilde{R}_\alpha (\mathcal{N} \otimes n) \leq k \tilde{R}_\alpha (\mathcal{N} \otimes l) + \alpha \frac{|A|^2}{\alpha - 1} \log k,
\]
where \(|A|^l\) is the dimension of the input to the channel \(\mathcal{N} \otimes l\).

**Proof.** We just apply the bound in Theorem 6 to the superchannel \(\mathcal{N} \otimes l\), i.e.,
\[
\tilde{R}_\alpha (\mathcal{N} \otimes n) = \tilde{R}_\alpha ((\mathcal{N} \otimes l)^{\otimes k}) \leq k \tilde{R}_\alpha (\mathcal{N} \otimes l) + \alpha \frac{|A|^2}{\alpha - 1} \log k.
\]
(7.28)

Another corollary 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 [38, 40] on a firm foundation. We provide its proof in Appendix B.

**Corollary 8** The Rains information of a quantum channel is weakly subadditive, in the sense that
\[
\limsup_{n \to \infty} \frac{1}{n} R(\mathcal{N}^\otimes n) \leq R(\mathcal{N})
\]
(7.29)

8 The Rains information of a channel is a strong converse bound for quantum communication

We are ready to establish our main result.

**Theorem 9 (Strong converse rate)** Suppose that the rate \(Q\) of an entanglement generation code exceeds the Rains information of a channel:
\[
Q > R(\mathcal{N})
\]
(8.1)
Then the fidelity of quantum communication decays exponentially fast to zero. In fact, this conclusion holds even if
\[
Q > \inf_{l \geq 1} \frac{1}{l} R(\mathcal{N} \otimes l).
\]
(8.2)
Proof. We only prove the latter, strictly stronger statement. Given the condition in (8.2), we can conclude that there exists a fixed $l$ for which

$$Q > \frac{1}{l} R(\mathcal{N}^{\otimes l}).$$  

(8.3)

We use the weak subadditivity result from Corollary 7 and then use the continuity of the $\alpha$-Rains information. Consider any $n$ for which $n = l \cdot k$. First, for any $\alpha > 1$, we have the following bound on the fidelity of any entanglement generation code by combining the bound in (6.3) with Theorem 6:

$$F \leq 2^{-\left(\frac{\alpha - 1}{\alpha}\right)\left(\log M - \bar{R}_\alpha(\mathcal{N}^{\otimes n})\right)},$$

(8.4)

$$= 2^{-\left(\frac{\alpha - 1}{\alpha}\right)\left(nQ - \bar{R}_\alpha(\mathcal{N}^{\otimes n})\right)}$$

(8.5)

$$\leq 2^{-\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{n}{\alpha} - \frac{n}{\alpha} \cdot A_{\bar{R}_\alpha}(\mathcal{N}^{\otimes n}) - \frac{n}{\alpha} \log k\right)},$$

(8.6)

$$= k^{|A|^2} 2^{-n\left(\frac{\alpha - 1}{\alpha}\right)\left(Q - \frac{1}{l} \bar{R}_\alpha(\mathcal{N}^{\otimes l})\right)}.$$  

(8.7)

By continuity of $\bar{R}_\alpha(\mathcal{N}^{\otimes l})$ as $\alpha \to 1^+$ and the condition (8.3), we find that there exists an $\alpha > 1$ such that $lQ > \bar{R}_\alpha(\mathcal{N}^{\otimes l}) > R(\mathcal{N}^{\otimes l})$. Thus, for this $\alpha$, we find

$$\left(\frac{\alpha - 1}{\alpha}\right) \left(Q - \frac{1}{l} \bar{R}_\alpha(\mathcal{N}^{\otimes l})\right) > 0.$$  

(8.8)

We can then conclude the statement of the theorem by taking $k$ (and thus $n$) to be large enough.

We say more about generalized dephasing channels.

9 Strong converse for the quantum capacity of generalized dephasing channels

Theorem 9 establishes the Rains information of a quantum channel as a strong converse rate for quantum communication. In this section, we show that the Rains information of a generalized dephasing channel is equal to the coherent information of this channel, thereby establishing that the quantum capacity of this class of channels obeys the strong converse property.

