Strong Converse for the Quantum Capacity of the Erasure Channel for Almost All Codes

Mark M. Wilde[1] and Andreas Winter[2]

1 **Hearne Institute for Theoretical Physics**
Department of Physics and Astronomy
Center for Computation and Technology
Louisiana State University
Baton Rouge, Louisiana 70808, USA
mwilde@lsu.edu

2 **ICREA & Física Teórica**
Informació i Fenomens Quántics
Universitat Autònoma de Barcelona
ES-08193 Bellaterra (Barcelona), Spain
andreas.winter@uab.cat

Abstract

A strong converse theorem for channel capacity establishes that the error probability in any communication scheme for a given channel necessarily tends to one if the rate of communication exceeds the channel's capacity. Establishing such a theorem for the quantum capacity of degradable channels has been an elusive task, with the strongest progress so far being a so-called "pretty strong converse." In this work, Morgan and Winter proved that the quantum error of any quantum communication scheme for a given degradable channel converges to a value larger than $1/\sqrt{2}$ in the limit of many channel uses if the quantum rate of communication exceeds the channel's quantum capacity. The present paper establishes a theorem that is a counterpart to this "pretty strong converse." We prove that the large fraction of codes having a rate exceeding the erasure channel's quantum capacity have a quantum error tending to one in the limit of many channel uses. Thus, our work adds to the body of evidence that a fully strong converse theorem should hold for the quantum capacity of the erasure channel. As a side result, we prove that the classical capacity of the quantum erasure channel obeys the strong converse property.

1998 ACM Subject Classification H.1.1 Systems and Information Theory, E.4 Coding and Information Theory, Error control codes

Keywords and phrases strong converse, quantum erasure channel, quantum capacity

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.52

1 Introduction

In his seminal paper on quantum error correction, Shor set out the task of determining the quantum capacity of a quantum channel [26], defined as the maximum rate at which it is possible to transmit qubits reliably over a noisy quantum communication channel. Subsequent to this, the coherent information was identified as being a relevant quantity for quantum capacity [23], a regularized upper bound on quantum capacity was established in terms of the coherent information [4, 5], and the coherent information lower bound on the quantum capacity was established by a sequence of works which are often said to bear

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 52–66

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

"increasing standards of rigor" [19, 27, 9]. [1] All of these works did not identify a tractable characterization of the quantum capacity in general, but Devetak and Shor subsequently proved that the coherent information is equal to the quantum capacity for a class of channels bearing the property of degradability [10]. Degradable channels are such that the receiver of the output of the channel can simulate the channel to the environment by applying a degrading map.

A particularly simple example of a degradable channel is the quantum erasure channel $\mathcal{N}_p$ [13], which has the following action on an input density operator ρ:

$$\mathcal{N}_p(\rho) \equiv (1-p)\rho + p|e\rangle\langle e|, \tag{1}$$

where $p \in [0,1]$ is the erasure probability and $|e\rangle$ is a state orthogonal to the input space (i.e., $\langle e|\rho|e\rangle = 0$ for all input ρ). One can readily show that the map to the environment is equivalent (up to isometry) to an erasure channel with the complementary probability:

$$\mathcal{N}_p(\rho) \equiv p\rho + (1-p)|e\rangle\langle e|. \tag{2}$$

The interpretation here is that if the receiver recovers the channel input, then the environment does not and instead receives the erasure flag, and vice versa.

The quantum capacity of the erasure channel was identified early on by employing a now well known "no-cloning" argument [7]. That is, when $p = 1/2$, the channels from input to the receiver and from input to the environment are the same, so that the quantum capacity of the original channel must vanish. If this were not the case, then it would be possible to send quantum data reliably to both the receiver and the environment of the channel, in violation of the no-cloning theorem. It is then possible to prove that the quantum capacity of the erasure channel in general is equal to $(1-2p)\log d$ for $p \geq 1/2$ and zero otherwise (in agreement with the aforementioned reasoning), where d is the dimension of the input space for the channel.

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 error-free quantum communication scheme.

However, the theorem stated as such still leaves more to be desired. For example, it has been known for a long time that the classical capacity of a classical channel obeys the strong converse property [33, 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 established that the strong converse property holds for the classical capacity of several quantum channels [32, 22, 18, 31, 30, 3] and for the entanglement-assisted classical capacity of all quantum channels [6, 8, 14].

Thus, we are left with the strong converse question for the quantum capacity, with the goal being to sharpen our understanding of quantum capacity. In general, the quantum capacity of arbitrary channels can exhibit rather exotic behavior [28], so it seems reasonable to restrict attention for now to the class of degradable channels since they are more well behaved. In this spirit, a recent work has proved that the quantum capacity of all degradable

[1] However, see the later works in [16] and [15], which respectively set [19] and [27] on a firm foundation.

channels exhibits a property dubbed the "pretty strong converse" [20]. These authors have proven that the quantum error[2] of any quantum communication scheme for a degradable channel experiences a sudden jump from zero to at least $1/\sqrt{2}$ when the communicate rate crosses the quantum capacity threshold (this statement is in the limit of many channel uses). At the very least, we now know that the quantum capacity experiences this jump, but the work of [20] left open the question of whether the jump in quantum error is actually from zero to one in the limit of many channel uses.

In this paper, we prove a statement that is similar in spirit to the pretty strong converse: for almost all codes having a rate exceeding the quantum capacity of the erasure channel, the error necessarily converges to one in the limit of many channel uses. We should clarify that we do not prove a strong converse for all codes, but instead show that the strong converse property holds for almost all codes. We will be more precise in what follows with clarifying what we mean by "almost all codes," but suffice it for now to say if anyone devises a communication scheme for quantum communication over the erasure channel whose rate exceeds capacity, then the chances are very good that, regardless of the scheme, it will fail with probability converging to one in the limit of many channel uses.

