

1-1-2014

Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy

Mark M. Wilde
Louisiana State University

Andreas Winter
Universitat Autònoma de Barcelona

Dong Yang
Universitat Autònoma de Barcelona

Follow this and additional works at: https://repository.lsu.edu/physics_astronomy_pubs

Recommended Citation

Wilde, M., Winter, A., & Yang, D. (2014). Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy. *Communications in Mathematical Physics*, 331 (2), 593-622. <https://doi.org/10.1007/s00220-014-2122-x>

This Article is brought to you for free and open access by the Department of Physics & Astronomy at LSU Scholarly Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Scholarly Repository. For more information, please contact ir@lsu.edu.

Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy

Mark M. Wilde*

Andreas Winter^{†‡}Dong Yang^{†§}

May 23, 2014

Abstract

A strong converse theorem for the classical capacity of a quantum channel states that the probability of correctly decoding a classical message converges exponentially fast to zero in the limit of many channel uses if the rate of communication exceeds the classical capacity of the channel. Along with a corresponding achievability statement for rates below the capacity, such a strong converse theorem enhances our understanding of the capacity as a very sharp dividing line between achievable and unachievable rates of communication. Here, we show that such a strong converse theorem holds for the classical capacity of all entanglement-breaking channels and all Hadamard channels (the complementary channels of the former). These results follow by bounding the success probability in terms of a “sandwiched” Rényi relative entropy, by showing that this quantity is subadditive for all entanglement-breaking and Hadamard channels, and by relating this quantity to the Holevo capacity. Prior results regarding strong converse theorems for particular covariant channels emerge as a special case of our results.

1 Introduction

One of the most fundamental tasks in quantum information theory is the transmission of classical data over many independent uses of a quantum channel, such that, for a fixed rate of communication, the error probability of the transmission decreases to zero in the limit of many channel uses. The maximum rate at which this is possible for a given channel is known as the classical capacity of the channel. Holevo, Schumacher, and Westmoreland (HSW) [30, 59] characterized the classical capacity of a quantum channel \mathcal{N} in terms of the following formula:

$$\chi(\mathcal{N}) \equiv \max_{\{p_X(x), \rho_x\}} I(X; B)_\rho, \quad (1)$$

where $\{p_X(x), \rho_x\}$ is an ensemble of quantum states, $I(X; B)_\rho \equiv H(X)_\rho + H(B)_\rho - H(XB)_\rho$ is the quantum mutual information, and $H(A)_\sigma \equiv -\text{Tr}\{\sigma \log \sigma\}$ is the von Neumann entropy of a

*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 Fenòmens, Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain

[‡]School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom

[§]Laboratory for Quantum Information, China Jiliang University, Hangzhou, Zhejiang 310018, China

state σ defined on system A .¹ In the above formula, the quantum mutual information $I(X; B)$ is computed with respect to the following classical-quantum state:

$$\rho_{XB} \equiv \sum_x p_X(x) |x\rangle\langle x|_X \otimes \mathcal{N}_{A \rightarrow B}(\rho_x), \quad (2)$$

for some orthonormal basis $\{|x\rangle\}$, and the notation $\mathcal{N}_{A \rightarrow B}$ indicates that the channel accepts an input on the system A and outputs to the system B .

For certain quantum channels, the HSW formula is equal to the classical capacity of the channel [4, 1, 36, 38, 62, 24, 17, 40]. These results follow because the Holevo formula was shown to be additive for these channels, in the sense that the following relation holds for these channels for any positive integer n :

$$\chi(\mathcal{N}^{\otimes n}) = n \chi(\mathcal{N}).$$

However, in general, if one cannot show that the Holevo formula is additive for a given channel, then our best characterization of the classical capacity is given by a regularized formula:

$$\chi_{\text{reg}}(\mathcal{N}) \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}).$$

The work of Hastings [28] suggests that the regularized limit is necessary unless we are able to find some better characterization of the classical capacity, other than the above one given by HSW. Also, an important implication of Hastings' result, which demonstrates a strong separation between the classical and quantum theories of information, is that using entangled quantum codewords between multiple channel uses can enhance the classical capacity of certain quantum channels, whereas it is known that classically correlated codewords do not [30, 59, 71, 52].

Given the above results, one worthwhile direction is to refine our understanding of the classical capacity of channels for which the HSW formula is additive. Indeed, the achievability part of the HSW coding theorem states that as long as the rate of communication is below the classical capacity of the channel, then there exists a coding scheme such that the error probability of the scheme decreases exponentially fast to zero. The converse part of the capacity theorem makes use of the well known Holevo bound [29], and it states that if the rate of communication exceeds the capacity, then the error probability of any coding scheme is bounded away from zero in the limit of many channel uses.

Such a converse statement as given above might suggest that there is room for a trade-off between error probability and communication rate. That is, such a “weak” converse suggests that it might be possible for one to increase communication rates by allowing for an increased error probability. A *strong converse theorem* leaves no such room for a trade-off—it states that if the rate of communication exceeds the capacity, then the error probability of any coding scheme converges to one in the limit of many channel uses. Importantly, a strong converse theorem establishes the capacity of a channel as a very sharp dividing line between which communication rates are achievable or unachievable in the limit of many channel uses.

Strong converse theorems hold for all discrete memoryless classical channels [76, 2]. Wolfowitz employed a combinatorial approach based on the theory of types in order to prove the strong converse theorem [75, 76]. Arimoto used Rényi entropies to bound the probability of successfully decoding in any communication scheme (hereafter referred to as “success probability”) [2], as a

¹Unless stated otherwise, log always denotes the base two logarithm.

counterpart to Gallager’s lower bounds on the success probability in terms of Rényi entropies [25]. Both the Wolfowitz and Arimoto approaches demonstrate that the success probability converges exponentially fast to zero if the rate of communication exceeds the capacity.² Much later, Polyanskiy and Verdú generalized the Arimoto approach in a very useful way, by showing how to obtain a bound on the success probability in terms of any relative-entropy-like quantity satisfying several natural properties [56].

Less is known about strong converses for quantum channels. However, Winter [71] and Ogawa and Nagaoka [52] independently proved a strong converse theorem for channels with classical inputs and quantum outputs. For such channels, the HSW formula in (1) is equal to the classical capacity. The proof of the strong converse in Ref. [71] used a combinatorial approach in the spirit of Wolfowitz. Ogawa and Nagaoka’s proof [52] is in the spirit of Arimoto. Both these proofs or proof techniques show that the strong converse holds for the Holevo capacity (HSW formula) when restricting to codes for which messages are encoded as product states (cf. [72]).

After this initial work, Koenig and Wehner proved that the strong converse holds for the classical capacity of particular covariant quantum channels [41]. Their proof is in the spirit of Arimoto—they considered a Holevo-like quantity derived from the quantum Rényi relative entropy and then showed that this quantity is additive for particular covariant channels. This reduction of the strong converse question to the additivity of an information quantity is similar to the approach of Arimoto, but the situation becomes more interesting for the case of quantum channels since entanglement between channel uses might lead to the quantity being non-additive.

2 Summary of results

In this paper, we prove that a strong converse theorem holds for the classical capacity of all entanglement-breaking channels [31, 62, 34] and their complementary channels, so-called Hadamard channels [37, 40].

Entanglement-breaking channels can be modeled as the following process:

1. The channel performs a quantum measurement on the incoming state.
2. The channel then prepares a particular quantum state at the output depending on the result of the measurement.

The channels are said to be entanglement-breaking because if one applies a channel in this class to a share of an entangled state, then the resulting bipartite state is a separable state, having no entanglement. An important subclass of the entanglement-breaking channels are quantum measurement channels, in which only the first step above occurs and the output is classical. A few authors have studied quantum measurement channels and their corresponding classical capacities in order to interpret the notion of the information gain of a quantum measurement [35, 33, 16, 54] (however, see also Refs. [73, 10, 70, 5] for different interpretations of the information gain of a quantum measurement).

Hadamard channels are the complementary channels of entanglement-breaking ones. That is, the map from the input to the environment of an entanglement-breaking channel is a Hadamard channel. Such channels are given the name “Hadamard” because their output is equal to the

²Note that the earlier approach of Wolfowitz [75] does not give such a bound, but his later approach does [76].

Hadamard (also known as Schur), i.e. entry-wise, multiplication of a representation of the input density matrix with a positive semi-definite matrix. Some interesting channels fall into this class: generalized dephasing channels [18, 77], cloning channels [6, 14], and the so-called Unruh channel [6, 7, 8]. The generalized dephasing channel represents a natural mechanism for decoherence in physical systems such as superconducting qubits [9], the cloning channel represents a natural process that occurs during stimulated emission [48, 64, 43], and the Unruh channel arises in relativistic quantum information theory [6, 7, 8], bearing connections to the process of black-hole stimulated emission [69].

Our result thus sharpens our understanding of the classical capacity for these two classes of channels, as motivated in the introduction. Also, there should be applications of our strong converse theorem in the setting of the noisy bounded storage model of cryptography as discussed in Ref. [42], but we do not specifically address this application here. Moreover, this paper introduces an information quantity, dubbed the “sandwiched Rényi relative entropy,” and we prove that it satisfies monotonicity under quantum operations. This quantity should be of independent interest for study in quantum information theory. It was independently defined in [51].

We now give a brief sketch of the proof of the strong converse for entanglement-breaking channels, as a guide for the details given in the rest of the paper. The proof for Hadamard channels follows some of the same steps, and it ultimately relies on their relation to entanglement-breaking channels along with some additional steps.

1. First, we recall the argument of Sharma and Warsi [61] (which in turn is based on Ref. [56]), in which they showed that any relative-entropy-like quantity that satisfies some natural requirements gives a bound on the success probability of any coding scheme. Let $\mathcal{D}(\rho\|\sigma)$ denote any generalized divergence that satisfies monotonicity (data processing). From this generalized divergence, one can define a Holevo-like quantity for a classical-quantum state of the form in (2), via

$$\chi_{\mathcal{D}}(\mathcal{N}) \equiv \max_{\{p_X(x), \rho_x\}} I_{\mathcal{D}}(X; B), \quad (3)$$

where

$$I_{\mathcal{D}}(X; B) \equiv \min_{\sigma_B} \mathcal{D}(\rho_{XB} \|\rho_X \otimes \sigma_B).$$

Such a quantity itself satisfies a data processing inequality, which we can then exploit to obtain a bound on the success probability for any (n, R, ε) code (a code that uses the channel n times at a fixed rate R and has an error probability no larger than ε).

2. We then introduce a “sandwiched” Rényi relative entropy, based on a parameter α and defined for quantum states ρ and σ as

$$\tilde{D}_{\alpha}(\rho\|\sigma) \equiv \frac{1}{\alpha - 1} \log \text{Tr} \left\{ \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^{\alpha} \right\}. \quad (4)$$

(See also Ref. [51]). This definition of the Rényi relative entropy is different from the traditional one employed in quantum information theory [55] (see Refs. [41, 49] for applications of this quantity). Recall that the Rényi relative entropy is defined as [55]

$$D_{\alpha}(\rho\|\sigma) \equiv \frac{1}{\alpha - 1} \log \text{Tr} \{ \rho^{\alpha} \sigma^{1-\alpha} \}.$$

However, it follows from the Lieb-Thirring trace inequality [45] that $\tilde{D}_\alpha(\rho\|\sigma) \leq D_\alpha(\rho\|\sigma)$ for all $\alpha > 1$. Also, one can easily see that the two quantities are equal to each other whenever ρ and σ commute (when the states are effectively classical).

We prove that $\tilde{D}_\alpha(\rho\|\sigma)$ is monotone under quantum operations for all $\alpha \in (1, 2]$ and that it reduces to the von Neumann relative entropy in the limit as $\alpha \rightarrow 1$. These properties establish $\tilde{D}_\alpha(\rho\|\sigma)$ as a relevant information quantity to consider in quantum information theory. In particular, it will be useful for us in establishing the strong converse for entanglement-breaking and Hadamard channels. We then define a Holevo-like quantity $\tilde{\chi}_\alpha(\mathcal{N})$ via the recipe given in (3).

3. Combining the above two results, we establish the following upper bound on the success probability of any rate R classical communication scheme that uses a channel n times:

$$p_{\text{succ}} \leq 2^{-n(\frac{\alpha-1}{\alpha})(R-\frac{1}{n}\tilde{\chi}_\alpha(\mathcal{N}^{\otimes n}))}.$$

One can realize by inspecting the above formula that subadditivity of $\tilde{\chi}_\alpha$ would be helpful in proving the strong converse, i.e., if the following holds

$$\tilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) \leq n\tilde{\chi}_\alpha(\mathcal{N}). \quad (5)$$

4. Our next step is to prove that the Holevo-like quantity $\tilde{\chi}_\alpha$ is equal to an “ α -information radius” [63, 15, 49]:

$$\tilde{\chi}_\alpha(\mathcal{N}) = \tilde{K}_\alpha(\mathcal{N}) \equiv \min_{\sigma} \max_{\rho} \tilde{D}_\alpha(\mathcal{N}(\rho)\|\sigma). \quad (6)$$

Proving this identity builds upon prior work in Refs. [60, 41].