A generalized dephasing channel has an isometric extension of the following form:

$$U_{N\rightarrow BE}^{\mathcal{N}_H} = \sum_{x=0}^{d-1} |x\rangle_B \langle x|_A \otimes |\psi_x\rangle_E,$$  

(9.1)

where the states $|\psi_x\rangle$ are arbitrary (not necessarily orthonormal). So the resulting CPTP map is as follows:

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

(9.2)

say more about generalized dephasing channels.

These channels are also known in the literature as “Hadamard diagonal” channels [21] and “Schur multiplier” channels [25], as well as “generalized dephasing” channels [14] [53] [20] [10].
Proposition 10 Let $N_{H}$ be a generalized dephasing channel. Then

$$I_c(N_{H}) = R(N_{H}).$$

(9.3)

Proof. This theorem follows by establishing (9.3) and applying Theorem 9. To begin with, note that the following inequality holds for any quantum channel $N$:

$$I_c(N) \leq R(N).$$

(9.4)

This result follows from the achievability of the coherent information for quantum communication [24, 37, 13] and the strong converse bound in Theorem 9.

We now establish that the other inequality holds for a generalized dephasing channel $N_{H}$:

$$I_c(N_{H}) \geq R(N_{H}).$$

(9.5)

Consider that any generalized dephasing channel $N_{H}$ obeys the following covariance property:

$$N_{H} \left(Z_A(z) \rho Z_A^\dagger(z)\right) = Z_B(z) N_{H}(\rho) Z_B^\dagger(z),$$

(9.6)

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

$$Z_A(z) |x\rangle_A = \exp \left\{2\pi i x z / d\right\} |x\rangle_A,$$

$$Z_B(z) |x\rangle_B = \exp \left\{2\pi i x z / d\right\} |x\rangle_B.$$ 

(9.7)

(9.8)

(The covariance in (9.6) in fact holds for any operators of the form $\sum_{x=0}^{d-1} \exp \left\{i \varphi_x\right\} |x\rangle \langle x|_A$ and $\sum_{x=0}^{d-1} \exp \left\{i \varphi_x\right\} |x\rangle \langle x|_B$ with $\varphi_x \in \mathbb{R}$, but it suffices to consider only the operators in (9.7)-(9.8) for our proof here.) Furthermore, a uniform mixing of these operators is equivalent to a “completely dephasing” channel:

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

(9.9)

with the same obviously true for the operators $\{Z_B(z)\}_{z=0, \ldots, d-1}$. Then we can apply Proposition 3 to conclude that the Rains information of a generalized dephasing channel is maximized by a state with a Schmidt decomposition of the following form:

$$|\varphi^p\rangle_{RA} \equiv \sum_x \sqrt{p_X(x)} |x\rangle_R |x\rangle_A,$$

(9.10)

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,

$$R(N_{H}) = \max_{\varphi^p_{RA}} \min_{\tau_{RB} \in T(R:B)} D \left((\text{id}_R \otimes N_{H}) \left(\varphi_{RA}^p\right) \| \tau_{RB}\right).$$

(9.11)

Let $P$ be the following projection and $\Delta_P$ a CPTP map constructed from it:

$$P \equiv \sum_x |x\rangle \langle x|_R \otimes |x\rangle \langle x|_B,$$

(9.12)

$$\Delta_P (\cdot) \equiv P (\cdot) P + (I - P) (\cdot) (I - P).$$

(9.13)
Then the following chain of inequalities holds

\[
I_c (N_H) = \max_{\varphi_{RA}} \min_{\sigma_B} D \left( (\text{id}_R \otimes N_H) (\varphi_{RA}) \| I_R \otimes \sigma_B \right) \quad (9.14) \\
\geq \max_{\varphi_{RA}} \min_{\sigma_B} D \left( (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \| I_R \otimes \sigma_B \right) \quad (9.15) \\
\geq \max_{\varphi_{RA}} \min_{\sigma_B} D \left( \Delta_P \left[ (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \right] \| \Delta_P \left( I_R \otimes \sigma_B \right) \right) \quad (9.16) \\
= \max_{\varphi_{RA}} \min_{\sigma_B} D \left( P \left[ (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \right] P \| P \left( I_R \otimes \sigma_B \right) P \right) \quad (9.17) \\
= \max_{\varphi_{RA}} \min_{\sigma_B} D \left( (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \| \sum_x q(x) |x\rangle_R \langle x|_R \otimes |x\rangle_B \langle x|_B \right) \quad (9.18) \\
\geq \max_{\varphi_{RA}} \min_{\tau_{RB} \in \mathcal{T}(R:B)} D \left( (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \| \tau_{RB} \right) \quad (9.19) \\
= R(N_H). \quad (9.20)
\]