In the absence of a proof that the strong converse holds, both the present paper and [20] are offering an increasing body of evidence that it should indeed hold for the class of quantum erasure channels. That is, both results allow us to conclude the following statement: all codes whose rate exceeds the quantum capacity of the erasure channel have a quantum error converging to $1/\sqrt{2}$ in the limit of many channel uses, and a large fraction of them in fact have quantum error converging to one.

This paper is organized as follows. The next section reviews the definition of an entanglement generation code. Section 3 then reviews the generalized divergence framework of Sharma and Warsi [25] for establishing bounds relating rate, error, and the channel of interest in any quantum communication protocol. Section 4 provides a proof for our main result: that the strong converse property holds for almost all codes used for quantum communication over the quantum erasure channel. We state some open directions in the conclusion. The appendix includes, as a side result, a proof that the strong converse holds for the classical capacity of the quantum erasure channel.

2 Entanglement generation codes

In this paper, we focus on entanglement generation codes, for which the goal is for the sender Alice to use the channel n times 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. However, we should emphasize again that our final statement is a bound that holds for almost all entanglement generation codes, so that we cannot conclude a full strong converse.

More formally, we now define an $(n, R, \varepsilon, \phi, D)$ entanglement generation code for a channel $\mathcal{N}$. Such a protocol begins with Alice preparing a state on $n+1$ systems, she sends n shares of the state through n instances of the channel, and then Bob decodes. That is,

[2] As quantified by the so-called "purified distance" (see Chapter 3 of [29], for example).

such a code begins with Alice preparing a state $|\phi\rangle_{AA_1\cdots A_n}$. The reduced state on system A has its rank equal to M, where $M = 2^{nR}$. Alice then transmits systems $A_1 \cdots A_n$ through n uses of the channel, leading to the state

$$\rho_{AB^n} \equiv \mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n}), \tag{3}$$

where $\mathcal{N}_{A^n\to B^n} \equiv \mathcal{N}^{\otimes n}$ and A^n is shorthand for $A_1 \cdots A_n$. Finally, Bob performs a decoding $D_{B^n\to\hat{B}}$, leading to the state

$$\omega_{A\hat{B}} \equiv D_{B^n\to\hat{B}}(\mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n})). \tag{4}$$

The fidelity of the code is given by

$$F \equiv \langle\Phi|_{A\hat{B}}\ \omega_{A\hat{B}}\ |\Phi\rangle_{A\hat{B}}, \tag{5}$$

where $|\Phi\rangle_{A\hat{B}}$ is the maximally entangled state

$$|\Phi\rangle_{A\hat{B}} \equiv \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} |i\rangle_A |i\rangle_{\hat{B}}, \tag{6}$$

so that the rate of entanglement generation is equal to $\frac{1}{n}\log_2 M$. An $(n, R, \varepsilon, \phi, D)$ code uses the state ϕ, the decoder D, the channel n times at rate R, 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 $\rho_{AA_1\cdots A_n}$ and the following condition holds

$$\langle\Phi|_{A\hat{B}} D_{B^n\to\hat{B}}(\mathcal{N}_{A^n\to B^n}(\rho_{AA_1\cdots A_n}))|\Phi\rangle_{A\hat{B}} \geq 1-\varepsilon, \tag{7}$$

then there always exists at least one pure state in the spectral decomposition of $\rho_{AA_1\cdots A_n}$ which meets the same fidelity constraint given above.

3 Generalized divergence framework for quantum communication

We now recall the Sharma-Warsi framework for bounding fidelities in quantum communication [25]. We say that $\mathcal{D}(X||Y)$ is a *generalized divergence* if it satisfies the following monotonicity inequality for all quantum channels $\mathcal{M}$ and positive operators X and Y:

$$\mathcal{D}(X||Y) \geq \mathcal{D}(\mathcal{M}(X)||\mathcal{M}(Y)). \tag{8}$$

Let $I_{\mathcal{D}}(A\rangle B)_\rho$ denote the generalized coherent information of a bipartite state ρ_{AB}:

$$I_{\mathcal{D}}(A\rangle B)_\rho \equiv \min_{\sigma_B} \mathcal{D}(\rho_{AB}||I_A \otimes \sigma_B). \tag{9}$$

Let $I_{\mathcal{D}}(\mathcal{N})$ denote the generalized coherent information of a quantum channel $\mathcal{N}$:

$$I_{\mathcal{D}}(\mathcal{N}) \equiv \max_{\phi_{AA'}} I_{\mathcal{D}}(A\rangle B)_{\mathcal{N}_{A'\to B}(\phi_{AA'})} \tag{10}$$

$$= \max_{\phi_{AA'}} \min_{\sigma_B} \mathcal{D}(\mathcal{N}_{A'\to B}(\phi_{AA'})||I_A \otimes \sigma_B). \tag{11}$$

If the generalized divergence is equal to the von Neumann relative entropy, then the above expressions are equal to the usual coherent information of a quantum state and coherent information of a quantum channel, respectively.

