This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

Bounding the forward classical capacity of bipartite quantum channels

Dawei Ding, Sumeet Khatri, Yihui Quek, Peter W. Shor, Xin Wang, Mark M. Wilde

***Abstract*—We introduce various measures of forward classical communication for bipartite quantum channels. Since a point-to-point channel is a special case of a bipartite channel, the measures reduce to measures of classical communication for point-to-point channels. As it turns out, these reduced measures have been reported in prior work of Wang *et al*. on bounding the classical capacity of a quantum channel. As applications, we show that the measures are upper bounds on the forward classical capacity of a bipartite channel. The reduced measures are upper bounds on the classical capacity of a point-to-point quantum channel assisted by a classical feedback channel. Some of the various measures can be computed by semi-definite programming.**

I. Introduction

The goal of quantum Shannon theory [1], [2], [3], [4] is to characterize the information-processing capabilities of quantum states and channels, for various tasks such as distillation of randomness, secret key, entanglement or communication of classical, private, and quantum information. With the goal of simplifying the theory or relating to practical communication scenarios, often assisting resources are allowed, such as shared entanglement [5], [6], [7] or public classical communication [8], [9], [10].

One of the earliest information-theoretic tasks studied in quantum Shannon theory is the capacity of a point-to-point quantum channel for transmitting classical information or generating shared randomness [11], [12], [13]. It is well known that the capacity for these two tasks is equal, and so we just refer to them both as the classical capacity of a quantum channel. A formal expression for the classical capacity of a quantum channel is known, given by what is called the regularized Holevo information of a quantum channel (see, e.g., [1], [2], [3], [4] for reviews). On the one hand, it is unclear whether this formal expression is computable for all quantum channels [14], and it is also known that the Holevo information formula is generally non-additive [15]. On the other hand, for some special classes of channels [16], [17], [18], the regularized Holevo information simplifies to what is known as the single-letter Holevo information. Even when this simplification happens, it is not necessarily guaranteed that the resulting capacity formula is efficiently computable [19].

This difficulty in calculating the classical capacity of a quantum channel has spurred the investigation of efficiently computable upper bounds on it [20], [21], [22]. The main idea driving these bounds [20] is to imagine a scenario in which the sender and receiver of a quantum channel could be supplemented by the help of a coding scheme that is simultaneously non-signaling and "positive-partial-transpose" (PPT) assisted [23], [20]. These latter coding scenarios could be considered fictitious from a physical perspective, but the perspective is actually extremely powerful when trying to provide an upper bound on the classical capacity, while being faced with the aforementioned difficulties. The reason is that the simultaneous constraints of being non-signaling and PPT-assisted are semi-definite constraints and ultimately lead to upper bounds that are efficiently computable by semi-definite programming.

Another thread that has been pursued on the topic of classical capacity, after the initial investigations of [11], [12], [13], is the classical capacity of a quantum channel assisted by a classical feedback channel. This direction started with [9], [10] and was ultimately inspired by Shannon's work on the feedback-assisted capacity of a classical channel [24], in which it was shown that feedback does not increase capacity.[1] For the quantum case, it is known that a classical feedback channel does not enhance the classical capacity of

1) an entanglement-breaking channel [10],
2) a pure-loss bosonic channel [25], and
3) a quantum erasure channel [25].

The first aforementioned result has been strengthened to a strong-converse statement [26]. However, due to the quantum effect of entanglement, it is also known that feedback can significantly increase the classical capacity of certain quantum channels [27]. More generally, [28] discussed several inequalities relating the classical capacity assisted by classical feedback to other capacities, and [29] established inequalities relating classical capacities assisted by classical communication to other notions of feedback-assisted capacities.

Dawei Ding is with Alibaba Quantum Laboratory, Alibaba Group, Bellevue, Washington, USA. Mark M. Wilde is with the Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, and Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA, and Sumeet Khatri had the same affiliation when this research was conducted. Yihui Quek was with the Information Systems Laboratory, Stanford University, Stanford, California 94305, USA, when this research was conducted. Sumeet Khatri and Yihui Quek are now with the Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin, Germany. Peter Shor is with the Center for Theoretical Physics and Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. Xin Wang is with the Institute for Quantum Computing, Baidu Research, Beijing 100193, China. Mark M. Wilde was also affiliated with the Stanford Institute for Theoretical Physics, Stanford University, Stanford, California 94305, USA, when this research was conducted. This work was presented in part at the 2021 IEEE International Symposium on Information Theory, Melbourne, Victoria, Australia.

[1] Note that the model of classical feedback considered in [9] is more restrictive than the general model considered in [10]—it is such that the decoding measurement of the receiver is restricted to one-way local operations and classical communication. See Section 4.2 of [1] for a review of the feedback scheme of [9].

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

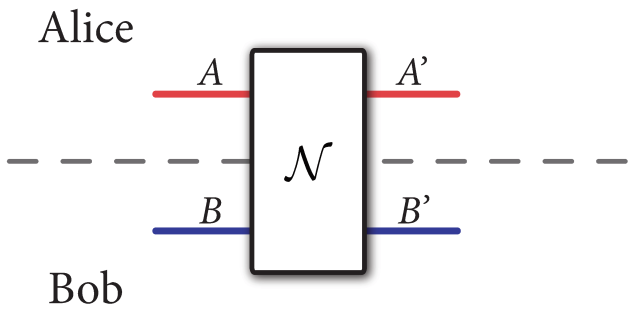

Fig. 1. Depiction of a bipartite channel. A bipartite channel is a four-terminal channel in which Alice has access to the input system A and the output system A', and Bob has access to the input system B and the output system B'.

In this paper, we generalize these tasks further by considering the forward classical capacity of a bipartite quantum channel, and we develop various upper bounds on this operational communication measure for a bipartite channel. This communication task has previously been studied for the special case of bipartite unitary channels [30] (see also [31], [32], [33] for studies of classical communication over bipartite unitary channels). To be clear, a bipartite channel is a four-terminal channel involving two parties, who each have access to one input port and one output port of the channel [34] (see Figure 1). The forward classical capacity of a bipartite channel is the optimal rate at which Alice can communicate classical bits to Bob, with error probability tending to zero as the number of channel uses becomes large.

A bipartite channel is an interesting communication scenario on its own, but it is also a generalization of a point-to-point channel, with the latter being a special case with one input port trivial and the output port for the other party trivial. In the same way, it is a generalization of a classical feedback channel. This relationship allows us to conclude upper bounds on the classical capacity of a point-to-point channel assisted by classical feedback. Interestingly, we prove that the most recent upper bound from [22], on the unassisted classical capacity, is actually an upper bound on the classical capacity assisted by classical feedback. As such, we now have an efficiently computable upper bound on this feedback-assisted capacity. By combining our results with [21, Section VI], we establish the strong converse for the classical capacity of the quantum erasure channel assisted by a classical feedback channel, thus improving the weak converse result of [25].

The rest of the paper is structured as follows. We begin in Section II by establishing some notation. In Section III, we introduce a measure of forward classical communication for a bipartite channel. Therein, we establish several of its properties, including the fact that it is equal to zero for a product of local channels, equal to zero for a classical feedback channel, and that it is subadditive under serial compositions of bipartite channels. We then introduce several variants of the basic measure and show how they reduce to previous measures from [20], [21], [22] for point-to-point channels. In Section IV, we detail several of the applications mentioned above: we establish that our measures of forward classical communication (in particular the ones based on geometric Rényi relative entropy) serve as upper bounds on the forward classical capacity of a bipartite channel and on the classical capacity of a point-to-point channel assisted by a classical feedback channel. In Section V, we explore the same applications but using the sandwiched Rényi relative entropy instead of the geometric Rényi relative entropy, and in Section VI, we show how these bounds simplify if the channels possess symmetry. In Section VII, we evaluate our bounds for several examples of bipartite and point-to-point channels. We finally conclude in Section VIII with a summary and some open questions for future work.

II. Notation

Here we list various notations and concepts that we use throughout the paper. A quantum channel is a completely positive and trace-preserving map. We denote the unnormalized maximally entangled operator by

$$\Gamma_{RA} := |\Gamma\rangle\langle\Gamma|_{RA}, \tag{1}$$

$$|\Gamma\rangle_{RA} := \sum_{i=0}^{d-1} |i\rangle_R |i\rangle_A, \tag{2}$$

where $R \simeq A$ with dimension d and $\{|i\rangle_R\}_{i=0}^{d-1}$ and $\{|i\rangle_A\}_{i=0}^{d-1}$ are orthonormal bases. The notation $R \simeq A$ means that the systems R and A are isomorphic. The maximally entangled state is denoted by

$$\Phi_{RA} := \frac{1}{d}\Gamma_{RA}, \tag{3}$$

and the maximally mixed state by

$$\pi_A := \frac{1}{d} I_A. \tag{4}$$

The Choi operator of a quantum channel $\mathcal{N}_{A\to B}$ (and more generally a linear map) is denoted by

$$\Gamma^{\mathcal{N}}_{RB} := \mathcal{N}_{A\to B}(\Gamma_{RA}). \tag{5}$$

A linear map $\mathcal{M}_{A\to B}$ is completely positive if and only if its Choi operator $\Gamma^{\mathcal{M}}_{RB}$ is positive semi-definite, and $\mathcal{M}_{A\to B}$ is trace preserving if and only if its Choi operator satisfies $\mathrm{Tr}_B[\Gamma^{\mathcal{M}}_{RB}] = I_R$.

We denote the transpose map acting on the quantum system A by

$$T_A(\cdot) := \sum_{i,j=0}^{d-1} |i\rangle\langle j|_A(\cdot)|i\rangle\langle j|_A. \tag{6}$$

A state ρ_{AB} is a positive partial transpose (PPT) state if $T_B(\rho_{AB})$ is positive semi-definite. The partial transpose is its own adjoint, in the sense that

$$\mathrm{Tr}[Y_{AB} T_A(X_{AB})] = \mathrm{Tr}[T_A(Y_{AB}) X_{AB}] \tag{7}$$

for all linear operators X_{AB} and Y_{AB}.

The following post-selected teleportation identity [35] plays a role in our analysis:

$$\mathcal{N}_{A\to B}(\rho_{SA}) = \langle\Gamma|_{AR}\rho_{SA} \otimes \Gamma^{\mathcal{N}}_{RB}|\Gamma\rangle_{AR}, \tag{8}$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

as it has in previous works on feedback-assisted capacities [36], [37], [38], [39], [40], [22]. We also make frequent use of the identities

$$\mathrm{Tr}_A[X_{AB}] = \langle\Gamma|_{RA}(I_R \otimes X_{AB})|\Gamma\rangle_{RA}, \tag{9}$$

$$X_{AB}|\Gamma\rangle_{AR} = T_R(X_{RB})|\Gamma\rangle_{AR}. \tag{10}$$

Given channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{B\to C}$, the Choi operator $\Gamma_{RC}^{\mathcal{M}\circ\mathcal{N}}$ of the serial composition $\mathcal{M}_{B\to C}\circ\mathcal{N}_{A\to B}$ is given by

$$\Gamma_{RC}^{\mathcal{M}\circ\mathcal{N}} = \langle\Gamma|_{BS}\Gamma_{RB}^{\mathcal{N}}\otimes\Gamma_{SC}^{\mathcal{M}}|\Gamma\rangle_{BS} \tag{11}$$

$$= \mathrm{Tr}_B[\Gamma_{RB}^{\mathcal{N}}T_B(\Gamma_{BC}^{\mathcal{M}})], \tag{12}$$

where $B \simeq S$, the operator $\Gamma_{RB}^{\mathcal{N}}$ is the Choi operator of $\mathcal{N}_{A\to B}$, and $\Gamma_{SC}^{\mathcal{M}}$ is the Choi operator of $\mathcal{M}_{B\to C}$.

III. Measures of forward classical communication for a bipartite channel

A. Basic measure

Before defining the basic measure of forward classical communication for a bipartite channel, let us recall some established concepts from quantum information theory.

A bipartite channel $\mathcal{M}_{AB\to A'B'}$ is a completely positive-partial-transpose preserving (C-PPT-P) channel if the output state $\omega_{R_AA'B'R_B} := \mathcal{M}_{AB\to A'B'}(\rho_{R_AABR_B})$ is a PPT state for every PPT input state $\rho_{R_AABR_B}$ [41], [42], [23], [43]. To be clear, the channel $\mathcal{M}_{AB\to A'B'}$ is defined to be C-PPT-P if $T_{B'R_B}(\omega_{R_AA'B'R_B}) \geq 0$ for every input state $\rho_{R_AABR_B}$ that satisfies $T_{BR_B}(\rho_{R_AABR_B}) \geq 0$. Let $\Gamma_{AA'BB'}^{\mathcal{M}}$ denote the Choi operator of $\mathcal{M}_{AB\to A'B'}$:

$$\Gamma_{AA'BB'}^{\mathcal{M}} := \mathcal{M}_{\hat{A}\hat{B}\to A'B'}(\Gamma_{A\hat{A}}\otimes\Gamma_{B\hat{B}}), \tag{13}$$

where $\hat{A}\simeq A$ and $\hat{B}\simeq B$. It is known that $\mathcal{M}_{AB\to A'B'}$ is C-PPT-P if and only if its Choi operator $\Gamma_{AA'BB'}^{\mathcal{M}}$ is PPT (i.e., $T_{BB'}(\Gamma_{AA'BB'}^{\mathcal{M}}) \geq 0$) [23], [43]. Equivalently, it is known that $\mathcal{M}_{AB\to A'B'}$ is C-PPT-P if and only if the map $T_{B'}\circ\mathcal{M}_{AB\to A'B'}\circ T_B$ is completely positive.

A C-PPT-P channel is not capable of generating entanglement shared between Alice and Bob at a non-trivial rate when used many times [37], [40]. As such, Alice cannot reliably communicate quantum information to Bob at a non-zero rate when using a C-PPT-P channel. This feature is helpful for us in devising a measure that serves as an upper bound on the forward classical capacity of a bipartite channel or on the classical capacity of a point-to-point channel assisted by classical feedback.

A bipartite channel $\mathcal{M}_{AB\to A'B'}$ is non-signaling from Alice to Bob [44], [45] if the following condition holds [46]

$$\mathrm{Tr}_{A'}\circ\mathcal{M}_{AB\to A'B'} = \mathrm{Tr}_{A'}\circ\mathcal{M}_{AB\to A'B'}\circ\mathcal{R}_A^{\pi}, \tag{14}$$

where $\mathcal{R}_A^{\pi}$ is a replacer channel, defined as $\mathcal{R}_A^{\pi}(\cdot) := \mathrm{Tr}_A[\cdot]\pi_A$, with $\pi_A := I_A/d_A$ the maximally mixed state on system A. To interpret this condition, consider the following. For Bob, the reduced state of his output system B' is obtained by tracing out Alice's output system A'. Note that the reduced state on B' is all that Bob can access at the output in this scenario. If the condition in (14) holds, then the reduced state on Bob's output system B' has no dependence on Alice's input system. Thus, if (14) holds, then Alice cannot use $\mathcal{M}_{AB\to A'B'}$ to send a signal to Bob.

One of our main interests in this paper is to bound the classical capacity of a quantum channel assisted by a classical feedback channel from Bob to Alice. In such a protocol, local channels are allowed for free, as well as the use of a classical feedback channel. Both of these actions can be considered as particular kinds of bipartite channels and both of them fall into the class of bipartite channels that are non-signaling from Alice to Bob and C-PPT-P (call this class $\mathrm{NS}_{A\not\to B}\cap\mathrm{PPT}$). As such, if we employ a measure of bipartite channels that involves a comparison between a bipartite channel of interest to all bipartite channels in $\mathrm{NS}_{A\not\to B}\cap\mathrm{PPT}$, then the two kinds of free channels would have zero value and the measure would indicate how different the channel of interest is from this set (i.e., how different it is from a channel that has no ability to send quantum information and no ability to signal from Alice to Bob). This is the main idea behind the measure that we propose below in Definition 1, but one should keep in mind that the measure below does not follow this reasoning precisely.

In Definition 1, although we motivated the measure for bipartite channels, we define it more generally for completely positive bipartite maps, as it turns out to be useful to do so when we later define other measures.

Definition 1: Let $\mathcal{M}_{AB\to A'B'}$ be a completely positive bipartite map. Then we define

$$C_\beta(\mathcal{M}_{AB\to A'B'}) := \log_2\beta(\mathcal{M}_{AB\to A'B'}), \tag{15}$$

$$\beta(\mathcal{M}_{AB\to A'B'}) := \inf_{\substack{S_{AA'BB'},\\ V_{AA'BB'}\in\mathrm{Herm}}}\left\{\begin{array}{c}\|\mathrm{Tr}_{A'B'}[S_{AA'BB'}]\|_\infty : \\ T_{BB'}(V_{AA'BB'}\pm\Gamma_{AA'BB'}^{\mathcal{M}})\geq 0, \\ S_{AA'BB'}\pm V_{AA'BB'}\geq 0, \\ \mathrm{Tr}_{A'}[S_{AA'BB'}] = \\ \pi_A\otimes\mathrm{Tr}_{AA'}[S_{AA'BB'}]\end{array}\right\}, \tag{16}$$

where Herm denotes the set of Hermitian operators and $\Gamma_{AA'BB'}^{\mathcal{M}}$ is the Choi operator of $\mathcal{M}_{AB\to A'B'}$:

$$\Gamma_{AA'BB'}^{\mathcal{M}} := \mathcal{M}_{\hat{A}\hat{B}\to A'B'}(\Gamma_{A\hat{A}}\otimes\Gamma_{B\hat{B}}). \tag{17}$$

In the above, $\hat{A}\simeq A$, $\hat{B}\simeq B$,

$$\Gamma_{A\hat{A}} := \sum_{i,j=0}^{d_A-1}|i\rangle\langle j|_A\otimes|i\rangle\langle j|_{\hat{A}},\quad \Gamma_{B\hat{B}} := \sum_{i,j=0}^{d_B-1}|i\rangle\langle j|_B\otimes|i\rangle\langle j|_{\hat{B}}, \tag{18}$$

and $\pi_A := I_A/d_A$.

Just before Definition 1, we discussed how the β measure incorporates PPT constraints, as well as non-signaling constraints. The constraint $T_{BB'}(V_{AA'BB'}\pm\Gamma_{AA'BB'}^{\mathcal{M}})\geq 0$ involves a PPT condition, and the constraint $\mathrm{Tr}_{A'}[S_{AA'BB'}] = \pi_A\otimes\mathrm{Tr}_{AA'}[S_{AA'BB'}]$ involves a non-signaling condition. Since $S_{AA'BB'}\pm V_{AA'BB'}\geq 0$, it follows that $S_{AA'BB'}\geq 0$, implying that the operator $S_{AA'BB'}$ corresponds to a completely positive map. The definition above becomes more transparent, but however does not decrease, if we simply set $V_{AA'BB'} = S_{AA'BB'}$. Then it is clear that there is just a

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

PPT constraint and non-signaling constraint corresponding to a single completely positive map. Furthermore, as we discuss below, the objective function $\|\mathrm{Tr}_{A'B'}[S_{AA'BB'}]\|_\infty$ measures how close $S_{AA'BB'}$ is to being a trace preserving map, and Proposition 2 states that the minimum value of the objective function is one, in which case $S_{AA'BB'}$ corresponds to a quantum channel.

In Appendix A, we prove that $\beta(\mathcal{M}_{AB\to A'B'})$ can alternatively be expressed as follows:

$$\beta(\mathcal{M}_{AB\to A'B'}) = \inf_{\substack{\mathcal{S}_{AB\to A'B'},\\ \mathcal{V}_{AB\to A'B'}\in \mathrm{HermP}}} \|\mathcal{S}_{AB\to A'B'}\|_1 \tag{19}$$

subject to

$$\begin{aligned} T_{B'}\circ(\mathcal{V}_{AB\to A'B'}\pm\mathcal{M}_{AB\to A'B'})\circ T_B \geq 0,\\ \mathcal{S}_{AB\to A'B'}\pm\mathcal{V}_{AB\to A'B'}\geq 0,\\ \mathrm{Tr}_{A'}\circ\mathcal{S}_{AB\to A'B'} = \mathrm{Tr}_{A'}\circ\mathcal{S}_{AB\to A'B'}\circ\mathcal{R}^{\pi}_{A} \end{aligned}, \tag{20}$$

where HermP is the set of Hermiticity preserving maps, $\|\mathcal{S}_{AB\to A'B'}\|_1$ is the trace norm of the bipartite map $\mathcal{S}_{AB\to A'B'}$, and the notation $\mathcal{L}_{AB\to A'B'}\geq 0$ means that the Hermiticity preserving map $\mathcal{L}_{AB\to A'B'}$ is completely positive. Related to how $S_{AA'BB'}\geq 0$ as discussed above, the constraint $\mathcal{S}_{AB\to A'B'}\pm\mathcal{V}_{AB\to A'B'}\geq 0$ implies that $\mathcal{S}_{AB\to A'B'}\geq 0$, which is the same as $\mathcal{S}_{AB\to A'B'}$ being a completely positive map. Thus, the trace norm objective function $\|\mathcal{S}_{AB\to A'B'}\|_1$ measures how close $\mathcal{S}_{AB\to A'B'}$ can be to a trace preserving map, i.e., a quantum channel, while satisfying the constraints given. With the expression in (20), it might become more clear that $\beta(\mathcal{M}_{AB\to A'B'})$ involves a comparison of $\mathcal{M}_{AB\to A'B'}$ to other Hermiticity-preserving bipartite maps, which involves the C-PPT-P condition and the non-signaling constraint. In Appendix A, not only do we prove the equality above, but we also explain these concepts in more detail.

We can also express $\beta(\mathcal{M}_{AB\to A'B'})$ as follows:

$$\inf_{\substack{\lambda\geq 0, S_{AA'BB'},\\ V_{AA'BB'}\in\mathrm{Herm}}} \left\{ \begin{array}{c} \lambda : \\ \mathrm{Tr}_{A'B'}[S_{AA'BB'}]\leq \lambda I_{AB} \\ T_{BB'}(V_{AA'BB'}\pm\Gamma^{\mathcal{M}}_{AA'BB'})\geq 0, \\ S_{AA'BB'}\pm V_{AA'BB'}\geq 0, \\ \mathrm{Tr}_{A'}[S_{AA'BB'}] = \\ \pi_A\otimes\mathrm{Tr}_{AA'}[S_{AA'BB'}] \end{array} \right\}. \tag{21}$$

By exploiting the equality constraint $\mathrm{Tr}_{A'}[S_{AA'BB'}] = \pi_A\otimes\mathrm{Tr}_{AA'}[S_{AA'BB'}]$, we find that

$$\begin{aligned} &\|\mathrm{Tr}_{A'B'}[S_{AA'BB'}]\|_\infty \\ &= \|\mathrm{Tr}_{B'}[\mathrm{Tr}_{A'}[S_{AA'BB'}]]\|_\infty && (22)\\ &= \|\mathrm{Tr}_{B'}[\pi_A\otimes\mathrm{Tr}_{AA'}[S_{AA'BB'}]]\|_\infty && (23)\\ &= \|\pi_A\otimes\mathrm{Tr}_{AA'B'}[S_{AA'BB'}]\|_\infty && (24)\\ &= \frac{1}{d_A}\|\mathrm{Tr}_{AA'B'}[S_{AA'BB'}]\|_\infty . && (25) \end{aligned}$$

Then we find that

$$\beta(\mathcal{M}_{AB\to A'B'}) = \inf_{\substack{S_{AA'BB'},\\ V_{AA'BB'}\in\mathrm{Herm}}} \left\{ \begin{array}{c} \frac{1}{d_A}\|\mathrm{Tr}_{AA'B'}[S_{AA'BB'}]\|_\infty : \\ T_{BB'}(V_{AA'BB'}\pm\Gamma^{\mathcal{M}}_{AA'BB'})\geq 0, \\ S_{AA'BB'}\pm V_{AA'BB'}\geq 0, \\ \mathrm{Tr}_{A'}[S_{AA'BB'}] = \\ \pi_A\otimes\mathrm{Tr}_{AA'}[S_{AA'BB'}] \end{array} \right\}. \tag{26}$$

Since $S_{AA'BB'}\pm V_{AA'BB'}\geq 0$ implies that $S_{AA'BB'}\geq 0$, we can also rewrite $\beta(\mathcal{M}_{AB\to A'B'})$ as

$$\beta(\mathcal{M}_{AB\to A'B'}) = \inf_{\substack{\lambda, S_{AA'BB'}\geq 0,\\ V_{AA'BB'}\in\mathrm{Herm}}} \left\{ \begin{array}{c} \lambda : \\ \frac{1}{d_A}\mathrm{Tr}_{AA'B'}[S_{AA'BB'}]\leq \lambda I_B, \\ T_{BB'}(V_{AA'BB'}\pm\Gamma^{\mathcal{M}}_{AA'BB'})\geq 0, \\ S_{AA'BB'}\pm V_{AA'BB'}\geq 0, \\ \mathrm{Tr}_{A'}[S_{AA'BB'}] = \\ \pi_A\otimes\mathrm{Tr}_{AA'}[S_{AA'BB'}] \end{array} \right\}. \tag{27}$$

B. Properties of the basic measure

We now establish several properties of $C_\beta(\mathcal{N}_{AB\to A'B'})$, which are basic properties that we might expect of a measure of forward classical communication for a bipartite channel. These include the following:

1) non-negativity (Proposition 2),
2) stability under tensoring with identity channels (Proposition 3),
3) zero value for classical feedback channels (Proposition 4),
4) zero value for a tensor product of local channels (Proposition 5),
5) subadditivity under serial composition (Proposition 6),
6) data processing under pre- and post-processing by local channels (Corollary 7),
7) invariance under local unitary channels (Corollary 8),
8) convexity of β (Proposition 9).

All of the properties above hold for bipartite channels, while the second and fifth through eighth hold more generally for completely positive bipartite maps.