The first equality is by definition. The first inequality follows by restricting the maximization to be over pure bipartite vectors of the form in (9.10). The second inequality follows from monotonicity of the relative entropy under the CPTP map \(\Delta_P\). The second equality follows because

\[
D \left( \Delta_P \left[ (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \right] \| \Delta_P \left( I_R \otimes \sigma_B \right) \right) = D \left( P \left[ (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \right] P \| P \left( I_R \otimes \sigma_B \right) P \right) \\
+ D \left( (I - P) \left[ (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \right] (I - P) \| (I - P) \left( I_R \otimes \sigma_B \right) (I - P) \right) \quad (9.21)
\]

since \(P \perp I - P\), and because the state \((\text{id}_R \otimes N_H) (\varphi_{RA}^P)\) has no support in the subspace onto which \(I - P\) projects, so that

\[
D \left( (I - P) \left[ (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \right] (I - P) \| (I - P) \left( I_R \otimes \sigma_B \right) (I - P) \right) = 0. \quad (9.22)
\]

The third equality follows because

\[
P \left[ (\text{id}_R \otimes N_H) (\varphi_{RA}^P) \right] P = (\text{id}_R \otimes N_H) (\varphi_{RA}^P), \quad (9.23)
\]

and

\[
P \left( I_R \otimes \sigma_B \right) P = \sum_x q(x) |x\rangle_R \langle x|_R \otimes |x\rangle_B \langle x|_B, \quad (9.24)
\]

for some distribution \(q(x) = \langle x|_B \sigma_B |x\rangle_B\). The final inequality follows because the state

\[
\sum_x q(x) |x\rangle_R \langle x|_R \otimes |x\rangle_B \langle x|_B \in \mathcal{T}(R:B), \quad (9.25)
\]

and the final equality follows from (9.11).

**Conclusion 11** As a result of Theorem 3, Proposition 10, and the fact that the quantum capacity of \(N_H\) is equal to \(I_c(N_H)\) [11], if the quantum communication rate of any entanglement generation code for a generalized dephasing channel \(N_H\) is strictly larger than its quantum capacity, then the fidelity of communication decays exponentially fast to zero with an increasing number of channel uses.
10 The Rains bound is a strong converse rate for entanglement distillation

In this final section, we briefly illustrate how our framework provides an answer to the open question posed by Rains [32], regarding a strong converse bound for entanglement distillation. To begin with, let us recall that an i.i.d. entanglement distillation protocol begins with \( n \) copies of an entangled state \( \rho_{AB} \) shared between Alice and Bob. They then engage in an LOCC protocol, described by an LOCC CPTP map \( P_{A^nB^n \rightarrow \hat{A}\hat{B}} \), in an attempt to distill a maximally entangled state \( \Phi_{\hat{A}\hat{B}} \) of dimension \( 2^nQ \). The protocol is characterized by the fidelity:

\[
F = \langle \Phi|_{\hat{A}\hat{B}} P_{A^nB^n \rightarrow \hat{A}\hat{B}} (\rho_{AB}^{\otimes n}) |\Phi\rangle_{\hat{A}\hat{B}}. \tag{10.1}
\]

**Theorem 12** Suppose that the rate \( Q \) of an entanglement distillation protocol exceeds the Rains relative entropy \( R(\rho_{AB}) \):

\[
Q > R(\rho_{AB}). \tag{10.2}
\]

Then the fidelity of entanglement distillation decays exponentially fast to zero with the number \( n \) of copies of the state \( \rho_{AB} \).