We now establish a bound relating the rate and error of any entanglement generation code for a quantum channel $\mathcal{N}$ to the generalized coherent information of the tensor-power channel $\mathcal{N}^{\otimes n}$. For our purposes here, we begin by considering the generalized divergence between the state ρ_{AB^n} defined in (3) that is output from n uses of the channel and any other operator of the form $I_A \otimes \sigma_{B^n}$, where σ_{B^n} is a density operator on the systems B^n:

$$\mathcal{D}(\rho_{AB^n} || I_A \otimes \sigma_{B^n}). \tag{12}$$

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

$$\mathcal{D}(\rho_{AB^n} || I_A \otimes \sigma_{B^n}) \geq \mathcal{D}\big(\omega_{A\hat{B}} || I_A \otimes D_{B^n \to \hat{B}}(\sigma_{B^n})\big). \tag{13}$$

Next, consider the following test (a completely positive trace-preserving map), which outputs a flag indicating whether a state is maximally entangled or not:

$$T_{A\hat{B} \to Z}(\cdot) \equiv \mathrm{Tr}\big\{\Phi_{A\hat{B}}(\cdot)\big\}|1\rangle\langle 1| + \mathrm{Tr}\big\{\big(I_{A\hat{B}} - \Phi_{A\hat{B}}\big)(\cdot)\big\}|0\rangle\langle 0|. \tag{14}$$

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

$$\mathcal{D}\big(\omega_{A\hat{B}} || I_A \otimes D_{B^n \to \hat{B}}(\sigma_{B^n})\big) \geq \mathcal{D}\big(T_{A\hat{B} \to Z}\big(\omega_{A\hat{B}}\big) || T_{A\hat{B} \to Z}\big(I_A \otimes D_{B^n \to \hat{B}}(\sigma_{B^n})\big)\big). \tag{15}$$

By defining

$$\rho_F \equiv F|1\rangle\langle 1| + (1-F)|0\rangle\langle 0|, \tag{16}$$

$$P_{\frac{1}{M}} \equiv \frac{1}{M}|1\rangle\langle 1| + \left(M - \frac{1}{M}\right)|0\rangle\langle 0|, \tag{17}$$

we see that

$$\mathcal{D}\big(T_{A\hat{B} \to Z}\big(\omega_{A\hat{B}}\big) || T_{A\hat{B} \to Z}\big(I_A \otimes D_{B^n \to \hat{B}}(\sigma_{B^n})\big)\big) = \mathcal{D}\Big(\rho_F || P_{\frac{1}{M}}\Big), \tag{18}$$

which follows from (5) and the fact that

$$\mathrm{Tr}\big\{\Phi_{A\hat{B}}\big(I_A \otimes D_{B^n \to \hat{B}}(\sigma_{B^n})\big)\big\} = \frac{1}{M}. \tag{19}$$

Thus, putting everything together, we obtain the following inequality

$$\mathcal{D}(\rho_{AB^n} || I_A \otimes \sigma_{B^n}) \geq \mathcal{D}\Big(\rho_F || P_{\frac{1}{M}}\Big). \tag{20}$$

This inequality holds for any choice of σ_{B^n}, so we can obtain the tightest upper bound on $\mathcal{D}(\rho_F || P_{\frac{1}{M}})$ for a particular entanglement generation code with initial state $\phi_{AA_1 \cdots A_n}$ by taking a minimization over all such σ_{B^n}:

$$\min_{\sigma_{B^n}} \mathcal{D}(\rho_{AB^n} || I_A \otimes \sigma_{B^n}) \geq \mathcal{D}\Big(\rho_F || P_{\frac{1}{M}}\Big). \tag{21}$$

We can then remove the dependence of the bound on any particular entanglement generation code by taking a maximization over all initial states $\phi_{AA_1 \cdots A_n}$:

$$\max_{\phi_{AA_1 \cdots A_n}} \min_{\sigma_{B^n}} \mathcal{D}(\rho_{AB^n} || I_A \otimes \sigma_{B^n}) \geq \mathcal{D}\Big(\rho_F || P_{\frac{1}{M}}\Big). \tag{22}$$

By employing the definition in (10), we find that the bound is equivalent to

$$I_{\mathcal{D}}\big(\mathcal{N}^{\otimes n}\big) \geq \mathcal{D}\Big(\rho_F || P_{\frac{1}{M}}\Big). \tag{23}$$

3.1 Specializing to Rényi relative entropies

The above development applies for any divergence satisfying monotonicity, and the Rényi relative entropy is a particular example of a generalized divergence, defined as

$$D_\alpha(\rho||\sigma) \equiv \frac{1}{\alpha-1}\log_2 \operatorname{Tr}\{\rho^\alpha \sigma^{1-\alpha}\}. \tag{24}$$

Monotonicity of $D_\alpha(\rho||\sigma)$ under quantum channels holds for all $\alpha \in [0,2]$ (see Appendix B of [29], for example). In the present paper, we are focused on $\alpha \in (1,2]$, especially when α is in a neighborhood near one in this interval. This is because the Rényi relative entropy converges to the von Neumann relative entropy as $\alpha \to 1$.

Now we can evaluate the bound in (21) for the case when the divergence is chosen to be the Rényi relative entropy:

$$\begin{aligned}
\min_{\sigma_{B^n}} \mathcal{D}(\rho_{AB^n}||I_A \otimes \sigma_{B^n}) &\geq D_\alpha\left(\rho_F||P_{\frac{1}{M}}\right) && (25)\\
&= \frac{1}{\alpha-1}\log_2\left[F^\alpha\left(\frac{1}{M}\right)^{1-\alpha} + (1-F)^\alpha\left(M-\frac{1}{M}\right)^{1-\alpha}\right] && (26)\\
&\geq \frac{1}{\alpha-1}\log_2\left[F^\alpha\left(\frac{1}{M}\right)^{1-\alpha}\right] && (27)\\
&= \frac{\alpha}{\alpha-1}\log_2[F] + \log_2 M && (28)\\
&= \frac{\alpha}{\alpha-1}\log_2[F] + nR && (29)
\end{aligned}$$