5. At this point, we exploit two observations. First, conjugating a completely positive entanglement-breaking map by a positive operator does not take it out of this class—i.e., if \mathcal{M}_{EB} is a completely positive entanglement-breaking map, then so is $\mathcal{X} \circ \mathcal{M}_{\text{EB}}$ for any positive operator X , where the action of $\mathcal{X} \circ \mathcal{M}_{\text{EB}}$ on a density operator ρ is defined by $X\mathcal{M}_{\text{EB}}(\rho)X$. Furthermore, if \mathcal{M} is an arbitrary completely positive map, then $\mathcal{X} \circ \mathcal{M}$ for any positive X is completely positive as well. Also, it is possible to interpret the α -information radius $\tilde{K}_\alpha(\mathcal{N})$ in terms of a “sandwiched” α -norm, defined as

$$\|A\|_{\alpha, X} \equiv \left\| X^{1/2} A X^{1/2} \right\|_{\alpha},$$

for any positive operator X and where

$$\|B\|_{\alpha} \equiv \text{Tr}\{(\sqrt{B^\dagger B})^\alpha\}^{1/\alpha}.$$

With these definitions and that in (4), one can see that

$$\tilde{K}_\alpha(\mathcal{N}) \equiv \min_{\sigma} \max_{\rho} \frac{\alpha}{\alpha-1} \log \|\mathcal{N}(\rho)\|_{\alpha, \sigma^{\frac{1-\alpha}{\alpha}}}.$$

King proved that the maximum output α -norm of an entanglement-breaking channel and any other channel is multiplicative [39] for $\alpha \geq 1$, and Holevo observed that King’s proof extends more generally to hold for a completely positive entanglement-breaking map and any other

completely positive map [32]. The following inequality then immediately results from these observations

$$\tilde{K}_\alpha(\mathcal{N}_{\text{EB}} \otimes \mathcal{N}) \leq \tilde{K}_\alpha(\mathcal{N}_{\text{EB}}) + \tilde{K}_\alpha(\mathcal{N}),$$

for \mathcal{N}_{EB} an entanglement-breaking channel and \mathcal{N} any other channel. With the identity in (6), it follows that

$$\tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}} \otimes \mathcal{N}) \leq \tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}}) + \tilde{\chi}_\alpha(\mathcal{N}),$$

and we can deduce the subadditivity relation in (5) for entanglement-breaking channels by an inductive argument.

6. The bound on the success probability for any coding scheme of rate R when using an entanglement-breaking channel then becomes as follows:

$$p_{\text{succ}} \leq 2^{-n(\frac{\alpha-1}{\alpha})(R-\tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}}))}.$$

Finally, by a standard argument [52, 61], we can choose $\varepsilon > 0$ such that $\tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}}) < \chi(\mathcal{N}_{\text{EB}}) + \varepsilon$ for all $\alpha \geq 1$ in some neighborhood of 1, so that the success probability decays exponentially fast to zero with n if $R > \chi(\mathcal{N}_{\text{EB}})$. The strong converse theorem for all entanglement-breaking channels then follows.

The next section reviews some preliminary material, and the rest of the paper proceeds in the order above, giving detailed proofs for each step. After this, we provide a proof of the strong converse for the classical capacity of Hadamard channels. We then conclude with a brief summary and a pointer to concurrent work in Refs. [50, 51, 23, 3].

3 Preliminaries

Operators, norms, states, maps, and channels. 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. The α -norm of an operator X is defined as

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

Let $\mathcal{B}(\mathcal{H})_+$ denote the subset of positive semidefinite operators (we often simply say that an operator is “positive” if it is positive semi-definite). We also write $X \geq 0$ if $X \in \mathcal{B}(\mathcal{H})_+$. An operator ρ is in the set $\mathcal{S}(\mathcal{H})$ of density operators if $\rho \in \mathcal{B}(\mathcal{H})_+$ and $\text{Tr}\{\rho\} = 1$. 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 multipartite 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 $\mathcal{N}_{A \rightarrow B}(\sigma_A) \in \mathcal{B}(\mathcal{H}_B)_+$ whenever $\sigma_A \in \mathcal{B}(\mathcal{H}_A)_+$. Let 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, we say that it is a quantum channel or quantum operation. A positive operator-valued measure (POVM) is a set $\{\Lambda_m\}$ of operators satisfying $\Lambda_m \geq 0 \forall m$ and $\sum_m \Lambda_m = I$.

Entanglement-breaking maps. Any linear map $\mathcal{M}_{A \rightarrow B}$ can be written in the following form:

$$\mathcal{M}_{A \rightarrow B}(X) = \sum_x N_x \text{Tr}\{M_x X\}, \quad (7)$$

for some sets of operators $\{N_x\}$ and $\{M_x\}$. If $N_x, M_x \geq 0$ for all x , then we say that the map is *entanglement-breaking* [31, 62, 34, 32], and one can also verify that it is completely positive as well. The following conditions are equivalent for an entanglement-breaking map \mathcal{M}_{EB} :

1. There is a representation of \mathcal{M}_{EB} of the form in (7) such that $N_x, M_x \geq 0$ for all x .
2. The map \mathcal{M}_{EB} is completely positive and has a Kraus representation with rank-one Kraus operators, so that

$$\mathcal{M}_{\text{EB}}(X) = \sum_y |\varphi_y\rangle\langle\phi_y| X |\phi_y\rangle\langle\varphi_y|,$$

for some sets of vectors $\{|\varphi_y\rangle\}$ and $\{|\phi_y\rangle\}$.

3. For any integer $d \geq 1$ and $\rho_{12} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_d)$, where \mathcal{H}_d is a d -dimensional Hilbert space,

$$(\mathcal{M}_{\text{EB}} \otimes \text{id}_d)(\rho_{12}) = \sum_z F_z \otimes G_z,$$

where $F_z, G_z \geq 0$ for all z .

Remark 1 *An important observation for the work presented here is that conjugating an entanglement-breaking map \mathcal{M}_{EB} by a positive operator Y does not take it out of the entanglement-breaking class. For example, by defining the map $\mathcal{Y}(X) = YXY$, one can easily see that*

$$(\mathcal{Y} \otimes \text{id}_d)(\mathcal{M}_{\text{EB}} \otimes \text{id}_d)(\rho_{12}) = \sum_z Y F_z Y \otimes G_z,$$

so that $Y F_z Y, G_z \geq 0$ for all z and thus $\mathcal{Y} \circ \mathcal{M}_{\text{EB}}$ is an entanglement-breaking map if \mathcal{M}_{EB} is. (One can check that the other equivalent conditions still hold as well.)

The above property is the main reason why our proof of the strong converse follows from King's proof of the multiplicativity of the maximum output α -norm for entanglement-breaking maps [39, 32]. King's proof in turn exploits the following Lieb-Thirring trace inequality [45] (see also [12]), which holds for $B \geq 0$, any operator C , and for $\alpha \geq 1$:

$$\text{Tr}\{(CBC^\dagger)^\alpha\} \leq \text{Tr}\{(C^\dagger C)^\alpha B^\alpha\}. \quad (8)$$

An entanglement-breaking map \mathcal{N}_{EB} is an *entanglement-breaking channel* if it is also trace-preserving. In this case, the above conditions are specialized, taking on a physical interpretation, so that

1. The set $\{M_x\}$ satisfies $\sum_x M_x = I$ and corresponds to a positive operator-valued measure. Each operator N_x is a density operator.

2. The sets of vectors $\{|\varphi_y\rangle\}$ and $\{|\phi_y\rangle\}$ satisfy the overcompleteness relation:

$$\sum_y |\phi_y\rangle\langle\varphi_y|\varphi_y\rangle\langle\phi_y| = I.$$

3. The output state $(\mathcal{M}_{\text{EB}} \otimes \text{id}_d)(\rho_{12})$ is a separable state (a convex combination of product states) for any input.³

Complementary maps and Hadamard maps. A completely positive map $\mathcal{M}_{A \rightarrow B}$ has a Kraus representation, so that its action on any input operator X is as follows:

$$\mathcal{M}_{A \rightarrow B}(X) = \sum_x A_x X A_x^\dagger,$$

for some set of operators $\{A_x\}$. Such a map is a quantum channel if it is also trace preserving, which is equivalent to the following condition on the Kraus operators: $\sum_x A_x^\dagger A_x = I$. We can define a linear operator $V_{A \rightarrow BE}$ as follows:

$$V_{A \rightarrow BE} \equiv \sum_x A_x \otimes |x\rangle_E,$$

for some orthonormal basis $\{|x\rangle\}$ for an environment system E . We recover the original map $\mathcal{M}_{A \rightarrow B}(X)$ by acting first with the linear operator $V_{A \rightarrow BE}$ on the input and then taking a partial trace over the environment system E :

$$\mathcal{M}_{A \rightarrow B}(X) = \text{Tr}_E \left\{ V_{A \rightarrow BE}(X) V_{A \rightarrow BE}^\dagger \right\}.$$

The map *complementary* to $\mathcal{M}_{A \rightarrow B}$, denoted by $\mathcal{M}_{A \rightarrow E}$ or \mathcal{M}^c , is recovered by instead taking a partial trace over the output system B :

$$\mathcal{M}_{A \rightarrow E}(X) = \text{Tr}_B \left\{ V_{A \rightarrow BE}(X) V_{A \rightarrow BE}^\dagger \right\}.$$

Such a map is unique up to a change of basis for the environment system E .

In the case that $\mathcal{M}_{A \rightarrow B}$ is a channel, we say that the linear operator $V_{A \rightarrow BE}$ is a Stinespring dilation of the channel $\mathcal{M}_{A \rightarrow B}$ [66] and one can see that it acts as an isometry. We also say that the map $\mathcal{M}_{A \rightarrow E}$ as defined above is the *channel* complementary to $\mathcal{M}_{A \rightarrow B}$ if $\mathcal{M}_{A \rightarrow B}$ is a channel.

Finally, we say that a map (channel) is Hadamard if it is complementary to an entanglement-breaking map (channel) [40, 32].

4 Bounding the success probability with a generalized divergence

For convenience of the reader, in this section we now review the Sharma-Warsi argument that bounds the success probability for any rate R classical communication scheme in terms of a generalized divergence [61]. This argument in turn is based on the classical argument in Ref. [56]. We include this review for completeness.

³This property is the reason why these channels are said to be “entanglement-breaking.”

A generalized divergence $\mathcal{D}(\rho\|\sigma)$ is a mapping from two quantum states ρ and σ to an extended real number.⁴ Intuitively, it should be some measure of distinguishability. A generalized divergence is useful for us if it is monotone under a quantum operation \mathcal{N} , in the sense that

$$\mathcal{D}(\rho\|\sigma) \geq \mathcal{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)).$$

Intuitively, one should not be able to increase the distinguishability of ρ and σ by processing with a noisy quantum operation \mathcal{N} .

From the above monotonicity property, we can conclude that $\mathcal{D}(\rho\|\sigma)$ is invariant under tensoring with another quantum state τ , in the sense that

$$\mathcal{D}(\rho \otimes \tau \|\sigma \otimes \tau) = \mathcal{D}(\rho\|\sigma). \quad (9)$$

This is because tensoring with another system is a CPTP map, so that $\mathcal{D}(\rho\|\sigma) \geq \mathcal{D}(\rho \otimes \tau \|\sigma \otimes \tau)$, while the partial trace is a CPTP map as well, so that $\mathcal{D}(\rho \otimes \tau \|\sigma \otimes \tau) \geq \mathcal{D}(\rho\|\sigma)$. The interpretation of (9) is that the distinguishability of ρ and σ should be the same if we append an additional quantum system in the state τ .

We can also conclude that it is invariant under the application of a unitary U , in the sense that

$$\mathcal{D}(\rho\|\sigma) = \mathcal{D}(U\rho U^\dagger\|U\sigma U^\dagger).$$

This follows because the maps $U(\cdot)U^\dagger$ and $U^\dagger(\cdot)U$ are CPTP, so that

$$\begin{aligned} \mathcal{D}(\rho\|\sigma) &\geq \mathcal{D}(U\rho U^\dagger\|U\sigma U^\dagger), \\ \mathcal{D}(U\rho U^\dagger\|U\sigma U^\dagger) &\geq \mathcal{D}(U^\dagger U\rho U^\dagger U\|U^\dagger U\sigma U^\dagger U) = \mathcal{D}(\rho\|\sigma). \end{aligned}$$

From this, we can conclude that the divergence reduces to a classical divergence (independent of any orthonormal basis) for the case of commuting, qubit states. Let

$$\begin{aligned} \rho_p &\equiv p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|, \\ \rho_q &\equiv q|0\rangle\langle 0| + (1-q)|1\rangle\langle 1|, \end{aligned}$$

for $0 \leq p, q \leq 1$ and some orthonormal basis $\{|0\rangle, |1\rangle\}$. Let

$$\delta(p\|q) \equiv \mathcal{D}(\rho_p\|\rho_q).$$

It follows that $\delta(p\|q)$ is independent of the choice of basis $\{|0\rangle, |1\rangle\}$.