**Proof.** This is just an application of ideas that we used throughout the paper. Recalling that

\[
\tilde{R}_\alpha (\rho_{AB}) = \min_{\tau_{AB} \in T(A:B)} \tilde{D}_\alpha (\rho_{AB} \| \tau_{AB}), \tag{10.3}
\]

\[
\tilde{R}_\alpha (\rho_{AB}^{\otimes n}) = \min_{\tau_{A^nB^n} \in T(A^n:B^n)} \tilde{D}_\alpha (\rho_{AB}^{\otimes n} \| \tau_{A^nB^n}), \tag{10.4}
\]

it follows for any \( \alpha > 1 \) that

\[
n\tilde{R}_\alpha (\rho_{AB}) \geq \tilde{R}_\alpha (\rho_{AB}^{\otimes n}) \geq \min_{\tau_{A^nB^n} \in T(A^n:B^n)} \tilde{D}_\alpha (\mathcal{B}_{\hat{A}\hat{B} \rightarrow Z} (P_{A^nB^n \rightarrow \hat{A}\hat{B}} (\rho_{AB}^{\otimes n})) \| \mathcal{B}_{\hat{A}\hat{B} \rightarrow Z} (P_{A^nB^n \rightarrow \hat{A}\hat{B}} (\tau_{A^nB^n})))) \geq \frac{\alpha}{\alpha - 1} \log F + nQ. \tag{10.5}
\]

The second inequality follows from monotonicity of \( \tilde{R}_\alpha \) under any LOCC map \( P_{A^nB^n \rightarrow \hat{A}\hat{B}} \) and the test \( \mathcal{B}_{\hat{A}\hat{B} \rightarrow Z} \) defined in (5.30). The last inequality follows from a line of reasoning similar to that in (6.5)-(6.9). Turning this around gives

\[
F \leq 2^{-n(\frac{\alpha - 1}{\alpha})(Q - \tilde{R}_\alpha (\rho_{AB}))}. \tag{10.6}
\]

Applying a line of reasoning similar to that in the proof of Theorem 9 then yields the result. ■

In fact, we obtain the following theorem by following the proof strategy of Theorem 9:

**Theorem 13** Suppose that the rate \( Q \) of an entanglement distillation protocol exceeds the regularized Rains relative entropy:

\[
Q > \lim_{l \to \infty} \frac{1}{l} R(\rho_{AB}^{\otimes l}). \tag{10.9}
\]

Then the fidelity of entanglement distillation decays exponentially fast to zero with the number \( n \) of copies of the state \( \rho_{AB} \).
11 Conclusion

This paper has established that the Rains information of a quantum channel is a strong converse rate for quantum communication. Similarly, we have shown that the Rains relative entropy of a bipartite state is a strong converse rate for entanglement distillation. The main application of the first result is to settle the strong converse question for the quantum capacity of all generalized dephasing channels.

Going forward from here, there are several questions to consider. First, is it possible to show weak subadditivity of a Rényi coherent information quantity for some class of channels, addressing the original question posed in [35]? To this end, the developments in [43] might be helpful. Next, is it possible to show that the Rains information of a quantum channel represents a strong converse rate for $Q_2$ (the quantum capacity assisted by forward and backward classical communication)? Similarly, can one show that the squashed entanglement of a channel [41] is a strong converse rate for $Q_2$? Recent work has proved that the squashed entanglement of a quantum channel is a weak converse upper bound on $Q_2$, and it could be that the quantities defined in [9] would be helpful for settling this question. 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 settle the strong converse question for them. For example, can we prove a strong converse theorem for the quantum capacity of general Hadamard channels? 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 [10] [47]? Can we establish second-order characterizations for this class of channels, in the sense of [45] [46] [12]?

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. 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.

A Proof of Lemma 4

Proof. The first statement follows because the underlying sandwiched relative entropy, $\tilde{D}_\alpha(\rho\|\sigma)$, is monotonically increasing in $\alpha$ [27 Theorem 7] for all $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, i.e.

$$\tilde{D}_\alpha(\rho\|\sigma) \geq \tilde{D}_\beta(\rho\|\sigma) \quad (A.1)$$
for all \( \alpha \geq \beta \geq 0 \). This already establishes that the limit exists and satisfies

\[
\lim_{\alpha \to 1^+} \tilde{R}_\alpha(\mathcal{N}) \geq R(\mathcal{N}). \tag{A.2}
\]

We would like to show the opposite inequality. Consider the following bound (from [42, Lemma 6.3], see also [50, Eq. (19)] and [44, Lemma 8])

\[
\tilde{D}_{1+\delta}(\rho||\sigma) \leq D(\rho||\sigma) + 4\delta \left( \log \nu(\rho,\sigma) \right)^2, \tag{A.3}
\]