If we optimize over all entanglement generation codes, then we have the bound

$$\max_{\phi_{AA_1\cdots A_n}} \min_{\sigma_{B^n}} D_\alpha(\rho_{AB^n}||I_A \otimes \sigma_{B^n}) \geq \frac{\alpha}{\alpha-1}\log_2[F] + nR. \tag{30}$$

This is equivalent to

$$I_\alpha\left(\mathcal{N}^{\otimes n}\right) \geq \frac{\alpha}{\alpha-1}\log_2[F] + nR, \tag{31}$$

where we define the Rényi coherent information I_α of a quantum channel according to the recipe in (10). Rewriting this, the bound is equivalent to

$$F \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{n}I_\alpha\left(\mathcal{N}^{\otimes n}\right)\right)}. \tag{32}$$

▶ Remark. It is worth noting at this point that if it is possible to prove that $\frac{1}{n}I_\alpha(\mathcal{N}^{\otimes n})$ is an additive function of the channel $\mathcal{N}$, in the sense that

$$\frac{1}{n}I_\alpha\left(\mathcal{N}^{\otimes n}\right) = I_\alpha(\mathcal{N}) \tag{33}$$

for any finite n, then this would be sufficient to prove that the strong converse holds according to the argument of [22] (which has since been repeated in different contexts in both [25] and [14]). (In fact, any subadditivity relation of the following form would suffice: $I_\alpha(\mathcal{N}^{\otimes n}) \leq nI_\alpha(\mathcal{N}) + o(n)$.) One could also consider using the recently developed sandwiched Rényi relative entropy [21, 31] in this context. So far, it is not clear to us whether either of the coherent information quantities derived from the traditional or sandwiched Rényi relative entropies are additive in the above sense for any degradable channel.

3.2 Application to the quantum erasure channel

We now specialize the above bounds to the case of the quantum erasure channel. Beginning from (25)-(29), we see that we can choose any state σ_{B^n} for establishing a bound relating rate and fidelity to an information quantity. So we choose $\sigma_{B^n} = [\mathcal{N}_p(\pi)]^{\otimes n} = ((1-p)\pi + p|e\rangle\langle e|)^{\otimes n}$, where $\pi = I/d$ is the maximally mixed qudit state on the input and $\mathcal{N}_p$ is the erasure channel defined in (1). This then leads to the following bound for any $(n, R, \varepsilon, \phi, D)$ entanglement generation code:

$$\frac{\alpha}{\alpha-1}\log_2[F(\phi)] + nR \leq \min_{\sigma_{B^n}} D_\alpha(\mathcal{N}_{A\to B^n}(\phi_{AA_1\cdots A_n})||I_A \otimes \sigma_{B^n}) \tag{34}$$

$$\leq D_\alpha\big(\mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n})||I_A \otimes [\mathcal{N}_p(\pi)]^{\otimes n}\big), \tag{35}$$

where $\mathcal{N}_{A^n\to B^n} = \mathcal{N}_p^{\otimes n}$ and $F(\phi)$ denotes the fidelity of an entanglement generation code with initial state ϕ.[3] Observe now that the output of n uses of the quantum erasure channel is rather special, in the sense that it can be written as a convex combination of 2^n density operators which are supported on orthogonal subspaces. We can index these by a binary string i (where ones in this string represent the systems that get erased and zeros represent systems that do not get erased), and we denote the density operators for $\mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n})$ by $\omega^i_{AB^n}$ and those for $[\mathcal{N}(\pi)]^{\otimes n}$ by $\tau^i_{B^n}$. Furthermore, let $\{i\}$ be the set of indices for the systems that get erased, so that we denote the systems that get erased by $A^{\{i\}}$ and those that do not by $A^{\{i\}^c}$. We then find that

$$D_\alpha\big(\mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n})||I_A \otimes [\mathcal{N}_p(\pi)]^{\otimes n}\big)$$

$$= \frac{1}{\alpha-1}\log \sum_{i\in\{0,1\}^n} (1-p)^{n-|i|}p^{|i|}\mathrm{Tr}\big\{\big[\omega^i_{AB^n}\big]^\alpha\big(I_A \otimes \big(\tau^i_{B^n}\big)^{1-\alpha}\big)\big\} \tag{36}$$

$$= \frac{1}{\alpha-1}\log \sum_{i\in\{0,1\}^n} \big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\mathrm{Tr}\{[\phi_{AA^{\{i\}^c}}]^\alpha\} \tag{37}$$

$$= \frac{1}{\alpha-1}\log \sum_{i\in\{0,1\}^n} \big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}, \tag{38}$$

where the last equality follows because the spectrum of $\phi_{AA^{\{i\}^c}}$ is equal to the spectrum of $\phi_{A^{\{i\}}}$ for a pure state. Rewriting (34)-(38), we obtain the following bound on the fidelity $F(\phi)$:

$$F(\phi) \leq \Big[2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\Big]\left[\sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}\right]^{\frac{1}{\alpha}}. \tag{39}$$

▶ Remark. By inspecting the above, we see that obtaining a general bound on the fidelity of an entanglement generation code for the quantum erasure channel is related to the quantum marginal problem [17], since the various terms $\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}$ in the sum are the α-purities of all of the 2^n marginals of the quantum state $\phi_{AA_1\cdots A_n}$.