From such a generalized divergence, we can then define a generalized Holevo information of a channel \mathcal{N} as

$$\chi_{\mathcal{D}}(\mathcal{N}) \equiv \max_{\{p_X(x), \rho_x\}} I_{\mathcal{D}}(X; B)_{\rho}, \quad (10)$$

where the optimization is over ensembles $\{p_X(x), \rho_x\}$ and

$$\begin{aligned} I_{\mathcal{D}}(X; B)_{\rho} &\equiv \min_{\sigma_B} \mathcal{D}(\rho_{XB}\|\rho_X \otimes \sigma_B), \\ \rho_{XB} &\equiv \sum_x p_X(x) |x\rangle\langle x| \otimes \mathcal{N}(\rho_x), \end{aligned}$$

⁴An extended real number can be finite or infinite.

where the optimization is over states σ_B . It is straightforward to show that the quantity $I_{\mathcal{D}}(X; B)$ obeys a data processing inequality by exploiting the fact that the generalized divergence \mathcal{D} does (see Lemma 1 of Ref. [61] for an explicit proof). In this case, a data processing inequality means that

$$I_{\mathcal{D}}(X; B)_{\rho} \geq I_{\mathcal{D}}(X; B')_{\omega},$$

for $\omega_{XB'} \equiv (\text{id}_X \otimes \mathcal{E}_{B \rightarrow B'}) (\rho_{XB})$, where $\mathcal{E}_{B \rightarrow B'}$ is a CPTP map.

4.1 Converse bound from a generalized divergence

We now review the converse argument from Refs. [61, 56] that gives a bound on the success probability for any rate R scheme for classical communication. Any (n, R, ε) protocol for communication has the following form: A sender chooses a message uniformly at random from a message set $\mathbb{M} \equiv \{1, \dots, |\mathbb{M}|\}$, where $|\mathbb{M}| = 2^{nR}$ (it suffices for our purposes to suppose that the choice is uniform). The sender transmits a quantum state ρ_m (a quantum codeword) through n uses of the channel \mathcal{N} . The overall state at this point is described by the following classical-quantum state:

$$\rho_{MB^n} \equiv \sum_m \frac{1}{|\mathbb{M}|} |m\rangle\langle m|_M \otimes \mathcal{N}^{\otimes n}(\rho_m).$$

The receiver applies a decoding POVM $\{\Lambda_m\}$ to the output of the channel to produce an estimate \hat{M} of message M . The resulting classical-quantum state is as follows:

$$\omega_{M\hat{M}} \equiv \sum_{m, m'} \frac{1}{|\mathbb{M}|} |m\rangle\langle m|_M \otimes \text{Tr}\{\Lambda_{m'} \mathcal{N}^{\otimes n}(\rho_m)\} |m'\rangle\langle m'|_{\hat{M}}.$$

The error probability of the scheme is ε if $\Pr\{\hat{M} \neq M\} \leq \varepsilon$. Also, without loss of generality, we can assume that $\varepsilon \leq 1 - 2^{-nR}$ (otherwise, the strong converse would already hold for rates above the capacity since the error probability would obey the bound $\varepsilon > 1 - 2^{-nR}$). We now show how to establish the following bound for any communication scheme as discussed above:

$$\delta(\varepsilon \| 1 - 2^{-nR}) \leq \chi_{\mathcal{D}}(\mathcal{N}^{\otimes n}). \quad (11)$$

Let σ_{B^n} denote an arbitrary density operator on the B^n systems. From the properties of a generalized divergence and the specification above, we can deduce that

$$\begin{aligned} \mathcal{D}(\rho_{MB^n} \| \rho_M \otimes \sigma_{B^n}) &\geq \mathcal{D}(\omega_{M\hat{M}} \| \omega_M \otimes \tau_{\hat{M}}) \\ &\geq \delta(\Pr\{\hat{M} \neq M\} \| 1 - 2^{-nR}) \\ &\geq \delta(\varepsilon \| 1 - 2^{-nR}). \end{aligned}$$

The first inequality follows from monotonicity of the generalized divergence under the decoding map $\sum_m \text{Tr}\{\Lambda_m(\cdot)\} |m\rangle\langle m|_{\hat{M}}$. Also, here, we are letting

$$\tau_{\hat{M}} \equiv \sum_m \text{Tr}\{\Lambda_m \sigma_{B^n}\} |m\rangle\langle m|_{\hat{M}}.$$

The second inequality follows from monotonicity of the generalized divergence under the ‘‘equality test,’’ which is a classical map testing if the value in M is equal to the value in \hat{M} , i.e., $(M, \hat{M}) \rightarrow$

$\delta_{M, \hat{M}}$ (with $\delta_{x,y}$ the Kronecker delta function). This test produces the distribution $(\Pr\{\hat{M} \neq M\}, \Pr\{\hat{M} = M\})$ when acting on the state $\omega_{M\hat{M}}$ and the distribution $(1 - 2^{-nR}, 2^{-nR})$ when acting on the product state $\omega_M \otimes \tau_{\hat{M}}$. The last inequality follows from the monotonicity $\delta(p' \| q) \leq \delta(p \| q)$ whenever $p \leq p' \leq q$ [56] (recall that we have $\Pr\{\hat{M} \neq M\} \leq \varepsilon \leq 1 - 2^{-nR}$). Given that σ_{B^n} is an arbitrary density operator, we can recover the tightest upper bound on $\delta(\varepsilon \| 1 - 2^{-nR})$ by minimizing \mathcal{D} with respect to all such σ_{B^n} :

$$\delta(\varepsilon \| 1 - 2^{-nR}) \leq \min_{\sigma_{B^n}} \mathcal{D}(\rho_{MB^n} \| \rho_M \otimes \sigma_{B^n}).$$

Finally, we can remove the dependence on the particular code by maximizing over all input ensembles:

$$\begin{aligned} \delta(\varepsilon \| 1 - 2^{-nR}) &\leq \max_{\{p_X(x), \rho_x\}} \min_{\sigma_{B^n}} \mathcal{D}(\rho_{XB^n} \| \rho_X \otimes \sigma_{B^n}) \\ &= \chi_{\mathcal{D}}(\mathcal{N}^{\otimes n}), \end{aligned}$$

where

$$\rho_{XB^n} \equiv \sum_x p_X(x) |x\rangle\langle x|_X \otimes \mathcal{N}^{\otimes n}(\rho_x)$$

and the second line follows from the definition of $\chi_{\mathcal{D}}$.

Remark 2 *In light of the above bound in terms of a generalized divergence, in hindsight, the approach of Arimoto [2] (and the follow-up work [52, 41]) appears to be somewhat ad hoc. This becomes amplified in the case of proving strong converse theorems for quantum channels, where one can choose from many different divergences that all reduce to the same classical divergence. In the next section, we define a divergence which gives bounds on the success probability that are tighter than those from Refs. [52, 41].*

Remark 3 *If one employs the von Neumann relative entropy as the divergence, then one arrives at the following weak converse bound:*

$$R \leq \frac{1}{n(1 - \varepsilon)} (\chi(\mathcal{N}^{\otimes n}) + h_2(\varepsilon)),$$

where $h_2(\varepsilon) \equiv -\varepsilon \log \varepsilon - (1 - \varepsilon) \log(1 - \varepsilon)$.

5 The sandwiched quantum Rényi relative entropy

We now define a “sandwiched” quantum Rényi relative entropy and prove several of its properties that establish its utility as an information measure. In particular, the sandwiched Rényi relative entropy is based on a parameter α , and its most important property is that it is monotone under quantum operations for $\alpha \in (1, 2]$. We define this quantity more generally on the space of positive operators, since it might find other applications in quantum information theory.

We begin by defining a quasi-relative entropy, in the spirit of [55], and from this, we obtain the sandwiched Rényi relative entropy.

Definition 4 The sandwiched quasi-relative entropy $\tilde{Q}_\alpha(A\|B)$ is defined for every $\alpha \in (1, \infty)$ and for $A, B \in \mathcal{B}(\mathcal{H})_+$ as

$$\tilde{Q}_\alpha(A\|B) \equiv \begin{cases} \text{Tr} \left\{ \left(B^{\frac{1-\alpha}{2\alpha}} A B^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} & \text{if } \text{supp}(A) \subseteq \text{supp}(B) \\ \infty & \text{otherwise} \end{cases}.$$

The sandwiched Rényi relative entropy is defined as

$$\tilde{D}_\alpha(A\|B) \equiv \frac{1}{\alpha - 1} \log \tilde{Q}_\alpha(A\|B).$$

The sandwiched Rényi relative entropy \tilde{D}_α was independently defined in [68, 22, 51]. One could certainly define these quantities for all non-negative α , but we only define it for the above range for simplicity since we use it just for $\alpha \in (1, 2]$.

One might suspect that there should be a relation between the sandwiched relative entropy and the traditional one. Recall that the quantum Rényi relative entropy is defined as

$$D_\alpha(A\|B) \equiv \frac{1}{\alpha - 1} \log \text{Tr} \{ A^\alpha B^{1-\alpha} \}. \quad (12)$$

By applying the Lieb-Thirring inequality from (8), we see that the following inequality holds for all $\alpha > 1$:

$$\tilde{D}_\alpha(A\|B) \leq D_\alpha(A\|B). \quad (13)$$

This relationship is the main reason why the sandwiched Rényi relative entropy allows us to obtain tighter upper bounds on the success probability of any rate R classical communication protocol. Furthermore, whenever A and B commute, both of these entropies are equal and reduce to the classical Rényi relative entropy. That is, suppose that $A = \sum_x a_x |x\rangle\langle x|$ and $B = \sum_x b_x |x\rangle\langle x|$. Then both quantities are equal to the classical Rényi relative entropy in such a case:

$$\tilde{D}_\alpha(A\|B) = D_\alpha(A\|B) = \frac{1}{\alpha - 1} \log \sum_x a_x^\alpha b_x^{1-\alpha}.$$

We now prove four different properties of the sandwiched quasi-relative entropy $\tilde{Q}_\alpha(A\|B)$: unitary invariance, multiplicativity under tensor-product operators, invariance under tensoring with another system, and joint convexity in its arguments. These four properties taken together then allow us to conclude that $\tilde{Q}_\alpha(A\|B)$ is monotone under noisy quantum operations. Monotonicity of $\tilde{Q}_\alpha(A\|B)$ then implies that $\tilde{D}_\alpha(A\|B)$ is monotone as well.

Theorem 5 The sandwiched quasi-relative entropy $\tilde{Q}_\alpha(A\|B)$ is invariant under all unitaries U , multiplicative under tensor-product operators $A_1 \otimes A_2$ and $B_1 \otimes B_2$, and invariant under tensoring A and B with another quantum system:

$$\begin{aligned} \tilde{Q}_\alpha(UAU^\dagger \| UBU^\dagger) &= \tilde{Q}_\alpha(A\|B), \\ \tilde{Q}_\alpha(A_1 \otimes A_2 \| B_1 \otimes B_2) &= \tilde{Q}_\alpha(A_1 \| B_1) \tilde{Q}_\alpha(A_2 \| B_2), \\ \tilde{Q}_\alpha(A \otimes \tau \| B \otimes \tau) &= \tilde{Q}_\alpha(A\|B). \end{aligned}$$

For all $\alpha \in (1, 2]$, the sandwiched quasi-relative entropy $\tilde{Q}_\alpha(A\|B)$ is jointly convex in its arguments

$$\sum_x p(x) \tilde{Q}_\alpha(A_x\|B_x) \geq \tilde{Q}_\alpha(A\|B).$$

where $A = \sum_x p(x) A_x$ and $B = \sum_x p(x) B_x$.

Proof. We establish unitary invariance by

$$\begin{aligned} \tilde{Q}_\alpha(UAU^\dagger\|UBU^\dagger) &= \text{Tr} \left\{ \left((UBU^\dagger)^{\frac{1-\alpha}{2\alpha}} (UAU^\dagger) (UBU^\dagger)^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} \\ &= \text{Tr} \left\{ \left(UB^{\frac{1-\alpha}{2\alpha}} U^\dagger (UAU^\dagger) UB^{\frac{1-\alpha}{2\alpha}} U^\dagger \right)^\alpha \right\} \\ &= \text{Tr} \left\{ \left(UB^{\frac{1-\alpha}{2\alpha}} AB^{\frac{1-\alpha}{2\alpha}} U^\dagger \right)^\alpha \right\} \\ &= \text{Tr} \left\{ U \left(B^{\frac{1-\alpha}{2\alpha}} AB^{\frac{1-\alpha}{2\alpha}} \right)^\alpha U^\dagger \right\} \\ &= \tilde{Q}_\alpha(A\|B). \end{aligned}$$

Multiplicativity under tensor-product operators follows because

$$\begin{aligned} \tilde{Q}_\alpha(A_1 \otimes A_2\|B_1 \otimes B_2) &= \text{Tr} \left\{ \left((B_1 \otimes B_2)^{\frac{1-\alpha}{2\alpha}} (A_1 \otimes A_2) (B_1 \otimes B_2)^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} \\ &= \text{Tr} \left\{ \left(\left(B_1^{\frac{1-\alpha}{2\alpha}} \otimes B_2^{\frac{1-\alpha}{2\alpha}} \right) (A_1 \otimes A_2) \left(B_1^{\frac{1-\alpha}{2\alpha}} \otimes B_2^{\frac{1-\alpha}{2\alpha}} \right) \right)^\alpha \right\} \\ &= \text{Tr} \left\{ \left(B_1^{\frac{1-\alpha}{2\alpha}} A_1 B_1^{\frac{1-\alpha}{2\alpha}} \otimes B_2^{\frac{1-\alpha}{2\alpha}} A_2 B_2^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} \\ &= \text{Tr} \left\{ \left(B_1^{\frac{1-\alpha}{2\alpha}} A_1 B_1^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \otimes \left(B_2^{\frac{1-\alpha}{2\alpha}} A_2 B_2^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} \\ &= \tilde{Q}_\alpha(A_1\|B_1) \tilde{Q}_\alpha(A_2\|B_2). \end{aligned}$$

Invariance under tensoring with another system then follows as a special case of multiplicativity since we assume that $\text{Tr}\{\tau\} = 1$.