Proposition 2 (Non-negativity): Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Then

$$C_\beta(\mathcal{N}_{AB\to A'B'})\geq 0. \tag{28}$$

Proof: We prove the equivalent statement $\beta(\mathcal{N}_{AB\to A'B'})\geq 1$. Let λ, $S_{AA'BB'}$, and $V_{AA'BB'}$ be arbitrary Hermitian operators satisfying the constraints in (27). Then consider that

$$\begin{aligned} \lambda d_B &= \lambda\,\mathrm{Tr}_B[I_B] && (29)\\ &\geq \frac{1}{d_A}\mathrm{Tr}_{AA'BB'}[S_{AA'BB'}] && (30)\\ &\geq \frac{1}{d_A}\mathrm{Tr}_{AA'BB'}[V_{AA'BB'}] && (31)\\ &= \frac{1}{d_A}\mathrm{Tr}_{AA'BB'}[T_{BB'}(V_{AA'BB'})] && (32)\\ &\geq \frac{1}{d_A}\mathrm{Tr}_{AA'BB'}[T_{BB'}(\Gamma^{\mathcal{N}}_{AA'BB'})] && (33) \end{aligned}$$

$$= \frac{1}{d_A} \operatorname{Tr}_{AA'BB'}[\Gamma^{\mathcal{N}}_{AA'BB'}] \tag{34}$$
$$= \frac{1}{d_A} \operatorname{Tr}_{AB}[I_{AB}] \tag{35}$$
$$= d_B. \tag{36}$$

This implies that $\lambda \geq 1$. Since the inequality holds for all λ, $S_{AA'BB'}$, and $V_{AA'BB'}$ satisfying the constraints in (27), we conclude the statement above. ∎

Proposition 3 (Stability): Let $\mathcal{M}_{AB\to A'B'}$ be a completely positive bipartite map. Then

$$C_\beta(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}) = C_\beta(\mathcal{M}_{AB\to A'B'}). \tag{37}$$

Proof: Let $S_{AA'BB'}$ and $V_{AA'BB'}$ be arbitrary Hermitian operators satisfying the constraints in (16) for $\mathcal{M}_{AB\to A'B'}$. The Choi operator of $\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}$ is given by

$$\Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}. \tag{38}$$

Let us show that $\Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}$ and $\Gamma_{\bar{A}\tilde{A}} \otimes V_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}$ satisfy the constraints in (16) for $\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}$. Consider that

$$T_{BB'}(V_{AA'BB'} \pm \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0 \tag{39}$$

$$\Leftrightarrow \quad T_{BB'}(\Gamma_{\bar{A}\tilde{A}} \otimes V_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}) \geq \pm T_{BB'}(\Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}) \tag{40}$$

$$\Leftrightarrow \quad T_{BB'\bar{B}\tilde{B}}(\Gamma_{\bar{A}\tilde{A}} \otimes V_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}) \geq \pm T_{BB'\bar{B}\tilde{B}}(\Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}) \tag{41}$$

and

$$S_{AA'BB'} \pm V_{AA'BB'} \geq 0 \tag{42}$$

$$\Leftrightarrow \quad \Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}} \geq \pm \Gamma_{\bar{A}\tilde{A}} \otimes V_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}} \tag{43}$$

and

$$\operatorname{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \operatorname{Tr}_{AA'}[S_{AA'BB'}] \tag{44}$$

$$\Leftrightarrow \quad \operatorname{Tr}_{A'\bar{A}}[\Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}]$$
$$= I_{\bar{A}} \otimes \pi_A \otimes \operatorname{Tr}_{AA'}[S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}] \tag{45}$$
$$= \pi_{\bar{A}} \otimes \pi_A \otimes \operatorname{Tr}_{AA'\bar{A}\tilde{A}}[\Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}]. \tag{46}$$

Also, consider that

$$\frac{1}{d_A d_{\bar{A}}} \left\| \operatorname{Tr}_{AA'\bar{A}\tilde{A}B'\tilde{B}}[\Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}] \right\|_\infty$$
$$= \frac{1}{d_A d_{\bar{A}}} \left\| d_{\bar{A}} \operatorname{Tr}_{AA'B'}[S_{AA'BB'} \otimes I_{\bar{B}}] \right\|_\infty \tag{47}$$
$$= \frac{1}{d_A} \left\| \operatorname{Tr}_{AA'B'}[S_{AA'BB'}] \otimes I_{\bar{B}} \right\|_\infty \tag{48}$$
$$= \frac{1}{d_A} \left\| \operatorname{Tr}_{AA'B'}[S_{AA'BB'}] \right\|_\infty. \tag{49}$$

Thus, it follows that

$$\beta(\mathcal{M}_{AB\to A'B'}) \geq \beta(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}). \tag{50}$$

Now let us show the opposite inequality. Let $S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}$ and $V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}$ be arbitrary Hermitian operators satisfying the constraints in (16) for $\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}$. Set

$$S'_{AA'BB'} := \frac{1}{d_{\bar{A}} d_{\bar{B}}} \operatorname{Tr}_{\bar{A}\tilde{A}\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}], \tag{51}$$
$$V'_{AA'BB'} := \frac{1}{d_{\bar{A}} d_{\bar{B}}} \operatorname{Tr}_{\bar{A}\tilde{A}\bar{B}\tilde{B}}[V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}]. \tag{52}$$

We show that $S'_{AA'BB'}$ and $V'_{AA'BB'}$ satisfy the constraints in (16) for $\mathcal{M}_{AB\to A'B'}$. Consider that

$$\Gamma^{\mathrm{id}\otimes\mathcal{N}\otimes\mathrm{id}}_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} = \Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}. \tag{53}$$

Then

$$T_{BB'\bar{B}\tilde{B}}(V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \pm \Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}) \geq 0 \tag{54}$$

$$\Rightarrow \quad \operatorname{Tr}_{\bar{A}\tilde{A}\bar{B}\tilde{B}}[T_{BB'\bar{B}\tilde{B}}(V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \pm \Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}})] \geq 0 \tag{55}$$

$$\Leftrightarrow \quad T_{BB'}(V_{AA'BB'} \pm d_{\bar{A}} d_{\bar{B}} \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0 \tag{56}$$
$$\Leftrightarrow \quad T_{BB'}(V'_{AA'BB'} \pm \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0. \tag{57}$$

Also

$$S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \pm V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \geq 0 \tag{58}$$
$$\Rightarrow \quad \operatorname{Tr}_{\bar{A}\tilde{A}\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \pm V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \geq 0 \tag{59}$$
$$\Leftrightarrow \quad S'_{AA'BB'} \pm V'_{AA'BB'} \geq 0, \tag{60}$$

and

$$\operatorname{Tr}_{\tilde{A}A'}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] = \pi_{\bar{A}A} \otimes \operatorname{Tr}_{\bar{A}\tilde{A}AA'}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \tag{61}$$

$$\Rightarrow \quad \operatorname{Tr}_{\bar{A}\tilde{A}A'\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}]$$
$$= \operatorname{Tr}_{\bar{A}\bar{B}\tilde{B}}[\pi_{\bar{A}A} \otimes \operatorname{Tr}_{\bar{A}\tilde{A}AA'}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}]] \tag{62}$$
$$= \pi_A \otimes \operatorname{Tr}_{\bar{A}\tilde{A}AA'\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \tag{63}$$

$$\Leftrightarrow \quad \operatorname{Tr}_{A'}[S'_{AA'BB'}] = \pi_A \otimes \operatorname{Tr}_{AA'}[S'_{AA'BB'}]. \tag{64}$$

Finally, let λ be such that

$$\frac{1}{d_A d_{\bar{A}}} \operatorname{Tr}_{\bar{A}\tilde{A}AA'B'\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \leq \lambda I_{B\bar{B}}. \tag{65}$$

Then it follows that

$$\operatorname{Tr}_{\bar{B}}\left[\frac{1}{d_A d_{\bar{A}}} \operatorname{Tr}_{\bar{A}\tilde{A}AA'B'\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}]\right] \leq \operatorname{Tr}_{\bar{B}}[\lambda I_{B\bar{B}}] \tag{66}$$

$$\Leftrightarrow \quad \frac{1}{d_A d_{\bar{A}}} \operatorname{Tr}_{\bar{A}\tilde{A}AA'B'\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \leq d_{\bar{B}} \lambda I_B \tag{67}$$
$$\Leftrightarrow \quad \frac{1}{d_A} \operatorname{Tr}_{AA'B'}[S'_{AA'BB'}] \leq \lambda I_B. \tag{68}$$

Thus, we conclude that

$$\beta(\mathcal{M}_{AB\to A'B'}) \leq \beta(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}). \tag{69}$$

This concludes the proof. ∎

Proposition 4 (Zero on classical feedback channels): Let $\overline{\Delta}_{B\to A'}$ be a classical feedback channel:

$$\overline{\Delta}_{B\to A'}(\cdot) \coloneqq \sum_{i=0}^{d-1} |i\rangle_{A'}\langle i|_B(\cdot)|i\rangle_B\langle i|_{A'}, \tag{70}$$

where $A' \simeq B$ and $d = d_{A'} = d_B$. Then

$$C_\beta(\overline{\Delta}_{B\to A'}) = 0. \tag{71}$$

Proof: We prove the equivalent statement that $\beta(\overline{\Delta}_{B\to A'}) = 1$. In this case, the A and B' systems are trivial, so that $d_A = 1$, and the Choi operator of $\overline{\Delta}_{B\to A'}$ is given by

$$\Gamma^{\overline{\Delta}}_{BA'} = \overline{\Gamma}_{BA'}, \tag{72}$$

where

$$\overline{\Gamma}_{BA'} \coloneqq \sum_{i=0}^{d_B-1} |i\rangle\!\langle i|_B \otimes |i\rangle\!\langle i|_{A'}. \tag{73}$$

Pick $S_{BA'} = V_{BA'} = \overline{\Gamma}_{BA'}$. Then we need to check that the constraints in (16) are satisfied for these choices. Consider that

$$\begin{aligned} & T_B(V_{BA'} \pm \Gamma^{\overline{\Delta}}_{BA'}) \geq 0 && (74)\\ \Leftrightarrow \quad & T_B(\overline{\Gamma}_{BA'} \pm \overline{\Gamma}_{BA'}) \geq 0 && (75)\\ \Leftrightarrow \quad & \overline{\Gamma}_{BA'} \pm \overline{\Gamma}_{BA'} \geq 0, && (76)\end{aligned}$$

and the last inequality is trivially satisfied. Also,

$$\begin{aligned} & S_{BA'} \pm V_{BA'} \geq 0 && (77)\\ \Leftrightarrow \quad & \overline{\Gamma}_{BA'} \pm \overline{\Gamma}_{BA'} \geq 0, && (78)\end{aligned}$$

and the no-signaling condition $\mathrm{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \mathrm{Tr}_{AA'}[S_{AA'BB'}]$ is trivially satisfied because the A system is trivial, having dimension equal to one. Finally, let us evaluate the objective function for these choices:

$$\begin{aligned} \frac{1}{d_A}\left\|\mathrm{Tr}_{AA'B'}[S_{AA'BB'}]\right\|_\infty &= \left\|\mathrm{Tr}_{A'}[S_{A'B}]\right\|_\infty && (79)\\ &= \left\|\mathrm{Tr}_{A'}[\overline{\Gamma}_{BA'}]\right\|_\infty && (80)\\ &= \left\|I_B\right\|_\infty && (81)\\ &= 1. && (82)\end{aligned}$$

Combined with the general lower bound from Proposition 2, we conclude (71). ∎

Proposition 5 (Zero on tensor product of local channels): Let $\mathcal{E}_{A\to A'}$ and $\mathcal{F}_{B\to B'}$ be quantum channels. Then

$$C_\beta(\mathcal{E}_{A\to A'} \otimes \mathcal{F}_{B\to B'}) = 0. \tag{83}$$

Proof: We prove the equivalent statement that $\beta(\mathcal{E}_{A\to A'} \otimes \mathcal{F}_{B\to B'}) = 1$. Set $S_{AA'BB'} = V_{AA'BB'} = \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}$, where $\Gamma^{\mathcal{E}}_{AA'}$ and $\Gamma^{\mathcal{F}}_{BB'}$ are the Choi operators of $\mathcal{E}_{A\to A'}$ and $\mathcal{F}_{B\to B'}$, respectively. We need to check that the constraints in (16) are satisfied for these choices. Consider that

$$\begin{aligned} & T_{BB'}(V_{AA'BB'} \pm \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}) \geq 0 && (84)\\ \Leftrightarrow\ & T_{BB'}(\Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'} \pm \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}) \geq 0 && (85)\\ \Leftrightarrow\ & \Gamma^{\mathcal{E}}_{AA'} \otimes T_{BB'}(\Gamma^{\mathcal{F}}_{BB'}) \pm \Gamma^{\mathcal{E}}_{AA'} \otimes T_{BB'}(\Gamma^{\mathcal{F}}_{BB'}) \geq 0, && (86)\end{aligned}$$

and the last inequality trivially holds because $T_{BB'}$ acts as a positive map on $\Gamma^{\mathcal{F}}_{BB'}$. Also,

$$\begin{aligned} & S_{AA'BB'} \pm V_{AA'BB'} \geq 0 && (87)\\ \Leftrightarrow \quad & \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'} \pm \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'} \geq 0, && (88)\end{aligned}$$

and

$$\begin{aligned} \mathrm{Tr}_{A'}[S_{AA'BB'}] &= \mathrm{Tr}_{A'}[\Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}] && (89)\\ &= I_A \otimes \Gamma^{\mathcal{F}}_{BB'} && (90)\\ &= \pi_A \otimes \mathrm{Tr}_{AA'}[\Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}] && (91)\\ &= \pi_A \otimes \mathrm{Tr}_{AA'}[S_{AA'BB'}]. && (92)\end{aligned}$$

Finally, consider that the objective function evaluates to

$$\begin{aligned} \left\|\mathrm{Tr}_{A'B'}[S_{AA'BB'}]\right\|_\infty &= \left\|\mathrm{Tr}_{A'B'}[\Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}]\right\|_\infty && (93)\\ &= \left\|I_{AB}\right\|_\infty && (94)\\ &= 1. && (95)\end{aligned}$$

Combined with the general lower bound from Proposition 2, we conclude (83). ∎

Proposition 6 (Subadditivity under composition): Let $\mathcal{M}^1_{AB\to A'B'}$ and $\mathcal{M}^2_{A'B'\to A''B''}$ be bipartite completely positive maps, and define

$$\mathcal{M}^3_{AB\to A''B''} \coloneqq \mathcal{M}^2_{A'B'\to A''B''} \circ \mathcal{M}^1_{AB\to A'B'}. \tag{96}$$

Then

$$C_\beta(\mathcal{M}^3_{AB\to A''B''}) \leq C_\beta(\mathcal{M}^2_{A'B'\to A''B''}) + C_\beta(\mathcal{M}^1_{AB\to A'B'}). \tag{97}$$

Proof: We prove the equivalent statement that

$$\beta(\mathcal{M}^3_{AB\to A''B''}) \leq \beta(\mathcal{M}^2_{A'B'\to A''B''}) \cdot \beta(\mathcal{M}^1_{AB\to A'B'}). \tag{98}$$

Let $S^1_{AA'BB'}$ and $V^1_{AA'BB'}$ satisfy

$$\begin{aligned} & T_{BB'}(V^1_{AA'BB'} \pm \Gamma^{\mathcal{M}^1}_{AA'BB'}) \geq 0, && (99)\\ & S^1_{AA'BB'} \pm V^1_{AA'BB'} \geq 0, && (100)\\ & \mathrm{Tr}_{A'}[S^1_{AA'BB'}] = \pi_A \otimes \mathrm{Tr}_{AA'}[S^1_{AA'BB'}], && (101)\end{aligned}$$

and let $S^2_{A'A''B'B''}$ and $V^2_{A'A''B'B''}$ satisfy

$$\begin{aligned} & T_{B'B''}(V^2_{A'A''B'B''} \pm \Gamma^{\mathcal{M}^2}_{A'A''B'B''}) \geq 0, && (102)\\ & S^2_{A'A''B'B''} \pm V^2_{A'A''B'B''} \geq 0, && (103)\\ & \mathrm{Tr}_{A''}[S^2_{A'A''B'B''}] = \pi_{A'} \otimes \mathrm{Tr}_{A'A''}[S^2_{A'A''B'B''}]. && (104)\end{aligned}$$

Then it follows that

$$\begin{aligned} & T_{BB'B'B''}(V^1_{AA'BB'} \otimes V^2_{A'A''B'B''} \\ & \qquad \pm \Gamma^{\mathcal{M}^1}_{AA'BB'} \otimes \Gamma^{\mathcal{M}^2}_{A'A''B'B''}) \geq 0, && (105)\\ & S^1_{AA'BB'} \otimes S^2_{A'A''B'B''} \pm V^1_{AA'BB'} \otimes V^2_{A'A''B'B''} \geq 0. && (106)\end{aligned}$$

This latter statement is a consequence of the general fact that if A, B, C, and D are Hermitian operators satisfying $A \pm B \geq 0$ and $C \pm D \geq 0$, then $A \otimes C \pm B \otimes D \geq 0$. To see this, consider that the original four operator inequalities imply the four operator inequalities $(A \pm B) \otimes (C \pm D) \geq 0$, and then summing these four different operator inequalities in various

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

ways leads to $A\otimes C \pm B\otimes D \geq 0$. See (377)–(385) for further clarification of this point.

Now apply the following positive map to (105)–(106):

$$(\cdot) \to (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(\cdot)(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}), \tag{107}$$

where

$$|\Gamma\rangle_{A'A'} \coloneqq \sum_i |i\rangle_{A'}|i\rangle_{A'}, \tag{108}$$

$$|\Gamma\rangle_{B'B'} \coloneqq \sum_i |i\rangle_{B'}|i\rangle_{B'}. \tag{109}$$

This gives

$$T_{BB''}(V^3_{AA''BB''} \pm \Gamma^{\mathcal{M}^2\circ\mathcal{M}^1}_{AA''BB''}) \geq 0, \tag{110}$$

$$S^3_{AA''BB''} \pm V^3_{AA''BB''} \geq 0, \tag{111}$$

where

$$V^3_{AA''BB''} \coloneqq (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) (V^1_{AA'BB'} \otimes V^2_{A'A''B'B''})(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}), \tag{112}$$

$$\Gamma^{\mathcal{M}^2\circ\mathcal{M}^1}_{AA''BB''} \coloneqq (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) (\Gamma^{\mathcal{M}^1}_{AA'BB'} \otimes \Gamma^{\mathcal{M}^2}_{A'A''B'B''})(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}), \tag{113}$$

$$S^3_{AA''BB''} \coloneqq (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) (S^1_{AA'BB'} \otimes S^2_{A'A''B'B''})(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}), \tag{114}$$

and we applied (11) to conclude that

$$(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(\Gamma^{\mathcal{M}^1}_{AA'BB'} \otimes \Gamma^{\mathcal{M}^2}_{A'A''B'B''}) (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) = \Gamma^{\mathcal{M}^2\circ\mathcal{M}^1}_{AA''BB''}. \tag{115}$$

Also, consider that

$$\begin{aligned}
&\mathrm{Tr}_{A''}[S^3_{AA''BB''}] \\
&= \mathrm{Tr}_{A''}[(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(S^1_{AA'BB'} \otimes S^2_{A'A''B'B''}) \\
&\qquad (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'})] && (116)\\
&= (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(S^1_{AA'BB'} \otimes \mathrm{Tr}_{A''}[S^2_{A'A''B'B''}]) \\
&\qquad (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) && (117)\\
&= (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) \\
&\qquad (S^1_{AA'BB'} \otimes \pi_{A'} \otimes \mathrm{Tr}_{A'A''}[S^2_{A'A''B'B''}]) \\
&\qquad (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) && (118)\\
&= \frac{1}{d_{A'}}(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) \\
&\qquad (S^1_{AA'BB'} \otimes I_{A'} \otimes \mathrm{Tr}_{A'A''}[S^2_{A'A''B'B''}]) \\
&\qquad (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) && (119)\\
&= \frac{1}{d_{A'}}\langle\Gamma|_{B'B'}(\mathrm{Tr}_{A'}[S^1_{AA'BB'}] \\
&\qquad \otimes \mathrm{Tr}_{A'A''}[S^2_{A'A''B'B''}])|\Gamma\rangle_{B'B'} && (120)\\
&= \frac{1}{d_{A'}}\langle\Gamma|_{B'B'}(\pi_A \otimes \mathrm{Tr}_{AA'}[S^1_{AA'BB'}] \\
&\qquad \otimes \mathrm{Tr}_{A'A''}[S^2_{A'A''B'B''}])|\Gamma\rangle_{B'B'} && (121)\\
&= \pi_A \otimes \frac{1}{d_{A'}}\langle\Gamma|_{B'B'}(\mathrm{Tr}_{AA'}[S^1_{AA'BB'}] \\
&\qquad \otimes \mathrm{Tr}_{A'A''}[S^2_{A'A''B'B''}])|\Gamma\rangle_{B'B'}. && (122)
\end{aligned}$$

Now consider that

$$\mathrm{Tr}_{AA''}[S^3_{AA''BB''}] = \frac{1}{d_{A'}}\langle\Gamma|_{B'B'}(\mathrm{Tr}_{AA'}[S^1_{AA'BB'}]\otimes\mathrm{Tr}_{A'A''}[S^2_{A'A''B'B''}])|\Gamma\rangle_{B'B'}. \tag{123}$$

So we conclude that

$$\mathrm{Tr}_{A''}[S^3_{AA''BB''}] = \pi_A \otimes \mathrm{Tr}_{AA''}[S^3_{AA''BB''}]. \tag{124}$$

Finally, consider that

$$\begin{aligned}
&\left\|\mathrm{Tr}_{A''B''}[S^3_{AA''BB''}]\right\|_\infty \\
&= \left\| \begin{array}{c} \mathrm{Tr}_{A''B''}[(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) \\ (S^1_{AA'BB'} \otimes S^2_{A'A''B'B''}) \\ (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'})] \end{array} \right\|_\infty && (125)\\
&= \left\| \begin{array}{c} (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) \\ (S^1_{AA'BB'} \otimes \mathrm{Tr}_{A''B''}[S^2_{A'A''B'B''}]) \\ (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) \end{array} \right\|_\infty && (126)\\
&\leq \left\|\mathrm{Tr}_{A''B''}[S^2_{A'A''B'B''}]\right\|_\infty \cdot \\
&\qquad \left\| \begin{array}{c} (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})\,(S^1_{AA'BB'} \otimes I_{A'B'}) \\ (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) \end{array} \right\|_\infty && (127)\\
&= \left\|\mathrm{Tr}_{A''B''}[S^2_{A'A''B'B''}]\right\|_\infty \left\|\mathrm{Tr}_{A'B'}[S^1_{AA'BB'}]\right\|_\infty . && (128)
\end{aligned}$$

Since $S^3_{AA''BB''}$ and $V^3_{AA''BB''}$ are particular choices that satisfy the constraints in (110), (111), and (124), we conclude that

$$\beta(\mathcal{M}^3_{AB\to A''B''}) \leq \left\|\mathrm{Tr}_{A''B''}[S^2_{A'A''B'B''}]\right\|_\infty \left\|\mathrm{Tr}_{A'B'}[S^1_{AA'BB'}]\right\|_\infty . \tag{129}$$

Since $S^1_{AA'BB'}$ and $V^1_{AA'BB'}$ are arbitrary Hermitian operators satisfying the constraints in (99)–(101) and $S^2_{A'A''B'B''}$ and $V^2_{A'A''B'B''}$ are arbitrary Hermitian operators satisfying the constraints in (102)–(104), we conclude (97). ∎

Corollary 7 (Data processing under local channels): Let $\mathcal{M}_{AB\to A'B'}$ be a completely positive bipartite map. Let $\mathcal{K}_{\hat{A}\to A}$, $\mathcal{L}_{\hat{B}\to B}$, $\mathcal{N}_{A'\to A''}$, and $\mathcal{P}_{B'\to B''}$ be local quantum channels, and define the bipartite completely positive map $\mathcal{F}_{\hat{A}\hat{B}\to A''B''}$ as follows:

$$\mathcal{F}_{\hat{A}\hat{B}\to A''B''} \coloneqq (\mathcal{N}_{A'\to A''} \otimes \mathcal{P}_{B'\to B''})\mathcal{M}_{AB\to A'B'}(\mathcal{K}_{\hat{A}\to A} \otimes \mathcal{L}_{\hat{B}\to B}). \tag{130}$$

Then

$$C_\beta(\mathcal{F}_{\hat{A}\hat{B}\to A''B''}) \leq C_\beta(\mathcal{M}_{AB\to A'B'}). \tag{131}$$

Proof: Apply Propositions 5 and 6 to find that

$$\begin{aligned}
&C_\beta(\mathcal{F}_{\hat{A}\hat{B}\to A''B''}) \\
&\leq C_\beta(\mathcal{N}_{A'\to A''} \otimes \mathcal{P}_{B'\to B''}) \\
&\qquad + C_\beta(\mathcal{M}_{AB\to A'B'}) + C_\beta(\mathcal{K}_{\hat{A}\to A} \otimes \mathcal{L}_{\hat{B}\to B}) && (132)\\
&= C_\beta(\mathcal{M}_{AB\to A'B'}). && (133)
\end{aligned}$$

This concludes the proof. ∎

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

Corollary 8 (Invariance under local unitary channels): Let $\mathcal{M}_{AB\to A'B'}$ be a completely positive bipartite map. Let $\mathcal{U}_A$, $\mathcal{V}_B$, $\mathcal{W}_{A'}$, and $\mathcal{Y}_{B'}$ be local unitary channels, and define the bipartite completely positive map $\mathcal{F}_{\hat{A}\hat{B}\to A''B''}$ as follows:

$$\mathcal{F}_{AB\to A'B'} \coloneqq (\mathcal{W}_{A'} \otimes \mathcal{Y}_{B'})\mathcal{M}_{AB\to A'B'}(\mathcal{U}_A \otimes \mathcal{V}_B). \quad (134)$$

Then

$$C_\beta(\mathcal{F}_{AB\to A'B'}) = C_\beta(\mathcal{M}_{AB\to A'B'}). \quad (135)$$

Proof: Apply Corollary 7 twice to conclude that $C_\beta(\mathcal{M}_{AB\to A'B'}) \geq C_\beta(\mathcal{F}_{AB\to A'B'})$ and $C_\beta(\mathcal{F}_{AB\to A'B'}) \geq C_\beta(\mathcal{M}_{AB\to A'B'})$. ∎

Proposition 9 (Convexity): The measure β is convex, in the following sense:

$$\beta(\mathcal{M}^\lambda_{AB\to A'B'}) \leq \lambda\beta(\mathcal{M}^1_{AB\to A'B'}) + (1-\lambda)\,\beta(\mathcal{M}^0_{AB\to A'B'}), \quad (136)$$

where $\mathcal{M}^0_{AB\to A'B'}$ and $\mathcal{M}^1_{AB\to A'B'}$ are completely positive bipartite maps, $\lambda \in [0,1]$, and

$$\mathcal{M}^\lambda_{AB\to A'B'} \coloneqq \lambda\mathcal{M}^1_{AB\to A'B'} + (1-\lambda)\,\mathcal{M}^0_{AB\to A'B'}. \quad (137)$$

Proof: Let $S^x_{AA'BB'}$ and $V^x_{AA'BB'}$ satisfy the constraints in (16) for $\mathcal{M}^x_{AB\to A'B'}$ for $x \in \{0,1\}$. Then

$$S^\lambda_{AA'BB'} \coloneqq \lambda S^1_{AA'BB'} + (1-\lambda)\,S^0_{AA'BB'}, \quad (138)$$
$$V^\lambda_{AA'BB'} \coloneqq \lambda V^1_{AA'BB'} + (1-\lambda)\,V^0_{AA'BB'}, \quad (139)$$

satisfy the constraints in (16) for $\mathcal{M}^\lambda_{AB\to A'B'}$. Then it follows that

$$\beta(\mathcal{M}^\lambda_{AB\to A'B'}) \leq \left\|\mathrm{Tr}_{A'B'}[S^\lambda_{AA'BB'}]\right\|_\infty \quad (140)$$
$$\leq \lambda\left\|\mathrm{Tr}_{A'B'}[S^1_{AA'BB'}]\right\|_\infty + (1-\lambda)\left\|\mathrm{Tr}_{A'B'}[S^0_{AA'BB'}]\right\|_\infty, \quad (141)$$

where the second inequality follows from convexity of the ∞-norm. Since the inequality holds for all $S^x_{AA'BB'}$ and $V^x_{AA'BB'}$ satisfying the constraints in (16) for $\mathcal{M}^x_{AB\to A'B'}$ for $x \in \{0,1\}$, we conclude (136). ∎

C. Related measures

We now define variations of the bipartite channel measure from (16). We employ generalized divergences to do so, and in doing so, we arrive at a large number of variations of the basic bipartite channel measure.