4 Strong converse for almost all codes

In the previous section, we established the bound (39) on the fidelity $F(\phi)$ of any $(n, R, \varepsilon, \phi, D)$ entanglement generation code. In this section, we prove our main result, i.e., that the large

[3] We could denote this fidelity as $F(\phi, D)$ because the fidelity of any code depends on the initial state ϕ and the decoder D, but the bound we find here is independent of the decoder D, so we suppress it from the notation.

fraction of capacity-exceeding entanglement generation codes satisfy the strong converse property. Before proving this result, we need to establish a measure on the set of all entanglement generation codes, in order to talk about the fraction of codes that satisfy the strong converse property. The most natural measure in this context is the unitarily invariant measure (Haar measure) on pure states, so that each possible initial state for an entanglement generation code is "receiving equal weight."

Now, suppose that we select the pure state ϕ_{AA^n} at random according to the Haar measure with $|A| = 2^{nR}$ and $|A_i| = d$ for all $i \in \{1, \ldots, n\}$. What makes the subsequent reasoning pertinent is the well-known fact that for $R < Q(\mathcal{N}_p) = (1-2p)\log d$, this choice results in a good code asymptotically with overwhelming probability. (Cf. for instance [15].)

We begin by analyzing the expectation of the fidelity $F(\phi)$:

$$\mathbb{E}_\phi\{F(\phi)\} \leq \mathbb{E}_\phi\left\{2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\left[\sum_{i\in\{0,1\}^n}\left[(1-p)d^{\alpha-1}\right]^{n-|i|}p^{|i|}\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}\right]^{\frac{1}{\alpha}}\right\} \tag{40}$$

$$\leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\left[\sum_{i\in\{0,1\}^n}\left[(1-p)d^{\alpha-1}\right]^{n-|i|}p^{|i|}\ \mathbb{E}_\phi\left\{\mathrm{Tr}\left\{\left[\phi_{A^{\{i\}}}\right]^\alpha\right\}\right\}\right]^{\frac{1}{\alpha}}, \tag{41}$$

with the first inequality following from the development in the previous section and the second inequality following from concavity of $x^{\frac{1}{\alpha}}$ for $\alpha \in (1,2]$. So it remains to analyze the term $\mathbb{E}\{\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}\}$. Let $M_i^\dagger M_i = \phi_{A^{\{i\}}}$ and consider that

$$\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\} = \mathrm{Tr}\left\{(M_i^\dagger M_i)^\alpha\right\} = \mathrm{Tr}\left\{(M_i^\dagger M_i)^{\alpha-1}\left(M_i^\dagger M_i\right)\right\} \tag{42}$$

$$\leq \left(\|M_i\|_\infty^2\right)^{\alpha-1}\mathrm{Tr}\left\{\left(M_i^\dagger M_i\right)\right\} = \left(\|M_i\|_\infty^2\right)^{\alpha-1} \tag{43}$$

By employing the above inequalities and concavity of $x^{\alpha-1}$ for $\alpha \in (1,2]$, we find that

$$\mathbb{E}\{\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}\} \leq \left[\mathbb{E}\left\{\|M_i\|_\infty^2\right\}\right]^{\alpha-1}. \tag{44}$$

For a randomly chosen pure state ψ_{RS} on systems R and S and such that $\psi_R = M^\dagger M$, we have the estimate

$$\mathbb{E}\left\{\|M\|_\infty^2\right\} \leq C d_R^{-1}, \tag{45}$$

where $d_R = \dim(\mathcal{H}_R)$ and C is a universal constant independent of d_R [2]. This then implies the following bound for our setting:

$$\mathbb{E}\{\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}\} \leq \left(C d^{-|i|}\right)^{\alpha-1} = C^{\alpha-1}d^{|i|(1-\alpha)}, \tag{46}$$

where we recall that d is the dimension of an individual input to the channel (so that the

support of $\psi_{A^{\{i\}}}$ has dimension $d^{|i|}$). Plugging back in to (41), we find the upper bound

$$\mathbb{E}_\phi\{F(\phi)\} \leq \left[2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\right]\left[\sum_{i\in\{0,1\}^n}\left[(1-p)d^{\alpha-1}\right]^{n-|i|}p^{|i|}\mathbb{E}\{\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}\}\right]^{\frac{1}{\alpha}} \tag{47}$$

$$\leq \left[2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\right]\left[\sum_{i\in\{0,1\}^n}\left[(1-p)d^{\alpha-1}\right]^{n-|i|}p^{|i|}C^{\alpha-1}d^{|i|(1-\alpha)}\right]^{\frac{1}{\alpha}} \tag{48}$$

$$= 2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\ C^{\alpha-1}\left[\sum_{i\in\{0,1\}^n}\left[(1-p)d^{\alpha-1}\right]^{n-|i|}\left[pd^{(1-\alpha)}\right]^{|i|}\right]^{\frac{1}{\alpha}} \tag{49}$$

$$= 2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\ C^{\alpha-1}\ \left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]^{\frac{n}{\alpha}} \tag{50}$$

$$= 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]-\frac{\alpha}{n}\log C\right)}. \tag{51}$$

We now argue that if the rate R of quantum communication is strictly larger than the quantum capacity $(1-2p)\log d$ of the erasure channel, then we can pick α as a constant near one and n large enough such that

$$\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]-\frac{\alpha}{n}\log C\right)>0. \tag{52}$$

So consider the term:

$$\frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]. \tag{53}$$

Let us set $\alpha = 1+t$, so that the above is

$$\frac{1}{t}\log\left((1-p)d^t+d^{-t}p\right). \tag{54}$$

The limit of this quantity as $t\to 0$ ($\alpha\to 1$) is given by

$$\left.\frac{(1-p)d^t\log d - pd^{-t}\log d}{(1-p)d^t+d^{-t}p}\right|_{t=0} = (1-2p)\log d. \tag{55}$$

The other term $-\frac{\alpha}{n}\log C$ in the exponent becomes arbitrarily small as n becomes larger. Thus, it is always possible to pick a constant α and n large enough so that (52) is satisfied, and we recover a strong converse property for the expectation of the fidelity under randomly chosen entanglement generation codes.