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

\[
\nu(\rho,\sigma) \equiv \text{Tr}(\rho^{3/2}\sigma^{-1/2}) + \text{Tr}(\rho^{1/2}\sigma^{1/2}) + 1. \tag{A.4}
\]

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

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

\[
\min_{\sigma_{RB} \in T(\rho_{RB}||\sigma_{RB})} \tilde{D}_{1+\delta}(\rho_{RB}||\sigma_{RB}) \leq \min_{\tau_{RB} \in T(\rho_{RB}||\sigma_{RB})} \tilde{D}_{1+\delta}(\rho_{RB}||\sigma_{RB} + \delta\tau_{RB}) \leq \min_{\tau_{RB} \in T(\rho_{RB}||\sigma_{RB})} D(\rho_{RB}||\sigma_{RB} + \delta\tau_{RB}) + 4\delta \left( \log \left( 2 + \frac{\sqrt{|A||B|}}{\sqrt{\delta}} \right) \right)^2 \tag{A.5}
\]

The first inequality follows by picking \( \sigma_{RB} \) to be of the form \( \sigma_{RB} = (1 - \delta)\tau_{RB} + \delta\pi_{RB} \) where \( \pi_{RB} \) is the fully mixed state on \( RB \). Also, note that \( (1 - \delta)\tau_{RB} + \delta\pi_{RB} \in T(\rho_{RB}||\sigma_{RB}) \). 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 \( D(\rho||\sigma) \leq D(\rho||\sigma') \) whenever \( \sigma' \leq \sigma \) to verify the last inequality.

Since, crucially, the upper bound in (A.5) is uniform in \( \rho_{RB} \), we can immediately conclude that

\[
\tilde{R}_{1+\delta}(\mathcal{N}) \leq R(\mathcal{N}) + \log \frac{1}{1 - \delta} + 4\delta \left( \log \left( 2 + \frac{\sqrt{|A||B|}}{\sqrt{\delta}} \right) \right)^2 \tag{A.6}
\]

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

B Proof of Corollary 8

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

\[
R(\mathcal{N} \otimes n) \leq \tilde{R}_{1+\delta}(\mathcal{N} \otimes n) \leq \frac{1 + \delta}{\delta} |A|^2 \log n + n\tilde{R}_{1+\delta}(\mathcal{N}). \tag{B.1}
\]

\[
\text{Proof.} \quad \text{We can focus only on operators } \tau_{RB} \text{ such that } \text{supp}(\rho_{RB}) \subseteq \text{supp}(\tau_{RB}), \text{ where } \rho_{RB} = \mathcal{N}_{A \rightarrow B}(\phi_{RA}). \text{ This is because the quantity of interest contains a minimization over all } \tau_{RB}. \text{ So consider that for any sufficiently small } \delta > 0
\]

\[
R(\mathcal{N} \otimes n) \leq \tilde{R}_{1+\delta}(\mathcal{N} \otimes n) \leq \frac{1 + \delta}{\delta} |A|^2 \log n + n\tilde{R}_{1+\delta}(\mathcal{N}). \tag{B.2}
\]
The first inequality is from the monotonicity of the sandwiched Rényi relative entropy in the Rényi parameter [27, Theorem 7]. The second inequality follows from Theorem 6.

This invites an application of (A.9), which gives

$$\frac{1}{n}R\left(\mathcal{N}^\otimes n\right) \leq R(\mathcal{N}) + \frac{1 + \delta}{\delta} |A|^2 \log \frac{n}{A} + \log \frac{1}{1 - \delta} + 4\delta \left(\log \left(2 + \frac{\sqrt{|A||B|}}{\sqrt{\delta}}\right)\right)^2$$ (B.3)

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

References

[1] Suguru Arimoto. On the converse to the coding theorem for discrete memoryless channels. IEEE Transactions on Information Theory, 19:357–359, May 1973.

[2] Koenraad Audenaert, Bart De Moor, Karl Gerd H. Vollbrecht, and Reinhard F. Werner. Asymptotic relative entropy of entanglement for orthogonally invariant states. Physical Review A, 66(3):032310, September 2002. arXiv:quant-ph/0204143.

[3] Howard Barnum, Emanuel Knill, and Michael A. Nielsen. On quantum fidelities and channel capacities. IEEE Transactions on Information Theory, 46(4):1317–1329, July 2000. arXiv:quant-ph/9809010.