Let $\boldsymbol{D}$ denote a generalized divergence [47], [48], which is a function that satisfies the following data-processing inequality, for every state ρ, positive semi-definite operator σ, and quantum channel $\mathcal{N}$:

$$\boldsymbol{D}(\rho\|\sigma) \geq \boldsymbol{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \quad (142)$$

In this paper, we make two additional minimal assumptions about a generalized divergence:

1) First, we assume that

$$\boldsymbol{D}(1\|c) \geq 0 \quad (143)$$

for $c \in (0,1]$. That is, if we plug in a trivial one-dimensional density operator ρ (i.e., the number 1) and a trivial positive semi-definite operator with trace less than or equal to one (i.e., $c \in (0,1]$), then the generalized divergence evaluates to a non-negative real.

2) Next, we assume that

$$\boldsymbol{D}(\rho\|\rho) = 0 \quad (144)$$

for every state ρ. We should clarify that this assumption is quite minimal. The reason is that it is essentially a direct consequence of (142) up to an inessential additive factor. That is, (142) implies that there exists a constant c such that

$$\boldsymbol{D}(\rho\|\rho) = c \quad (145)$$

for every state ρ. To see this, consider that one can get from the state ρ to another state ω by means of a trace and replace channel $\mathrm{Tr}[\cdot]\omega$, so that (142) implies that

$$\boldsymbol{D}(\rho\|\rho) \geq \boldsymbol{D}(\omega\|\omega). \quad (146)$$

However, by the same argument, $\boldsymbol{D}(\omega\|\omega) \geq \boldsymbol{D}(\rho\|\rho)$, so that the claim holds. So the assumption in (144) amounts to a redefinition of the generalized divergence as

$$\boldsymbol{D}'(\rho\|\sigma) \coloneqq \boldsymbol{D}(\rho\|\sigma) - c. \quad (147)$$

Let us list particular choices of interest for a generalized divergence. The quantum relative entropy [49] is defined as

$$D(\rho\|\sigma) \coloneqq \mathrm{Tr}[\rho(\log_2\rho - \log_2\sigma)] \quad (148)$$

if $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ and it is equal to $+\infty$ otherwise. The sandwiched Rényi relative entropy is defined for all $\alpha \in (0,1)\cup(1,\infty)$ as [50], [51]

$$\widetilde{D}_\alpha(\rho\|\sigma) \coloneqq \lim_{\varepsilon\to 0^+}\frac{1}{\alpha-1}\log_2\mathrm{Tr}[(\sigma_\varepsilon^{-(1-\alpha)/2\alpha}\rho\sigma_\varepsilon^{-(1-\alpha)/2\alpha})^\alpha], \quad (149)$$

where $\sigma_\varepsilon \coloneqq \sigma + \varepsilon I$. In the case that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, we have the following simplification:

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log_2\mathrm{Tr}[(\sigma^{-(1-\alpha)/2\alpha}\rho\sigma^{-(1-\alpha)/2\alpha})^\alpha]. \quad (150)$$

Note that $\widetilde{D}_\alpha(\rho\|\sigma) = +\infty$ if $\alpha > 1$ and $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$. The sandwiched Rényi relative entropy obeys the data-processing inequality in (142) for $\alpha \in [1/2,1)\cup(1,\infty)$ [52], [53]. Some basic properties of the sandwiched Rényi relative entropy are as follows [50], [51]: for all $\alpha > \beta > 0$

$$\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\beta(\rho\|\sigma), \quad (151)$$

and

$$\lim_{\alpha\to 1}\widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma). \quad (152)$$

The Belavkin–Staszewski relative entropy [54] is defined as

$$\widehat{D}(\rho\|\sigma) \coloneqq \mathrm{Tr}[\rho\log_2(\rho^{1/2}\sigma^{-1}\rho^{1/2})] \quad (153)$$

if $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ and it is equal to $+\infty$ otherwise. The geometric Rényi relative entropy is defined for all $\alpha \in (0,1) \cup (1,\infty)$ as [55], [56], [57], [58]

$$\widehat{D}_\alpha(\rho\|\sigma) \coloneqq \lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1} \log_2 \operatorname{Tr}[\sigma_\varepsilon(\sigma_\varepsilon^{-1/2}\rho\sigma_\varepsilon^{-1/2})^\alpha], \tag{154}$$

where $\sigma_\varepsilon \coloneqq \sigma + \varepsilon I$. In the case that $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, we have the following simplification:

$$\widehat{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1} \log_2 \operatorname{Tr}[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha]. \tag{155}$$

The geometric Rényi relative entropy obeys the data-processing inequality in (142) for $\alpha \in (0,1) \cup (1,2]$. Some basic properties of the geometric Rényi relative entropy are as follows [58]: for all $\alpha > \beta > 0$

$$\widehat{D}_\alpha(\rho\|\sigma) \geq \widehat{D}_\beta(\rho\|\sigma), \tag{156}$$

and

$$\lim_{\alpha\to 1} \widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}(\rho\|\sigma). \tag{157}$$

We are also interested in the hypothesis testing relative entropy [59], [60], [61], defined for $\varepsilon \in [0,1]$ as

$$D_H^\varepsilon(\rho\|\sigma) \coloneqq -\log_2 \inf_{\Lambda\geq 0} \left\{\operatorname{Tr}[\Lambda\sigma] : \operatorname{Tr}[\Lambda\rho] \geq 1-\varepsilon,\ \Lambda \leq I\right\}. \tag{158}$$

The property in (143) holds for all of the relative entropies that we have listed above, while the property in (144) holds for all of them except for the hypothesis testing relative entropy. For the hypothesis testing relative entropy, the constant c in (145) is equal to $-\log_2(1-\varepsilon)$, and the alternative definition in (147) is sometimes used [62].

A generalized channel divergence between a quantum channel $\mathcal{N}_{A\to B}$ and a completely positive map $\mathcal{M}_{A\to B}$ is defined from a generalized divergence as follows [63]:

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) \coloneqq \sup_{\rho_{RA}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})), \tag{159}$$

where the optimization is with respect to every bipartite state ρ_{RA}, with the system R arbitrarily large. By a standard argument (detailed in [63]), the following simplification occurs

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) \coloneqq \sup_{\psi_{RA}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{M}_{A\to B}(\psi_{RA})), \tag{160}$$

where the optimization is with respect to all pure bipartite states with $R \simeq A$. Using this, we define the following:

Definition 10: For a bipartite channel $\mathcal{N}_{AB\to A'B'}$, we define the following measure of forward classical communication:

$$\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) \coloneqq \inf_{\mathcal{M}_{AB\to A'B'}:\beta(\mathcal{M}_{AB\to A'B'})\leq 1} \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}\|\mathcal{M}_{AB\to A'B'}), \tag{161}$$

where the optimization is with respect to every completely positive bipartite map $\mathcal{M}_{AB\to A'B'}$.

Using the quantum relative entropy, the sandwiched Rényi relative entropy, the Belavkin–Staszewski relative entropy, and the geometric Rényi relative entropy, we then obtain the following respective channel measures:

$$\Upsilon(\mathcal{N}_{AB\to A'B'}), \tag{162}$$

$$\widetilde{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'}), \tag{163}$$

$$\widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'}), \tag{164}$$

$$\widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'}), \tag{165}$$

defined by substituting $\boldsymbol{D}$ with D, $\widetilde{D}_\alpha$, $\widehat{D}$, and $\widehat{D}_\alpha$ in (161).

We now establish some properties of $\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'})$, analogous to those established earlier for $C_\beta(\mathcal{N}_{AB\to A'B'})$ in Section III-B.

Proposition 11 (Non-negativity): Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Then

$$\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) \geq 0. \tag{166}$$

Proof: Let $\mathcal{M}_{AB\to A'B'}$ be an arbitrary completely positive bipartite map satisfying $\beta(\mathcal{M}_{AB\to A'B'}) \leq 1$. Then consider that

$$\begin{aligned}
&\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}\|\mathcal{M}_{AB\to A'B'}) \\
&\geq \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})\|\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})) \\
&\geq \boldsymbol{D}(\operatorname{Tr}[\mathcal{N}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})]\| \\
&\qquad \operatorname{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})]) &(167)\\
&= \boldsymbol{D}(1\|\operatorname{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})]). &(168)
\end{aligned}$$

The first inequality follows because the quantity $\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}\|\mathcal{M}_{AB\to A'B'})$ involves an optimization over all possible input states, and we have chosen the product of maximally entangled states. The second inequality follows from the data-processing inequality for the generalized divergence. It then follows from Definition 10 that

$$\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) \geq \inf_{\substack{\mathcal{M}_{AB\to A'B'}:\\ \beta(\mathcal{M}_{AB\to A'B'})\leq 1}} \boldsymbol{D}(1\|\operatorname{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})]). \tag{169}$$

Thus, the inequality follows if we can show that

$$\operatorname{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})] \leq 1. \tag{170}$$

Let λ, $S_{AA'BB'}$, and $V_{AA'BB'}$ be arbitrary Hermitian operators satisfying the constraints in (27) for $\mathcal{M}_{AB\to A'B'}$. Then, we find that

$$\begin{aligned}
\lambda d_A d_B &= \lambda \operatorname{Tr}_{AB}[I_{AB}] &(171)\\
&\geq \operatorname{Tr}_{AA'BB'}[S_{AA'BB'}] &(172)\\
&\geq \operatorname{Tr}_{AA'BB'}[V_{AA'BB'}] &(173)\\
&= \operatorname{Tr}_{AA'BB'}[T_{BB'}(V_{AA'BB'})] &(174)\\
&\geq \operatorname{Tr}_{AA'BB'}[T_{BB'}(\Gamma^{\mathcal{M}}_{AA'BB'})] &(175)\\
&= \operatorname{Tr}_{AA'BB'}[\Gamma^{\mathcal{M}}_{AA'BB'}] &(176)\\
&= \operatorname{Tr}[\Gamma^{\mathcal{M}}_{AA'BB'}], &(177)
\end{aligned}$$

which is equivalent to

$$\lambda \geq \operatorname{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})]. \tag{178}$$

Taking an infimum over λ, $S_{AA'BB'}$, and $V_{AA'BB'}$ satisfying the constraints in (27) for $\mathcal{M}_{AB\to A'B'}$ and applying the assumption $\beta(\mathcal{M}_{AB\to A'B'}) \leq 1$, we conclude (170). ∎

Proposition 12 (Stability): Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Then

$$\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) = \boldsymbol{\Upsilon}(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{N}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}). \quad (179)$$

Proof: The definition of the generalized channel divergence in (159) implies that it is stable, in the sense that

$$\begin{aligned}\boldsymbol{D}(\mathcal{N}_{AB\to A'B'} \| \mathcal{M}_{AB\to A'B'}) = \\ \boldsymbol{D}(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{N}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}} \| \\ \mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}),\end{aligned} \quad (180)$$

for every channel $\mathcal{N}_{AB\to A'B'}$ and completely positive map $\mathcal{M}_{AB\to A'B'}$. Combining with Proposition 3 and the definition in (161), we conclude (179). ∎

Proposition 13 (Zero on classical feedback channels): Let $\overline{\Delta}_{B\to A'}$ be a classical feedback channel:

$$\overline{\Delta}_{B\to A'}(\cdot) := \sum_{i=0}^{d-1} |i\rangle_{A'} \langle i|_B (\cdot) |i\rangle_B \langle i|_{A'}, \quad (181)$$

where $A' \simeq B$ and $d = d_{A'} = d_B$. Then

$$\boldsymbol{\Upsilon}(\overline{\Delta}_{B\to A'}) = 0. \quad (182)$$

Proof: This follows from Proposition 4. Since $\beta(\overline{\Delta}_{B\to A'}) = 1$, we can pick $\mathcal{M}_{B\to A'} = \overline{\Delta}_{B\to A'}$, and then

$$\boldsymbol{D}(\overline{\Delta}_{B\to A'} \| \mathcal{M}_{B\to A'}) = \boldsymbol{D}(\overline{\Delta}_{B\to A'} \| \overline{\Delta}_{B\to A'}) = 0. \quad (183)$$

So this establishes that $\boldsymbol{\Upsilon}(\overline{\Delta}_{B\to A'}) \leq 0$, and the other inequality $\boldsymbol{\Upsilon}(\overline{\Delta}_{B\to A'}) \geq 0$ follows from Proposition 11. ∎

Proposition 14 (Zero on tensor product of local channels): Let $\mathcal{E}_{A\to A'}$ and $\mathcal{F}_{B\to B'}$ be quantum channels. Then

$$\boldsymbol{\Upsilon}(\mathcal{E}_{A\to A'} \otimes \mathcal{F}_{B\to B'}) = 0. \quad (184)$$

Proof: Same argument as given for Proposition 13, but use Proposition 5 instead. ∎

We now establish some properties that are more specific to the Belavkin–Staszewski and geometric Rényi relative entropies (however the first actually holds for quantum relative entropy and other quantum Rényi relative entropies).

Proposition 15: Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Then for all $\alpha \in (1, 2]$,

$$\widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'}) \leq C_\beta(\mathcal{N}_{AB\to A'B'}). \quad (185)$$

Proof: Pick $\mathcal{M}_{AB\to A'B'} = \frac{1}{\beta(\mathcal{N}_{AB\to A'B'})} \mathcal{N}_{AB\to A'B'}$ in the definition of $\widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'})$ and $\widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'})$ and use the fact that, for $c > 0$, $\widehat{D}(\rho\|c\sigma) = \widehat{D}(\rho\|\sigma) - \log_2 c$ and $\widehat{D}_\alpha(\rho\|c\sigma) = \widehat{D}_\alpha(\rho\|\sigma) - \log_2 c$ for all $\alpha \in (1, 2]$. We also require (156). ∎

Proposition 16 (Subadditivity): For bipartite channels $\mathcal{N}^1_{AB\to A'B'}$ and $\mathcal{N}^2_{A'B'\to A''B''}$, the following inequality holds for all $\alpha \in (0, 1) \cup (1, 2]$:

$$\widehat{\Upsilon}_\alpha(\mathcal{N}^2_{A'B'\to A''B''} \circ \mathcal{N}^1_{AB\to A'B'}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}^2_{A'B'\to A''B''}) + \widehat{\Upsilon}_\alpha(\mathcal{N}^1_{AB\to A'B'}). \quad (186)$$

Proof: This inequality is a direct consequence of the subadditivity inequality in Eq. (18) of [22], Proposition 47 of [58], and the fact that if $\mathcal{M}^1$ and $\mathcal{M}^2$ are completely positive bipartite maps satisfying $\beta(\mathcal{M}^1), \beta(\mathcal{M}^2) \leq 1$, then $\beta(\mathcal{M}^2 \circ \mathcal{M}^1) \leq 1$ (see Proposition 6). ∎

D. Measure of classical communication for a point-to-point channel

Let $\mathcal{M}_{A\to B'}$ be a point-to-point completely positive map, which is a special case of a completely positive bipartite map with the Bob input B trivial and the Alice output A' trivial. We first show that β in (16) reduces to the measure from [20].

Proposition 17: Let $\mathcal{M}_{A\to B'}$ be a point-to-point completely positive map. Then

$$\beta(\mathcal{M}_{A\to B'}) = \inf_{S_{B'}, V_{AB'} \in \mathrm{Herm}} \left\{ \begin{array}{c} \mathrm{Tr}[S_{B'}] : \\ T_{B'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ I_A \otimes S_{B'} \pm V_{AB'} \geq 0 \end{array} \right\}. \quad (187)$$

Proof: In this case, the systems A' and B are trivial. So then the definition in (16) reduces to

$$\beta(\mathcal{M}_{A\to B'}) = \inf_{S_{AB'}, V_{AB'} \in \mathrm{Herm}} \left\{ \begin{array}{c} \|\mathrm{Tr}_{B'}[S_{AB'}]\|_\infty : \\ T_{B'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ S_{AB'} \pm V_{AB'} \geq 0, \\ S_{AB'} = \pi_A \otimes \mathrm{Tr}_A[S_{AB'}] \end{array} \right\}. \quad (188)$$

The last constraint implies that the optimization simplifies to

$$\begin{aligned}&\beta(\mathcal{M}_{A\to B'}) \\ &= \inf_{S_{AB'}, V_{AB'} \in \mathrm{Herm}} \left\{ \begin{array}{c} \|\mathrm{Tr}_{B'}[\pi_A \otimes \mathrm{Tr}_A[S_{AB'}]]\|_\infty : \\ T_{B'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes \mathrm{Tr}_A[S_{AB'}] \pm V_{AB'} \geq 0 \end{array} \right\} && (189) \\ &= \inf_{S'_{B'}, V_{AB'} \in \mathrm{Herm}} \left\{ \begin{array}{c} \|\mathrm{Tr}_{B'}[\pi_A \otimes S'_{B'}]\|_\infty : \\ T_{B'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes S'_{B'} \pm V_{AB'} \geq 0 \end{array} \right\} && (190) \\ &= \inf_{S'_{B'}, V_{AB'} \in \mathrm{Herm}} \left\{ \begin{array}{c} \mathrm{Tr}[S'_{B'}] \, \|\pi_A\|_\infty : \\ T_{B'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes S'_{B'} \pm V_{AB'} \geq 0 \end{array} \right\} && (191) \\ &= \inf_{S'_{B'}, V_{AB'} \in \mathrm{Herm}} \left\{ \begin{array}{c} \frac{1}{d_A} \mathrm{Tr}[S'_{B'}] : \\ T_{B'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes S'_{B'} \pm V_{AB'} \geq 0 \end{array} \right\} && (192) \\ &= \inf_{S_{B'}, V_{AB'} \in \mathrm{Herm}} \left\{ \begin{array}{c} \mathrm{Tr}[S_{B'}] : \\ T_{B'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ I_A \otimes S_{B'} \pm V_{AB'} \geq 0 \end{array} \right\}. && (193)\end{aligned}$$

This concludes the proof. ∎

More generally, consider that the definition in (161) becomes as follows for a point-to-point channel $\mathcal{N}_{A\to B'}$:

$$\boldsymbol{\Upsilon}(\mathcal{N}_{A\to B'}) := \inf_{\mathcal{M}_{A\to B'} : \beta(\mathcal{M}_{A\to B'}) \leq 1} \boldsymbol{D}(\mathcal{N}_{A\to B'} \| \mathcal{M}_{A\to B'}), \quad (194)$$

which leads to the quantities $\widehat{\Upsilon}(\mathcal{N}_{A\to B'})$ and $\widehat{\Upsilon}_\alpha(\mathcal{N}_{A\to B'})$, for which we have the following bounds for $\alpha \in (1,2]$:

$$\widehat{\Upsilon}(\mathcal{N}_{A\to B'}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_{A\to B'}) \leq C_\beta(\mathcal{N}_{A\to B'}). \tag{195}$$

Note that the quantities given just above were defined in [21], [22], and our observation here is that the definition in (161) reduces to them.

The next proposition is critical for establishing our upper bound proofs in Section IV. It states that if one share of a maximally classically correlated state passes through a completely positive map $\mathcal{M}_{A\to B'}$ for which $\beta(\mathcal{M}_{A\to B'}) \leq 1$, then the resulting operator has a very small chance of passing the comparator test, as defined in (198).

Proposition 18 (Bound for comparator test success probability): Let

$$\overline{\Phi}_{\hat{A}A} \coloneqq \frac{1}{d}\sum_{i=0}^{d-1} |i\rangle\langle i|_{\hat{A}} \otimes |i\rangle\langle i|_A \tag{196}$$

denote the maximally classically correlated state, and let $\mathcal{M}_{A\to B'}$ be a completely positive map for which $\beta(\mathcal{M}_{A\to B'}) \leq 1$. Then

$$\operatorname{Tr}[\Pi_{\hat{A}B'}\mathcal{M}_{A\to B'}(\overline{\Phi}_{\hat{A}A})] \leq \frac{1}{d}, \tag{197}$$

where $\Pi_{\hat{A}B'}$ is the comparator test:

$$\Pi_{\hat{A}B'} \coloneqq \sum_{i=0}^{d-1} |i\rangle\langle i|_{\hat{A}} \otimes |i\rangle\langle i|_{B'}, \tag{198}$$

and $\hat{A} \simeq A \simeq B'$.

Proof: Recall the expression for $\beta(\mathcal{M}_{A\to B'})$ in (187). Let $S_{B'}$ and $V_{AB'}$ be arbitrary Hermitian operators satisfying the constraints for $\beta(\mathcal{M}_{A\to B'})$. An application of (8) implies that

$$\mathcal{M}_{A\to B'}(\overline{\Phi}_{\hat{A}A}) = \langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes \Gamma^{\mathcal{M}}_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}, \tag{199}$$

where $\tilde{A} \simeq A$. This means that

$$\begin{aligned}
&\operatorname{Tr}[\Pi_{\hat{A}B'}\mathcal{M}_{A\to B'}(\overline{\Phi}_{\hat{A}A})] \\
&= \operatorname{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes \Gamma^{\mathcal{M}}_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}] && (200)\\
&= \operatorname{Tr}[T_{B'}(\Pi_{\hat{A}B'})\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes \Gamma^{\mathcal{M}}_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}] && (201)\\
&= \operatorname{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes T_{B'}(\Gamma^{\mathcal{M}}_{\tilde{A}B'})|\Gamma\rangle_{A\tilde{A}}] && (202)\\
&\leq \operatorname{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes T_{B'}(V_{\tilde{A}B'})|\Gamma\rangle_{A\tilde{A}}] && (203)\\
&= \operatorname{Tr}[T_{B'}(\Pi_{\hat{A}B'})\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes V_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}] && (204)\\
&= \operatorname{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes V_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}] && (205)\\
&\leq \operatorname{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes I_{\tilde{A}} \otimes S_{B'}|\Gamma\rangle_{A\tilde{A}}] && (206)\\
&= \operatorname{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes I_{\tilde{A}}|\Gamma\rangle_{A\tilde{A}} \otimes S_{B'}] && (207)\\
&= \operatorname{Tr}[\Pi_{\hat{A}B'}\operatorname{Tr}_A[\overline{\Phi}_{\hat{A}A}] \otimes S_{B'}] && (208)\\
&= \frac{1}{d}\operatorname{Tr}[\Pi_{\hat{A}B'}I_{\hat{A}} \otimes S_{B'}] && (209)\\
&= \frac{1}{d}\operatorname{Tr}[S_{B'}]. && (210)
\end{aligned}$$

Since this holds for all $S_{B'}$ and $V_{AB'}$ satisfying the constraints for $\beta(\mathcal{M}_{A\to B'})$, we conclude that

$$\operatorname{Tr}[\Pi_{\hat{A}B'}\mathcal{M}_{A\to B'}(\overline{\Phi}_{\hat{A}A})] \leq \frac{1}{d}. \tag{211}$$

This concludes the proof. ∎

We finally state another proposition that plays an essential role in our upper bound proofs in Section IV.

Proposition 19: Suppose that $\mathcal{N}_{A\to B}$ is a channel with $A \simeq B$ that satisfies

$$\frac{1}{2}\left\|\mathcal{N}_{A\to B}(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1 \leq \varepsilon, \tag{212}$$

for $\varepsilon \in [0,1)$ and where $\overline{\Phi}_{RB} \coloneqq \frac{1}{d}\sum_i |i\rangle\langle i|_R \otimes |i\rangle\langle i|_B$ and $d = d_R = d_A = d_B$. Then

$$\log_2 d \leq \inf_{\mathcal{M}_{A\to B}:\beta(\mathcal{M}_{A\to B})\leq 1} D_H^\varepsilon(\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})\|\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})), \tag{213}$$

and for all $\alpha \in (1,2]$,

$$\log_2 d \leq \inf_{\mathcal{M}_{A\to B}:\beta(\mathcal{M}_{A\to B})\leq 1} \widehat{D}_\alpha(\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})\|\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right). \tag{214}$$

Proof: We begin by proving (213). The condition

$$\frac{1}{2}\left\|\mathcal{N}_{A\to B}(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1 \leq \varepsilon \tag{215}$$

implies that

$$\operatorname{Tr}[\Pi_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \geq 1-\varepsilon, \tag{216}$$

where $\Pi_{RB} \coloneqq \sum_i |i\rangle\langle i|_R \otimes |i\rangle\langle i|_B$ is the comparator test. Indeed, applying a completely dephasing channel $\overline{\Delta}_B(\cdot) \coloneqq \sum_i |i\rangle\langle i|(\cdot)|i\rangle\langle i|$ to the output of the channel $\mathcal{N}_{A\to B}$ and applying the data-processing inequality for trace distance, we conclude that

$$\begin{aligned}
\varepsilon &\geq \frac{1}{2}\left\|\mathcal{N}_{A\to B}(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1 && (217)\\
&\geq \frac{1}{2}\left\|(\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA}) - \overline{\Delta}_B(\overline{\Phi}_{RB})\right\|_1 && (218)\\
&= \frac{1}{2}\left\|(\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1. && (219)
\end{aligned}$$

Let $\omega_{RB} \coloneqq (\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA})$ and observe that it can be written as

$$\omega_{RB} = \frac{1}{d}\sum_{i,j} p(j|i)|i\rangle\langle i|_R \otimes |j\rangle\langle j|_B \tag{220}$$

for some conditional probability distribution $p(j|i)$. Then

$$\begin{aligned}
&\frac{1}{2}\left\|(\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1 \\
&= \frac{1}{2}\left\| \begin{array}{c} \frac{1}{d}\sum_{i,j} p(j|i)|i\rangle\langle i|_R \otimes |j\rangle\langle j|_B \\ -\frac{1}{d}\sum_{i,j}\delta_{i,j}|i\rangle\langle i|_R \otimes |j\rangle\langle j|_B \end{array} \right\|_1 && (221)\\
&= \frac{1}{2d}\sum_i \left\|\sum_j (p(j|i) - \delta_{i,j})|j\rangle\langle j|_B\right\|_1 && (222)\\
&= \frac{1}{2d}\sum_i \left[(1-p(i|i)) + \sum_{j\neq i} p(j|i)\right] && (223)
\end{aligned}$$

$$= \frac{1}{d}\sum_i (1 - p(i|i)) \tag{224}$$

$$= 1 - \sum_i \frac{1}{d} p(i|i). \tag{225}$$

This implies that

$$\sum_i \frac{1}{d} p(i|i) \geq 1 - \varepsilon. \tag{226}$$

Now consider that

$$\begin{aligned}
&\operatorname{Tr}[\Pi_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \\
&= \operatorname{Tr}[\overline{\Delta}_B(\Pi_{RB})\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] && (227)\\
&= \operatorname{Tr}[\Pi_{RB}(\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA})] && (228)\\
&= \operatorname{Tr}[\Pi_{RB}\omega_{RB}] && (229)\\
&= \sum_i \frac{1}{d} p(i|i). && (230)
\end{aligned}$$

So we conclude that

$$\operatorname{Tr}[\Pi_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \geq 1 - \varepsilon. \tag{231}$$

Applying the definition of the hypothesis testing relative entropy from (158), we conclude that

$$\inf_{\mathcal{M}_{A\to B}:\beta(\mathcal{M}_{A\to B})\leq 1} D_H^\varepsilon(\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})\|\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})) =$$

$$-\log_2 \sup_{\substack{\mathcal{M}_{A\to B}:\\ \beta(\mathcal{M}_{A\to B})\leq 1}} \inf_{\Lambda_{RB}\geq 0} \left\{ \begin{array}{c} \operatorname{Tr}[\Lambda_{RB}\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})] : \\ \operatorname{Tr}[\Lambda_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \\ \geq 1-\varepsilon, \\ \Lambda_{RB} \leq I_{RB} \end{array} \right\}.$$

Now consider that

$$\begin{aligned}
&\sup_{\substack{\mathcal{M}_{A\to B}:\\ \beta(\mathcal{M}_{A\to B})\leq 1}} \inf_{\Lambda_{RB}\geq 0} \left\{ \begin{array}{c} \operatorname{Tr}[\Lambda_{RB}\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})] : \\ \operatorname{Tr}[\Lambda_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \geq 1-\varepsilon, \\ \Lambda_{RB} \leq I_{RB} \end{array} \right\} \\
&\leq \sup_{\mathcal{M}_{A\to B}:\beta(\mathcal{M}_{A\to B})\leq 1} \operatorname{Tr}[\Pi_{RB}\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})] && (232)\\
&\leq \frac{1}{d}, && (233)
\end{aligned}$$

where the first inequality follows from (231) and the definition in (158) and the last inequality follows from Proposition 18. Then applying a negative logarithm gives (213).