Finally, we prove that this quantity is jointly convex in its arguments $A = \sum_x p(x) A_x$ and $B = \sum_x p(x) B_x$ whenever $\alpha \in (1, 2]$:

$$\sum_x p(x) \tilde{Q}_\alpha(A_x\|B_x) \geq \tilde{Q}_\alpha(A\|B).$$

Taking $|\gamma\rangle = \sum_i |i\rangle|i\rangle$, we can rewrite $\tilde{Q}_\alpha(A\|B)$ as

$$\tilde{Q}_\alpha(A\|B) = \text{Tr} \left\{ |\gamma\rangle\langle\gamma| \sqrt{g(B)} f\left(g(B)^{-1/2} h(A) g(B)^{-1/2}\right) \sqrt{g(B)} \right\},$$

where

$$\begin{aligned} f(x) &\equiv x^\alpha, \\ g(B) &\equiv B^{\frac{\alpha-1}{\alpha}} \otimes (B^T)^{\frac{1}{\alpha}}, \\ h(A) &\equiv A \otimes I. \end{aligned}$$

The function $f(x)$ is operator convex for $\alpha \in (1, 2]$. Also, $g(B)$ is operator concave for $\alpha \in (1, 2]$ because $(L, R) \mapsto L^x \otimes R^y$ is jointly operator concave on positive operators for $x, y \geq 0$ and $x + y \leq 1$ (see Corollary 5.5 of [74]). Also, $h(A)$ is clearly affine. With all of this, it follows from Theorem 5.14 of [74] that

$$\sqrt{g(B)}f\left(g(B)^{-1/2}h(A)g(B)^{-1/2}\right)\sqrt{g(B)}$$

is jointly operator convex. This then implies that the functional $\tilde{Q}_\alpha(A\|B)$ is jointly convex in its arguments. ■

Monotonicity of $\tilde{Q}_\alpha(A\|B)$ then follows by using the above properties and a standard argument detailed in Theorem 5.16 of [74]. Also, by inspecting the definition of $\tilde{D}_\alpha(A\|B)$, it follows that $\tilde{D}_\alpha(A\|B)$ is monotone given that $\tilde{Q}_\alpha(A\|B)$ is.

For convenience of the reader, this paper's appendix reproduces the statements of Theorem 5.14, Corollary 5.5, and Theorem 5.16 from [74].

Corollary 6 (Monotonicity) *For all $\alpha \in (1, 2]$, the sandwiched quasi-relative entropy \tilde{Q}_α and the sandwiched Rényi relative entropy \tilde{D}_α are monotone under a quantum operation \mathcal{N} :*

$$\begin{aligned}\tilde{Q}_\alpha(A\|B) &\geq \tilde{Q}_\alpha(\mathcal{N}(A)\|\mathcal{N}(B)), \\ \tilde{D}_\alpha(A\|B) &\geq \tilde{D}_\alpha(\mathcal{N}(A)\|\mathcal{N}(B)).\end{aligned}$$

We note that this corollary generalizes Theorem 21 of Ref. [19] beyond $\alpha = 2$ (the above proof of joint convexity of $\tilde{Q}_\alpha(A\|B)$ is in fact a straightforward generalization of the proof of Theorem 21 in Ref. [19]).

Corollary 7 (Positivity) *The sandwiched Rényi relative entropy $\tilde{D}_\alpha(\rho\|\sigma)$ is non-negative for density operators ρ and σ and for $\alpha \in (1, 2]$.*

Proof. Writing a spectral decomposition for ρ as $\rho = \sum_x p(x)|\phi_x\rangle\langle\phi_x|$, we can apply a “dephasing” or “pinching” map $\Delta(\cdot) \equiv \sum_x |\phi_x\rangle\langle\phi_x|(\cdot)|\phi_x\rangle\langle\phi_x|$ to both states. From monotonicity, we find that

$$\tilde{D}_\alpha(\rho\|\sigma) \geq \tilde{D}_\alpha(\Delta(\rho)\|\Delta(\sigma)) \geq 0,$$

where the second inequality follows because the sandwiched Rényi relative entropy reduces to the classical one, which we know is non-negative for probability distributions. ■

Corollary 8 (Equality conditions) *For density operators ρ and σ and $\alpha \in (1, 2]$, the sandwiched Rényi relative entropy satisfies $\tilde{D}_\alpha(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.*

Proof. If $\rho = \sigma$, then $\tilde{D}_\alpha(\rho\|\sigma) = 0$ simply by inspecting the definition of the sandwiched Rényi relative entropy. Now suppose that $\tilde{D}_\alpha(\rho\|\sigma) = 0$. In this case, we can perform an informationally-complete measurement map on the states ρ and σ [57, 11, 58]. Such a measurement map has the following form:

$$\mathcal{M}(\omega) = \sum_x \text{Tr}\{M_x \omega\} |x\rangle\langle x|,$$

for some orthonormal basis $\{|x\rangle\}$ and operators M_x such that $M_x \geq 0$ for all x and $\sum_x M_x = I$, and it is informationally complete in the sense that all the parameters of the density operator ω

are encoded in the distribution $\text{Tr}\{M_x\omega\}$ of the outcomes. From monotonicity and positivity of the sandwiched Rényi relative entropy under quantum operations, it follows that $\tilde{D}_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) = 0$. But this Rényi relative entropy is with respect to classical states, and it is known that the equality conditions for the classical Rényi relative entropies are that $\tilde{D}_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) = 0$ if and only if $\text{Tr}\{M_x\rho\} = \text{Tr}\{M_x\sigma\}$ for all x [15]. Since we chose the measurement to be informationally complete, it follows that $\rho = \sigma$.

An alternate proof of the implication $\tilde{D}_\alpha(\rho\|\sigma) = 0 \implies \rho = \sigma$, suggested by an anonymous referee, is as follows. Let U be any unitary and let Δ be the dephasing or pinching map given above. Then we have

$$0 = \tilde{D}_\alpha(\rho\|\sigma) = \tilde{D}_\alpha(U\rho U^\dagger\|U\sigma U^\dagger) \geq \tilde{D}_\alpha(\Delta(U\rho U^\dagger)\|\Delta(U\sigma U^\dagger)) = 0.$$

By the classical conditions for equality, it follows that $\Delta(U(\rho - \sigma)U^\dagger) = 0$ for any unitary U . But then it immediately follows that $\text{Tr}\{B(\rho - \sigma)\} = 0$ for any Hermitian B , from which we can conclude that $\rho = \sigma$. ■

Corollary 9 (Joint quasi-convexity) *The sandwiched relative Rényi entropy $\tilde{D}_\alpha(A\|B)$ is jointly quasi-convex in its arguments for $\alpha \in (1, 2]$, in the sense that*

$$\tilde{D}_\alpha(A\|B) \leq \max_x \tilde{D}_\alpha(A_x\|B_x),$$

where $A = \sum_x p(x)A_x$ and $B = \sum_x p(x)B_x$.

Proof. This follows by employing joint convexity of $\tilde{Q}_\alpha(A\|B)$:

$$\begin{aligned} \tilde{D}_\alpha(A\|B) &= \frac{1}{\alpha - 1} \log \tilde{Q}_\alpha(A\|B) \\ &\leq \frac{1}{\alpha - 1} \log \sum_x p(x) \tilde{Q}_\alpha(A_x\|B_x) \\ &\leq \frac{1}{\alpha - 1} \log \max_x \tilde{Q}_\alpha(A_x\|B_x) \\ &= \max_x \tilde{D}_\alpha(A_x\|B_x). \end{aligned}$$

■

Definition 10 *The von Neumann relative entropy for $A, B \in \mathcal{B}(\mathcal{H})_+$ is defined as*

$$D(A\|B) \equiv \begin{cases} \text{Tr}\{A \log A\} - \text{Tr}\{A \log B\} & \text{if } \text{supp}(A) \subseteq \text{supp}(B) \\ \infty & \text{otherwise} \end{cases}.$$

Proposition 11 *In the limit as α approaches one, the sandwiched relative Rényi entropy $\tilde{D}_\alpha(A\|B)$ converges to the von Neumann relative entropy $D(A\|B)$ if $\text{Tr}\{A\} = 1$:*

$$\lim_{\alpha \rightarrow 1} \tilde{D}_\alpha(A\|B) = D(A\|B).$$

Proof. A proof follows by exploiting some ideas of Carlen and Lieb [13] and Ogawa and Nagaoka [52]. It suffices to show that

$$\frac{\partial}{\partial \alpha} \text{Tr} \left\{ \left(B^{\frac{1-\alpha}{2\alpha}} A B^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} \Big|_{\alpha=1} = \text{Tr} \{ A \log A \} - \text{Tr} \{ A \log B \}.$$

This is because, in order to evaluate the limit, we require L'Hôpital's rule, so that

$$\begin{aligned} \lim_{\alpha \rightarrow 1} \tilde{D}_\alpha(A \| B) &= \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \log \tilde{Q}_\alpha(A \| B) \\ &= \lim_{\alpha \rightarrow 1} \frac{1}{\tilde{Q}_\alpha(A \| B)} \frac{\partial}{\partial \alpha} \tilde{Q}_\alpha(A \| B) \\ &= \frac{\partial}{\partial \alpha} \text{Tr} \left\{ \left(B^{\frac{1-\alpha}{2\alpha}} A B^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} \Big|_{\alpha=1}. \end{aligned}$$

(In this proof, we will take \log to denote the natural logarithm, but note that the result follows simply by replacing the natural logarithm in both definitions with the binary logarithm.) We assume that the support of A is contained in the support of B . Otherwise, there is nothing to prove since both quantities are infinite.

Let us rewrite the expression inside the trace, using $\alpha = 1 + \varepsilon$, as

$$\text{Tr} \left\{ \left(B^{\frac{-\varepsilon}{2(1+\varepsilon)}} A B^{\frac{-\varepsilon}{2(1+\varepsilon)}} \right)^{1+\varepsilon} \right\}.$$

Furthermore, we can use two parameters ε_1 and ε_2 so that the above expression is a special case of

$$f(\varepsilon_1, \varepsilon_2) \equiv \text{Tr} \left\{ \left(B^{\frac{-\varepsilon_1}{2(1+\varepsilon_1)}} A B^{\frac{-\varepsilon_1}{2(1+\varepsilon_1)}} \right)^{1+\varepsilon_2} \right\}.$$

We then have that

$$\begin{aligned} \frac{\partial}{\partial \alpha} \text{Tr} \left\{ \left(B^{\frac{1-\alpha}{2\alpha}} A B^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} \Big|_{\alpha=1} &= \frac{\partial}{\partial \varepsilon} f(\varepsilon, \varepsilon) \Big|_{\varepsilon=0} \\ &= \frac{\partial}{\partial \varepsilon_1} f(\varepsilon_1, 0) \Big|_{\varepsilon_1=0} + \frac{\partial}{\partial \varepsilon_2} f(0, \varepsilon_2) \Big|_{\varepsilon_2=0}. \end{aligned}$$

Consider the following Taylor expansions around $\varepsilon = 0$

$$\begin{aligned} X^{1+\varepsilon} &= X + \varepsilon X \log X + O(\varepsilon^2), \\ X^{\frac{-\varepsilon}{1+\varepsilon}} &= I - \varepsilon \log X + O(\varepsilon^2). \end{aligned}$$

From these, we calculate $f(\varepsilon_1, 0)$ as

$$\begin{aligned} f(\varepsilon_1, 0) &= \text{Tr} \left\{ B^{\frac{-\varepsilon_1}{2(1+\varepsilon_1)}} A B^{\frac{-\varepsilon_1}{2(1+\varepsilon_1)}} \right\} \\ &= \text{Tr} \left\{ A B^{\frac{-\varepsilon_1}{1+\varepsilon_1}} \right\} \\ &= \text{Tr} \{ A(I - \varepsilon_1 \log B) \} + O(\varepsilon_1^2) \\ &= \text{Tr} \{ A \} - \varepsilon_1 \text{Tr} \{ A \log B \} + O(\varepsilon_1^2). \end{aligned}$$

It then follows that

$$\left. \frac{\partial}{\partial \varepsilon_1} f(\varepsilon_1, 0) \right|_{\varepsilon_1=0} = -\text{Tr}\{A \log B\}.$$

Assuming that the support of A is contained in the support of B , we then calculate $f(0, \varepsilon_2)$ as

$$\begin{aligned} f(0, \varepsilon_2) &= \text{Tr}\{A^{1+\varepsilon_2}\} \\ &= \text{Tr}\{A\} + \varepsilon_2 \text{Tr}\{A \log A\} + O(\varepsilon_2^2). \end{aligned}$$

It then follows that

$$\left. \frac{\partial}{\partial \varepsilon_2} f(0, \varepsilon_2) \right|_{\varepsilon_2=0} = \text{Tr}\{A \log A\}.$$

Putting these together, we find that

$$\left. \frac{\partial}{\partial \varepsilon} f(\varepsilon, \varepsilon) \right|_{\varepsilon=0} = \text{Tr}\{A \log A\} - \text{Tr}\{A \log B\} = D(A\|B).$$

■

5.1 Holevo-like quantity from the sandwiched Rényi relative entropy

This section establishes a relation between $\tilde{\chi}_\alpha(\mathcal{N})$ and an α -information radius quantity, defined below. The development here gives an improvement to Lemma I.3 in [41], such that we establish an equality rather than two inequalities, as seen by comparing our Lemma 14 to Lemma I.3 in [41].