[4] Howard Barnum, M. A. Nielsen, and Benjamin Schumacher. Information transmission through a noisy quantum channel. Physical Review A, 57(6):4153–4175, June 1998. arXiv:quant-ph/9702049.

[5] Salman Beigi. Sandwiched Rényi divergence satisfies data processing inequality. Journal of Mathematical Physics, 54(12):122202, December 2013. arXiv:1306.5920.

[6] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. Quantum reverse Shannon theorem. IEEE Transactions on Information Theory, 60(5):2926–2959, May 2014. arXiv:0912.5537.

[7] Mario Berta, Fernando G. S. L. Brandao, Matthias Christandl, and Stephanie Wehner. Entanglement cost of quantum channels. IEEE Transactions on Information Theory, 59(10):6779–6795, October 2013. arXiv:1108.5357.

[8] Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse Shannon theorem based on one-shot information theory. Communications in Mathematical Physics, 306(3):579–615, August 2011. arXiv:0912.3805.

[9] Mario Berta, Kaushik Seshadreesan, and Mark M. Wilde. Rényi generalizations of the conditional quantum mutual information. March 2014. arXiv:1403.6102.

[10] Kamil Brádler, Patrick Hayden, Dave Touchette, and Mark M. Wilde. Trade-off capacities of the quantum Hadamard channels. Physical Review A, 81(6):062312, June 2010. arXiv:1001.1732.
[11] Matthias Christandl, Robert König, and Renato Renner. Postselection technique for quantum channels with applications to quantum cryptography. *Physical Review Letters*, 102(2):020504, January 2009. arXiv:0809.3019.

[12] Nilanjana Datta, Marco Tomamichel, and Mark M. Wilde. Second-order coding rates for entanglement-assisted communication. May 2014. arXiv:1405.1797.

[13] Igor Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51(1):44–55, January 2005. arXiv:quant-ph/0304127.

[14] Igor Devetak and Peter W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Communications in Mathematical Physics*, 256(2):287–303, June 2005. arXiv:quant-ph/0311131.

[15] Rupert L. Frank and Elliott H. Lieb. Monotonicity of a relative Rényi entropy. *Journal of Mathematical Physics*, 54(12):122201, December 2013. arXiv:1306.5358.

[16] Manish K. Gupta and Mark M. Wilde. Multiplicativity of completely bounded $p$-norms implies a strong converse for entanglement-assisted capacity. October 2013. arXiv:1310.7028.

[17] Patrick Hayden, Peter W. Shor, and Andreas Winter. Random quantum codes from Gaussian ensembles and an uncertainty relation. *Open Systems & Information Dynamics*, 15(1):71–89, March 2008. arXiv:0712.0975.

[18] Michal Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1-2):1–8, November 1996. arXiv:quant-ph/9605038.

[19] Paweł Horodecki, Michal Horodecki, and Ryszard Horodecki. Binding entanglement channels. *Journal of Modern Optics*, 47(2-3):347–354, July 2000. arXiv:quant-ph/9904092.

[20] Min-Hsiu Hsieh, Igor Devetak, and Andreas Winter. Entanglement-assisted capacity of quantum multiple-access channels. *IEEE Transactions on Information Theory*, 54(7):3078–3090, July 2008. arXiv:quant-ph/0511228.

[21] Christopher King, Keiji Matsumoto, Michael Nathanson, and Mary Beth Ruskai. Properties of conjugate channels with applications to additivity and multiplicativity. *Markov Processes and Related Fields*, 13(2):391–423, 2007. J. T. Lewis memorial issue, arXiv:quant-ph/0509126.

[22] Rochus Klesse. A random coding based proof for the quantum coding theorem. *Open Systems & Information Dynamics*, 15(1):21–45, March 2008. arXiv:0712.2558.

[23] Robert Koenig and Stephanie Wehner. A strong converse for classical channel coding using entangled inputs. *Physical Review Letters*, 103:070504, August 2009. arXiv:0903.2838.

[24] Seth Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55(3):1613–1622, March 1997. arXiv:quant-ph/9604015.

[25] Ciara Morgan and Andreas Winter. “Pretty strong” converse for the quantum capacity of degradable channels. *IEEE Transactions on Information Theory*, 60(1):317–333, January 2014. arXiv:1301.4927.
[26] Milán Mosonyi and Tomohiro Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. September 2013. arXiv:1309.3228.