The inequality in (214) follows as a direct application of the following relationship between the hypothesis testing relative entropy and the geometric Rényi relative entropy:

$$D_H^\varepsilon(\rho\|\sigma) \leq \widehat{D}_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha - 1}\log_2\left(\frac{1}{1-\varepsilon}\right), \tag{234}$$

as well as the previous proposition. The proof of (234) follows the same proof given for [64, Lemma 5]. ∎

IV. Applications

A. Bounding the forward classical capacity of a bipartite channel

We now apply the bipartite channel measure in (161) to obtain an upper bound on the forward classical capacity of a bipartite channel $\mathcal{N}_{AB\to A'B'}$. We begin by describing a forward classical communication protocol for a bipartite channel and then define the associated capacities.

Fix $n, M \in \mathbb{N}$ and $\varepsilon \in [0,1]$. An (n, M, ε) protocol for forward classical communication using a bipartite channel $\mathcal{N}_{AB\to A'B'}$ begins with a reference party preparing the state $\overline{\Phi}^p_{R\hat{A}}$ and sending the $\hat{A}$ system to Alice, where $\overline{\Phi}^p_{R\hat{A}}$ is the following classically correlated state:

$$\overline{\Phi}^p_{R\hat{A}} := \sum_{m=1}^M p(m)|m\rangle\langle m|_R \otimes |m\rangle\langle m|_{\hat{A}}, \tag{235}$$

and $p(m)$ is a probability distribution over the messages. Alice acts on the system $\hat{A}$ with a local encoding channel $\mathcal{E}^{(0)}_{\hat{A}\to A_1''A_1}$, resulting in the following state:

$$\sigma_{RA_1''A_1} := \mathcal{E}^{(0)}_{\hat{A}\to A_1''A_1}(\overline{\Phi}^p_{R\hat{A}}). \tag{236}$$

Bob prepares the local state $\tau_{B_1''B_1}$, so that the initial global state of the reference, Alice, and Bob is

$$\sigma_{RA_1''A_1} \otimes \tau_{B_1''B_1}. \tag{237}$$

The systems A_1B_1 are then fed into the first use of the channel, producing the output state

$$\rho^{(1)}_{RA_1'A_1'B_1''B_1'} := \mathcal{N}_{A_1B_1\to A_1'B_1'}(\sigma^{(1)}_{RA_1''A_1B_1''B_1}), \tag{238}$$

$$\sigma^{(1)}_{RA_1'A_1B_1''B_1} := \sigma_{A_1'A_1} \otimes \tau_{B_1''B_1}. \tag{239}$$

Then Alice applies the local channel $\mathcal{E}^{(1)}_{A_1'A_1'\to A_2''A_2}$ to her systems, and Bob applies the local channel $\mathcal{F}^{(1)}_{B_1''B_1'\to B_2''B_2}$ to his systems. The systems A_2B_2 are fed into the next channel use, leading to the state

$$\rho^{(2)}_{RA_2'A_2'B_2''B_2'} := \mathcal{N}_{A_2B_2\to A_2'B_2'}(\sigma^{(2)}_{RA_2''A_2B_2''B_2}), \tag{240}$$

$$\sigma^{(2)}_{RA_2''A_2B_2''B_2} := (\mathcal{E}^{(1)}_{A_1''A_1'\to A_2''A_2} \otimes \mathcal{F}^{(1)}_{B_1''B_1'\to B_2''B_2})(\rho^{(1)}_{RA_1''A_1'B_1''B_1'}). \tag{241}$$

This process iterates $n-2$ more times, and we define

$$\rho^{(i)}_{RA_i'A_i'B_i''B_i'} := \mathcal{N}_{A_iB_i\to A_i'B_i'}(\sigma^{(i)}_{RA_i''A_iB_i''B_i}), \tag{242}$$

$$\sigma^{(i)}_{RA_i''A_iB_i''B_i} := (\mathcal{E}^{(i-1)}_{A_{i-1}''A_{i-1}'\to A_i''A_i} \otimes \mathcal{F}^{(i-1)}_{B_{i-1}''B_{i-1}'\to B_i''B_i})(\rho^{(i-1)}_{RA_{i-1}''A_{i-1}'B_{i-1}''B_{i-1}'}), \tag{243}$$

for $i \in \{3, \ldots, n\}$. The final channel output state $\rho^{(n)}_{RA_n''A_n'B_n''B_n'}$ is processed a final time with the local channels $\mathcal{E}^{(n)}_{A_n''A_n'\to\emptyset} := \operatorname{Tr}_{A_n''A_n'}$ and $\mathcal{F}^{(n)}_{B_n''B_n'\to\hat{B}}$, where $\emptyset$ indicates a trivial system, to produce the final protocol state

$$\omega^p_{R\hat{B}} := (\mathcal{E}^{(n)}_{A_n''A_n'\to\emptyset} \otimes \mathcal{F}^{(n)}_{B_n''B_n'\to\hat{B}})(\rho^{(n)}_{RA_n''A_n'B_n''B_n'}). \tag{244}$$

For an (n, M, ε) protocol, the final state $\omega^p_{R\hat{B}}$ satisfies the following condition

$$\max_p \frac{1}{2}\left\|\omega^p_{R\hat{B}} - \overline{\Phi}^p_{R\hat{B}}\right\|_1 \leq \varepsilon, \tag{245}$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

where the maximization is over all probability distributions $p(m)$ and

$$\overline{\Phi}^{p}_{R\hat{B}} := \sum_{m=1}^{M} p(m)|m\rangle\langle m|_{\hat{A}} \otimes |m\rangle\langle m|_{\hat{B}}. \tag{246}$$

Note that the condition in (245) is equivalent to the traditional condition on the decoding error probability (see [65, Lemma 6.2]). Figure 2 depicts such a protocol with $n = 4$.

Let us denote the set consisting of the initially prepared state and the sequence of local channels as the protocol $\mathcal{P}^{(n)}$:

$$\mathcal{P}^{(n)} := \left\{ \mathcal{E}^{(0)}_{\hat{A}\to A_1''A_1} \otimes \tau_{B_1''B_1}, \left\{ \mathcal{E}^{(i-1)}_{A_{i-1}''A_{i-1}'\to A_i''A_i} \otimes \mathcal{F}^{(i-1)}_{B_{i-1}''B_{i-1}'\to B_i''B_i} \right\}_{i=2}^{n}, \mathcal{E}^{(n)}_{A_n''A_n'\to\emptyset} \otimes \mathcal{F}^{(n)}_{B_n''B_n'\to\hat{B}} \right\}. \tag{247}$$

Then we can write the final state $\omega_{R\hat{B}}$ as

$$\omega^{p}_{R\hat{B}} = \mathcal{C}_{\hat{A}\to\hat{B}}(\overline{\Phi}^{p}_{R\hat{A}}), \tag{248}$$

where

$$\mathcal{C}_{\hat{A}\to\hat{B}} := \mathcal{L}^{(n)} \circ \mathcal{N} \circ \mathcal{L}^{(n-1)} \cdots \circ \mathcal{L}^{(2)} \circ \mathcal{N} \circ \mathcal{L}^{(1)} \circ \mathcal{N} \circ \mathcal{L}^{(0)}, \tag{249}$$

$\mathcal{L}^{(0)}$ acts on system $\hat{A}$ of $\overline{\Phi}_{R\hat{A}}$ to prepare the state $\sigma^{(1)}_{RA_1''A_1B_1''B_1}$,

$$\mathcal{L}^{(i-1)} \equiv \mathcal{E}^{(i-1)}_{A_{i-1}''A_{i-1}'\to A_i''A_i} \otimes \mathcal{F}^{(i-1)}_{B_{i-1}''B_{i-1}'\to B_i''B_i}, \tag{250}$$

for $i \in \{2, \ldots, n\}$, and

$$\mathcal{L}^{(n)} \equiv \mathcal{E}^{(n)}_{A_n''A_n'\to\emptyset} \otimes \mathcal{F}^{(n)}_{B_n''B_n'\to\hat{B}}. \tag{251}$$

The n-shot forward classical capacity of a bipartite channel $\mathcal{N}_{AB\to A'B'}$ is then defined as follows:

$$C^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}) := \sup_{M\in\mathbb{N},\ \mathcal{P}^{(n)}} \left\{ \frac{1}{n}\log_2 M : \exists(n, M, \varepsilon) \text{ protocol } \mathcal{P}^{(n)} \right\}. \tag{252}$$

The forward classical capacity and strong converse forward classical capacity of the bipartite channel $\mathcal{N}_{AB\to A'B'}$ are defined as

$$C(\mathcal{N}_{AB\to A'B'}) := \inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} C^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}), \tag{253}$$

$$\widetilde{C}(\mathcal{N}_{AB\to A'B'}) := \sup_{\varepsilon\in(0,1)} \limsup_{n\to\infty} C^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}). \tag{254}$$

From the definitions, it is clear that

$$C(\mathcal{N}_{AB\to A'B'}) \leq \widetilde{C}(\mathcal{N}_{AB\to A'B'}). \tag{255}$$

An (n, M, ε) randomness transmission protocol is exactly as specified above, but with $p(m) = 1/M$ (i.e., the uniform distribution) in (235). Let us define

$$\overline{\Phi}_{R\hat{A}} := \frac{1}{M}\sum_{m=1}^{M} |m\rangle\langle m|_R \otimes |m\rangle\langle m|_{\hat{A}}. \tag{256}$$

Then the error criterion for such a protocol is

$$\frac{1}{2}\left\|\omega_{R\hat{B}} - \overline{\Phi}_{R\hat{B}}\right\|_1 \leq \varepsilon, \tag{257}$$

where $\omega_{R\hat{B}}$ is defined as in (248) but with $\overline{\Phi}^{p}_{R\hat{A}}$ replaced by $\overline{\Phi}_{R\hat{A}}$. Also,

$$\overline{\Phi}_{R\hat{B}} := \frac{1}{M}\sum_{m=1}^{M} |m\rangle\langle m|_{\hat{A}} \otimes |m\rangle\langle m|_{\hat{B}}. \tag{258}$$

Note that the condition in (257) is equivalent to the traditional condition on the average decoding error probability (see [65, Lemma 6.2]).

We define the following quantities for the randomness transmission capacity of $\mathcal{N}_{AB\to A'B'}$:

$$R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}) = \sup_{M\in\mathbb{N},\ \mathcal{P}^{(n)}} \left\{ \frac{1}{n}\log_2 M : \exists(n, M, \varepsilon) \text{ RT protocol } \mathcal{P}^{(n)} \right\}, \tag{259}$$

where RT is an abbreviation for "randomness transmission." The randomness transmission capacity and strong converse randomness transmission capacity of the bipartite channel $\mathcal{N}_{AB\to A'B'}$ are defined as

$$R(\mathcal{N}_{AB\to A'B'}) := \inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}), \tag{260}$$

$$\widetilde{R}(\mathcal{N}_{AB\to A'B'}) := \sup_{\varepsilon\in(0,1)} \limsup_{n\to\infty} R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}). \tag{261}$$

From the definitions, it is clear that

$$R(\mathcal{N}_{AB\to A'B'}) \leq \widetilde{R}(\mathcal{N}_{AB\to A'B'}). \tag{262}$$

Since every (n, M, ε) forward classical communication protocol is an (n, M, ε) randomness transmission protocol, the following inequality holds

$$C^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}) \leq R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}). \tag{263}$$

By the standard expurgation argument (throwing away the worst half of the codewords to give maximal error probability $\leq 2\varepsilon$; see, e.g., [4, Exercise 2.2.1]), the following inequality holds

$$R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}) - \frac{1}{n} \leq C^{n,2\varepsilon}(\mathcal{N}_{AB\to A'B'}). \tag{264}$$

By employing definitions, we conclude that

$$C(\mathcal{N}_{AB\to A'B'}) = R(\mathcal{N}_{AB\to A'B'}) \tag{265}$$

$$\leq \widetilde{C}(\mathcal{N}_{AB\to A'B'}) \tag{266}$$

$$\leq \widetilde{R}(\mathcal{N}_{AB\to A'B'}). \tag{267}$$

In what follows, we establish an upper bound on the strong converse randomness transmission capacity of $\mathcal{N}_{AB\to A'B'}$, and by the inequalities above, this gives an upper bound on the forward classical capacity and strong converse forward classical capacity of $\mathcal{N}_{AB\to A'B'}$.

Theorem 20: The following upper bound holds for the n-shot randomness transmission capacity of a bipartite channel $\mathcal{N}_{AB\to A'B'}$:

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

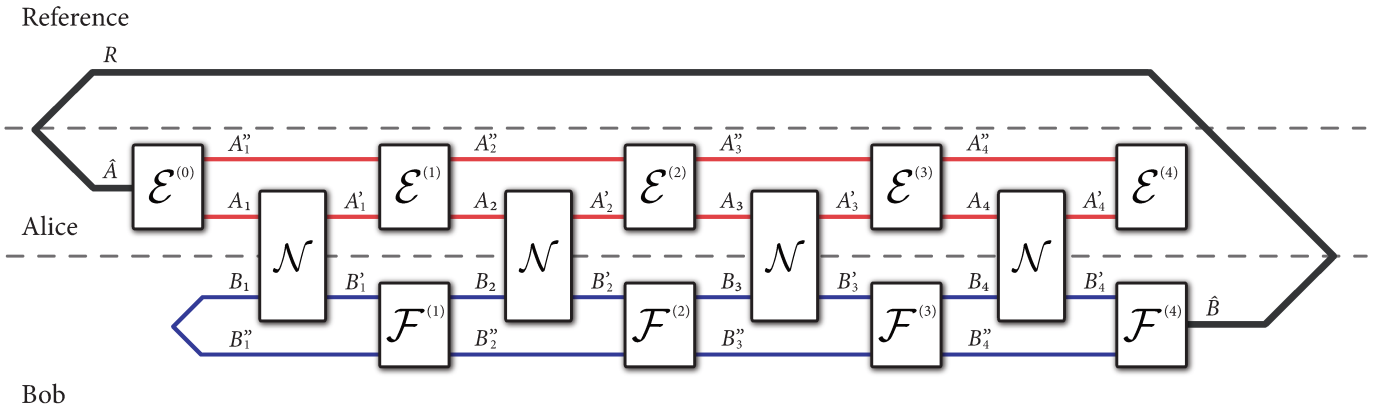

Fig. 2. Depiction of a protocol for forward classical communication over a bipartite channel, by making $n = 4$ uses of the bipartite channel $\mathcal{N}_{AB\to A'B'}$.

$$R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'}) + \frac{\alpha}{n\,(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right), \quad (268)$$

for all $\alpha \in (1,2]$ and $\varepsilon \in [0,1)$.

Proof: Consider an arbitrary n-shot randomness transmission protocol of the form described above. Focusing in particular on (245) and (248), we apply (214) of Proposition 19 to conclude that

$$\begin{aligned}
&\log_2 M \\
&\leq \inf_{\mathcal{M}_{\hat{A}\to\hat{B}}:\beta(\mathcal{M}_{\hat{A}\to\hat{B}})\leq 1} \widehat{D}_\alpha(\mathcal{C}_{\hat{A}\to\hat{B}}(\overline{\Phi}_{R\hat{A}})\|\mathcal{M}_{\hat{A}\to\hat{B}}(\overline{\Phi}_{R\hat{A}})) \\
&\quad + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right) && (269)\\
&\leq \widehat{\Upsilon}_\alpha(\mathcal{C}_{\hat{A}\to\hat{B}}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right), && (270)
\end{aligned}$$

where the inequality follows from the definition in (194) with $\boldsymbol{D}$ set to $\widehat{D}_\alpha$. Eq. (248) indicates that the whole protocol is a serial composition of bipartite channels. Then we find that

$$\begin{aligned}
&\widehat{\Upsilon}_\alpha(\mathcal{C}_{\hat{A}\to\hat{B}}) \\
&= \widehat{\Upsilon}_\alpha(\mathcal{L}^{(n)}\circ\mathcal{N}\circ\mathcal{L}^{(n-1)}\cdots\circ\mathcal{L}^{(2)}\circ\mathcal{N}\circ\mathcal{L}^{(1)}\circ\mathcal{N}\circ\mathcal{L}^{(0)}) && (271)\\
&\leq n\widehat{\Upsilon}_\alpha(\mathcal{N}) + \sum_{i=0}^{n}\widehat{\Upsilon}_\alpha(\mathcal{L}^{(i)}) && (272)\\
&= n\widehat{\Upsilon}_\alpha(\mathcal{N}). && (273)
\end{aligned}$$

The inequality follows from Proposition 6 and the last equality from Proposition 14. We also implicitly used the stability property in Proposition 12. Then we find that

$$\frac{1}{n}\log_2 M \leq \widehat{\Upsilon}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right). \quad (274)$$

Since the upper bound holds for an arbitrary protocol, this concludes the proof. ∎

Theorem 21: The following upper bound holds for the strong converse randomness transmission capacity of a bipartite channel $\mathcal{N}_{AB\to A'B'}$:

$$\widetilde{R}(\mathcal{N}_{AB\to A'B'}) \leq \widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'}), \quad (275)$$

where $\widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'})$ is defined from (161) using the Belavkin–Staszewski relative entropy.

Proof: Applying the bound in (268) and taking the $n\to\infty$ limit, we find that the following holds for all $\alpha\in(1,2]$ and $\varepsilon\in[0,1)$:

$$\begin{aligned}
&\limsup_{n\to\infty} R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}) \\
&\leq \limsup_{n\to\infty}\left[\widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'}) + \frac{\alpha}{n\,(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right)\right] && (276)\\
&= \widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'}). && (277)
\end{aligned}$$

Since the upper bound holds for all $\alpha\in(1,2]$, we can take the infimum over all these values, and we conclude that the following holds for all $\varepsilon\in[0,1)$:

$$\limsup_{n\to\infty} R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}) \leq \widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'}). \quad (278)$$

Here we applied the definitions of $\widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'})$ and $\widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'})$ and Proposition 36 in Appendix B. The upper bound holds for all $\varepsilon\in(0,1)$ and so we conclude the statement of the theorem. ∎

B. Bounding the classical capacity of a point-to-point quantum channel assisted by a classical feedback channel

One of the main applications in our paper is an upper bound on the classical capacity of a point-to-point quantum channel assisted by classical feedback. For a point-to-point channel $\mathcal{N}_{A\to B'}$, this capacity is denoted by $C_{\leftarrow}(\mathcal{N}_{A\to B'})$.

In what follows, we briefly define the classical capacity of a point-to-point quantum channel $\mathcal{N}_{A\to B'}$ assisted by classical feedback. Before doing so, let us first expand the notion of an n-shot protocol for forward classical communication from the previous section, such that each use of the bipartite channel is no longer constrained to be identical. The final state of such a protocol is then a generalization of that in (248):

$$\begin{aligned}
\omega^p_{R\hat{B}} = (&\mathcal{L}^{(n)}\circ\mathcal{N}^{(n)}\circ\mathcal{L}^{(n-1)}\circ\cdots \\
&\circ\mathcal{L}^{(2)}\circ\mathcal{N}^{(2)}\circ\mathcal{L}^{(1)}\circ\mathcal{N}^{(1)}\circ\mathcal{L}^{(0)})(\overline{\Phi}^p_{R\hat{A}}), \quad (279)
\end{aligned}$$

and the protocol is an (n, M, ε) protocol if the inequality

$$\max_p \frac{1}{2} \left\| \omega^p_{R\hat{B}} - \overline{\Phi}^p_{R\hat{B}} \right\|_1 \leq \varepsilon \tag{280}$$

holds with $\overline{\Phi}^p_{R\hat{B}}$ the classically correlated state as defined in (246). Note that the following bound holds for all (n, M, ε) forward classical communication protocols, with $n, M \in \mathbb{N}$ and $\varepsilon \in (0, 1]$, and $\alpha \in (1, 2]$:

$$\log_2 M \leq \sum_{i=1}^{n} \widehat{\Upsilon}_\alpha(\mathcal{N}^{(i)}_{AB\to A'B'}) + \frac{\alpha}{\alpha - 1} \log_2 \left(\frac{1}{1-\varepsilon} \right), \tag{281}$$

by following the same steps given in the proof of Theorem 20.

With the more general definition in hand, we define an (n, M, ε) protocol for classical communication over a point-to-point channel $\mathcal{N}_{A\to B'}$ assisted by a classical feedback channel as a special case of a $(2n, M, \varepsilon)$ protocol of the form above, in which every $\mathcal{N}^{(i)}$ with i odd is replaced by a classical feedback channel $\overline{\Delta}_{B_i\to A'_i}$ (with $d_i := d_{B_i} = d_{A'_i}$ and trivial input system A_i and trivial output system B'_i) and every $\mathcal{N}^{(i)}$ with i even is replaced by the forward point-to-point channel $\mathcal{N}_{A\to B'}$ (such that the input system B_i and the output system A'_i are trivial). The final state of the protocol is given by

$$\begin{aligned} &\omega_{R\hat{B}} := \\ &(\mathcal{L}^{(2n)} \circ \mathcal{N}_{A_{2n}\to B'_{2n}} \circ \mathcal{L}^{(2n-1)} \circ \overline{\Delta}_{B_{2n-1}\to A'_{2n-1}} \circ \mathcal{L}^{(2n-2)} \circ \\ &\quad \cdots \circ \mathcal{L}^{(2)} \circ \mathcal{N}_{A_2\to B'_2} \circ \mathcal{L}^{(1)} \circ \overline{\Delta}_{B_1\to A'_1} \circ \mathcal{L}^{(1)})(\overline{\Phi}_{R\hat{A}}). \end{aligned} \tag{282}$$

Let $\mathcal{P}^{(2n)}$ denote the protocol, which consists of $\mathcal{L}^{(0)}$, $\mathcal{L}^{(1)}$, $\ldots$, $\mathcal{L}^{(2n)}$. This protocol is depicted in Figure 3.

The n-shot classical capacity of the point-to-point channel $\mathcal{N}_{A\to B'}$ assisted by classical feedback is defined as

$$C^{n,\varepsilon}_{\leftarrow}(\mathcal{N}_{A\to B'}) = \sup_{M\in\mathbb{N},\ \mathcal{P}^{(2n)}} \left\{ \frac{1}{n} \log_2 M : \exists (n, M, \varepsilon) \text{ protocol } \mathcal{P}^{(2n)} \right\}. \tag{283}$$

That is, it is the largest rate at which messages can be transmitted up to an ε error probability. The classical capacity of the point-to-point channel $\mathcal{N}_{A\to B'}$ assisted by classical feedback is defined as the following limit:

$$C_{\leftarrow}(\mathcal{N}_{A\to B'}) := \inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} C^{n,\varepsilon}_{\leftarrow}(\mathcal{N}_{A\to B'}), \tag{284}$$

and the strong converse classical capacity as

$$\widetilde{C}_{\leftarrow}(\mathcal{N}_{A\to B'}) := \sup_{\varepsilon\in(0,1)} \limsup_{n\to\infty} C^{n,\varepsilon}_{\leftarrow}(\mathcal{N}_{A\to B'}). \tag{285}$$

The following inequality is an immediate consequence of definitions:

$$C_{\leftarrow}(\mathcal{N}_{A\to B'}) \leq \widetilde{C}_{\leftarrow}(\mathcal{N}_{A\to B'}). \tag{286}$$

Theorem 22: Fix $n \in \mathbb{N}$ and $\varepsilon \in [0, 1)$. The n-shot classical capacity $C^{n,\varepsilon}_{\leftarrow}(\mathcal{N}_{A\to B'})$ of the point-to-point channel $\mathcal{N}_{A\to B'}$ assisted by classical feedback is bounded from above as follows:

$$C^{n,\varepsilon}_{\leftarrow}(\mathcal{N}_{A\to B'}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_{A\to B'}) + \frac{\alpha}{n(\alpha - 1)} \log_2 \left(\frac{1}{1-\varepsilon} \right), \tag{287}$$

for all $\alpha \in (1, 2]$.