Since the fidelity $F(\phi)$ is a non-negative random variable between zero and one, we can appeal to Markov's inequality to recover the following bound:

$$\begin{aligned}&\Pr_\phi\left\{F(\phi) > 2^{-\frac{1}{2}n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]-\frac{\alpha}{n}\log C\right)}\right\}\\ &\leq \frac{\mathbb{E}_\phi\{F(\phi)\}}{2^{-\frac{1}{2}n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log[(1-p)d^{\alpha-1}+d^{1-\alpha}p]-\frac{\alpha}{n}\log C\right)}}\\ &\leq 2^{-\frac{1}{2}n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]-\frac{\alpha}{n}\log C\right)},\end{aligned} \tag{56}$$

where we used the bound in (51) for the second inequality. Thus, our conclusion is that if $R>(1-2p)\log d$, then we can choose α a constant and n large enough so that (52) holds, with the fraction of codes satisfying the strong converse property rapidly approaching one as the number of channel uses increases.

We can obtain an even sharper statement about the convergence by appealing to Levy's Lemma (see [12], for example):

▶ **Lemma 1** (Levy's Lemma). *Let $f : \mathbb{C}^d \to \mathbb{R}$ and $\eta > 0$ be such that for all pure states $|\varphi_1\rangle$ and $|\varphi_2\rangle$ in $\mathbb{C}^d$*

$$|f(|\varphi_1\rangle) - f(|\varphi_1\rangle)| \leq \eta \||\varphi_1\rangle - |\varphi_2\rangle\|_2.$$

Let $|\varphi\rangle$ be a random pure state in $\mathbb{C}^d$. Then for all $\delta \in [0, \eta]$, the following bound holds

$$\Pr\{|f(|\varphi\rangle) - \mathbb{E}\{f(|\varphi\rangle)\}| \geq \delta\} \leq 4 \exp\left\{-\frac{d\delta^2}{c\eta}\right\},$$

where c is a positive constant.

We obtain a Lipschitz constant for the fidelity as a function of pure input states as follows:

$$\begin{aligned} |F(\varphi_1) - F(\varphi_2)| &\leq |F(\varphi_1) - F(\varphi_2)| + |[1 - F(\varphi_1)] - [1 - F(\varphi_2)]| && (57)\\ &\leq \|\varphi_1 - \varphi_2\|_1 && (58)\\ &\leq 2\||\varphi_1\rangle - |\varphi_2\rangle\|_2. && (59)\end{aligned}$$

The first inequality is obvious, the second follows from monotonicity of trace distance under quantum operations (with these operations being a test for the maximally entangled state, the decoder, the channel and the encoder), and the third inequality is straightforward (see Lemma I.4 in [11], for example).

Since we have the bound

$$0 \leq \mathbb{E}_\phi\{F(\phi)\} \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1} + d^{1-\alpha}p\right] - \frac{\alpha}{n}\log C\right)} \equiv g, \tag{60}$$

it follows from Levy's lemma that

$$\begin{aligned} \Pr\{F(\phi) \geq g + \delta\} &\leq \Pr\{F(\phi) \geq \mathbb{E}_\phi\{F(\phi)\} + \delta\} && (61)\\ &\leq 4 \exp\left\{-\frac{2^{n[R+\log d]}\delta^2}{2c}\right\} && (62)\end{aligned}$$

We can take $\delta = g$, to find that

$$\begin{aligned} &\Pr\left\{F(\phi) \geq 2 \cdot 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1} + d^{1-\alpha}p\right] - \frac{\alpha}{n}\log C\right)}\right\}\\ &\qquad \leq 4 \exp\left\{-\frac{2^{n[R+\log d]}\left[2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1} + d^{1-\alpha}p\right] - \frac{\alpha}{n}\log C\right)}\right]^2}{2c}\right\}. \end{aligned} \tag{63}$$

Now, without loss of generality, we can take $R \leq \log d$ (otherwise the strong converse already holds for all codes), so that $R + \log d \geq 2\left(\frac{\alpha-1}{\alpha}\right)R$. Thus, we see that the fraction of codes with $R > (1 - 2p)\log d$ and obeying the strong converse approaches one doubly exponentially fast in the number of channel uses.

5 Conclusion

The main result of the present paper is a proof that the large fraction of codes with a quantum communication rate exceeding the quantum capacity of the erasure channel satisfy the strong converse. We view this result as adding to the evidence from [20] that a strong converse should hold for the quantum capacity of these channels. The main open question going forward from here is to prove that a fully strong converse holds for the quantum capacity of the erasure channel (i.e., that if the rate of any quantum communication scheme exceeds the

quantum capacity of the erasure channel, then the quantum error necessarily converges to one).