Definition 12 (α -Holevo information) *By following the recipe given in (10), we define the α -Holevo information of a channel \mathcal{N} as follows:*

$$\tilde{\chi}_\alpha(\mathcal{N}) \equiv \max_{\{p_X(x), \rho_x\}} \tilde{\chi}_\alpha(\{p_X(x), \mathcal{N}(\rho_x)\}),$$

where

$$\begin{aligned} \tilde{\chi}_\alpha(\{p_X(x), \rho_x\}) &\equiv \min_{\sigma_Q} \tilde{D}_\alpha(\rho_{XQ} \| \rho_X \otimes \sigma_Q), \\ \rho_{XQ} &\equiv \sum_x p_X(x) |x\rangle\langle x|_X \otimes (\rho_x)_Q. \end{aligned}$$

By exploiting the above definition and Definition 4, it follows that

$$\tilde{\chi}_\alpha(\{p_X(x), \rho_x\}) = \min_{\sigma} \frac{1}{\alpha - 1} \log \left[\sum_x p_X(x) \text{Tr} \left\{ \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho_x \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\} \right]. \quad (14)$$

Definition 13 (α -information radius) *The α -information radius of a channel \mathcal{N} [49, 15, 63] is defined as*

$$\tilde{K}_\alpha(\mathcal{N}) \equiv \min_{\sigma} \max_{\rho} \tilde{D}_\alpha(\mathcal{N}(\rho) \| \sigma).$$

The reason that quantities like $\tilde{K}_\alpha(\mathcal{N})$ are often referred to as an “information radius” is that if we think of \tilde{D}_α as a distance measure (even though it is only a pseudo-distance), then it quantifies the “radius” of the possible channel outputs $\mathcal{N}(\rho)$ with respect to the distance measure \tilde{D}_α .

The following lemma is very helpful in analyzing whether $\tilde{\chi}_\alpha(\mathcal{N})$ is additive for a given channel:

Lemma 14 *The α -Holevo information $\tilde{\chi}_\alpha(\mathcal{N})$ is equal to the α -information radius $\tilde{K}_\alpha(\mathcal{N})$ for $\alpha \in (1, 2]$:*

$$\tilde{\chi}_\alpha(\mathcal{N}) = \tilde{K}_\alpha(\mathcal{N}).$$

Proof. We first prove the inequality $\tilde{K}_\alpha(\mathcal{N}) \leq \tilde{\chi}_\alpha(\mathcal{N})$. Recalling the definition of \tilde{Q}_α from Definition 4, consider that

$$\begin{aligned} \tilde{K}_\alpha(\mathcal{N}) &= \min_{\sigma} \max_{\rho} \tilde{D}_\alpha(\mathcal{N}(\rho) \parallel \sigma) \\ &= \min_{\sigma} \max_{\rho} \frac{1}{\alpha - 1} \log \tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma) \\ &= \frac{1}{\alpha - 1} \log \min_{\sigma} \max_{\rho} \tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma) \end{aligned}$$

So now we focus on the \tilde{Q}_α quantity and find that

$$\begin{aligned} \min_{\sigma} \max_{\rho} \tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma) &\leq \min_{\sigma} \sup_{\mu} \int d\mu(\rho) \tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma) \\ &= \sup_{\mu} \min_{\sigma} \int d\mu(\rho) \tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma) \\ &= \max_{\{p_X(x), \rho_x\}} \min_{\sigma} \sum_x p_X(x) \tilde{Q}_\alpha(\mathcal{N}(\rho_x) \parallel \sigma) \\ &= \max_{\{p_X(x), \rho_x\}} \min_{\sigma_B} \tilde{Q}_\alpha(\rho_{XB} \parallel \rho_X \otimes \sigma_B) \end{aligned} \tag{15}$$

The first inequality follows by taking a supremum over all probability measures μ on the set of all states ρ . The first equality is a result of applying the Sion minimax theorem [65]—we can do so because the function $\int d\mu(\rho) \tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma)$ is linear in the probability measure μ and convex in states σ . Convexity of $\tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma)$ in σ follows because

$$\tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma) = \text{Tr} \left\{ \left([\mathcal{N}(\rho)]^{1/2} \sigma^{(1-\alpha)/\alpha} [\mathcal{N}(\rho)]^{1/2} \right)^\alpha \right\},$$

$x^{(1-\alpha)/\alpha}$ is operator convex for $\alpha \in (1, 2]$ and x^α is operator convex for $\alpha \in (1, 2]$. The second equality follows by an application of the Fenchel-Eggleston-Caratheodory theorem (see [21], for example): the function $\tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma)$ is continuous in ρ , which is a density operator acting on a d -dimensional Hilbert space, so that to each μ , there exists a probability distribution $p_X(x)$ on no more than d^2 letters such that

$$\int d\mu(\rho) \tilde{Q}_\alpha(\mathcal{N}(\rho) \parallel \sigma) = \sum_x p_X(x) \tilde{Q}_\alpha(\mathcal{N}(\rho_x) \parallel \sigma).$$

The last equality in (15) follows from the properties of \tilde{Q}_α and by defining

$$\rho_{XB} \equiv \sum_x p_X(x) |x\rangle\langle x|_X \otimes [\mathcal{N}(\rho_x)]_B.$$

So we can then conclude that

$$\begin{aligned}
\tilde{K}_\alpha(\mathcal{N}) &\leq \frac{1}{\alpha-1} \log \max_{\{p_X(x), \rho_x\}} \min_{\sigma_B} \tilde{Q}_\alpha(\rho_{XB} \| \rho_X \otimes \sigma_B) \\
&= \max_{\{p_X(x), \rho_x\}} \min_{\sigma_B} \frac{1}{\alpha-1} \log \tilde{Q}_\alpha(\rho_{XB} \| \rho_X \otimes \sigma_B) \\
&= \tilde{\chi}_\alpha(\mathcal{N}).
\end{aligned}$$

The proof of the other inequality $\tilde{K}_\alpha(\mathcal{N}) \geq \tilde{\chi}_\alpha(\mathcal{N})$ is simpler. Consider that

$$\begin{aligned}
\tilde{\chi}_\alpha(\mathcal{N}) &= \max_{\{p_X(x), \rho_x\}} \min_{\sigma} \tilde{D}_\alpha(\rho_{XB} \| \rho_X \otimes \sigma) \\
&\leq \max_{\{p_X(x), \rho_x\}} \tilde{D}_\alpha(\rho_{XB} \| \rho_X \otimes \sigma) \\
&\leq \max_x \tilde{D}_\alpha(|x\rangle\langle x| \otimes \mathcal{N}(\rho_x) \| |x\rangle\langle x| \otimes \sigma) \\
&= \max_x \tilde{D}_\alpha(\mathcal{N}(\rho_x) \| \sigma) \\
&\leq \max_{\rho} \tilde{D}_\alpha(\mathcal{N}(\rho) \| \sigma).
\end{aligned}$$

The second inequality follows from joint quasi-convexity of \tilde{D}_α (Lemma 9). Since the above inequality holds for all states σ , we can conclude that $\tilde{K}_\alpha(\mathcal{N}) \geq \tilde{\chi}_\alpha(\mathcal{N})$. (This last realization is what allows for the improvement over Lemma I.3 in [41].) ■

Remark 15 *The above proof unchanged demonstrates that*

$$\chi_\alpha(\mathcal{N}) = K_\alpha(\mathcal{N}),$$

where these quantities are defined in the same way as $\tilde{\chi}_\alpha(\mathcal{N})$ and $\tilde{K}_\alpha(\mathcal{N})$, except through the traditional Rényi relative entropy defined in (12).

5.2 The sandwiched Rényi relative entropy is induced by a norm

We define the sandwiched α -norm of an operator A by

$$\|A\|_{\alpha, X} \equiv \left\| X^{1/2} A X^{1/2} \right\|_{\alpha},$$

for any positive operator X and where

$$\|B\|_{\alpha} \equiv \text{Tr}\{(\sqrt{B^\dagger B})^\alpha\}^{1/\alpha}.$$

With these definitions and that in (4), it is easy to see that for $\alpha > 1$

$$\begin{aligned}
\tilde{K}_\alpha(\mathcal{N}) &\equiv \min_{\sigma} \max_{\rho} \frac{\alpha}{\alpha-1} \log \|\mathcal{N}(\rho)\|_{\alpha, \sigma^{\frac{1-\alpha}{\alpha}}} \\
&= \min_{\sigma} \frac{\alpha}{\alpha-1} \log \max_{\rho} \|\mathcal{N}(\rho)\|_{\alpha, \sigma^{\frac{1-\alpha}{\alpha}}}.
\end{aligned} \tag{16}$$

This reformulation in terms of the sandwiched α -norm will make it easier to see that $\tilde{\chi}_\alpha$ is subadditive for the class of entanglement-breaking channels.

6 Bounding the success probability with the sandwiched relative Rényi entropy

Combining the results of the previous two sections (i.e., the bound in (11) and the fact that the sandwiched Rényi relative entropy is a generalized divergence), we find the following bound on the success probability for any rate R scheme for classical communication over n uses of a quantum channel \mathcal{N} :

$$p_{\text{succ}} \leq 2^{-n \left(\frac{\alpha-1}{\alpha} \right) \left(R - \frac{1}{n} \tilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) \right)}, \quad (17)$$

for all $\alpha \in (1, 2]$. Indeed, since the divergence \tilde{D}_α satisfies all of the requirements from Section 4, we find the following bound

$$\tilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) \geq \tilde{\delta}_\alpha(\varepsilon \| 1 - 2^{-nR}),$$

where $\tilde{\delta}_\alpha$ is the classical divergence induced from \tilde{D}_α . Since the following inequality holds for $\alpha > 1$

$$\begin{aligned} \tilde{\delta}_\alpha(\varepsilon \| 1 - 2^{-nR}) &= \frac{1}{\alpha - 1} \log \left(\varepsilon^\alpha (1 - 2^{-nR})^{1-\alpha} + (1 - \varepsilon)^\alpha (2^{-nR})^{1-\alpha} \right) \\ &\geq \frac{1}{\alpha - 1} \log \left((1 - \varepsilon)^\alpha (2^{-nR})^{1-\alpha} \right) \\ &= \frac{\alpha}{\alpha - 1} \log(1 - \varepsilon) + nR, \end{aligned}$$

we arrive at (17). Thus, we have now reduced the proof of the strong converse to the subadditivity of the quantity $\tilde{\chi}_\alpha(\mathcal{N}^{\otimes n})$.