[27] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: a new definition and some properties. *Journal of Mathematical Physics*, 54(12):122203, December 2013. arXiv:1306.3142.

[28] Tomohiro Ogawa and Hiroshi Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Transactions on Information Theory*, 45:2486–2489, November 1999. arXiv:quant-ph/9808063.

[29] Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77(8):1413–1415, August 1996. arXiv:quant-ph/9604005.

[30] Yury Polyanskiy and Sergio Verdú. Arimoto channel coding converse and Rényi divergence. In *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computation*, pages 1327–1333, September 2010.

[31] Eric M. Rains. Bound on distillable entanglement. *Physical Review A*, 60(1):179–184, July 1999. arXiv:quant-ph/9809082.

[32] Eric M. Rains. A semidefinite program for distillable entanglement. *IEEE Transactions on Information Theory*, 47(7):2921–2933, November 2001. arXiv:quant-ph/0008047.

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

[34] Benjamin Schumacher and Michael A. Nielsen. Quantum data processing and error correction. *Physical Review A*, 54(4):2629–2635, October 1996. arXiv:quant-ph/9604022.

[35] Naresh Sharma and Naqueeb Ahmad Warsi. On the strong converses for the quantum channel capacity theorems. June 2012. arXiv:1205.1712.

[36] Peter W. Shor. Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52(4):R2493–R2496, October 1995.

[37] Peter W. Shor. The quantum channel capacity and coherent information. In *Lecture Notes, MSRI Workshop on Quantum Computation*, 2002.

[38] Graeme Smith and John A. Smolin. Additive extensions of a quantum channel. *Proceedings of the IEEE Information Theory Workshop*, pages 368–372, May 2008. arXiv:0712.2471.

[39] Graeme Smith and John A. Smolin. Detecting incapacity of a quantum channel. *Physical Review Letters*, 108:230507, June 2012. arXiv:1108.1807.

[40] Graeme Smith, John A. Smolin, and Andreas Winter. The quantum capacity with symmetric side channels. *IEEE Transactions on Information Theory*, 54(9):4208–4217, September 2008. arXiv:quant-ph/0607039.

[41] Masahiro Takeoka, Saikat Guha, and Mark M. Wilde. The squashed entanglement of a quantum channel. *Accepted for publication in IEEE Transactions on Information Theory*, 2014. arXiv:1310.0129.
[42] Marco Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory.* PhD thesis, ETH Zurich, March 2012. arXiv:1203.2142.

[43] Marco Tomamichel, Mario Berta, and Masahito Hayashi. A duality relation connecting different quantum generalizations of the conditional Rényi entropy. November 2013. arXiv:1311.3887.

[44] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, December 2009. arXiv:0811.1221.

[45] Marco Tomamichel and Masahito Hayashi. A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Transactions on Information Theory*, 59(11):7693–7710, November 2013. arXiv:1208.1478.

[46] Marco Tomamichel and Vincent Y. F. Tan. On the Gaussian approximation for the classical capacity of quantum channels. August 2013. arXiv:1308.6503.

[47] Mark M. Wilde and Min-Hsiu Hsieh. The quantum dynamic capacity formula of a quantum channel. *Quantum Information Processing*, 11(6):1431–1463, December 2012. arXiv:1004.0458.

[48] Mark M. Wilde and Andreas Winter. Strong converse for the classical capacity of the pure-loss bosonic channel. *To appear in Problems of Information Transmission*, August 2013. arXiv:1308.6732.

[49] Mark M. Wilde and Andreas Winter. Strong converse for the quantum capacity of the erasure channel for almost all codes. *To appear in the Proceedings of the 9th Conference on the Theory of Quantum Computation, Communication and Cryptography*, February 2014. arXiv:1402.3626.

[50] Mark M. Wilde, Andreas Winter, and Dong Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Accepted for publication in Communications in Mathematical Physics*, June 2013. arXiv:1306.1586.

[51] Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, November 1999.

[52] Jacob Wolfowitz. *Coding Theorems of Information Theory*, volume 31. Springer, 1964.

[53] Jon Yard, Patrick Hayden, and Igor Devetak. Capacity theorems for quantum multiple-access channels: Classical-quantum and quantum-quantum capacity regions. *IEEE Transactions on Information Theory*, 54(7):3091–3113, July 2008. arXiv:quant-ph/0501045.