Proof: Applying the bound in (281) with the choices in (282), we conclude that the following bound holds for all $\alpha \in (1, 2]$ and for an arbitrary (n, M, ε) classical communication protocol assisted by a classical feedback channel:

$$\begin{aligned} &\log_2 M \\ &\leq n\widehat{\Upsilon}_\alpha(\mathcal{N}_{A\to B'}) + \sum_{i=1}^{n} \widehat{\Upsilon}_\alpha(\overline{\Delta}_{B_{2i-1}\to A'_{2i-1}}) \\ &\quad + \frac{\alpha}{\alpha - 1} \log_2 \left(\frac{1}{1-\varepsilon} \right) \qquad (288) \\ &= n\widehat{\Upsilon}_\alpha(\mathcal{N}_{A\to B'}) + \frac{\alpha}{\alpha - 1} \log_2 \left(\frac{1}{1-\varepsilon} \right). \qquad (289) \end{aligned}$$

The equality follows from Proposition 13. This then implies the following bound

$$\frac{1}{n} \log_2 M \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_{A\to B'}) + \frac{\alpha}{n(\alpha - 1)} \log_2 \left(\frac{1}{1-\varepsilon} \right). \tag{290}$$

Since the bound in (290) holds for an arbitrary protocol, we conclude the statement of the theorem after applying the definition in (283). ∎

Theorem 23: The strong converse classical capacity of a point-to-point quantum channel $\mathcal{N}_{A\to B'}$ assisted by a classical feedback channel is bounded from above as follows:

$$\widetilde{C}_{\leftarrow}(\mathcal{N}_{A\to B'}) \leq \widehat{\Upsilon}(\mathcal{N}_{A\to B'}). \tag{291}$$

Proof: The reasoning here is the same as that given in the proof of Theorem 21. ∎

Before proceeding to the next section, let us finally note that our bounds above apply in a more general setting in which the classical feedback channel is replaced by an entanglement-breaking channel [66]. This follows because every entanglement-breaking channel can be written as a composition of a general pre-processing quantum channel, followed by a classical channel, which is in turn followed by a general post-processing quantum channel [66]. It is then clear that the pre-processing channel can be absorbed into a local operation of the receiver Bob, while the post-processing channel can be absorbed into a local operation of Alice, so that this is essentially just assistance by a classical feedback channel again, and our result thus applies.

V. Employing the sandwiched Rényi relative entropy

In this section, we explore what kinds of bounds we can obtain on the previously defined capacities, by making use of the sandwiched Rényi relative entropy.

Let us define the amortized sandwiched Rényi divergence of the completely positive maps $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ as follows [67]:

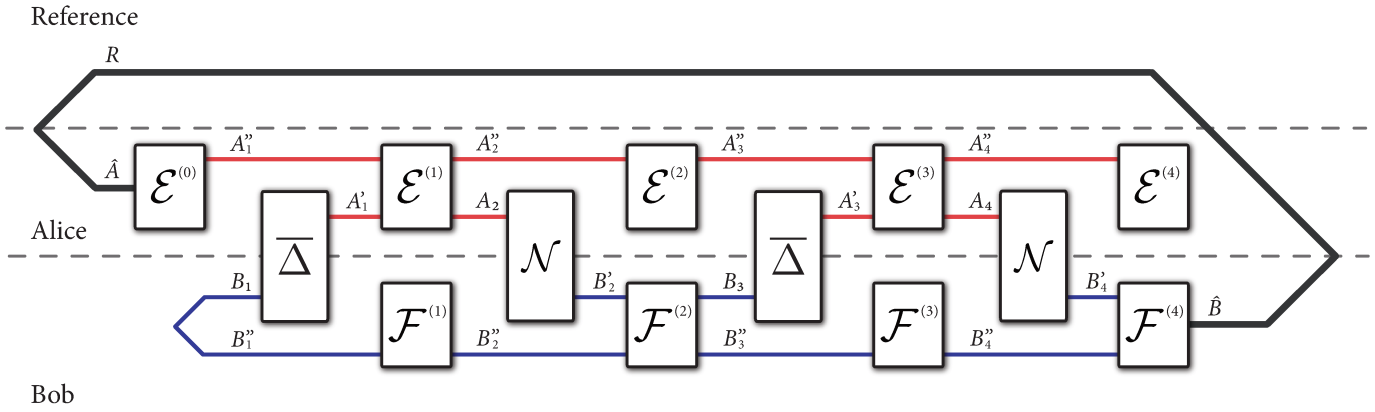

Fig. 3. Depiction of a protocol for classical communication over a point-to-point channel with classical feedback, by making $n = 2$ uses of the point-to-point channel $\mathcal{N}_{A\to B'}$. This protocol is related to the one in Figure 2, by replacing every odd use of $\mathcal{N}_{AB\to A'B'}$ with the classical feedback channel $\overline{\Delta}_{B\to A'}$ and every even use of $\mathcal{N}_{AB\to A'B'}$ with the point-to-point channel $\mathcal{N}_{A\to B'}$.

$$\widetilde{D}^{\mathcal{A}}_{\alpha}(\mathcal{N}_{A\to B}\|\mathcal{M}_{A\to B}) := \sup_{\rho_{RA},\sigma_{RA}} \widetilde{D}_{\alpha}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\sigma_{RA})) - \widetilde{D}_{\alpha}(\rho_{RA}\|\sigma_{RA}), \tag{292}$$

for $\alpha \in (0,1)\cup(1,\infty)$, where the optimization is over all density operators ρ_{RA} and σ_{RA}. By exploiting the definition of the sandwiched Rényi relative entropy, it follows that the quantity above does not change if we optimize more generally over positive semi-definite operators ρ_{RA} and σ_{RA} with strictly positive trace.

The amortized sandwiched Rényi divergence is subadditive in the following sense:

Proposition 24 (Subadditivity): Let $\mathcal{N}^1_{A\to B}$, $\mathcal{N}^2_{B\to C}$, $\mathcal{M}^1_{A\to B}$, and $\mathcal{M}^2_{B\to C}$ be completely positive maps. Then

$$\widetilde{D}^{\mathcal{A}}_{\alpha}(\mathcal{N}^2\circ\mathcal{N}^1\|\mathcal{M}^2\circ\mathcal{M}^1) \leq \widetilde{D}^{\mathcal{A}}_{\alpha}(\mathcal{N}^1\|\mathcal{M}^1) + \widetilde{D}^{\mathcal{A}}_{\alpha}(\mathcal{N}^2\|\mathcal{M}^2), \tag{293}$$

for all $\alpha\in(0,1)\cup(1,\infty)$.

Proof: Let ρ_{RA} and σ_{RA} be arbitrary positive semi-definite operators. Then

$$\begin{aligned}
&\widetilde{D}_{\alpha}(\mathcal{N}^2_{B\to C}(\mathcal{N}^1_{A\to B}(\rho_{RA}))\|\mathcal{M}^2_{B\to C}(\mathcal{M}^1_{A\to B}(\sigma_{RA}))) \\
&\quad - \widetilde{D}_{\alpha}(\rho_{RA}\|\sigma_{RA}) \qquad (294)\\
&= \widetilde{D}_{\alpha}(\mathcal{N}^2_{B\to C}(\mathcal{N}^1_{A\to B}(\rho_{RA}))\|\mathcal{M}^2_{B\to C}(\mathcal{M}^1_{A\to B}(\sigma_{RA}))) \\
&\quad - \widetilde{D}_{\alpha}(\mathcal{N}^1_{A\to B}(\rho_{RA})\|\mathcal{M}^1_{A\to B}(\sigma_{RA})) \\
&\quad + \widetilde{D}_{\alpha}(\mathcal{N}^1_{A\to B}(\rho_{RA})\|\mathcal{M}^1_{A\to B}(\sigma_{RA})) - \widetilde{D}_{\alpha}(\rho_{RA}\|\sigma_{RA}) \qquad (295)\\
&\leq \widetilde{D}^{\mathcal{A}}_{\alpha}(\mathcal{N}^1\|\mathcal{M}^1) + \widetilde{D}^{\mathcal{A}}_{\alpha}(\mathcal{N}^2\|\mathcal{M}^2). \qquad (296)
\end{aligned}$$

The desired inequality follows because ρ_{RA} and σ_{RA} are arbitrary. ∎

The regularized sandwiched Rényi divergence of the completely positive maps $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ is defined for $\alpha\in(0,1)\cup(1,\infty)$ as follows:

$$\widetilde{D}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{A\to B}\|\mathcal{M}_{A\to B}) := \lim_{n\to\infty}\frac{1}{n}\widetilde{D}_{\alpha}(\mathcal{N}^{\otimes n}_{A\to B}\|\mathcal{M}^{\otimes n}_{A\to B}), \tag{297}$$

and the limit exists, as argued in [68, Theorem 5.4].

The following equality holds for all $\alpha > 1$ [68, Theorem 5.4]:

$$\widetilde{D}^{\mathcal{A}}_{\alpha}(\mathcal{N}_{A\to B}\|\mathcal{M}_{A\to B}) = \widetilde{D}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{A\to B}\|\mathcal{M}_{A\to B}). \tag{298}$$

As such, by applying (298) and Proposition 24, it follows that

$$\widetilde{D}^{\mathrm{reg}}_{\alpha}(\mathcal{N}^2\circ\mathcal{N}^1\|\mathcal{M}^2\circ\mathcal{M}^1) \leq \widetilde{D}^{\mathrm{reg}}_{\alpha}(\mathcal{N}^1\|\mathcal{M}^1) + \widetilde{D}^{\mathrm{reg}}_{\alpha}(\mathcal{N}^2\|\mathcal{M}^2) \tag{299}$$

for all $\alpha > 1$.

We can then replace the use of the geometric Rényi relative entropy in Theorems 20 and 22 with the regularized sandwiched Rényi relative entropy and arrive at the following statements:

Corollary 25: The following upper bound holds for the n-shot randomness transmission capacity of a bipartite channel $\mathcal{N}_{AB\to A'B'}$:

$$R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'}) \leq \widetilde{\Upsilon}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{AB\to A'B'}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right), \tag{300}$$

for all $\alpha > 1$ and $\varepsilon\in[0,1)$, where

$$\widetilde{\Upsilon}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{AB\to A'B'}) := \inf_{\substack{\mathcal{M}_{AB\to A'B'}:\\ \beta(\mathcal{M}_{AB\to A'B'})\leq 1}} \widetilde{D}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{AB\to A'B'}\|\mathcal{M}_{AB\to A'B'}). \tag{301}$$

Corollary 26: The following upper bound holds for the n-shot randomness transmission capacity of a point-to-point channel $\mathcal{N}_{A\to B'}$ assisted by a classical feedback channel:

$$C^{n,\varepsilon}(\mathcal{N}_{A\to B'}) \leq \widetilde{\Upsilon}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{A\to B'}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right), \tag{302}$$

for all $\alpha > 1$ and $\varepsilon\in[0,1)$, where

$$\widetilde{\Upsilon}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{A\to B'}) := \inf_{\substack{\mathcal{M}_{A\to B'}:\\ \beta(\mathcal{M}_{A\to B'})\leq 1}} \widetilde{D}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{A\to B'}\|\mathcal{M}_{A\to B'}). \tag{303}$$

These bounds are not particularly useful, because the quantities $\widetilde{\Upsilon}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{AB\to A'B'})$ and $\widetilde{\Upsilon}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{A\to B'})$ may be difficult to compute in practice. However, see the discussions in [68, Section 5.1] for progress on algorithms for computing $\widetilde{D}^{\mathrm{reg}}_{\alpha}(\mathcal{N}\|\mathcal{M})$. In the next section, we show how these bounds simplify when the channels of interest possess symmetry.

VI. Exploiting symmetries

In this section, we discuss how to improve the upper bounds in Corollaries 25 and 26 when a bipartite channel and point-to-point channel possess symmetries, respectively.

We begin by recalling the definition of a bicovariant bipartite channel [40]. Let G and H be finite groups, and for $g \in G$ and $h \in H$, let $g \to U_A(g)$ and $h \to V_B(h)$ be unitary representations. Also, let $(g,h) \to W_{A'}(g,h)$ and $(g,h) \to Y_{B'}(g,h)$ be unitary representations. A bipartite channel $\mathcal{N}_{AB\to A'B'}$ is bicovariant with respect to these representations if the following equality holds for all group elements $g \in G$ and $h \in H$:

$$\mathcal{N}_{AB\to A'B'} \circ (\mathcal{U}_A(g) \otimes \mathcal{V}_B(h)) = (\mathcal{W}_{A'}(g,h) \otimes \mathcal{Y}_{B'}(g,h)) \circ \mathcal{N}_{AB\to A'B'}, \quad (304)$$

where $\mathcal{U}_A(g)(\cdot) := U_A(g)(\cdot)U_A(g)^\dagger$, with similar definitions for $\mathcal{V}_B(h)$, $\mathcal{W}_{A'}(g,h)$, and $\mathcal{Y}_{B'}(g,h)$. A bipartite channel is bicovariant if it is bicovariant with respect to groups that have representations as unitary one-designs, i.e.,

$$\frac{1}{|G|}\sum_{g\in G} \mathcal{U}_A(g)(X) = \mathrm{Tr}[X]\pi_A, \quad (305)$$

$$\frac{1}{|H|}\sum_{h\in H} \mathcal{V}_B(h)(Y) = \mathrm{Tr}[Y]\pi_B. \quad (306)$$

Two bipartite maps $\mathcal{N}_{AB\to A'B'}$ and $\mathcal{M}_{AB\to A'B'}$ are jointly bicovariant if they are bicovariant with respect to the same representations, i.e., if (304) holds for both $\mathcal{N}_{AB\to A'B'}$ and $\mathcal{M}_{AB\to A'B'}$.

Proposition 27: Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel that is bicovariant with respect to unitary representations as defined above. Then

$$\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) = \inf_{\substack{\mathcal{M}\in\mathcal{C}:\\ \beta(\mathcal{M})\leq 1}} \sup_{\psi\in\mathcal{S}} \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\psi_{RAB})\|\mathcal{M}_{AB\to A'B'}(\psi_{RAB})), \quad (307)$$

where $\mathcal{S}$ is the set consisting of every pure state ψ_{RAB} such that the reduced state ψ_{AB} satisfies

$$\psi_{AB} = \frac{1}{|G|\,|H|}\sum_{g\in G,h\in H} (\mathcal{U}_A(g)\otimes\mathcal{V}_B(h))(\psi_{AB}), \quad (308)$$

and $\mathcal{C}$ is the set of all completely positive bipartite maps that are bicovariant with respect to the unitary representations defined above. In the case that $\mathcal{N}_{AB\to A'B'}$ is bicovariant, then

$$\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) = \inf_{\substack{\mathcal{M}\in\mathcal{C}^1:\\ \beta(\mathcal{M})\leq 1}} \boldsymbol{D}(\mathcal{N}(\Phi_{\hat{A}A}\otimes\Phi_{B\hat{B}})\|\mathcal{M}(\Phi_{\hat{A}A}\otimes\Phi_{B\hat{B}})), \quad (309)$$

where $\Phi_{\hat{A}A}\otimes\Phi_{B\hat{B}}$ is a tensor product of maximally entangled states and $\mathcal{C}^1$ is the set of all completely positive bicovariant maps $\mathcal{M}_{AB\to A'B'}$ (i.e., covariant with respect to one-designs).

Proof: Let ψ_{RAB} be an arbitrary pure state. Define

$$\overline{\rho}_{AB} := \frac{1}{|G|\,|H|}\sum_{g\in G,h\in H} (\mathcal{U}_A(g)\otimes\mathcal{V}_B(h))(\psi_{AB}). \quad (310)$$

Let $\phi^{\overline{\rho}}_{SAB} \in \mathcal{S}$ be a purification of $\overline{\rho}_{AB}$. Another purification of $\overline{\rho}_{AB}$ is given by

$$\psi^{\overline{\rho}}_{GHRAB} := |\psi^{\overline{\rho}}\rangle\langle\psi^{\overline{\rho}}|_{GHRAB}, \quad (311)$$

where

$$|\psi^{\overline{\rho}}\rangle_{GHRAB} := \frac{1}{\sqrt{|G|\,|H|}}\sum_{g\in G,h\in H} |g\rangle_G|h\rangle_H (U_A(g)\otimes V_B(h))|\psi\rangle_{RAB}. \quad (312)$$

Let $\mathcal{M}_{AB\to A'B'}$ be an arbitrary completely positive map satisfying $\beta(\mathcal{M}_{AB\to A'B'}) \leq 1$. Define

$$\overline{\mathcal{M}}_{AB\to A'B'} := \frac{1}{|G|\,|H|}\sum_{g\in G,h\in H} (\mathcal{W}_{A'}(g,h)\otimes\mathcal{Y}_{B'}(g,h))^\dagger \circ \mathcal{M}_{AB\to A'B'} \circ (\mathcal{U}_A(g)\otimes\mathcal{V}_B(h)), \quad (313)$$

and observe that $\overline{\mathcal{M}}_{AB\to A'B'} \in \mathcal{C}$. Consider the development in (314)–(318). The first equality in (314) holds because all purifications are related by an isometric channel acting on the purifying system, the channel $\mathcal{N}_{AB\to A'B'}$ commutes with the action of this isometric channel because they act on different systems, and the generalized divergence is invariant under the action of isometric channels. The first inequality in (315) follows by acting with a completely dephasing channel on the systems GH and then applying the data-processing inequality. The second equality in (316) follows from the bicovariance of $\mathcal{N}_{AB\to A'B'}$ with respect to the given representations. The third equality in (317) follows by applying the unitary

$$\sum_{g\in G,h\in H} |g,h\rangle\langle g,h|_{GH}\otimes W^\dagger_{A'}(g,h)\otimes Y^\dagger_{B'}(g,h), \quad (319)$$

and from the unitary invariance of the generalized divergence. We have also defined

$$\mathcal{M}^{g,h}_{AB\to A'B'} := (\mathcal{W}_{A'}(g,h)\otimes\mathcal{Y}_{B'}(g,h))^\dagger \circ \mathcal{M}_{AB\to A'B'} \circ (\mathcal{U}_A(g)\otimes\mathcal{V}_B(h)). \quad (320)$$

The last inequality in (318) follows from tracing over the registers GH and from the data-processing inequality. Since the inequality holds for all pure states, we conclude that

$$\sup_{\phi_{SAB}\in\mathcal{S}} \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi_{SAB})\|\mathcal{M}_{AB\to A'B'}(\phi_{SAB})) \geq \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}\|\overline{\mathcal{M}}_{AB\to A'B'}) \quad (321)$$

$$\geq \boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}). \quad (322)$$

The second inequality follows because $\overline{\mathcal{M}}_{AB\to A'B'}$ satisfies $\beta(\overline{\mathcal{M}}_{AB\to A'B'}) \leq 1$ if $\mathcal{M}_{AB\to A'B'}$ does. This in turn is

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

$$
\begin{aligned}
&\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi^{\overline{\rho}}_{SAB})\|\mathcal{M}_{AB\to A'B'}(\phi^{\overline{\rho}}_{SAB})) \\
&= \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\psi^{\overline{\rho}}_{GHRAB})\|\mathcal{M}_{AB\to A'B'}(\psi^{\overline{\rho}}_{GHRAB})) && (314)\\
&\geq \boldsymbol{D}\Bigg(\frac{1}{|G|\,|H|}\sum_{g\in G,h\in H}|g,h\rangle\langle g,h|_{GH}\otimes(\mathcal{N}_{AB\to A'B'}\circ(\mathcal{U}_A(g)\otimes\mathcal{V}_B(h)))(\psi_{RAB})\Bigg\| \\
&\qquad \frac{1}{|G|\,|H|}\sum_{g\in G,h\in H}|g,h\rangle\langle g,h|_{GH}\otimes(\mathcal{M}_{AB\to A'B'}\circ(\mathcal{U}_A(g)\otimes\mathcal{V}_B(h)))(\psi_{RAB})\Bigg) && (315)\\
&= \boldsymbol{D}\Bigg(\frac{1}{|G|\,|H|}\sum_{g\in G,h\in H}|g,h\rangle\langle g,h|_{GH}\otimes((\mathcal{W}_{A'}(g,h)\otimes\mathcal{Y}_{B'}(g,h))\circ\mathcal{N}_{AB\to A'B'})(\psi_{RAB})\Bigg\| \\
&\qquad \frac{1}{|G|\,|H|}\sum_{g\in G,h\in H}|g,h\rangle\langle g,h|_{GH}\otimes(\mathcal{M}_{AB\to A'B'}\circ(\mathcal{U}_A(g)\otimes\mathcal{V}_B(h)))(\psi_{RAB})\Bigg) && (316)\\
&= \boldsymbol{D}\Bigg(\frac{1}{|G|\,|H|}\sum_{g\in G,h\in H}|g,h\rangle\langle g,h|_{GH}\otimes\mathcal{N}_{AB\to A'B'}(\psi_{RAB})\Bigg\| \\
&\qquad \frac{1}{|G|\,|H|}\sum_{g\in G,h\in H}|g,h\rangle\langle g,h|_{GH}\otimes\mathcal{M}^{g,h}_{AB\to A'B'}(\psi_{RAB})\Bigg) && (317)\\
&\geq \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\psi_{RAB})\|\overline{\mathcal{M}}_{AB\to A'B'}(\psi_{RAB})). && (318)
\end{aligned}
$$

a consequence of the convexity of β (Proposition 9) and its invariance under local unitary channels (Corollary 8). Since the inequality holds for all $\mathcal{M}_{AB\to A'B'}$ satisfying $\beta(\mathcal{M}_{AB\to A'B'})\leq 1$, we conclude that

$$
\begin{aligned}
&\inf_{\substack{\mathcal{M}:\\ \beta(\mathcal{M})\leq 1}}\sup_{\phi\in\mathcal{S}}\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi_{SAB})\|\mathcal{M}_{AB\to A'B'}(\phi_{SAB})) \\
&\qquad\qquad \geq \boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}). \qquad (323)
\end{aligned}
$$

However, the definition of $\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'})$ implies that

$$
\begin{aligned}
&\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'})\geq \\
&\inf_{\substack{\mathcal{M}:\\ \beta(\mathcal{M})\leq 1}}\sup_{\phi\in\mathcal{S}}\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi_{SAB})\|\mathcal{M}_{AB\to A'B'}(\phi_{SAB})). \qquad (324)
\end{aligned}
$$

So we conclude the equality

$$
\begin{aligned}
&\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) = \\
&\inf_{\substack{\mathcal{M}:\\ \beta(\mathcal{M})\leq 1}}\sup_{\phi\in\mathcal{S}}\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi_{SAB})\|\mathcal{M}_{AB\to A'B'}(\phi_{SAB})). \qquad (325)
\end{aligned}
$$

Now suppose that $\mathcal{M}_{AB\to A'B'}$ is an arbitrary completely positive map satisfying $\beta(\mathcal{M}_{AB\to A'B'})\leq 1$, and let $\phi_{SAB}\in\mathcal{S}$. Then by (321), we conclude that

$$
\begin{aligned}
&\sup_{\phi_{SAB}\in\mathcal{S}}\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi_{SAB})\|\mathcal{M}_{AB\to A'B'}(\phi_{SAB})) \\
&\geq \inf_{\substack{\overline{\mathcal{M}}\in\mathcal{C}:\\ \beta(\overline{\mathcal{M}})\leq 1}}\sup_{\phi\in\mathcal{S}}\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi_{SAB})\|\overline{\mathcal{M}}_{AB\to A'B'}(\phi_{SAB})). \qquad (326)
\end{aligned}
$$

Since this holds for every completely positive map $\mathcal{M}_{AB\to A'B'}$ satisfying $\beta(\mathcal{M}_{AB\to A'B'})\leq 1$, we conclude that

$$
\begin{aligned}
&\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'})\geq \\
&\inf_{\substack{\overline{\mathcal{M}}\in\mathcal{C}:\\ \beta(\overline{\mathcal{M}})\leq 1}}\sup_{\phi\in\mathcal{S}}\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi_{SAB})\|\overline{\mathcal{M}}_{AB\to A'B'}(\phi_{SAB})). \qquad (327)
\end{aligned}
$$

However, from the definition of $\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'})$, we have the inequality

$$
\begin{aligned}
&\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'})\leq \\
&\inf_{\substack{\overline{\mathcal{M}}\in\mathcal{C}:\\ \beta(\overline{\mathcal{M}})\leq 1}}\sup_{\phi\in\mathcal{S}}\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\phi_{SAB})\|\overline{\mathcal{M}}_{AB\to A'B'}(\phi_{SAB})). \qquad (328)
\end{aligned}
$$

Thus, the equality in (307) follows. The equality in (309) follows because the only state satisfying (308) for one-designs is the tensor product of maximally mixed states, and the tensor product of maximally entangled states purifies this state. ■

Recall from Section V that the bounds in Corollaries 25 and 26 are not particularly useful on their own because $\widetilde{\Upsilon}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{AB\to A'B'})$ may be difficult to compute in practice. However, if the bipartite channel is bicovariant, then (309) implies that the regularized quantity is bounded from above by a single-letter quantity:

$$
\widetilde{\Upsilon}^{\mathrm{reg}}_{\alpha}(\mathcal{N}_{AB\to A'B'})\leq\widetilde{\Upsilon}_{\alpha}(\mathcal{N}_{AB\to A'B'}). \qquad (329)
$$

We then obtain the following:

Corollary 28: The following upper bound holds for the n-shot randomness transmission capacity of a bicovariant bipartite channel $\mathcal{N}_{AB\to A'B'}$:

$$
\begin{aligned}
&R^{n,\varepsilon}(\mathcal{N}_{AB\to A'B'})\leq \\
&\qquad \widetilde{\Upsilon}_{\alpha}(\mathcal{N}_{AB\to A'B'})+\frac{\alpha}{n\,(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right), \qquad (330)
\end{aligned}
$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

for all $\alpha > 1$ and $\varepsilon \in [0, 1)$.

By applying the same reasoning in the proof of Theorem 21, we conclude the following:

Corollary 29: The following upper bound holds for the strong converse randomness transmission capacity of a bicovariant bipartite channel $\mathcal{N}_{AB\to A'B'}$:

$$\widetilde{R}(\mathcal{N}_{AB\to A'B'}) \leq \Upsilon(\mathcal{N}_{AB\to A'B'}). \tag{331}$$

Let G be a group and let $U_A(g)$ and $V_{B'}(g)$ be unitary representations of g. A point-to-point channel $\mathcal{N}_{A\to B'}$ is covariant with respect to these representations if the following equality holds for all $g \in G$ [69]:

$$\mathcal{N}_{A\to B'} \circ \mathcal{U}_A(g) = \mathcal{V}_{B'}(g) \circ \mathcal{N}_{A\to B'}. \tag{332}$$

A point-to-point channel is covariant if it is covariant with respect to a one-design.