7 Subadditivity of the α -information radius for entanglement-breaking channels

The main result of this section is that $\tilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) \leq n \tilde{\chi}_\alpha(\mathcal{N})$ whenever \mathcal{N} is an entanglement-breaking channel. We start by recalling a definition and a theorem:

Definition 16 *The maximum output α -norm of a completely positive map \mathcal{M} is defined as*

$$\nu_\alpha(\mathcal{M}) \equiv \max_{\rho} \|\mathcal{M}(\rho)\|_\alpha.$$

Theorem 17 ([39, 32]) *The maximum output α -norm is multiplicative for a completely-positive entanglement-breaking map \mathcal{M}_{EB} and an arbitrary completely positive map \mathcal{M} for all $\alpha \geq 1$:*

$$\nu_\alpha(\mathcal{M}_{\text{EB}} \otimes \mathcal{M}) = \nu_\alpha(\mathcal{M}_{\text{EB}}) \nu_\alpha(\mathcal{M}).$$

The following subadditivity relation then results from the above theorem:

Theorem 18 *For an entanglement-breaking channel \mathcal{N}_{EB} and any other channel \mathcal{N} and for all $\alpha \in (1, 2]$, the following subadditivity relation holds*

$$\tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}} \otimes \mathcal{N}) \leq \tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}}) + \tilde{\chi}_\alpha(\mathcal{N}).$$

Proof. We proceed with just a few steps:

$$\begin{aligned}
\tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}} \otimes \mathcal{N}) &= \tilde{K}_\alpha(\mathcal{N}_{\text{EB}} \otimes \mathcal{N}) \\
&= \min_{\sigma_{B_1 B_2}} \frac{\alpha}{\alpha - 1} \log \max_{\rho_{A_1 A_2}} \|(\mathcal{N}_{\text{EB}} \otimes \mathcal{N})(\rho_{A_1 A_2})\|_{\alpha, \sigma_{B_1 B_2}^{(1-\alpha)/\alpha}} \\
&\leq \min_{\sigma_{B_1} \otimes \sigma_{B_2}} \frac{\alpha}{\alpha - 1} \log \max_{\rho_{A_1 A_2}} \|(\mathcal{N}_{\text{EB}} \otimes \mathcal{N})(\rho_{A_1 A_2})\|_{\alpha, \sigma_{B_1}^{(1-\alpha)/\alpha} \otimes \sigma_{B_2}^{(1-\alpha)/\alpha}} \\
&\leq \min_{\sigma_{B_1} \otimes \sigma_{B_2}} \frac{\alpha}{\alpha - 1} \log \left[\max_{\rho_{A_1}} \|\mathcal{N}_{\text{EB}}(\rho_{A_1})\|_{\alpha, \sigma_{B_1}^{(1-\alpha)/\alpha}} \max_{\rho_{A_2}} \|\mathcal{N}(\rho_{A_2})\|_{\alpha, \sigma_{B_2}^{(1-\alpha)/\alpha}} \right] \\
&= \min_{\sigma_{B_1} \otimes \sigma_{B_2}} \frac{\alpha}{\alpha - 1} \left[\log \max_{\rho_{A_1}} \|\mathcal{N}_{\text{EB}}(\rho_{A_1})\|_{\alpha, \sigma_{B_1}^{(1-\alpha)/\alpha}} + \log \max_{\rho_{A_2}} \|\mathcal{N}(\rho_{A_2})\|_{\alpha, \sigma_{B_2}^{(1-\alpha)/\alpha}} \right] \\
&= \tilde{K}_\alpha(\mathcal{N}_{\text{EB}}) + \tilde{K}_\alpha(\mathcal{N}) \\
&= \tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}}) + \tilde{\chi}_\alpha(\mathcal{N}).
\end{aligned}$$

The first equality follows from Lemma 14. The second equality follows from the observation in (16). The first inequality follows by minimizing over tensor-product states rather than general states. The second inequality follows from the observation in Remark 1 (that an entanglement-breaking map conjugated by a positive operator $\sigma_{B_1}^{(1-\alpha)/2\alpha}$ is still an entanglement-breaking map) and from Theorem 17. The last few equalities follow by applying the logarithm and from definitions. ■

The above subadditivity relation and an inductive argument are sufficient for us to conclude the following corollary:

Corollary 19 *For an entanglement-breaking channel \mathcal{N}_{EB} , for all $\alpha \in (1, 2]$, and for any positive integer n , we have the following subadditivity relation:*

$$\tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}}^{\otimes n}) \leq n \tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}}).$$

8 Final steps for the strong converse for entanglement-breaking channels

Returning to (17), the subadditivity relation from Corollary 19 allows us to conclude the following upper bound on the success probability when communicating over an entanglement-breaking channel \mathcal{N}_{EB} :

$$p_{\text{succ}} \leq 2^{-n \left(\frac{\alpha-1}{\alpha} \right) (R - \tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}}))}. \quad (18)$$

It follows by a standard argument [52, 41] that if $R > \chi(\mathcal{N}_{\text{EB}})$, then the success probability decreases exponentially fast in n to zero. That is, we can analyze the derivative of $K_\alpha(\mathcal{N}_{\text{EB}})$ with respect to α and as $\alpha \rightarrow 1$, $K_\alpha(\mathcal{N}_{\text{EB}})$ approaches $\min_\sigma \max_\rho D(\mathcal{N}_{\text{EB}}(\rho) \parallel \sigma)$ which we know is equal to $\chi(\mathcal{N}_{\text{EB}})$ [53, 60]. If $R > \chi(\mathcal{N}_{\text{EB}})$, one can always find an α close enough to one such that the exponent

$$\left(\frac{\alpha - 1}{\alpha} \right) (R - \tilde{\chi}_\alpha(\mathcal{N}_{\text{EB}})) > 0.$$

One could then take a supremum over all $\alpha \in (1, 2]$ to optimize the exponent. We point the reader to Section 6 of [26] for additional details of this standard argument. From this line of reasoning, we can conclude the strong converse for entanglement-breaking channels.

However, we can also prove this result with a different approach. The resulting bound still gives an exponential decay of the success probability, but the approach above gives a stronger decay since it includes an optimization over the Rényi parameter α . Consider the following inequality from Lemma 6.3 of Ref. [67]:

$$D_\alpha(\rho\|\sigma) \leq D(\rho\|\sigma) + 4(\alpha - 1)(\log \nu)^2, \quad (19)$$

where

$$\begin{aligned} 1 < \alpha < 1 + \frac{\log 3}{4 \log \nu}, \\ \nu &= 2^{\frac{1}{2}D_{3/2}(\rho\|\sigma)} + 2^{-\frac{1}{2}D_{1/2}(\rho\|\sigma)} + 1. \end{aligned} \quad (20)$$

Combining the inequality above and in (13), we find that

$$\tilde{D}_\alpha(\rho\|\sigma) \leq D(\rho\|\sigma) + 4(\alpha - 1)(\log \nu)^2. \quad (21)$$

We can use this bound to deduce the strong converse.

Consider the information radius [53, 60]:

$$\min_\sigma \max_\rho D(\mathcal{N}_{\text{EB}}(\rho)\|\sigma) = \chi(\mathcal{N}_{\text{EB}}).$$

We know that there is an optimal value of σ for the above quantity, and let us call it σ^* . Furthermore, we know that

$$\max_\rho D(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*)$$

is a finite number (because it is equal to $\chi(\mathcal{N}_{\text{EB}})$). Thus, the support of $\mathcal{N}_{\text{EB}}(\rho)$ is contained in the support of σ^* for all ρ —otherwise, there would be some ρ that could make the above quantity infinite. So using (19), we have the following inequality holding for all ρ :

$$\tilde{D}_\alpha(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*) \leq D(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*) + 4(\alpha - 1)(\log \nu)^2, \quad (22)$$

where

$$\nu = 2^{\frac{1}{2}D_{3/2}(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*)} + 2^{-\frac{1}{2}D_{1/2}(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*)} + 1.$$

Since

$$2^{-\frac{1}{2}D_{1/2}(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*)} = \text{Tr}\left\{\sqrt{\mathcal{N}_{\text{EB}}(\rho)}\sqrt{\sigma^*}\right\} \leq 1,$$

it follows that

$$\nu \leq 2^{\frac{1}{2}D_{3/2}(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*)} + 2.$$

Also, since the support of $\mathcal{N}_{\text{EB}}(\rho)$ is contained in the support of σ^* for all ρ , it follows that $D_{3/2}(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*) < \infty$, so that

$$\nu \leq c(\mathcal{N}_{\text{EB}}) < \infty,$$

where $c(\mathcal{N}_{\text{EB}})$ is some constant that depends on the channel \mathcal{N}_{EB} (we can pick it to be independent of ρ as well). Combining with (22), we find that

$$\max_\rho \tilde{D}_\alpha(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*) \leq \max_\rho D(\mathcal{N}_{\text{EB}}(\rho)\|\sigma^*) + 4(\alpha - 1)(\log c(\mathcal{N}_{\text{EB}}))^2.$$

Taking one more minimization and recalling the choice of σ^* finally gives that

$$\min_{\sigma} \max_{\rho} \tilde{D}_{\alpha}(\mathcal{N}_{\text{EB}}(\rho) \parallel \sigma) \leq \min_{\sigma} \max_{\rho} D(\mathcal{N}_{\text{EB}}(\rho) \parallel \sigma) + 4(\alpha - 1)(\log c(\mathcal{N}_{\text{EB}}))^2,$$

which is equivalent to

$$\tilde{K}_{\alpha}(\mathcal{N}_{\text{EB}}) \leq \chi(\mathcal{N}_{\text{EB}}) + 4(\alpha - 1)(\log c(\mathcal{N}_{\text{EB}}))^2. \quad (23)$$

Finally, assume that $R > \chi(\mathcal{N}_{\text{EB}})$. We choose α as follows:

$$\alpha = 1 + \min \left\{ \frac{\log 3}{4 \log c(\mathcal{N}_{\text{EB}})}, \frac{R - \chi(\mathcal{N}_{\text{EB}})}{8(\log c(\mathcal{N}_{\text{EB}}))^2}, 1 \right\},$$

so that the following inequality holds

$$\chi(\mathcal{N}_{\text{EB}}) + (\alpha - 1)(\log c(\mathcal{N}_{\text{EB}}))^2 \leq \frac{1}{2}(R + \chi(\mathcal{N}_{\text{EB}})).$$

(Furthermore, it is reasonable for us to assume that R is close enough to $\chi(\mathcal{N}_{\text{EB}})$ so that α is actually equal to $1 + [R - \chi(\mathcal{N}_{\text{EB}})]/8(\log c(\mathcal{N}_{\text{EB}}))^2$.) Using the bounds in (18) and (23), we then obtain the following bound on the success probability for any classical communication protocol over an entanglement-breaking channel:

$$\begin{aligned} p_{\text{succ}} &\leq 2^{-n \left(\frac{\alpha-1}{\alpha} \right) (R - \tilde{\chi}_{\alpha}(\mathcal{N}_{\text{EB}}))} \\ &= 2^{-n \left(\frac{\alpha-1}{\alpha} \right) [R - \tilde{K}_{\alpha}(\mathcal{N}_{\text{EB}})]} \\ &\leq 2^{-n \left(\frac{\alpha-1}{\alpha} \right) [R - [\chi(\mathcal{N}_{\text{EB}}) + 4(\alpha-1)(\log c(\mathcal{N}_{\text{EB}}))^2]]} \\ &\leq 2^{-n \left(\frac{\alpha-1}{2} \right) [R - [\frac{1}{2}(R + \chi(\mathcal{N}_{\text{EB}}))]]} \\ &= 2^{-n \left(\frac{\alpha-1}{4} \right) [R - \chi(\mathcal{N}_{\text{EB}})]} \\ &\leq 2^{-n [R - \chi(\mathcal{N}_{\text{EB}})]^2 / 32 (\log c(\mathcal{N}_{\text{EB}}))^2}. \end{aligned} \quad (24)$$

Thus, in the case that $R > \chi(\mathcal{N}_{\text{EB}})$, the success probability converges exponentially fast to zero.

One might be concerned about our restriction to rates near $\chi(\mathcal{N}_{\text{EB}})$, but it is also easy to see that choosing $\alpha = 1 + \frac{1}{\sqrt{n}}$ recovers the bound

$$p_{\text{succ}} \leq 2^{-\sqrt{n} \left(\frac{1}{1+1/\sqrt{n}} \right) [R - [\chi(\mathcal{N}_{\text{EB}}) + \frac{4}{\sqrt{n}} (\log c(\mathcal{N}_{\text{EB}}))^2]]},$$

which decays to zero exponentially fast in \sqrt{n} for any rate $R > \chi(\mathcal{N}_{\text{EB}})$.

8.1 Prior results on particular covariant channels follow as a special case

We remark briefly on how the prior results in Ref. [41] follow as a special case of our approach. There, Koenig and Wehner showed that the strong converse theorem holds for all covariant channels with an additive minimum output Rényi entropy. For these channels, they proved that

$$\chi_{\alpha}(\mathcal{N}^{\otimes n}) = n[\log d - H_{\alpha}^{\min}(\mathcal{N})],$$

where the minimum output Rényi entropy of a channel is defined as

$$H_\alpha^{\min}(\mathcal{N}) \equiv \min_{\rho} H_\alpha(\mathcal{N}(\rho)),$$

$$H_\alpha(\sigma) \equiv \frac{1}{1-\alpha} \log \text{Tr}\{\sigma^\alpha\}.$$

By following a development similar to that in the previous section, the strong converse for these channels follows.