The focus on the erasure channel of the present discussion may be justified by the simplicity of the channel (including its additivity). It also allowed us to give an illustration of the power of the Rényi divergence approach. At the same time, it seems to be true for all currently known random code ensembles achieving the coherent information for a channel $\mathcal{N}$ with Stinespring isometry $V : A' \hookrightarrow B \otimes E$ (with respect to a given input density ρ_A), that at rates above the same coherent information they have fidelity going to zero, with overwhelming probability. Of course this has to be verified for each ensemble separately, but rests on two properties that hold for most codes in the ensemble. Namely, with respect to the pure state $|\psi\rangle_{AB^nE^n} = (I \otimes V^{\otimes n})|\phi\rangle_{AA'^n}$:

1. **Typicality of B**. The channel output ψ_{B^n} is largely in the typical subspace of $\mathcal{N}(\rho_A)^{\otimes n}$ in the sense that $H^\delta_{\max}(B^n) \leq nS(\mathcal{N}(\rho_A)) + o(n)$.
2. **Saturation of E**. The complementary channel output ψ_{E^n} covers essentially uniformly the typical subspace of $\mathcal{N}^c(\rho_A)^{\otimes n}$ in the sense that $H^\delta_{\min}(E^n) \geq nS(\mathcal{N}^c(\rho_A)) - o(n)$.

[In fact, in practice the latter property tends to be true for most states in most code subspaces.] We refer to [29] (cf. [20]) for the definitions and necessary properties of (smooth) min- and max-entropies used in the following.

Now, if our code is supposed to generate entanglement at rate R with fidelity F, then by the decoupling principle,

$$H^{\sqrt{1-F^2}}_{\min}(A|E^n) \geq nR. \tag{64}$$

On the other hand, using relations between min- and max-entropies as well as chain rules,

$$\begin{aligned} H^{\sqrt{1-F^2}}_{\min}(A|E^n) &\lesssim H^\epsilon_{\max}(A|E^n) \\ &\lesssim H^\delta_{\max}(AE^n) - H^\delta_{\min}(E^n) \\ &= H^\delta_{\max}(B^n) - H^\delta_{\min}(E^n), \end{aligned} \tag{65}$$

where $\epsilon = \frac{1}{2}(1 - \sqrt{1-F^2})$ and $\delta = \frac{1}{4}\epsilon$, the inequalities are true up to terms of order $\log \frac{1}{\delta}$. By the typicality and saturation properties, (64) and (65) bound the rate as desired,

$$R \leq S(\mathcal{N}(\rho_A)) - S(\mathcal{N}^c(\rho_A)) + o(1) = I(A\rangle B) + o(1). \tag{66}$$

Acknowledgements. We are grateful to Naresh Sharma for many conversations from which the ideas in this paper arose. We thank the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge for organizing the semester "Mathematical Challenges in Quantum Information," at which we had an opportunity to discuss this research. MMW is grateful to the Department of Physics and Astronomy at Louisiana State University for startup funds that supported this research and acknowledges support from the DARPA Quiness Program through US Army Research Office award W31P4Q-12-1-0019. 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.

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 Guillaume Aubrun, Stanislaw Szarek, and Elisabeth Werner. Non-additivity of Rényi entropy and Dvoretzky's theorem. October 2009. arXiv:0910.1189.

3 Bhaskar Roy Bardhan, Raul Garcia-Patron, Mark M. Wilde, and Andreas Winter. Strong converse for the classical capacity of all phase-insensitive bosonic Gaussian channels. January 2014. arXiv:1401.4161.

4 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.

5 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.

6 Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. Quantum reverse Shannon theorem. December 2012. arXiv:0912.5537.

7 Charles H. Bennett, David P. DiVincenzo, and John A. Smolin. Capacities of quantum erasure channels. *Physical Review Letters*, 78(16):3217–3220, April 1997. arXiv:quant-ph/9701015.

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 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.

10 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.

11 Frederic Dupuis. *The decoupling approach to quantum information theory*. PhD thesis, Université de Montréal, November 2009. arXiv:1004.1641.

12 Omar Fawzi. *Uncertainty relations for multiple measurements with applications*. PhD thesis, McGill University, August 2012. arXiv:1208.5918.

13 Markus Grassl, Thomas Beth, and Thomas Pellizzari. Codes for the quantum erasure channel. *Physical Review A*, 56(1):33–38, July 1997. arXiv:quant-ph/9610042.

14 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.

15 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.

16 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.

17 Alexander Klyachko. Quantum marginal problem and representations of the symmetric group. September 2004. arXiv:quant-ph/0409113.

18 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.

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

20 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.

21 Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: a new definition and some properties. June 2013. arXiv:1306.3142.

22 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.

23 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.

24 Benjamin Schumacher and Michael D. Westmoreland. Optimal signal ensembles. *Physical Review A*, 63:022308, January 2001.

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

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

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

28 Graeme Smith and Jon Yard. Quantum communication with zero-capacity channels. *Science*, 321:1812–1815, September 2008. arXiv:0807.4935.

29 Marco Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory*. PhD thesis, ETH Zurich, 2012. arXiv:1203.2142.

30 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.

31 Mark M. Wilde, Andreas Winter, and Dong Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels. June 2013. arXiv:1306.1586.

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

33 Jacob Wolfowitz. *Coding Theorems of Information Theory*, volume 31. Springer, 1964.

A Strong converse for the classical capacity of the quantum erasure channel

In this appendix, we detail a proof that the strong converse holds for the classical capacity of the quantum erasure channel. To our knowledge, a proof of this statement has not yet appeared in the literature. This result was obtained in collaboration with Naresh Sharma.