By applying the same reasoning as given above, we have the following results:

Corollary 30: The following upper bound holds for the n-shot classical capacity of a covariant point-to-point channel $\mathcal{N}_{A\to B'}$ assisted by a classical feedback channel:

$$C_{\leftarrow}^{n,\varepsilon}(\mathcal{N}_{A\to B'}) \leq \widetilde{\Upsilon}_\alpha(\mathcal{N}_{A\to B'}) + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right), \tag{333}$$

for all $\alpha > 1$ and $\varepsilon \in [0, 1)$.

Corollary 31: The following upper bound holds for the strong converse classical capacity of a covariant point-to-point channel $\mathcal{N}_{A\to B'}$:

$$\widetilde{C}_{\leftarrow}(\mathcal{N}_{A\to B'}) \leq \Upsilon(\mathcal{N}_{A\to B'}). \tag{334}$$

VII. Examples

In this section, we apply the bounds to some key examples of bipartite and point-to-point channels. The Matlab code used to generate the plots below is available with the arXiv posting of our paper.

A. Partial swap bipartite channel

The partial swap unitary is defined for $p \in [0, 1]$ as [70], [71]

$$S_{AB}^p := \sqrt{1-p} I_{AB} + i\sqrt{p} S_{AB}, \tag{335}$$

$$S_{AB} := \sum_{i,j=0}^{d-1} |i\rangle\langle j|_A \otimes |j\rangle\langle i|_B, \tag{336}$$

where $A \simeq B$ and $d = d_A = d_B$. The following identity holds

$$S_{AB}^p = e^{itS_{AB}} = \cos(t) I_{AB} + i \sin(t) S_{AB}, \tag{337}$$

where $\sqrt{1-p} = \cos t$. Thus, we can understand the unitary operator S_{AB}^p as arising from time evolution according to the Hamiltonian S_{AB}. We then define the bipartite partial swap channel as

$$\mathcal{S}_{AB}^p(\cdot) := S_{AB}^p(\cdot)(S_{AB}^p)^\dagger. \tag{338}$$

Suppose that $p = 1$. Then the channel $\mathcal{S}_{AB}^p$ is equivalent to a swap channel. In this case, the forward classical capacity is equal to $2\log_2 d$. This follows by an argument given in [30]. To see that the rate $2\log_2 d$ is achievable, consider the following strategy. On the first use of the channel, Alice inputs one classical dit to her input and Bob inputs one share of a maximally entangled state. Bob can decode the classical dit, and after the first channel use, they share a maximally entangled state Φ^d. Before the second channel use, Alice can employ a super-dense coding strategy [5]. She applies one of the d^2 Heisenberg–Weyl unitaries to her share of Φ^d and transmits it through her input to the channel. Bob again prepares Φ^d and sends one share through his channel input. Bob can then decode the message Alice sent, by performing a Bell measurement, and they again share Φ^d. They then repeat this procedure many times. Even though the first channel use allows for only $\log_2 d$ bits to be transmitted, all of the other channel uses allow for $2\log_2 d$ bits to be transmitted. So in the limit of many channel uses, the rate $2\log_2 d$ is achievable. An upper bound of $2\log_2 d$ is argued in [30] by employing a simulation argument. Alternatively, it can be seen from our approach by employing Theorem 21 and Proposition 15, and picking $S_{AA'BB'} = V_{AA'BB'} = I_{AA'BB'}$ in the definition of C_β. These choices satisfy the constraints and $\log_2 \left\| \mathrm{Tr}_{A'B'}[S_{AA'BB'}] \right\|_\infty = 2\log_2 d$.

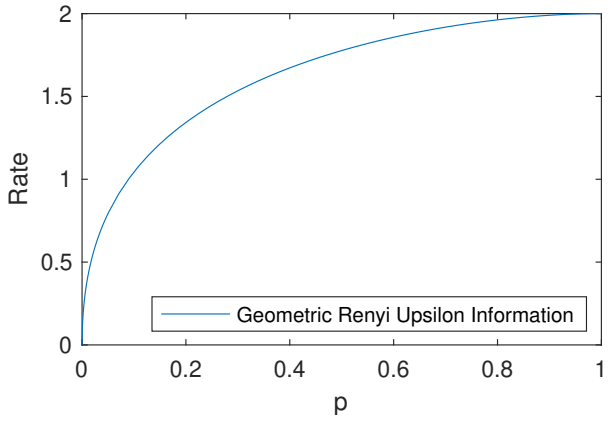

Fig. 4. Upper bound on the forward classical capacity of the partial swap bipartite channel in (338), with $d = 2$.

At the other extreme, when $p = 0$, the channel $\mathcal{S}_{AB}^p$ reduces to the tensor product of identity channels. Since this channel is a product of local channels, Theorem 21 and Proposition 14 imply that $C(\mathcal{S}_{AB}^p) = 0$. Thus, the partial swap unitary interpolates between these two extremes.

Interestingly, the partial swap unitary is not bicovariant for $p \in (0, 1)$ because the general definition involves both the identity and the swap. As such, our bound from Theorem 21 is useful in such a case. By employing a semi-definite program to calculate $\widehat{\Upsilon}_\alpha(\mathcal{S}_{AB}^p)$ for $d = 2$ and $\alpha = 1 + 2^{-\ell}$, with $\ell = 4$, we arrive at the plot given in Figure 4. The semi-definite program is included in the arXiv posting of this paper and is based on the methods mentioned in [22, Remark 4].

We remark that the partial swap channel is bicovariant with respect to all unitaries of the form $U \otimes U$. As such, by applying Proposition 27, we conclude that it suffices to maximize $\widehat{\Upsilon}_\alpha(\mathcal{S}_{AB}^p)$ over input states ψ_{RAB} possessing the

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

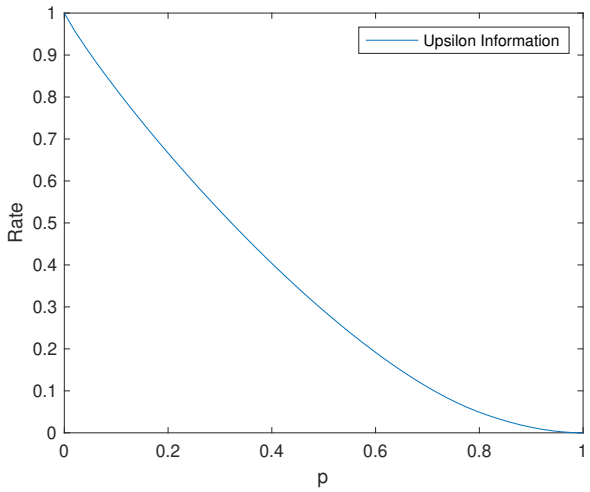

Fig. 5. Upper bound on the forward classical capacity of the noisy CNOT bipartite channel in (343), with $d = 2$.

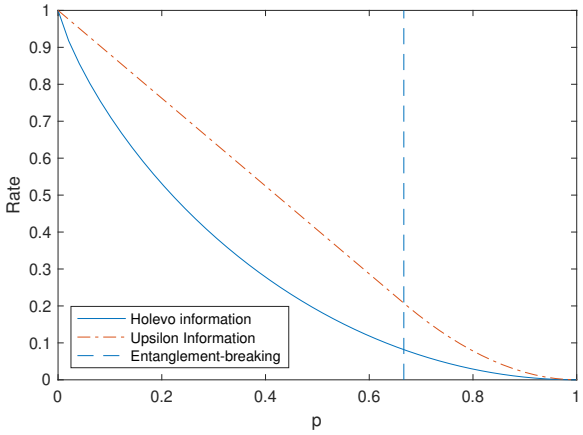

Fig. 6. Lower and upper bounds on the classical-feedback-assisted classical capacity of the qubit depolarizing channel in (348), with $d = 2$. The dashed vertical line indicates that the qubit depolarizing channel is entanglement breaking for $p \geq 2/3$, so that the Holevo information is equal to the feedback-assisted capacity for these values [10], [26].

following symmetry:

$$\psi_{AB} = \int dU \ (U_A \otimes U_B)\psi_{AB}(U_A \otimes U_B)^\dagger, \tag{339}$$

where dU denotes the Haar measure. States possessing this symmetry are known as Werner states [72] and can be written in terms of a single parameter $q \in [0, 1]$ as follows:

$$W_{AB}^{(q,d)} \coloneqq (1-q)\,\frac{2}{d\,(d+1)}\Pi_{AB}^{+} + q\frac{2}{d\,(d-1)}\Pi_{AB}^{-}, \tag{340}$$

where $\Pi_{AB}^{\pm} \coloneqq (I_{AB} \pm S_{AB})\,/2$ are the projections onto the symmetric and antisymmetric subspaces of A and B, with S_{AB} defined in (336). Additionally, by exploiting the same symmetry, it suffices to minimize over completely positive bipartite maps $\mathcal{M}_{AB\to A'B'}$ such that

$$\mathcal{M}_{AB\to A'B'} = \int dU \ (\mathcal{U}_{A'} \otimes \mathcal{U}_{B'})^\dagger \circ \mathcal{M}_{AB\to A'B'} \circ (\mathcal{U}_A \otimes \mathcal{U}_B). \tag{341}$$

This is equivalent to their Choi operators satisfying

$$\Gamma^{\mathcal{M}}_{AA'BB'} = \int dU \ (\mathcal{U}_{A'} \otimes \mathcal{U}_{B'} \otimes \overline{\mathcal{U}}_A \otimes \overline{\mathcal{U}}_B)(\Gamma^{\mathcal{M}}_{AA'BB'}), \tag{342}$$

where $\overline{\mathcal{U}}$ denotes the complex conjugate. This further reduces the number of parameters needed in the optimization task, which is useful for computing $\widehat{\Upsilon}_\alpha(\mathcal{S}^p_{AB})$ for higher-dimensional partial swap bipartite channels. We note that Haar integrals of the form in (342) can be computed by generalizing the methods of [73], [74] (see also [75, Section VII]).

B. Noisy CNOT gate

Another example of a bipartite channel of interest is a noisy CNOT gate, defined as follows:

$$\mathcal{D}^p_{AB}(\cdot) \coloneqq (1-p)\,\mathrm{CNOT}_{AB}(\cdot)\mathrm{CNOT}_{AB} + p\mathcal{R}^\pi_{AB}(\cdot), \tag{343}$$

where

$$\mathrm{CNOT}_{AB} \coloneqq \sum_{i=0}^{d-1} |i\rangle\langle i|_A \otimes X(i)_B, \tag{344}$$

$$X(i)_B \coloneqq \sum_{j=0}^{d-1} |i \oplus j\rangle\langle j|, \tag{345}$$

$$\mathcal{R}^\pi_{AB}(\cdot) \coloneqq \mathrm{Tr}_{AB}[\cdot]\pi_{AB}, \tag{346}$$

$$\pi_{AB} \coloneqq \frac{I_{AB}}{d_A d_B}. \tag{347}$$

When $p = 0$, the channel is a perfect CNOT gate, and when $p = 1$, it is a replacer channel. Thus, when $p = 0$, the result from [30] applies, implying that $C(\mathcal{D}^{p=0}_{AB}) = \log_2 d$, and when $p = 1$, the forward classical capacity $C(\mathcal{D}^{p=0}_{AB}) = 0$.

This channel is bicovariant, as argued in [40], and so Corollary 29 applies. Evaluating the Υ-information of $\mathcal{D}^p_{AB}$, we obtain the plot in Figure 5.

C. Point-to-point depolarizing channel

Here we consider the point-to-point depolarizing channel, defined as

$$\mathcal{D}^p(X) \coloneqq (1-p)\,X + p\,\mathrm{Tr}[X]\pi, \tag{348}$$

$$\pi \coloneqq I/d. \tag{349}$$

It was already established in [22] that $\Upsilon(\mathcal{D}^p)$ is an upper bound on its (unassisted) classical capacity, and the Holevo information is equal to its classical capacity [17]. Our contribution here is that $\Upsilon(\mathcal{D}^p)$ is an upper bound on its classical capacity assisted by a classical feedback channel. Figure 6 plots this upper bound and also plots the Holevo information lower bound when $d = 2$. The latter is given by $1 - h_2(p/2)$, where h_2 is the binary entropy function. Note that the depolarizing channel is entanglement breaking for $p \geq \frac{d}{d+1}$. As such, the bounds from [10], [26] apply, so that, for $p \geq \frac{d}{d+1}$, the Holevo information $1 - h_2(p/2)$ is equal to the classical capacity assisted by classical feedback.

D. Point-to-point erasure channel

The point-to-point quantum erasure channel is defined for $p \in [0,1]$ and integer $d \geq 2$ as [76]

$$\mathcal{E}_{p,d}(\rho) \coloneqq (1-p)\rho + p|e\rangle\langle e|, \tag{350}$$

where $|e\rangle\langle e|$ is a quantum erasure symbol, orthogonal to every d-dimensional input state ρ, so that the channel output has dimension $d+1$. This channel is covariant, as defined just after (332). Thus, by combining [21, Lemma 12] with Corollary 31, we conclude that the strong converse holds for the classical capacity of the erasure channel $\mathcal{E}_{p,d}$ assisted by classical feedback; i.e.,

$$C_{\leftarrow}(\mathcal{E}_{p,d}(\rho)) = \widetilde{C}_{\leftarrow}(\mathcal{E}_{p,d}(\rho)) = (1-p)\log_2 d. \tag{351}$$

E. Other point-to-point channels

We note here that the reader can consult [22, Section 6.4] to find other examples of channels for which the Υ-information has been calculated, including dephrasure and generalized amplitude damping channels. In all cases, our results strengthen those findings, because our results imply that these quantities are upper bounds on the classical-feedback-assisted classical capacity, rather than just the unassisted classical capacity.

VIII. Conclusion

In this paper, we established several measures of classical communication and proved that they are useful as upper bounds on the classical capacity of bipartite quantum channels. We did so by establishing several key properties of these measures, which played essential roles in the upper bound proofs. One of the most critical properties is that the measures are subadditive under serial composition of bipartite channels, which is a property that is useful in the analysis of feedback-assisted protocols. One important application of our results is improved upper bounds on the classical capacity of a quantum channel assisted by classical feedback, which is a problem that has been analyzed in the literature for some time now [10], [28], [26], [29], [25].

Going forward from here, an open question is whether our bounds could be improved in any way. The recent techniques of [68] might be helpful in obtaining refined non-asymptotic bounds, but the main result of [77] implies that the sharp Rényi divergence of [68] will not be helpful in the asymptotic case. As a key example, we wonder whether classical feedback could increase the classical capacity of the depolarizing channel. We also wonder whether our new bounds on the classical capacity assisted by classical feedback generally improve upon the entropy bound from [25].

Acknowledgements—We thank Andreas Winter for helpful discussions about this project and about [29], and we thank Masahito Hayashi for notifying us about [9]. We are grateful to Bjarne Bergh, Nilanjana Datta, and Robert Salzmann for pointing out an issue with the proof of Lemma 10 of [64], as well as Milan Mosonyi for discussions about this issue. We additionally thank Milan Mosonyi for extensive editorial feedback on and corrections for our paper, which helped improve it, as well as for the observation in Remark 38. We also thank Bjarne Bergh for a clarification about the proof of Lemma 8 of [78]. DD and MMW thank God for all of His provisions. SK acknowledges support from NSERC and the NSF and under Grant No. 1714215. YQ acknowledges support from a Stanford QFARM fellowship and from an NUS Overseas Graduate Scholarship. PWS is supported by the National Science Foundation under Grant No. CCF-1729369 and through the NSF Science and Technology Center for Science of Information under Grant No. CCF-0939370. MMW acknowledges support from NSF Grant No. 1907615, as well as Stanford QFARM and AFOSR (FA9550-19-1-0369).

References

[1] M. Hayashi, *Quantum Information Theory: Mathematical Foundation*, 2nd ed. Springer, 2017.
[2] A. S. Holevo, *Quantum Systems, Channels, Information: A Mathematical Introduction*, 2nd ed. Walter de Gruyter, 2019.
[3] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018.
[4] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge University Press, 2017, arXiv:1106.1445. [Online]. Available: https://doi.org/10.1017/CBO9781139525343
[5] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states," *Physical Review Letters*, vol. 69, no. 20, pp. 2881–2884, Nov. 1992.
[6] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002, arXiv:quant-ph/0106052.
[7] A. S. Holevo, "On entanglement assisted classical capacity," *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4326–4333, Sep. 2002, arXiv:quant-ph/0106075.
[8] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-state entanglement and quantum error correction," *Physical Review A*, vol. 54, no. 5, pp. 3824–3851, Nov. 1996, arXiv:quant-ph/9604024.
[9] A. Fujiwara and H. Nagaoka, "Operational capacity and pseudoclassicality of a quantum channel," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1071–1086, May 1998.
[10] G. Bowen and R. Nagarajan, "On feedback and the classical capacity of a noisy quantum channel," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 320–324, Jan. 2005, arXiv:quant-ph/0305176.
[11] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Problems of Information Transmission*, vol. 9, pp. 177–183, 1973.
[12] ——, "The capacity of the quantum channel with general signal states," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 269–273, Jan. 1998, arXiv:quant-ph/9611023.
[13] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Physical Review A*, vol. 56, no. 1, pp. 131–138, Jul. 1997.
[14] M. M. Wolf, T. S. Cubitt, and D. Perez-Garcia, "Are problems in quantum information theory (un)decidable?" Nov. 2011, arXiv:1111.5425.
[15] M. B. Hastings, "Superadditivity of communication capacity using entangled inputs," *Nature Physics*, vol. 5, pp. 255–257, Apr. 2009, arXiv:0809.3972.
[16] P. W. Shor, "Additivity of the classical capacity of entanglement-breaking quantum channels," *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4334–4340, 2002, arXiv:quant-ph/0201149.
[17] C. King, "The capacity of the quantum depolarizing channel," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 221–229, Jan. 2003, arXiv:quant-ph/0204172.
[18] C. King, K. Matsumoto, M. Nathanson, and M. B. Ruskai, "Properties of conjugate channels with applications to additivity and multiplicativity," *Markov Processes and Related Fields*, vol. 13, no. 2, pp. 391–423, 2007, J. T. Lewis memorial issue, arXiv:quant-ph/0509126.
[19] S. Beigi and P. W. Shor, "On the complexity of computing zero-error and Holevo capacity of quantum channels," Sep. 2007, arXiv:0709.2090.
[20] X. Wang, W. Xie, and R. Duan, "Semidefinite programming strong converse bounds for classical capacity," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 640–653, Jan. 2018, arXiv:1610.06381.

[21] X. Wang, K. Fang, and M. Tomamichel, "On converse bounds for classical communication over quantum channels," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4609–4619, Jul. 2019, arXiv:1709.05258.

[22] K. Fang and H. Fawzi, "Geometric Rényi divergence and its applications in quantum channel capacities," *Communications in Mathematical Physics*, vol. 384, pp. 1615–1677, Jun. 2021, arXiv:1909.05758.

[23] D. Leung and W. Matthews, "On the power of PPT-preserving and non-signalling codes," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4486–4499, Aug. 2015, arXiv:1406.7142.

[24] C. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[25] D. Ding, Y. Quek, P. W. Shor, and M. M. Wilde, "Entropy bound for the classical capacity of a quantum channel assisted by classical feedback," in *Proceedings of the 2019 IEEE International Symposium on Information Theory*, Paris, France, Jul. 2019, pp. 250–254, arXiv:1902.02490.

[26] D. Ding and M. M. Wilde, "Strong converse exponents for the feedback-assisted classical capacity of entanglement-breaking channels," *Problems of Information Transmission*, vol. 54, no. 1, pp. 1–19, Jan. 2018, arXiv:1506.02228.

[27] G. Smith and J. A. Smolin, "Extensive nonadditivity of privacy," *Physical Review Letters*, vol. 103, no. 12, p. 120503, Sep. 2009, arXiv:0904.4050.

[28] C. H. Bennett, I. Devetak, P. W. Shor, and J. A. Smolin, "Inequalities and separations among assisted capacities of quantum channels," *Physical Review Letters*, vol. 96, no. 15, p. 150502, Apr. 2006, arXiv:quant-ph/0406086.

[29] R. García-Patrón, W. Matthews, and A. Winter, "Quantum enhancement of randomness distribution," *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4664–4673, Jun. 2018, arXiv:1610.05190.

[30] C. H. Bennett, A. W. Harrow, D. W. Leung, and J. A. Smolin, "On the capacities of bipartite Hamiltonians and unitary gates," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1895–1911, Aug. 2003, arXiv:quant-ph/0205057.

[31] A. W. Harrow and D. W. Leung, "Bidirectional coherent classical communication," *Quantum Information and Computation*, vol. 5, no. 4-5, pp. 380–395, Jul. 2005, arXiv:quant-ph/0412126.

[32] A. W. Harrow and P. W. Shor, "Time reversal and exchange symmetries of unitary gate capacities," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 462–475, Jan. 2010, arXiv:quant-ph/0511219.

[33] A. W. Harrow and D. W. Leung, "A communication-efficient nonlocal measurement with application to communication complexity and bipartite gate capacities," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5504–5508, Aug. 2011, arXiv:0803.3066.

[34] A. M. Childs, D. W. Leung, and H.-K. Lo, "Two-way quantum communication channels," *International Journal of Quantum Information*, vol. 04, no. 01, pp. 63–83, Feb. 2006, arXiv:quant-ph/0506039.

[35] C. H. Bennett, "Simulated time travel, teleportation without communication, and how to conduct a romance with someone who has fallen into a black hole," https://www.research.ibm.com/people/b/bennetc/QUPONBshort.pdf, May 2005.

[36] M. Berta and M. M. Wilde, "Amortization does not enhance the max-Rains information of a quantum channel," *New Journal of Physics*, vol. 20, no. 5, p. 053044, May 2018, arXiv:1709.00200.

[37] S. Bäuml, S. Das, and M. M. Wilde, "Fundamental limits on the capacities of bipartite quantum interactions," *Physical Review Letters*, vol. 121, no. 25, p. 250504, Dec. 2018, arXiv:1812.08223.

[38] S. Das, "Bipartite quantum interactions: Entangling and information processing abilities," Ph.D. dissertation, Louisiana State University, Oct. 2018, available at https://digitalcommons.lsu.edu/gradschool_dissertations/4717/.

[39] E. Kaur, S. Das, M. M. Wilde, and A. Winter, "Extendibility limits the performance of quantum processors," *Physical Review Letters*, vol. 123, no. 7, p. 070502, Aug. 2019, arXiv:1803.10710.

[40] S. Das, S. Bäuml, and M. M. Wilde, "Entanglement and secret-key-agreement capacities of bipartite quantum interactions and read-only memory devices," *Physical Review A*, vol. 101, no. 1, p. 012344, Jan. 2020, arXiv:1712.00827.

[41] E. M. Rains, "Bound on distillable entanglement," *Physical Review A*, vol. 60, no. 1, pp. 179–184, Jul. 1999, arXiv:quant-ph/9809082.

[42] ——, "A semidefinite program for distillable entanglement," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2921–2933, Nov. 2001, arXiv:quant-ph/0008047.

[43] E. Chitambar, J. I. de Vicente, M. W. Girard, and G. Gour, "Entanglement manipulation and distillability beyond LOCC," *Journal of Mathematical Physics*, vol. 61, no. 4, p. 042201, Apr. 2020, arXiv:1711.03835.

[44] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, "Causal and localizable quantum operations," *Physical Review A*, vol. 64, no. 5, p. 052309, Oct. 2001, arXiv:quant-ph/0102043.

[45] T. Eggeling, D. Schlingemann, and R. F. Werner, "Semicausal operations are semilocalizable," *Europhysics Letters*, vol. 57, no. 6, pp. 782–788, Mar. 2002, arXiv:quant-ph/0104027.

[46] M. Piani, M. Horodecki, P. Horodecki, and R. Horodecki, "Properties of quantum nonsignaling boxes," *Physical Review A*, vol. 74, no. 1, p. 012305, Jul. 2006, arXiv:quant-ph/0505110.

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

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

[49] H. Umegaki, "Conditional expectations in an operator algebra IV (entropy and information)," *Kodai Mathematical Seminar Reports*, vol. 14, no. 2, pp. 59–85, 1962.

[50] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: a new generalization and some properties," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122203, Dec. 2013, arXiv:1306.3142.

[51] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy," *Communications in Mathematical Physics*, vol. 331, no. 2, pp. 593–622, Oct. 2014, arXiv:1306.1586.

[52] R. L. Frank and E. H. Lieb, "Monotonicity of a relative Rényi entropy," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122201, Dec. 2013, arXiv:1306.5358.

[53] M. M. Wilde, "Optimized quantum f-divergences and data processing," *Journal of Physics A*, vol. 51, no. 37, p. 374002, Sep. 2018, arXiv:1710.10252.

[54] V. P. Belavkin and P. Staszewski, "C*-algebraic generalization of relative entropy and entropy," *Annales de l'I.H.P. Physique théorique*, vol. 37, no. 1, pp. 51–58, 1982.

[55] D. Petz and M. B. Ruskai, "Contraction of generalized relative entropy under stochastic mappings on matrices," *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, vol. 1, no. 1, pp. 83–89, Jan. 1998.

[56] K. Matsumoto, "A new quantum version of f-divergence," 2013, arXiv:1311.4722.

[57] ——, "A new quantum version of f-divergence," in *Reality and Measurement in Algebraic Quantum Theory*, M. Ozawa, J. Butterfield, H. Halvorson, M. Rédei, Y. Kitajima, and F. Buscemi, Eds., vol. 261. Singapore: Springer Singapore, 2018, pp. 229–273, series Title: Springer Proceedings in Mathematics & Statistics.

[58] V. Katariya and M. M. Wilde, "Geometric distinguishability measures limit quantum channel estimation and discrimination," *Quantum Information Processing*, vol. 20, p. 78, Feb. 2021, arXiv:2004.10708.

[59] F. Buscemi and N. Datta, "The quantum capacity of channels with arbitrarily correlated noise," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1447–1460, Mar. 2010, arXiv:0902.0158.

[60] F. G. S. L. Brandao and N. Datta, "One-shot rates for entanglement manipulation under non-entangling maps," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1754–1760, Mar. 2011, arXiv:0905.2673.

[61] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Physical Review Letters*, vol. 108, no. 20, p. 200501, May 2012, arXiv:1007.5456.

[62] F. Dupuis, L. Kraemer, P. Faist, J. M. Renes, and R. Renner, "Generalized entropies," *XVIIth International Congress on Mathematical Physics*, pp. 134–153, 2013, arXiv:1211.3141.

[63] F. Leditzky, E. Kaur, N. Datta, and M. M. Wilde, "Approaches for approximate additivity of the Holevo information of quantum channels," *Physical Review A*, vol. 97, no. 1, p. 012332, Jan. 2018, arXiv:1709.01111.

[64] T. Cooney, M. Mosonyi, and M. M. Wilde, "Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication," *Communications in Mathematical Physics*, vol. 344, no. 3, pp. 797–829, Jun. 2016, arXiv:1408.3373.