To recover their result, we can modify the proof of Theorem 18 as follows:

$$\begin{aligned} \tilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) &= \tilde{K}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) \\ &= \min_{\sigma_{B_1 B_2}} \frac{\alpha}{\alpha-1} \log \max_{\rho_{A_1 A_2}} \|(\mathcal{N}_1 \otimes \mathcal{N}_2)(\rho_{A_1 A_2})\|_{\alpha, \sigma_{B_1 B_2}^{(1-\alpha)/\alpha}} \\ &\leq \frac{\alpha}{\alpha-1} \log \max_{\rho_{A_1 A_2}} \|(\mathcal{N}_1 \otimes \mathcal{N}_2)(\rho_{A_1 A_2})\|_{\alpha, \pi_{B_1}^{(1-\alpha)/\alpha} \otimes \pi_{B_2}^{(1-\alpha)/\alpha}} \\ &= \log d_1 + \log d_2 - H_\alpha^{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2), \end{aligned}$$

where we denote the maximally mixed state by π . The inequality follows simply by making the suboptimal choice of setting $\sigma_{B_1 B_2}$ to be the maximally mixed state. Thus, if $H_\alpha^{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2) = H_\alpha^{\min}(\mathcal{N}_1) + H_\alpha^{\min}(\mathcal{N}_2)$ for some particular channels \mathcal{N}_1 and \mathcal{N}_2 , we can then conclude additivity of $\tilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2)$. All the classes of channels considered by Koenig and Wehner have the property that the minimum output entropy of the channel and any other channel is additive. Thus, one can conclude additivity of $H_\alpha^{\min}(\mathcal{N}^{\otimes n})$ by an inductive argument that is the same as what we used in Corollary 19. The rest of the proof follows easily after establishing subadditivity of $\tilde{\chi}_\alpha(\mathcal{N}^{\otimes n})$.

The above development in fact shows that we obtain a strong converse rate of $\log d - H^{\min}(\mathcal{N})$ for any channel for which its minimum output Rényi entropy is additive for all $\alpha \geq 1$. (In the above, $H^{\min}(\mathcal{N})$ denotes the minimum output von Neumann entropy of the channel.)

9 Strong converse for the classical capacity of Hadamard channels

We now prove that the strong converse holds for the classical capacity of Hadamard channels. This result follows from the following theorem, along with some additional arguments:

Theorem 20 ([40, 32]) *If the maximum output α -norm is multiplicative for one pair of completely positive maps \mathcal{M}_1 and \mathcal{M}_2 :*

$$\nu_\alpha(\mathcal{M}_1 \otimes \mathcal{M}_2) = \nu_\alpha(\mathcal{M}_1) \nu_\alpha(\mathcal{M}_2),$$

then the same is true for their respective complementary maps \mathcal{M}_1^c and \mathcal{M}_2^c :

$$\nu_\alpha(\mathcal{M}_1^c \otimes \mathcal{M}_2^c) = \nu_\alpha(\mathcal{M}_1^c) \nu_\alpha(\mathcal{M}_2^c).$$

Definition 21 *Given a given channel \mathcal{N} and a state σ on the output space of \mathcal{N} , let $\tilde{K}_\alpha^{[\sigma]}(\mathcal{N})$ denote the α -information radius of the channel around σ :*

$$\tilde{K}_\alpha^{[\sigma]}(\mathcal{N}) \equiv \max_{\rho} \tilde{D}_\alpha(\mathcal{N}(\rho) \parallel \sigma). \quad (25)$$

Note that by definition, $\tilde{K}_\alpha(\mathcal{N}) = \min_\sigma \tilde{K}_\alpha^{[\sigma]}(\mathcal{N})$.

By a similar development as in Section 6, we find that the following inequality holds for any code of rate R with success probability $1 - \varepsilon$ that uses the channel n times:

$$\begin{aligned} \frac{1}{\alpha - 1} \log \left((1 - \varepsilon)^\alpha (2^{-nR})^{1-\alpha} \right) &\leq \tilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) \\ &= \tilde{K}_\alpha(\mathcal{N}^{\otimes n}) \\ &\leq \tilde{K}_\alpha^{[\sigma^{\otimes n}]}(\mathcal{N}^{\otimes n}). \end{aligned}$$

where σ is an arbitrary state on the output system of a single channel. We now choose σ as the optimal state in the Schumacher-Westmoreland characterization of $\chi(\mathcal{N})$ [60]:

$$\chi(\mathcal{N}) = \min_\sigma \max_\rho D(\mathcal{N}(\rho) \parallel \sigma).$$

For this, note also the previously used fact

$$\tilde{K}_\alpha^{[\sigma]}(\mathcal{N}) \leq \chi(\mathcal{N}) + 4(\alpha - 1)(\log \nu)^2.$$

Thus, we find the following bound on the success probability:

$$p_{\text{succ}} = 1 - \varepsilon \leq 2^{-n \left(\frac{\alpha-1}{\alpha} \right) \left(R - \frac{1}{n} \tilde{K}_\alpha^{[\sigma^{\otimes n}]}(\mathcal{N}^{\otimes n}) \right)}. \quad (26)$$

The crucial observation, which in fact we also used to prove the strong converse for entanglement-breaking channels, is that

$$\tilde{K}_\alpha^{[\sigma]}(\mathcal{N}) = \max_\rho \frac{1}{\alpha - 1} \log \text{Tr} \left\{ \left(\sigma^{\frac{1-\alpha}{2\alpha}} \mathcal{N}(\rho) \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right\},$$

which is $\frac{\alpha}{\alpha-1}$ times the logarithm of the maximum output α -norm of the sandwiched map

$$(\mathcal{X} \circ \mathcal{N})(\rho) \equiv X \mathcal{N}(\rho) X,$$

with $X = \sigma^{\frac{1-\alpha}{2\alpha}}$.

Now, we first prove that the strong converse holds for a Hadamard channel \mathcal{N}_H whose complementary channel \mathcal{N}_H^c is in the interior of the set of entanglement breaking channels.⁵ In such a case, $X = \sigma^{\frac{1-\alpha}{2\alpha}}$ becomes arbitrarily close to the identity operator I for α sufficiently close to one. (Without loss of generality, we can assume that σ has full rank—otherwise either $\tilde{K}_\alpha^{[\sigma]}(\mathcal{N}_H) = +\infty$, or we can reduce the size of the output system without affecting the performance of a given code.) But then, the complementary map $(\mathcal{X} \circ \mathcal{N}_H)^c$ is arbitrarily close to \mathcal{N}_H^c , and hence (always for sufficiently small $\alpha > 1$) it is arbitrarily close to a completely positive entanglement-breaking map. So it follows that $(\mathcal{X} \circ \mathcal{N}_H)$ is a Hadamard map for α sufficiently close to one, and Theorem 20 implies that its maximum output α -norm is multiplicative, so that the α -information radius around σ is subadditive:

$$\frac{1}{n} \tilde{K}_\alpha^{[\sigma^{\otimes n}]}(\mathcal{N}_H^{\otimes n}) \leq \tilde{K}_\alpha^{[\sigma]}(\mathcal{N}_H).$$

⁵Such channels have the property that their Choi matrix is in the interior of the set of separable states. That the interior of the set of entanglement-breaking channels is non-empty then follows from [27].

Hence, from (26), we find the following upper bound on the success probability:

$$1 - \varepsilon \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)(R - \tilde{K}_\alpha^{[\sigma]}(\mathcal{N}_H))}.$$

By following the same steps as in Section 8 (always choosing α sufficiently close to one), the strong converse follows, with a bound on the success probability that converges exponentially fast to zero.

For a Hadamard channel \mathcal{N}_H whose complement \mathcal{N}_H^c is on the boundary of the set of entanglement-breaking channels, the argument above does not apply, since the perturbation inflicted by sandwiching with $X \approx I$ might take the complementary channel outside the set of entanglement-breaking maps. However, we can use the following continuity argument: For $p \geq 0$, consider the depolarizing channel on the environment system E :

$$\mathcal{D}_p(\rho) = (1 - p)\rho + p\frac{I}{|E|}\text{Tr}\rho,$$

with a suitable Stinespring isometry $W_p : E \rightarrow E \otimes F$, where $|F| = |E|^2$. Then, not only is $\mathcal{M}_p^c \equiv \mathcal{D}_p \circ \mathcal{N}_H^c$ entanglement-breaking, but it is in the interior of the set of entanglement-breaking channels whenever $p > 0$. Furthermore, in the limit as $p \rightarrow 0$, \mathcal{M}_p^c converges to $\mathcal{M}_0^c = \mathcal{N}_H^c$. Hence, a similar limiting argument applies for the map \mathcal{M}_p :

$$\mathcal{M}_p \rightarrow \mathcal{M}_0 = \mathcal{N}_H \otimes |0\rangle\langle 0|,$$

where \mathcal{M} maps A to $B \otimes F$, via $\mathcal{M}(\rho) = \text{Tr}_E\{W_p V \rho V^\dagger W_p^\dagger\}$. By the continuity of the Holevo information χ in the channel [44], we observe that $\chi(\mathcal{M}_p) \rightarrow \chi(\mathcal{N}_H)$.

Furthermore, $\mathcal{N}_H = \text{Tr}_F \circ \mathcal{M}_p$, so that every code for \mathcal{N}_H is immediately a code with the same rate and error parameters for \mathcal{M}_p . Now we can choose, for an n -block code of rate $R > \chi(\mathcal{N}_H)$ and error ε , a $p > 0$ such that $R > \chi(\mathcal{M}_p)$. At this point the strong converse follows for \mathcal{M}_p by the previous argument, and hence also for \mathcal{N}_H .

10 Conclusion

We have proven a strong converse theorem for the classical capacity of all entanglement-breaking and Hadamard channels, and these results strengthen the interpretation of the classical capacity for these channels. Our result follows by obtaining tighter bounds on the success probability in terms of a “sandwiched” Rényi relative entropy. This information measure should find other applications in quantum information theory, given that many other information measures can be obtained from a relative entropy.

We have left the superadditivity of $\tilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2)$ for two channels as an open question, but Beigi has recently provided a solution to this problem [3]. That is, Beigi has proved that the following inequality holds for any two channels:

$$\tilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \tilde{\chi}_\alpha(\mathcal{N}_1) + \tilde{\chi}_\alpha(\mathcal{N}_2).$$

Such an inequality for χ_α easily follows—one can employ the Sibson identity to find an explicit form for χ_α and then the inequality follows by simply choosing a suboptimal tensor product ensemble for $\chi_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2)$ (see Ref. [41]). However, it is not clear to us that a Sibson identity holds for $\tilde{D}_\alpha(\rho||\sigma)$ except for when the states ρ and σ are commuting. So the proof of the above inequality is more advanced than the usual approach.

Finally, it might be possible to use the tools developed in Refs. [47, 46] in order to prove strong converse theorems, but this remains an open question.

Note: After completing the work for the first version of this paper, we discovered that other authors had already defined [68, 22] and proved [50, 20] some of the properties of the sandwiched Rényi relative entropy. However, only the definition of the sandwiched Rényi relative entropy was publicly available at the time when we completed this work. These authors have posted details of their work, now published in Ref. [51].

Since our original arXiv post, there has been more activity in developing the sandwiched Rényi relative entropy. In particular, Müller-Lennert *et al.* have been able to prove many of their conjectures concerning this quantity in a second version of their paper, while Frank and Lieb have proved that it is monotone under quantum operations for all $\alpha \in [1/2, \infty]$ [23]. Simultaneously, Beigi provided a different proof that it is monotone for all $\alpha \in (1, \infty)$ [3].

Acknowledgements—We are grateful to Min-Hsiu Hsieh, Joe Renes, and Graeme Smith for helpful discussions and to Marco Tomamichel for carefully reading our paper and pointing out a correction to a previous version of Proposition 11. We thank Frédéric Dupuis, Marco Tomamichel, and Serge Fehr for passing along Refs. [68, 22, 50, 20]. MMW is grateful to the quantum information theory group at the Universitat Autònoma de Barcelona for hosting him for a research visit during April-May 2013. AW’s work is supported by the European Commission (STREP “QCS”), the European Research Council (Advanced Grant “IRQUAT”) and the Philip Leverhulme Trust. DY’s work is supported by the ERC (Advanced Grant “IRQUAT”) and the NSFC (Grant No. 11375165).

A Appendix

We reproduce here, for convenience of the reader, the statements of Theorem 5.14, Corollary 5.5, and Theorem 5.16 from [74].

Theorem 22 (Theorem 5.16 [74]) *Let $g : \mathcal{D} \subseteq \mathcal{M}_{d_1} \times \cdots \times \mathcal{M}_{d_n} \rightarrow \mathcal{M}_d$ be a map on the direct product \mathcal{D} of n positive operators, and similarly $h : \mathcal{D}' \subseteq \mathcal{M}_{d'_1} \times \cdots \times \mathcal{M}_{d'_n} \rightarrow \mathcal{M}_d$. Suppose that g is jointly operator concave and positive and h is semi-definite. Let $I \ni 0$ be the positive/negative real half line depending on whether h is positive or negative semi-definite. For any function $f : I \rightarrow \mathbb{R}$ with $f(0) \leq 0$, define $F : \mathcal{D}' \times \mathcal{D} \rightarrow \mathcal{M}_d$ as*

$$F(L, R) \equiv \sqrt{g(R)} f\left(g(R)^{-1/2} h(L) g(R)^{-1/2}\right) \sqrt{g(R)}.$$

We consider joint operator convexity of F in its $n + m$ arguments. F is jointly operator convex on positive operators for which g is invertible if at least one of the following holds: 1) h is jointly operator concave and f is operator anti-monotone. 2) h is affine and f is operator convex.