Using the generalized divergence framework established in [25] and reviewed in [31] (or even the method of Koenig-Wehner [18]), we obtain the following bound on the success probability when transmitting a classical message through the quantum erasure channel

$$p_{\text{succ}} \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{n}\chi_\alpha\left(\mathcal{N}^{\otimes n}\right)\right)}, \tag{67}$$

where

$$\frac{1}{n}\chi_\alpha\left(\mathcal{N}^{\otimes n}\right) \tag{68}$$

is the regularized Rényi-Holevo information of the erasure channel. So our goal is to prove that this quantity is additive as a function of the quantum erasure channel. First recall that this quantity can be written as an information radius [24, 31]:

$$\chi_\alpha\left(\mathcal{N}^{\otimes n}\right) = \min_{\sigma_{B^n}} \max_{\rho_{A^n}} D_\alpha\left(\mathcal{N}^{\otimes n}(\rho_{A^n})||\sigma_{B^n}\right). \tag{69}$$

With this, we see that we can upper bound this quantity simply by choosing σ_{B^n} to be the output of the erasure channel when the tensor-power maximally mixed state is input:

$$\chi_\alpha\big(\mathcal{N}^{\otimes n}\big) \leq \max_{\rho_{A^n}} D_\alpha\big(\mathcal{N}^{\otimes n}(\rho_{A^n})\|[\mathcal{N}(\pi)]^{\otimes n}\big). \tag{70}$$

As discussed in Section 3.2, the output of the quantum erasure channel is rather special, in the sense that it can be written as a linear combination of 2^n density operators which are supported on orthogonal subspaces. We can index these by a binary string i (where ones in this string represent the systems that get erased and zeros represent systems that do not get erased), and we denote the density operators for $\mathcal{N}^{\otimes n}(\rho_{A^n})$ by $\omega^i_{B^n}$ and those for $[\mathcal{N}(\pi)]^{\otimes n}$ by $\tau^i_{B^n}$. Furthermore, let $\{i\}$ be the set of indices for the systems that get erased, so that we denote the systems that get erased by $A^{\{i\}}$ and those that do not by $A^{\{i\}^c}$. We then find that

$$\begin{aligned}
&\max_{\rho_{A^n}} D_\alpha\big(\mathcal{N}^{\otimes n}(\rho_{A^n})\|[\mathcal{N}(\pi)]^{\otimes n}\big) \\
&= \frac{1}{\alpha-1}\log \max_{\rho_{A^n}} \operatorname{Tr}\big\{\big[\mathcal{N}^{\otimes n}(\rho_{A^n})\big]^\alpha\big([\mathcal{N}(\pi)]^{\otimes n}\big)^{1-\alpha}\big\} && (71)\\
&= \frac{1}{\alpha-1}\log \max_{\rho_{A^n}} \sum_{i\in\{0,1\}^n} (1-p)^{n-|i|}p^{|i|}\operatorname{Tr}\big\{\big[\omega^i_{B^n}\big]^\alpha\big[\tau^i_{B^n}\big]^{1-\alpha}\big\} && (72)\\
&= \frac{1}{\alpha-1}\log \max_{\rho_{A^n}} \sum_{i\in\{0,1\}^n} (1-p)^{n-|i|}p^{|i|}\operatorname{Tr}\big\{[\rho_{A^{\{i\}^c}}]^\alpha[\pi_{A^{\{i\}^c}}]^{1-\alpha}\big\} && (73)
\end{aligned}$$

The above equalities follow simply by substitution and some algebra. Continuing, the last line above is equal to

$$\begin{aligned}
&= \frac{1}{\alpha-1}\log \max_{\rho_{A^n}} \sum_{i\in\{0,1\}^n} \big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\operatorname{Tr}\{[\rho_{A^{\{i\}^c}}]^\alpha\} && (74)\\
&\leq \frac{1}{\alpha-1}\log \sum_{i\in\{0,1\}^n} \big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|} && (75)\\
&= \frac{1}{\alpha-1}\log \sum_{k=0}^{n}\big[(1-p)d^{\alpha-1}\big]^{n-k}p^k\binom{n}{k} && (76)\\
&= \frac{1}{\alpha-1}\log\Big((1-p)d^{(\alpha-1)}+p\Big)^n && (77)\\
&= n\left[\frac{1}{\alpha-1}\log\Big((1-p)d^{(\alpha-1)}+p\Big)\right] && (78)
\end{aligned}$$

The inequality follows because $\operatorname{Tr}\{[\rho_{A^{\{i\}^c}}]^\alpha\} \leq 1$ for all $\alpha \geq 1$ (and we are considering $\alpha \in (1,2]$ here). The next few equalities are straightforward. Returning to (67), all of this development implies that we get the following upper bound on success probability

$$p_{\text{succ}} \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\left[\frac{1}{\alpha-1}\log\left((1-p)d^{(\alpha-1)}+p\right)\right]\right)} \tag{79}$$

The last line above is a single-letter upper bound. Now, let us set $\alpha = 1+t$, so that the above is

$$\frac{1}{t}\log\big((1-p)d^t+p\big). \tag{80}$$

The limit of this quantity as $t \to 0$ is given by

$$\left.\frac{(1-p)d^t\log d}{(1-p)d^t+p}\right|_{\varepsilon=0} = (1-p)\log d, \tag{81}$$

which is exactly the classical capacity of the quantum erasure channel. Thus, whenever the classical communication rate $R > (1-p)\log d$, we can always find a value of α in a neighborhood of one such that

$$\left(\frac{\alpha-1}{\alpha}\right)\left(R - \left[\frac{1}{\alpha-1}\log\Big((1-p)d^{(\alpha-1)}+p\Big)\right]\right) > 0. \tag{82}$$

This concludes the proof.

Interestingly, the proof above demonstrates that tensor-product pure-state codewords are the optimal choice in order to saturate the bound given above. That is, for pure-state codewords, we have the equality $\mathrm{Tr}\{[\rho_{A^{\{i\}^c}}]^\alpha\} = 1$, so that the upper bound is saturated by this choice.