[65] S. Khatri and M. M. Wilde, *Principles of Quantum Communication Theory: A Modern Approach*, Nov. 2020, arXiv:2011.04672v1.

[66] M. Horodecki, P. W. Shor, and M. B. Ruskai, "Entanglement breaking channels," *Reviews in Mathematical Physics*, vol. 15, pp. 629–641, 2003, arXiv:quant-ph/0302031. [Online]. Available: https://doi.org/10.1142/S0129055X03001709

[67] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, "Amortized channel divergence for asymptotic quantum channel discrimination," *Letters in Mathematical Physics*, vol. 110, no. 8, pp. 2277–2336, August 2020, arXiv:1808.01498.

[68] H. Fawzi and O. Fawzi, "Defining quantum divergences via convex optimization," *Quantum*, vol. 5, p. 387, Jan. 2021, arXiv:2007.12576.

[69] A. S. Holevo, "Remarks on the classical capacity of quantum channel," Dec. 2002, quant-ph/0212025.

[70] O. Fawzi, P. Hayden, I. Savov, P. Sen, and M. M. Wilde, "Classical communication over a quantum interference channel," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3670–3691, Jun. 2012, arXiv:1102.2624.

[71] K. Audenaert, N. Datta, and M. Ozols, "Entropy power inequalities for qudits," *Journal of Mathematical Physics*, vol. 57, no. 5, p. 052202, May 2016, arXiv:1503.04213.

[72] R. F. Werner, "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model," *Physical Review A*, vol. 40, no. 8, pp. 4277–4281, Oct. 1989.

[73] T. Eggeling and R. F. Werner, "Separability properties of tripartite states with $u \bigotimes u \bigotimes u$ symmetry," *Physical Review A*, vol. 63, no. 4, p. 042111, Mar. 2001, arXiv:quant-ph/0010096. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.63.042111

[74] P. D. Johnson and L. Viola, "Compatible quantum correlations: Extension problems for Werner and isotropic states," *Physical Review A*, vol. 88, no. 3, p. 032323, September 2013, arXiv:1305.1342.

[75] M. Studzinski, M. Mozrzymas, P. Kopszak, and M. Horodecki, "Efficient multi port-based teleportation schemes," August 2020, arXiv:2008.00984.

[76] M. Grassl, T. Beth, and T. Pellizzari, "Codes for the quantum erasure channel," *Physical Review A*, vol. 56, no. 1, pp. 33–38, July 1997, arXiv:quant-ph/9610042.

[77] B. Bergh, R. Salzmann, and N. Datta, "The $\alpha \to 1$ limit of the sharp quantum Rényi divergence," *Journal of Mathematical Physics*, vol. 62, no. 9, p. 092205, Sep. 2021, arXiv:2102.06576. [Online]. Available: https://doi.org/10.1063/5.0049791

[78] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Transactions on Information Theory*, vol. 55, pp. 5840–5847, Dec. 2009, arXiv:0811.1221. [Online]. Available: https://ieeexplore.ieee.org/document/5319753

[79] M. Mosonyi and F. Hiai, "Some continuity properties of quantum Rényi divergences," Sep. 2022, arXiv:2209.00646.

[80] D. Petz, "Quasi-entropies for states of a von Neumann algebra," *Publications of the Research Institute for Mathematical Sciences*, vol. 21, no. 4, pp. 787–800, 1985. [Online]. Available: https://doi.org/10.2977/prims/1195178929

[81] ——, "Quasi-entropies for finite quantum systems," *Reports in Mathematical Physics*, vol. 23, pp. 57–65, 1986.

[82] M. Tomamichel, "A framework for non-asymptotic quantum information theory," Ph.D. dissertation, ETH Zurich, 2012, arXiv:1203.2142.

[83] A. Kitaev, "Quantum computations: algorithms and error correction," *Russian Mathematical Surveys*, vol. 52, pp. 1191–1249, 1997. [Online]. Available: https://doi.org/10.1070/rm1997v052n06abeh002155

[84] M. Mosonyi and F. Hiai, "On the quantum Rényi relative entropies and related capacity formulas," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2474–2487, Apr. 2011, arXiv:0912.1286.

[85] F. Hiai, "Quantum f-divergences in von Neumann algebras. II. Maximal f-divergences," *Journal of Mathematical Physics*, vol. 60, no. 1, p. 012203, Jan. 2019, arXiv:1807.03118. [Online]. Available: https://doi.org/10.1063/1.5051427

APPENDIX A
ALTERNATIVE FORMULATION OF MEASURE OF FORWARD CLASSICAL COMMUNICATION FOR A BIPARTITE CHANNEL

In this appendix, we prove the equality in (20), and we also provide the background needed to understand it. We also provide an alternate proof of Proposition 6, in order to showcase the utility of the expression in (20).

By definition, $\mathcal{P}_{A\to B}$ is Hermiticity preserving if $\mathcal{P}_{A\to B}(X_A)$ is Hermitian for every Hermitian X_A.

A linear map $\mathcal{P}_{A\to B}$ is Hermiticity preserving if and only if its Choi operator is Hermitian. Suppose that the Choi operator $\Gamma^{\mathcal{P}}_{RB}$ is Hermitian. Then by the standard construction,

$$\mathcal{P}_{A\to B}(X_A) = \langle\Gamma|_{AR} X_A \otimes \Gamma^{\mathcal{P}}_{RB} |\Gamma\rangle_{AR}. \tag{352}$$

Since $\Gamma^{\mathcal{P}}_{RB}$ is Hermitian and X_A is also, it follows that $\mathcal{P}_{A\to B}(X_A)$ is Hermitian. Now suppose that $\mathcal{P}_{A\to B}$ is Hermiticity preserving. Then

$$\begin{aligned}(\Gamma^{\mathcal{P}}_{RB})^\dagger = (\mathcal{P}_{A\to B}(\Gamma_{RA}))^\dagger &= \mathcal{P}_{A\to B}(\Gamma^\dagger_{RA}) \\ &= \mathcal{P}_{A\to B}(\Gamma_{RA}) = \Gamma^{\mathcal{P}}_{RB}.\end{aligned} \tag{353}$$

To every Hermitian operator $R_{AA'BB'}$, there is an associated Hermiticity-preserving map, defined as

$$\begin{aligned}&\mathcal{R}_{AB\to A'B'}(X_{AB}) = \\ &\quad(\langle\Gamma|_{A\hat{A}} \otimes \langle\Gamma|_{B\hat{B}})(X_{AB} \otimes R_{\hat{A}A'\hat{B}B'})(|\Gamma\rangle_{A\hat{A}} \otimes |\Gamma\rangle_{B\hat{B}}).\end{aligned} \tag{354}$$

Consider that

$$\begin{aligned}&\mathcal{R}_{AB\to A'B'}(X_{AB}) \\ &= (\langle\Gamma|_{A\hat{A}} \otimes \langle\Gamma|_{B\hat{B}})(X_{AB} \otimes R_{\hat{A}A'\hat{B}B'})(|\Gamma\rangle_{A\hat{A}} \otimes |\Gamma\rangle_{B\hat{B}})\end{aligned} \tag{355}$$

$$= (\langle\Gamma|_{A\hat{A}} \otimes \langle\Gamma|_{B\hat{B}})(T_{\hat{A}\hat{B}}(X_{\hat{A}\hat{B}})R_{\hat{A}A'\hat{B}B'})(|\Gamma\rangle_{A\hat{A}} \otimes |\Gamma\rangle_{B\hat{B}}) \tag{356}$$

$$= \mathrm{Tr}_{\hat{A}\hat{B}}[T_{\hat{A}\hat{B}}(X_{\hat{A}\hat{B}})R_{\hat{A}A'\hat{B}B'}]. \tag{357}$$

Also, if $R_{AA'BB'} \geq 0$, then $\mathcal{R}_{AB\to A'B'}$ is completely positive. We also make the abbreviation

$$\mathcal{R}_{AB\to A'B'} \geq 0 \qquad \Leftrightarrow \qquad \mathcal{R}_{AB\to A'B'} \in \mathrm{CP}. \tag{358}$$

Then consider that, for positive semi-definite $R_{AA'BB'}$,

$$\begin{aligned}&\|\mathrm{Tr}_{A'B'}[R_{AA'BB'}]\|_\infty \\ &= \sup_{\rho_{AB}\geq 0, \mathrm{Tr}[\rho_{AB}]=1} \mathrm{Tr}[\rho_{AB}\,\mathrm{Tr}_{A'B'}[R_{AA'BB'}]] &(359)\\ &= \sup_{\rho_{AB}\geq 0, \mathrm{Tr}[\rho_{AB}]=1} \mathrm{Tr}[(\rho_{AB}\otimes I_{A'B'})R_{AA'BB'}] &(360)\\ &= \sup_{\rho_{AB}\geq 0, \mathrm{Tr}[\rho_{AB}]=1} \mathrm{Tr}[\mathcal{R}_{AB\to A'B'}(\rho_{AB})] &(361)\\ &=: \|\mathcal{R}_{AB\to A'B'}\|_1 . &(362)\end{aligned}$$

Thus, the function $\beta(\mathcal{M}_{AB\to A'B'})$ for a completely positive map $\mathcal{M}_{AB\to A'B'}$ can be written as

$$\beta(\mathcal{M}_{AB\to A'B'}) = \inf_{\substack{\mathcal{S}_{AB\to A'B'},\\ \mathcal{V}_{AB\to A'B'}\in \mathrm{HermP}}} \|\mathcal{S}_{AB\to A'B'}\|_1 \tag{363}$$

subject to

$$\begin{aligned}T_{B'}\circ(\mathcal{V}_{AB\to A'B'} \pm \mathcal{M}_{AB\to A'B'})\circ T_B \geq 0,\\ \mathcal{S}_{AB\to A'B'} \pm \mathcal{V}_{AB\to A'B'} \geq 0,\\ \mathrm{Tr}_{A'}\circ\mathcal{S}_{AB\to A'B'} = \mathrm{Tr}_{A'}\circ\mathcal{S}_{AB\to A'B'}\circ\mathcal{R}^\pi_A\end{aligned}, \tag{364}$$

where $\mathcal{R}^\pi_A(\cdot) := \mathrm{Tr}_A[\cdot]\pi_A$ and $\pi_A := I_A/d_A$ is the maximally mixed state. Note that $\mathcal{S}_{AB\to A'B'} \geq 0$ follows because $\mathcal{S}_{AB\to A'B'} \pm \mathcal{V}_{AB\to A'B'} \geq 0$ and adding these allows us to conclude that $\mathcal{S}_{AB\to A'B'} \geq 0$.

For a point-to-point channel $\mathcal{N}_{A\to B'}$, this translates to

$$\begin{aligned}
&\beta(\mathcal{N}_{A\to B'}) \\
&:= \inf_{\substack{\mathcal{S}_{A\to B'},\\ \mathcal{V}_{A\to B'}\in \text{HermP}}} \left\{ \begin{array}{c} \left\|\mathcal{S}_{A\to B'}\right\|_1 : \\ T_{B'}\circ(\mathcal{V}_{A\to B'}\pm\mathcal{M}_{A\to B'})\geq 0, \\ \mathcal{S}_{A\to B'}\pm\mathcal{V}_{A\to B'}\geq 0, \\ \mathcal{S}_{A\to B'}=\mathcal{S}_{A\to B'}\circ\mathcal{R}^{\pi}_{A} \end{array} \right\} && (365)\\
&= \inf_{\substack{S_{B'}\in \text{PSD},\\ \mathcal{V}_{A\to B'}\in \text{HermP}}} \left\{ \begin{array}{c} \left\|\mathcal{R}^{S}_{A\to B'}\right\|_1 : \\ T_{B'}\circ(\mathcal{V}_{A\to B'}\pm\mathcal{M}_{A\to B'})\geq 0, \\ \mathcal{R}^{S}_{A\to B'}\pm\mathcal{V}_{A\to B'}\geq 0 \end{array} \right\} && (366)\\
&= \inf_{\substack{S_{B'}\in \text{PSD},\\ \mathcal{V}_{A\to B'}\in \text{HermP}}} \left\{ \begin{array}{c} \text{Tr}[S_{B'}] : \\ T_{B'}\circ(\mathcal{V}_{A\to B'}\pm\mathcal{M}_{A\to B'})\geq 0, \\ \mathcal{R}^{S}_{A\to B'}\pm\mathcal{V}_{A\to B'}\geq 0 \end{array} \right\}, && (367)
\end{aligned}$$

where $\mathcal{R}^{S}_{A\to B'}(\cdot)=\text{Tr}_A[\cdot]S_{B'}$ is a replacer map. Thus,

$$\begin{aligned}
&\beta(\mathcal{N}_{A\to B'}) = \\
&\inf_{\substack{S_{B'}\in \text{PSD},\\ \mathcal{V}_{A\to B'}\in \text{HermP}}} \left\{ \begin{array}{c} \text{Tr}[S_{B'}] : \\ T_{B'}\circ(\mathcal{V}_{A\to B'}\pm\mathcal{M}_{A\to B'})\geq 0, \\ \mathcal{R}^{S}_{A\to B'}\pm\mathcal{V}_{A\to B'}\geq 0 \end{array} \right\}. && (368)
\end{aligned}$$

This kind of formulation is general. For example, consider the following SDP for the diamond norm:

$$\begin{aligned}
&\frac{1}{2}\left\|\mathcal{N}-\mathcal{M}\right\|_{\diamond} = \\
&\quad\inf_{Z_{RB}\geq 0}\left\{\left\|\text{Tr}_B[Z_{RB}]\right\|_{\infty} : Z_{RB}\geq \Gamma^{\mathcal{N}}_{RB}-\Gamma^{\mathcal{M}}_{RB}\right\}. && (369)
\end{aligned}$$

Using the above rephrasing, we can rewrite this optimization as

$$\begin{aligned}
&\frac{1}{2}\left\|\mathcal{N}-\mathcal{M}\right\|_{\diamond} = \\
&\quad\inf_{\mathcal{Z}_{A\to B}\in \text{CP}}\left\{\left\|\mathcal{Z}_{A\to B}\right\|_1 : \mathcal{Z}_{A\to B}\geq \mathcal{N}_{A\to B}-\mathcal{M}_{A\to B}\right\}. && (370)
\end{aligned}$$

We can use the expression in (363) to provide an alternate proof of Proposition 6:

Proposition 32: Let $\mathcal{M}^1_{AB\to A'B'}$ and $\mathcal{M}^2_{A'B'\to A''B''}$ be completely positive maps. Then

$$\begin{aligned}
&\beta(\mathcal{M}^2_{A'B'\to A''B''}\circ\mathcal{M}^1_{AB\to A'B'})\leq \\
&\qquad\beta(\mathcal{M}^2_{A'B'\to A''B''})\cdot\beta(\mathcal{M}^1_{AB\to A'B'}). && (371)
\end{aligned}$$

Proof: Let $\mathcal{S}^1_{AB\to A'B'}$ and $\mathcal{V}^1_{AB\to A'B'}$ be Hermiticity preserving maps satisfying the constraints in (363) for $\mathcal{M}^1_{AB\to A'B'}$, and let $\mathcal{S}^2_{A'B'\to A''B''}$ and $\mathcal{V}^2_{A'B'\to A''B''}$ be Hermiticity preserving maps satisfying the constraints in (363) for $\mathcal{M}^2_{A'B'\to A''B''}$. Then pick

$$\begin{aligned}
\mathcal{S}^3_{AB\to A''B''} &= \mathcal{S}^2_{A'B'\to A''B''}\circ\mathcal{S}^1_{AB\to A'B'}, && (372)\\
\mathcal{V}^3_{AB\to A''B''} &= \mathcal{V}^2_{A'B'\to A''B''}\circ\mathcal{V}^1_{AB\to A'B'}. && (373)
\end{aligned}$$

Also, set

$$\mathcal{M}^3_{AB\to A''B''} = \mathcal{M}^2_{A'B'\to A''B''}\circ\mathcal{M}^1_{AB\to A'B'} \quad (374)$$

Then it follows that

$$\begin{aligned}
T_{B''}\circ(\mathcal{V}^3_{AB\to A''B''}\pm\mathcal{M}^3_{AB\to A''B''})\circ T_B &\geq 0, && (375)\\
\mathcal{S}^3_{AB\to A''B''}\pm\mathcal{V}^3_{AB\to A''B''} &\geq 0. && (376)
\end{aligned}$$

This follows from the general observation that if $\mathcal{A}\pm\mathcal{B}\geq 0$ and $\mathcal{C}\pm\mathcal{D}\geq 0$, then $\mathcal{A}\circ\mathcal{C}\pm\mathcal{B}\circ\mathcal{D}\geq 0$. This in turn follows because

$$\mathcal{A}+\mathcal{B}\geq 0,\quad \mathcal{A}-\mathcal{B}\geq 0,\quad \mathcal{C}+\mathcal{D}\geq 0,\quad \mathcal{C}-\mathcal{D}\geq 0, \quad (377)$$

implies that

$$\begin{aligned}
0 &\leq (\mathcal{A}+\mathcal{B})\circ(\mathcal{C}+\mathcal{D}) && (378)\\
&= \mathcal{A}\circ\mathcal{C}+\mathcal{A}\circ\mathcal{D}+\mathcal{B}\circ\mathcal{C}+\mathcal{B}\circ\mathcal{D}, && (379)\\
0 &\leq (\mathcal{A}+\mathcal{B})\circ(\mathcal{C}-\mathcal{D}) && (380)\\
&= \mathcal{A}\circ\mathcal{C}-\mathcal{A}\circ\mathcal{D}+\mathcal{B}\circ\mathcal{C}-\mathcal{B}\circ\mathcal{D}, && (381)\\
0 &\leq (\mathcal{A}-\mathcal{B})\circ(\mathcal{C}+\mathcal{D}) && (382)\\
&= \mathcal{A}\circ\mathcal{C}+\mathcal{A}\circ\mathcal{D}-\mathcal{B}\circ\mathcal{C}-\mathcal{B}\circ\mathcal{D}, && (383)\\
0 &\leq (\mathcal{A}-\mathcal{B})\circ(\mathcal{C}-\mathcal{D}) && (384)\\
&= \mathcal{A}\circ\mathcal{C}-\mathcal{A}\circ\mathcal{D}-\mathcal{B}\circ\mathcal{C}+\mathcal{B}\circ\mathcal{D}. && (385)
\end{aligned}$$

Now add the first and last to get $\mathcal{A}\circ\mathcal{C}+\mathcal{B}\circ\mathcal{D}\geq 0$ and the second and third to get $\mathcal{A}\circ\mathcal{C}-\mathcal{B}\circ\mathcal{D}\geq 0$.

Now consider that

$$\begin{aligned}
&\text{Tr}_{A''}\circ\mathcal{S}^3_{AB\to A''B''} \\
&= \text{Tr}_{A''}\circ\mathcal{S}^2_{A'B'\to A''B''}\circ\mathcal{S}^1_{AB\to A'B'} && (386)\\
&= \text{Tr}_{A''}\circ\mathcal{S}^2_{A'B'\to A''B''}\circ\mathcal{R}^{\pi}_{A'}\circ\mathcal{S}^1_{AB\to A'B'} && (387)\\
&= \text{Tr}_{A''}\circ\mathcal{S}^2_{A'B'\to A''B''}\circ\mathcal{P}^{\pi}_{A'}\circ\text{Tr}_{A'}\circ\mathcal{S}^1_{AB\to A'B'} && (388)\\
&= \text{Tr}_{A''}\circ\mathcal{S}^2_{A'B'\to A''B''}\circ\mathcal{P}^{\pi}_{A'}\circ\text{Tr}_{A'}\circ\mathcal{S}^1_{AB\to A'B'}\circ\mathcal{R}^{\pi}_{A} && (389)\\
&= \text{Tr}_{A''}\circ\mathcal{S}^2_{A'B'\to A''B''}\circ\mathcal{R}^{\pi}_{A'}\circ\mathcal{S}^1_{AB\to A'B'}\circ\mathcal{R}^{\pi}_{A}. && (390)
\end{aligned}$$

Since the first two lines show that $\text{Tr}_{A''}\circ\mathcal{S}^3_{AB\to A''B''} = \text{Tr}_{A''}\circ\mathcal{S}^2_{A'B'\to A''B''}\circ\mathcal{R}^{\pi}_{A'}\circ\mathcal{S}^1_{AB\to A'B'}$, we conclude that

$$\text{Tr}_{A''}\circ\mathcal{S}^3_{AB\to A''B''} = \text{Tr}_{A''}\circ\mathcal{S}^3_{AB\to A''B''}\circ\mathcal{R}^{\pi}_{A}. \quad (391)$$

Finally, consider that

$$\begin{aligned}
\left\|\mathcal{S}^3_{AB\to A''B''}\right\|_1 &= \left\|\mathcal{S}^2_{A'B'\to A''B''}\circ\mathcal{S}^1_{AB\to A'B'}\right\|_1 && (392)\\
&\leq \left\|\mathcal{S}^2_{A'B'\to A''B''}\right\|_1\cdot\left\|\mathcal{S}^1_{AB\to A'B'}\right\|_1. && (393)
\end{aligned}$$

The inequality follows because the trace norm on superoperators is submultiplicative. So $\mathcal{S}^3_{AB\to A''B''}$ and $\mathcal{V}^3_{AB\to A''B''}$ satisfy the constraints in (363) for $\mathcal{M}^3_{AB\to A''B''}$, so we conclude that

$$\beta(\mathcal{M}^3_{AB\to A''B''})\leq\left\|\mathcal{S}^2_{A'B'\to A''B''}\right\|_1\cdot\left\|\mathcal{S}^1_{AB\to A'B'}\right\|_1. \quad (394)$$

Since the argument holds for all $\mathcal{S}^1_{AB\to A'B'}$ and $\mathcal{V}^1_{AB\to A'B'}$ satisfying the constraints in (363) for $\mathcal{M}^1_{AB\to A'B'}$, and for all $\mathcal{S}^2_{A'B'\to A''B''}$ and $\mathcal{V}^2_{A'B'\to A''B''}$ satisfying the constraints in (363) for $\mathcal{M}^2_{A'B'\to A''B''}$, we conclude the statement of the proposition. ∎

APPENDIX B
THE $\alpha \to 1$ LIMIT OF RÉNYI CHANNEL DIVERGENCES

The following lemma was claimed in [64], but the proof there is not correct. Here, for completeness, we provide a proof, and we note that a different proof has been derived as well [79].

Lemma 33: Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\mathcal{M}_{A\to B}$ be a completely positive map. The following limits hold

$$\lim_{\alpha\to 1} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) = D(\mathcal{N}\|\mathcal{M}), \tag{395}$$
$$\lim_{\alpha\to 1} D_\alpha(\mathcal{N}\|\mathcal{M}) = D(\mathcal{N}\|\mathcal{M}), \tag{396}$$

where $\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M})$ is the sandwiched Rényi channel divergence, $D_\alpha(\mathcal{N}\|\mathcal{M})$ is the Petz–Rényi channel divergence, and $D(\mathcal{N}\|\mathcal{M})$ is the channel relative entropy. Specifically,

$$\begin{aligned}
\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) &:= \sup_{\rho_{RA}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})),\\
D_\alpha(\mathcal{N}\|\mathcal{M}) &:= \sup_{\rho_{RA}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})),\\
D(\mathcal{N}\|\mathcal{M}) &:= \sup_{\rho_{RA}} D(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})),
\end{aligned} \tag{397}$$

where the optimizations are over every state ρ_{RA}, with the reference system R arbitrarily large. The sandwiched Rényi relative entropy $\widetilde{D}_\alpha(\rho\|\sigma)$ is defined in (149) and the quantum relative entropy $D(\rho\|\sigma)$ in (148). The Petz–Rényi relative entropy is defined for all $\alpha \in (0,1)\cup(1,\infty)$ as [80], [81]

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log_2 \operatorname{Tr}[\rho^\alpha \sigma^{1-\alpha}] \tag{398}$$

if $\alpha \in (0,1)$ or, $\alpha \in (1,\infty)$ and $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$. Otherwise, we define $D_\alpha(\rho\|\sigma) = +\infty$.

Proof: We first prove (395) and then argue that similar reasoning establishes (396).

If there exists a state ρ_{RA} such that $\operatorname{supp}(\mathcal{N}_{A\to B}(\rho_{RA})) \not\subseteq \operatorname{supp}(\mathcal{M}_{A\to B}(\rho_{RA}))$, then it follows that the limit on the left-hand side of (395) and the quantity on the right are both equal to $+\infty$. The same is true for (396).

So let us instead consider the case when $\operatorname{supp}(\mathcal{N}_{A\to B}(\rho_{RA})) \subseteq \operatorname{supp}(\mathcal{M}_{A\to B}(\rho_{RA}))$ for every state ρ_{RA}, which is equivalent to the single condition $\operatorname{supp}(\Gamma^{\mathcal{N}}_{RB}) \subseteq \operatorname{supp}(\Gamma^{\mathcal{M}}_{RB})$, where $\Gamma^{\mathcal{N}}_{RB}$ and $\Gamma^{\mathcal{M}}_{RB}$ are the Choi operators of $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$, respectively. If this is the case, then it follows that

$$D_{\max}(\mathcal{N}\|\mathcal{M}) < \infty, \tag{399}$$

where

$$\begin{aligned}
&D_{\max}(\mathcal{N}\|\mathcal{M})\\
&:= \sup_{\rho_{RA}} D_{\max}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) &(400)\\
&= D_{\max}(\Gamma^{\mathcal{N}}\|\Gamma^{\mathcal{M}}), &(401)
\end{aligned}$$

with the latter equality established in [67].

First, we show the following equality, by a straightforward argument:

$$\lim_{\alpha\to 1^-} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) = D(\mathcal{N}\|\mathcal{M}). \tag{402}$$

Indeed, consider that

$$\begin{aligned}
&\lim_{\alpha\to 1^-} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M})\\
&= \sup_{\alpha\in(0,1)} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) &(403)\\
&= \sup_{\alpha\in(0,1)} \sup_{\rho_{RA}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) &(404)\\
&= \sup_{\rho_{RA}} \sup_{\alpha\in(0,1)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) &(405)\\
&= \sup_{\rho_{RA}} D(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) &(406)\\
&= D(\mathcal{N}\|\mathcal{M}). &(407)
\end{aligned}$$

The first equality is a consequence of the α-monotonicity of $\widetilde{D}_\alpha$. The fourth equality is a consequence of (152) and the α-monotonicity of $\widetilde{D}_\alpha$.