Corollary 23 (Corollary 5.5 [74]) $\mathcal{M}_d \times \mathcal{M}_d \ni (L, R) \rightarrow L^x \otimes R^y$ *is jointly operator concave on positive operators for $x, y \geq 0$ with $x + y \leq 1$.*

Theorem 24 (Theorem 5.16 [74]) *Consider a functional $F : \mathcal{D} \subseteq \mathcal{M}_d \times \cdots \times \mathcal{M}_d \rightarrow \mathbb{R}$ which is defined for all dimensions $d \in \mathbb{N}$. Suppose that F satisfies 1) joint convexity in \mathcal{D} , 2) unitary invariance, i.e., for all $A \in \mathcal{D}$ and all unitaries $U \in \mathcal{M}_d(\mathbb{C})$, it holds that $F(UA_1U^\dagger, \dots, UA_nU^\dagger) = F(A_1, \dots, A_n)$, and 3) invariance under tensor products, meaning that for all $A \in \mathcal{D}$ and all density*

operators $\tau \in \mathcal{M}_d(\mathbb{C})$, we have $F(A_1 \otimes \tau, \dots, A_n \otimes \tau) = F(A_1, \dots, A_n)$. Then F is monotone with respect to all CPTP maps $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_{d'}(\mathbb{C})$, in the sense that for all $A \in \mathcal{D}$,

$$F(T(A_1), \dots, T(A_n)) \leq F(A_1, \dots, A_n).$$

References

- [1] Grigori G. Amosov, Alexander S. Holevo, and R. F. Werner. On some additivity problems in quantum information theory. *Problems of Information Transmission*, 36(4):25, 2000. arXiv:math-ph/0003002.
- [2] Suguru Arimoto. On the converse to the coding theorem for discrete memoryless channels. *IEEE Transactions on Information Theory*, 19:357–359, May 1973.
- [3] Salman Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics*, 54(12):122202, December 2013. arXiv:1306.5920.
- [4] 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.
- [5] Mario Berta, Joseph M. Renes, and Mark M. Wilde. Identifying the information gain of a quantum measurement. January 2013. arXiv:1301.1594.
- [6] Kamil Brádler. An infinite sequence of additive channels: the classical capacity of cloning channels. *IEEE Transactions on Information Theory*, 57(8):5497–5503, August 2011. arXiv:0903.1638.
- [7] Kamil Brádler, Nicolas Dutil, Patrick Hayden, and Abubakr Muhammad. Conjugate degradability and the quantum capacity of cloning channels. *Journal of Mathematical Physics*, 51(7):072201, 2010. arXiv:0909.3297.
- [8] Kamil Brádler, Patrick Hayden, and Prakash Panangaden. Private information via the Unruh effect. *Journal of High Energy Physics*, 2009(08):074, 2009. arXiv:0807.4536.
- [9] Frederico Brito, David P. DiVincenzo, Roger H. Koch, and Matthias Steffen. Efficient one- and two-qubit pulsed gates for an oscillator-stabilized Josephson qubit. *New Journal of Physics*, 10(3):033027 (33pp), 2008.
- [10] Francesco Buscemi, Masahito Hayashi, and Michał Horodecki. Global information balance in quantum measurements. *Physical Review Letters*, 100:210504, 2008. arXiv:quant-ph/0702166.
- [11] Paul Busch. Informationally complete sets of physical quantities. *International Journal of Theoretical Physics*, 30(9):1217–1227, September 1991.
- [12] Eric A. Carlen. Trace inequalities and quantum entropy: An introductory course. *Contemporary Mathematics*, 529:73–140, 2010.
- [13] Eric A. Carlen and Elliott H. Lieb. A Minkowski type trace inequality and strong subadditivity of the quantum entropy II. *Letters in Mathematical Physics*, 83(2):107–126, 2008. arXiv:0710.4167.

- [14] Giulio Chiribella. On quantum estimation, quantum cloning and finite quantum de finetti theorems. *Theory of Quantum Computation, Communication, and Cryptography, Lecture Notes in Computer Science*, 6519:9–25, 2011. arXiv:1010.1875.
- [15] Imre Csiszár. Generalized cutoff rates and Rényi’s information measures. *IEEE Transactions on Information Theory*, 41(1):26–34, January 1995.
- [16] Michele Dall’Arno, Giacomo Mauro D’Ariano, and Massimiliano F. Sacchi. Informational power of quantum measurements. *Physical Review A*, 83:062304, June 2011. arXiv:1103.1972.
- [17] Nilanjana Datta, Alexander S. Holevo, and Yuri Suhov. Additivity for transpose depolarizing channels. *International Journal of Quantum Information*, 4(1):85–98, 2006. arXiv:quant-ph/0412034.
- [18] Igor Devetak and Peter W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Communications in Mathematical Physics*, 256:287–303, 2005. arXiv:quant-ph/0311131.
- [19] Frédéric Dupuis, Omar Fawzi, and Stephanie Wehner. Entanglement sampling and applications. May 2013. arXiv:1305.1316.
- [20] Frédéric Dupuis, Oleg Szehr, and Marco Tomamichel. unpublished notes, 2013.
- [21] Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge University Press, January 2012.
- [22] Serge Fehr. On the conditional Rényi entropy. Lecture at the Beyond IID Workshop at the University of Cambridge, January 2013.
- [23] 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.
- [24] Motohisa Fukuda. Extending additivity from symmetric to asymmetric channels. *Journal of Physics A: Mathematical and General*, 38(45):L753–L758, 2005. arXiv:quant-ph/0505022.
- [25] Robert G. Gallager. *Information theory and reliable communication*. John Wiley and Sons, 1968.
- [26] 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.
- [27] Leonid Gurvits and Howard Barnum. Largest separable balls around the maximally mixed bipartite quantum state. *Physical Review A*, 66(6):062311, December 2002. arXiv:quant-ph/0204159.
- [28] Matthew B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5:255–257, April 2009. arXiv:0809.3972.
- [29] Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9:177–183, 1973.

- [30] Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, January 1998. arXiv:quant-ph/9611023.
- [31] Alexander S. Holevo. Quantum coding theorems. *Russian Mathematical Surveys*, 53:1295–1331, 1999.
- [32] Alexander S. Holevo. Multiplicativity of p-norms of completely positive maps and the additivity problem in quantum information theory. *Russian Mathematical Surveys*, 61(2):301–339, 2006.
- [33] Alexander S. Holevo. Information capacity of quantum observable. *Problems of Information Transmission*, 48:1, 2012. arXiv:1103.2615.
- [34] Michał Horodecki, Peter W. Shor, and Mary Beth Ruskai. Entanglement breaking channels. *Reviews in Mathematical Physics*, 15(6):629–641, 2003. arXiv:quant-ph/0302031.
- [35] Kurt Jacobs. On the properties of information gathering in quantum and classical measurements. 2003. arXiv:quant-ph/0304200v1.
- [36] Christopher King. Additivity for unital qubit channels. *Journal of Mathematical Physics*, 43(10):4641–4653, 2002. arXiv:quant-ph/0103156.
- [37] Christopher King. An application of the Lieb-Thirring inequality in quantum information theory. *Fourteenth International Congress on Mathematical Physics*, pages 486–490, 2003. arXiv:quant-ph/0412046.
- [38] Christopher King. The capacity of the quantum depolarizing channel. *IEEE Transactions on Information Theory*, 49(1):221–229, January 2003. arXiv:quant-ph/0204172.
- [39] Christopher King. Maximal p-norms of entanglement breaking channels. *Quantum Information and Computation*, 3(2):186–190, 2003. arXiv:quant-ph/0212057.
- [40] 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.
- [41] 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.
- [42] Robert Koenig, Stephanie Wehner, and Jürg Wullschleger. Unconditional security from noisy quantum storage. *IEEE Transactions on Information Theory*, 58(3):1962–1984, March 2012. arXiv:0906.1030.
- [43] Antía Lamas-Linares, Christoph Simon, John C. Howell, and Dik Bouwmeester. Experimental quantum cloning of single photons. *Science*, 296:712–714, 2002.
- [44] Debbie Leung and Graeme Smith. Continuity of quantum channel capacities. *Communications in Mathematical Physics*, 292(1):201–215, November 2009. arXiv:0810.4931.
- [45] Elliott H. Lieb and Walter Thirring. *Studies in mathematical physics*, chapter Inequalities for the moments of the eigenvalues of the Schroedinger Hamiltonian and their relation to Sobolev inequalities, pages 269–297. Princeton University Press, Princeton, 1976.

- [46] William Matthews. A linear program for the finite block length converse of polyanskiy-poorverdú via nonsignaling codes. *IEEE Transactions on Information Theory*, 58(12):7036–7044, December 2012. arXiv:1109.5417.
- [47] William Matthews and Stephanie Wehner. Finite blocklength converse bounds for quantum channels. October 2012. arXiv:1210.4722.
- [48] P. W. Milonni and M. L. Hardies. Photons cannot always be replicated. *Physics Letters A*, 92(7):321–322, November 1982.
- [49] Milán Mosonyi and Fumio Hiai. On the quantum Rényi relative entropies and related capacity formulas. *IEEE Transactions on Information Theory*, 57(4):2474–2487, April 2011. arXiv:0912.1286.
- [50] Martin Müller-Lennert. Quantum relative Rényi entropies. Master’s thesis, ETH Zurich, April 2013.
- [51] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: a new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, December 2013. arXiv:1306.3142.
- [52] Tomohiro Ogawa and Hiroshi Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Transactions on Information Theory*, 45(7):2486–2489, November 1999. arXiv:quant-ph/9808063.
- [53] Masanori Ohya, Dénes Petz, and Noboru Watanabe. On capacities of quantum channels. *Probability and Mathematical Statistics—Wrocław University*, 17:179–196, 1997.
- [54] Ognjan Oreshkov, John Calsamiglia, Ramon Muñoz-Tapia, and Emili Bagan. Optimal signal states for quantum detectors. *New Journal of Physics*, 13(7):073032, 2011. arXiv:1103.2365.
- [55] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports in Mathematical Physics*, 23:57–65, 1986.
- [56] 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.
- [57] Eduard Prugovečki. Information-theoretical aspects of quantum measurement. *International Journal of Theoretical Physics*, 16:321–331, May 1977.
- [58] Joseph M. Renes, Robin Blume-Kohout, A. J. Scott, and Carlton M. Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45:21712180, 2004. arXiv:quant-ph/0310075.
- [59] Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.
- [60] Benjamin Schumacher and Michael D. Westmoreland. Optimal signal ensembles. *Physical Review A*, 63:022308, January 2001.

- [61] Naresh Sharma and Naqeeb Ahmad Warsi. On the strong converses for the quantum channel capacity theorems. June 2012. arXiv:1205.1712.
- [62] Peter W. Shor. Additivity of the classical capacity of entanglement-breaking quantum channels. *Journal of Mathematical Physics*, 43(9):4334–4340, 2002. arXiv:quant-ph/0201149.
- [63] Robin Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(2):149160, 1969.
- [64] Christoph Simon, Gregor Weihs, and Anton Zeilinger. Optimal quantum cloning via stimulated emission. *Physical Review Letters*, 84(13):2993–2996, March 2000.
- [65] Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.
- [66] W. F. Stinespring. Positive functions on C^* -algebras. *Proceedings of the American Mathematical Society*, 6:211–216, 1955.
- [67] Marco Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory*. PhD thesis, ETH Zurich, 2012. arXiv:1203.2142.
- [68] Marco Tomamichel. Smooth entropies—a tutorial: With focus on applications in cryptography. Tutorial at QCRYPT 2012, slides available at <http://2012.qcrypt.net/docs/slides/Marco.pdf>, September 2012.
- [69] William G. Unruh. Notes on black-hole evaporation. *Physical Review D*, 14(4):870–892, August 1976.
- [70] Mark M. Wilde, Patrick Hayden, Francesco Buscemi, and Min-Hsiu Hsieh. The information-theoretic costs of simulating quantum measurements. *Journal of Physics A: Mathematical and Theoretical*, 45(45):453001, November 2012. arXiv:1206.4121.
- [71] Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999.
- [72] Andreas Winter. *Coding Theorems of Quantum Information Theory*. PhD thesis, Universität Bielefeld, July 1999. arXiv:quant-ph/9907077.
- [73] Andreas Winter. “Extrinsic” and “intrinsic” data in quantum measurements: Asymptotic convex decomposition of positive operator valued measures. *Communications in Mathematical Physics*, 244:157, 2004. arXiv:quant-ph/0109050.
- [74] Michael M. Wolf. Quantum channels & operations: Guided tour. Lecture notes available at <http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf>, July 2012.
- [75] Jacob Wolfowitz. The coding of messages subject to chance errors. *Illinois Journal of Mathematics*, 1:591–606, 1957.
- [76] Jacob Wolfowitz. *Coding Theorems of Information Theory*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1962.

- [77] 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.