The following inequality

$$\lim_{\alpha\to 1^+} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) \geq D(\mathcal{N}\|\mathcal{M}) \tag{408}$$

is straightforward, being a consequence of monotonicity in α of the sandwiched Rényi relative entropies [50], as well as the $\alpha \to 1$ limit [50], [51]. To see it, let ρ_{RA} be an arbitrary state. Then it follows from α-monotonicity that the following inequality holds for all $\alpha > 1$:

$$\begin{aligned}
&\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) \geq\\
&\qquad D(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})).
\end{aligned} \tag{409}$$

Then

$$\begin{aligned}
&\lim_{\alpha\to 1^+} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M})\\
&= \inf_{\alpha>1} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) &(410)\\
&= \inf_{\alpha>1} \sup_{\rho_{RA}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) &(411)\\
&\geq \sup_{\rho_{RA}} \inf_{\alpha>1} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) &(412)\\
&\geq \sup_{\rho_{RA}} D(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) &(413)\\
&= D(\mathcal{N}\|\mathcal{M}). &(414)
\end{aligned}$$

The first equality is a consequence of α-monotonicity.

Let us then establish the opposite inequality. Let ρ be a state and σ a positive semi-definite operator. Recall the following bound from Lemma 8 of [78] (see also Lemma 6.3 of [82]):

$$D_{1+\delta}(\rho\|\sigma) \leq D(\rho\|\sigma) + 4\delta\left[\log_2 \nu(\rho,\sigma)\right]^2, \tag{415}$$

which holds when $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ and for $\delta \in \left(0, \frac{\ln 3}{4\ln \nu(\rho,\sigma)}\right)$, where

$$\begin{aligned}
\nu(\rho,\sigma) &:= \operatorname{Tr}[\rho^{\frac{3}{2}}\sigma^{-\frac{1}{2}}] + \operatorname{Tr}[\rho^{\frac{1}{2}}\sigma^{\frac{1}{2}}] + 1 &(416)\\
&= 2^{\frac{1}{2}D_{\frac{3}{2}}(\rho\|\sigma)} + 2^{-\frac{1}{2}D_{\frac{1}{2}}(\rho\|\sigma)} + 1. &(417)
\end{aligned}$$

Note that $\nu(\rho,\sigma) \geq 3$ (as argued just after [78, Eq. (22)]), as well as

$$\nu(\rho,\sigma) \leq 2^{\frac{1}{2}D_{\max}(\rho\|\sigma)} + \sqrt{\operatorname{Tr}[\sigma]} + 1, \tag{418}$$

which follows because $D_{\frac{3}{2}}(\rho\|\sigma) \leq D_2(\rho\|\sigma) \leq D_{\max}(\rho\|\sigma)$, which in turn follows from the α-monotonicity of the Petz–Rényi relative entropy and the latter inequality was proven

in [59, Lemma 7]. Also, we applied the Cauchy–Schwarz inequality to conclude that $\mathrm{Tr}[\rho^{\frac{1}{2}}\sigma^{\frac{1}{2}}] \leq \sqrt{\mathrm{Tr}[\sigma]}$. From the fact that $\widetilde{D}_{1+\delta}(\rho\|\sigma) \leq D_{1+\delta}(\rho\|\sigma)$ for all $\delta > 0$ [51], we conclude that

$$\widetilde{D}_{1+\delta}(\rho\|\sigma) \leq D(\rho\|\sigma) + 4\delta \left[\log_2(\nu(\rho,\sigma))\right]^2. \tag{419}$$

By picking $\delta \in (0, c)$, where

$$c := \frac{\ln 3}{4 \ln \nu(\mathcal{N},\mathcal{M})}, \tag{420}$$

where

$$\nu(\mathcal{N},\mathcal{M}) := \sup_{\rho_{RA}} \nu(\mathcal{N}_{A\to B}(\rho_{RA}), \mathcal{M}_{A\to B}(\rho_{RA})), \tag{421}$$

with the optimization over every state ρ_{RA}, we find that the following inequality holds for every input state ρ_{RA}:

$$\begin{aligned}\widetilde{D}_{1+\delta}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) \leq \\ D(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) \\ + 4\delta \left[\log_2 \nu(\mathcal{N}_{A\to B}(\rho_{RA}), \mathcal{M}_{A\to B}(\rho_{RA}))\right]^2.\end{aligned} \tag{422}$$

Note that $\nu(\mathcal{N},\mathcal{M}) < \infty$ because $D_{\max}(\mathcal{N}\|\mathcal{M}) < \infty$. Indeed, from (418), we conclude that

$$\nu(\mathcal{N},\mathcal{M}) \leq 2^{\frac{1}{2}D_{\max}(\mathcal{N}\|\mathcal{M})} + \sqrt{\|\mathcal{M}\|_\diamond} + 1, \tag{423}$$

where the diamond norm of $\mathcal{M}$ is defined as [83]

$$\|\mathcal{M}\|_\diamond := \sup_{\rho_{RA}\geq 0, \mathrm{Tr}[\rho_{RA}]=1} \|\mathcal{M}_{A\to B}(\rho_{RA})\|_1. \tag{424}$$

(We could also set

$$c := \frac{\ln 3}{4 \ln(2^{\frac{1}{2}D_{\max}(\mathcal{N}\|\mathcal{M})} + \sqrt{\|\mathcal{M}\|_\diamond} + 1)} \tag{425}$$

if desired, and we note here that an advantage of doing so is that both $D_{\max}(\mathcal{N}\|\mathcal{M})$ and $\|\mathcal{M}\|_\diamond$ are efficiently computable by semi-definite programming.) Now taking a supremum over every input state ρ_{RA}, we conclude that

$$\widetilde{D}_{1+\delta}(\mathcal{N}\|\mathcal{M}) \leq D(\mathcal{N}\|\mathcal{M}) + 4\delta \left[\log_2 \nu(\mathcal{N},\mathcal{M})\right]^2. \tag{426}$$

Thus, by taking the limit of (426) as $\delta \to 0$, we conclude that

$$\lim_{\alpha\to 1^+} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) \leq D(\mathcal{N}\|\mathcal{M}). \tag{427}$$

Putting together (408) and (427), we conclude (395).

A proof of (396) follows exactly the same approach, but we finally use (415) directly instead and similar reasoning as above to establish that $\lim_{\alpha\to 1^+} D_\alpha(\mathcal{N}\|\mathcal{M}) \leq D(\mathcal{N}\|\mathcal{M})$. ■

Now we discuss how to generalize this development to the geometric Rényi and Belavkin–Staszewski relative entropies. We first begin with the following simple extension of Lemma 8 of [78]:

Lemma 34: Let ρ be a quantum state and σ a positive semi-definite operator. Then

$$\widehat{D}_{1+\delta}(\rho\|\sigma) \leq \widehat{D}(\rho\|\sigma) + 4\delta \left[\log_2 \widehat{v}(\rho,\sigma)\right]^2, \tag{428}$$

holds for $\delta \in \left(0, \frac{\ln 3}{4 \ln \widehat{v}(\rho,\sigma)}\right)$, where

$$\widehat{v}(\rho,\sigma) := 2^{\frac{1}{2}\widehat{D}_{3/2}(\rho\|\sigma)} + 2^{-\frac{1}{2}\widehat{D}_{1/2}(\rho\|\sigma)} + 1 \geq 3. \tag{429}$$

Proof: The proof below follows the proof of Lemma 8 of [78] quite closely, with some slight differences to account for the different entropies involved. We provide a detailed proof for completeness. First, suppose that $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$. Then both the left-hand side and right-hand side of (428) are equal to $+\infty$, so that there is nothing to prove in this case.

Now suppose that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, which implies that we can restrict the development to the support of σ, and on this space, σ is invertible. Let us suppose furthermore for now that $\mathrm{supp}(\rho) = \mathrm{supp}(\sigma)$, and then we apply a limit at the end of the proof. As observed in [58, Proposition 72], we can write

$$\widehat{D}_{1+\delta}(\rho\|\sigma) = \frac{1}{\delta}\log_2 \langle\varphi^\rho|X^\delta|\varphi^\rho\rangle, \tag{430}$$

where

$$X := \rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}} \otimes I, \tag{431}$$

$$|\varphi^\rho\rangle := (\rho^{\frac{1}{2}} \otimes I)|\Gamma\rangle, \tag{432}$$

and $|\Gamma\rangle$ is defined in (2). Then consider that

$$\frac{1}{\delta}\log_2 \langle\varphi^\rho|X^\delta|\varphi^\rho\rangle \leq \frac{1}{\delta \ln 2}\left(\langle\varphi^\rho|X^\delta|\varphi^\rho\rangle - 1\right), \tag{433}$$

where we have applied the inequality $\ln x \leq x - 1$, which holds for all $x > 0$. Now expand X^δ as

$$X^\delta = I + \delta \ln X + r_\delta(X), \tag{434}$$

where $r_\delta(X) := X^\delta - \delta \ln X - I$. Then it follows that

$$\begin{aligned}&\widehat{D}_{1+\delta}(\rho\|\sigma) \\ &\leq \frac{1}{\delta \ln 2}\left(\delta\langle\varphi^\rho|\ln X|\varphi^\rho\rangle + \langle\varphi^\rho|r_\delta(X)|\varphi^\rho\rangle\right) &(435)\\ &= \widehat{D}(\rho\|\sigma) + \frac{1}{\delta \ln 2}\langle\varphi^\rho|r_\delta(X)|\varphi^\rho\rangle. &(436)\end{aligned}$$

Now, again applying the inequality $\ln x \leq x - 1$ for $x > 0$, consider that

$$\begin{aligned}r_\delta(x) &= e^{\delta \ln x} - \delta \ln x - 1 &(437)\\ &= e^{\delta \ln x} + \ln\left(\frac{1}{x^\delta}\right) - 1 &(438)\\ &\leq e^{\delta \ln x} + \frac{1}{x^\delta} - 2 &(439)\\ &= e^{\delta \ln x} + e^{-\delta \ln x} - 2 &(440)\\ &= 2\left(\cosh(\delta \ln x) - 1\right) &(441)\\ &=: s_\delta(x). &(442)\end{aligned}$$

Since $\frac{\partial}{\partial x}s_\delta(x) = 2\delta \sinh(\delta \ln x)/x$ and thus $\frac{\partial}{\partial x}s_\delta(x) \geq 0$ for $x \geq 1$, it follows that $s_\delta(x)$ is monotonically increasing in x for $x \geq 1$. Also, since $\frac{\partial^2}{\partial x^2}s_\delta(x) = 2\delta(\delta\cosh(\delta \ln x) - \sinh(\delta \ln x))/x^2$ and thus $\frac{\partial^2}{\partial x^2}s_\delta(x) \leq 0$ for all $\delta \leq 1/2$ and $x \geq 3$, it follows that $s_\delta(x)$ is concave in x for all $\delta \leq 1/2$ and $x \geq 3$. Furthermore, we have that

$$s_\delta(x) = s_\delta(1/x), \tag{443}$$

$$s_\delta(x^2) = s_{2\delta}(x). \tag{444}$$

Then we find, for all $x > 0$, that

$$s_\delta(x) \leq s_\delta\left(x + \frac{1}{x} + 2\right) \tag{445}$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3233924

$$= s_\delta\left(\left(\sqrt{x}+\frac{1}{\sqrt{x}}\right)^2\right) \tag{446}$$

$$= s_{2\delta}\left(\sqrt{x}+\frac{1}{\sqrt{x}}\right) \tag{447}$$

$$\leq s_{2\delta}\left(\sqrt{x}+\frac{1}{\sqrt{x}}+1\right). \tag{448}$$

The first inequality follows from monotonicity of $s_\delta(x)$ in x for $x \geq 1$, as well as $s_\delta(x) = s_\delta(1/x)$. Indeed, for $x \geq 1$, we apply monotonicity to conclude that $s_\delta(x) \leq s_\delta\big(x+\frac{1}{x}+2\big)$. For $x \in (0,1)$, it follows that $1/x > 1$, and so (443) and monotonicity imply that $s_\delta(x) = s_\delta(1/x) \leq s_\delta\big(x+\frac{1}{x}+2\big)$. The second equality follows from applying (444). The last inequality again follows from the facts that $\sqrt{x}+\frac{1}{\sqrt{x}} \geq 1$ for $x > 0$ and from applying monotonicity of $s_\delta(x)$ in x for $x \geq 1$. Now consider that

$$\langle\varphi^\rho|r_\delta(X)|\varphi^\rho\rangle$$

$$\leq \langle\varphi^\rho|s_\delta(X)|\varphi^\rho\rangle \tag{449}$$

$$\leq \langle\varphi^\rho|s_{2\delta}\left(\sqrt{X}+\frac{1}{\sqrt{X}}+I\right)|\varphi^\rho\rangle \tag{450}$$

$$\leq s_{2\delta}(\widehat{v}(\rho,\sigma)). \tag{451}$$

The first inequality follows because the scalar inequality $r_\delta(x) \leq s_\delta(x)$ extends to the operator inequality $r_\delta(X) \leq s_\delta(X)$, holding for all positive definite X. The second inequality follows for a similar reason, but using the scalar inequality $s_\delta(x) \leq s_{2\delta}\left(\sqrt{x}+\frac{1}{\sqrt{x}}+1\right)$. The final inequality follows from Jensen's inequality (see [78, Lemma 11]) and the fact that [58, Eq. (H.172)]

$$\widehat{v}(\rho,\sigma) = \langle\varphi^\rho|(\sqrt{X}+1/\sqrt{X}+I)|\varphi^\rho\rangle, \tag{452}$$

and also because $\sqrt{X}+\frac{1}{\sqrt{X}}+I$ has its eigenvalues in $[3,\infty)$. Note that this latter statement justifies the inequality $\widehat{v}(\rho,\sigma) \geq 3$, which implies that $2\delta < \frac{\ln 3}{2\ln\widehat{v}(\rho,\sigma)} \leq \frac{1}{2}$. Letting $f(y) := 2(\cosh(y)-1)$, Taylor's theorem implies that there exists a constant $c \in [0,y]$ such that

$$f(y) = f(0) + f'(0)y + \frac{f''(c)}{2}y^2 \tag{453}$$

$$= \frac{f''(c)}{2}y^2 \tag{454}$$

$$= \cosh(c)y^2 \tag{455}$$

$$\leq \cosh(y)y^2. \tag{456}$$

Using this and the fact that $s_{2\delta}(x) = f(2\delta\ln x)$, we find that

$$\frac{1}{\delta\ln 2}s_{2\delta}(\widehat{v}(\rho,\sigma))$$

$$\leq \frac{1}{\delta\ln 2}\cosh(2\delta\ln\widehat{v}(\rho,\sigma))\,(2\delta\ln\widehat{v}(\rho,\sigma))^2 \tag{457}$$

$$= 4\delta\left(\log_2\widehat{v}(\rho,\sigma)\right)^2\ln 2\cosh(2\delta\ln\widehat{v}(\rho,\sigma)) \tag{458}$$

$$\leq 4\delta\left(\log_2\widehat{v}(\rho,\sigma)\right)^2. \tag{459}$$

The last inequality follows from the assumption that $\delta \leq \frac{\ln 3}{4\ln\widehat{v}(\rho,\sigma)}$, so that

$$\ln 2\cosh(2\delta\ln\widehat{v}(\rho,\sigma)) \leq \ln 2\cosh\left(\frac{\ln 3}{2}\right) \leq 1. \tag{460}$$

In the case that $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, we define $\rho_\lambda = (1-\lambda)\rho + \lambda\pi_\sigma$, where $\lambda \in [0,1]$ and $\pi_\sigma := \Pi_\sigma/\operatorname{Tr}[\Pi_\sigma]$. Then applying the above development we find that

$$\widehat{D}_{1+\delta}(\rho_\lambda\|\sigma) \leq \widehat{D}(\rho_\lambda\|\sigma) + 4\delta\left[\log_2\widehat{v}(\rho_\lambda,\sigma)\right]^2. \tag{461}$$

The inequality in (428) then follows by taking the limit $\lambda \to 0$. ∎

Now by applying Lemma 34, and an argument similar to that given for Lemma 33, so that

$$\widehat{D}_{1+\delta}(\mathcal{N}\|\mathcal{M}) \leq \widehat{D}(\mathcal{N}\|\mathcal{M}) + 4\delta\left[\log_2\widehat{\nu}(\mathcal{N},\mathcal{M})\right]^2, \tag{462}$$

where

$$\widehat{\nu}(\mathcal{N},\mathcal{M}) := \sup_{\rho_{RA}}\widehat{\nu}(\mathcal{N}_{A\to B}(\rho_{RA}),\mathcal{M}_{A\to B}(\rho_{RA})), \tag{463}$$

we conclude the following:

Lemma 35: Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\mathcal{M}_{A\to B}$ be a completely positive map. The following limit holds

$$\lim_{\alpha\to 1}\widehat{D}_\alpha(\mathcal{N}\|\mathcal{M}) = \widehat{D}(\mathcal{N}\|\mathcal{M}), \tag{464}$$

where $\widehat{D}_\alpha(\mathcal{N}\|\mathcal{M})$ is the geometric Rényi channel divergence and $\widehat{D}(\mathcal{N}\|\mathcal{M})$ is the Belavkin–Staszewski channel relative entropy, both defined from (159).

Finally, we have the following:

Proposition 36: Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\mathcal{C}$ be a compact set of completely positive maps. Then

$$\lim_{\alpha\to 1}\inf_{\mathcal{M}\in\mathcal{C}}D_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}}D(\mathcal{N}\|\mathcal{M}), \tag{465}$$

$$\lim_{\alpha\to 1}\inf_{\mathcal{M}\in\mathcal{C}}\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}}\widetilde{D}(\mathcal{N}\|\mathcal{M}), \tag{466}$$

$$\lim_{\alpha\to 1}\inf_{\mathcal{M}\in\mathcal{C}}\widehat{D}_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}}\widehat{D}(\mathcal{N}\|\mathcal{M}). \tag{467}$$

Proof: First, if there does not exist $\mathcal{M} \in \mathcal{C}$ such that $D_{\max}(\mathcal{N}\|\mathcal{M}) < \infty$, then all quantities are equal to $+\infty$. This is because the condition $D_{\max}(\mathcal{N}\|\mathcal{M}) < \infty$ holds if and only if $\operatorname{supp}(\Gamma^{\mathcal{N}}_{RB}) \subseteq \operatorname{supp}(\Gamma^{\mathcal{M}}_{RB})$, where $\Gamma^{\mathcal{N}}_{RB}$ and $\Gamma^{\mathcal{M}}_{RB}$ are the Choi operators of $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$, respectively, and all of the underlying quantities are equal to $+\infty$ if this condition does not hold (this is the case for D_α, $\widetilde{D}_\alpha$, and $\widehat{D}_\alpha$ for $\alpha > 1$ and it is also the case for these quantities in the limit $\alpha \to 1^-$).

So let us suppose that there is such an $\mathcal{M} \in \mathcal{C}$. We conclude that

$$\lim_{\alpha\to 1^-}\inf_{\mathcal{M}\in\mathcal{C}}D_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}}D(\mathcal{N}\|\mathcal{M}), \tag{468}$$

$$\lim_{\alpha\to 1^-}\inf_{\mathcal{M}\in\mathcal{C}}\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}}\widetilde{D}(\mathcal{N}\|\mathcal{M}), \tag{469}$$

$$\lim_{\alpha\to 1^-}\inf_{\mathcal{M}\in\mathcal{C}}\widehat{D}_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}}\widehat{D}(\mathcal{N}\|\mathcal{M}), \tag{470}$$

by applying Lemmas 33 and 35, the α-monotonicity of the underlying Rényi divergences, as well as [84, Corollary A2], along with the facts that $D_\alpha(\mathcal{N}\|\mathcal{M})$, $\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M})$, and $\widehat{D}_\alpha(\mathcal{N}\|\mathcal{M})$ are lower semi-continuous in $\mathcal{M}$ (see Lemma 37 below).

By employing the fact that the channel relative entropies are ordered with respect to α, so that the limit as $\alpha \to 1^+$ is the

same as the infimum over $\alpha > 1$, and applying Lemmas 33 and 35, we conclude that

$$\lim_{\alpha\to 1^+} \inf_{\mathcal{M}\in\mathcal{C}} D_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}} D(\mathcal{N}\|\mathcal{M}), \tag{471}$$

$$\lim_{\alpha\to 1^+} \inf_{\mathcal{M}\in\mathcal{C}} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}} \widetilde{D}(\mathcal{N}\|\mathcal{M}), \tag{472}$$

$$\lim_{\alpha\to 1^+} \inf_{\mathcal{M}\in\mathcal{C}} \widehat{D}_\alpha(\mathcal{N}\|\mathcal{M}) = \inf_{\mathcal{M}\in\mathcal{C}} \widehat{D}(\mathcal{N}\|\mathcal{M}). \tag{473}$$

This concludes the proof. ∎

Lemma 37: Let $\mathcal{N}$ be a quantum channel and $\mathcal{M}$ a completely positive map. The channel divergences $D_\alpha(\mathcal{N}\|\mathcal{M})$, $\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{M})$, and $\widehat{D}_\alpha(\mathcal{N}\|\mathcal{M})$ are lower semi-continuous in $\mathcal{N}$ and $\mathcal{M}$ for the values of α for which the data-processing inequality holds.

Proof: For a state ρ and a positive semi-definite operator σ, it is known that the underlying divergences $D_\alpha(\rho\|\sigma)$, $\widetilde{D}_\alpha(\rho\|\sigma)$, and $\widehat{D}_\alpha(\rho\|\sigma)$ are lower semi-continuous in ρ and σ for the values of α for which the data-processing inequality holds. This follows from the reasoning in [68, Lemma A.3]. We can then use this prove the desired statement for the channel divergences, and we show the proof explicitly for $D_\alpha(\mathcal{N}\|\mathcal{M})$, with the proofs for the other quantities following the same line of reasoning. Let $\mathcal{N}_n$ be a sequence of channels that converge to $\mathcal{N}$, and let $\mathcal{M}_n$ be a sequence of completely positive maps that converge to $\mathcal{M}$ (we can take the convergence to be in the diamond norm, but it is not so relevant since we are in the finite-dimensional case). Then the desired statement is equivalent to proving that

$$\liminf_{n\to\infty} D_\alpha(\mathcal{N}_n\|\mathcal{M}_n) \geq D_\alpha(\mathcal{N}\|\mathcal{M}). \tag{474}$$

To this end, let ρ_{RA} be an arbitrary state. It then follows that $(\mathrm{id}_R\otimes\mathcal{N}_n)(\rho_{RA}) \to (\mathrm{id}_R\otimes\mathcal{N})(\rho_{RA})$ and $(\mathrm{id}_R\otimes\mathcal{M}_n)(\rho_{RA}) \to (\mathrm{id}_R\otimes\mathcal{M})(\rho_{RA})$. From the lower semi-continuity of D_α, we conclude that

$$\begin{aligned}\liminf_{n\to\infty} &D_\alpha((\mathrm{id}_R\otimes\mathcal{N}_n)(\rho_{RA})\|(\mathrm{id}_R\otimes\mathcal{M}_n)(\rho_{RA})) \\ &\geq D_\alpha((\mathrm{id}_R\otimes\mathcal{N})(\rho_{RA})\|(\mathrm{id}_R\otimes\mathcal{M})(\rho_{RA})).\end{aligned} \tag{475}$$

Since this holds for every state ρ_{RA}, we conclude that

$$D_\alpha(\mathcal{N}\|\mathcal{M})$$

$$= \sup_{\rho_{RA}} D_\alpha((\mathrm{id}_R\otimes\mathcal{N})(\rho_{RA})\|(\mathrm{id}_R\otimes\mathcal{M})(\rho_{RA})) \tag{476}$$

$$\leq \sup_{\rho_{RA}} \liminf_{n\to\infty} D_\alpha((\mathrm{id}_R\otimes\mathcal{N}_n)(\rho_{RA})\|(\mathrm{id}_R\otimes\mathcal{M}_n)(\rho_{RA})) \tag{477}$$

$$\leq \liminf_{n\to\infty} \sup_{\rho_{RA}} D_\alpha((\mathrm{id}_R\otimes\mathcal{N}_n)(\rho_{RA})\|(\mathrm{id}_R\otimes\mathcal{M}_n)(\rho_{RA})) \tag{478}$$

$$= \liminf_{n\to\infty} D_\alpha(\mathcal{N}_n\|\mathcal{M}_n). \tag{479}$$

The second inequality follows because the quantity can only increase with the supremum on the inside. ∎

Remark 38: One can extend the statement of Lemma 37 to values of α beyond those for which data processing holds, by the following argument. For all $\alpha \in (0,\infty)$ and $\varepsilon > 0$, the relative entropies $D_\alpha(\rho\|\sigma+\varepsilon I)$ and $\widetilde{D}_\alpha(\rho\|\sigma+\varepsilon I)$ are continuous in (ρ,σ) and monotone decreasing in ε. Furthermore,

$$D_\alpha(\rho\|\sigma) = \sup_{\varepsilon>0} D_\alpha(\rho\|\sigma+\varepsilon I), \tag{480}$$

$$\widetilde{D}_\alpha(\rho\|\sigma) = \sup_{\varepsilon>0} \widetilde{D}_\alpha(\rho\|\sigma+\varepsilon I). \tag{481}$$

Since the supremum of a set of lower semi-continuous functions is lower semi-continuous, it follows that $D_\alpha(\rho\|\sigma)$ and $\widetilde{D}_\alpha(\rho\|\sigma)$ are lower semi-continuous in (ρ,σ). Since this is all that Lemma 37 relies upon, the desired statement follows for D_α and $\widetilde{D}_\alpha$. A similar conclusion can be made for $\widehat{D}_\alpha$ by invoking Theorem 5.5 of [85], where it was shown that the maximal f-divergence is lower semi-continuous for an arbitrary operator convex function f, and also noting that $\widehat{D}_\alpha$ is an example of a maximal f-divergence.

Dawei Ding received the Ph.D. degree from Stanford University in 2019. He is currently a Research Scientist with the Alibaba Quantum Laboratory in Sunnyvale, California. His research interests span quantum Shannon theory, superconducting quantum computing, quantum computer systems, and quantum gravity.

Sumeet Khatri received the PhD degree from Louisiana State University in 2021. He is currently a postdoctoral researcher at Freie Universitat Berlin. His research interests are quantum communication, quantum computing, and machine learning.

Yihui Quek received her PhD from Stanford University in 2021. She is currently a Humboldt fellow at the Free University of Berlin, and will begin a Harvard Quantum Initiative Prize Postdoctoral Fellowship at Harvard University in late 2022. She is interested in quantum Shannon theory, quantum algorithms and applying learning theory to quantum settings, and is excited to contribute to building a new theory of physical computer science.

Peter W. Shor received a B.S. in Mathematics from Caltech in 1981. He then got a Ph.D. in Applied Mathematics from MIT in 1985. After a one year post-doctoral fellowship at the Mathematical Sciences Research Institute in Berkeley, he took a job at AT&T Bell Laboratories. He worked there until 2003, when he went to MIT as the Morss Professor of Applied Mathematics. In 1994, he discovered an algorithm for factoring large integers into primes on a (still hypothetical) quantum computer. Since then, he has spent most of his research time investigating quantum computing and quantum information theory.

Xin Wang received the Ph.D. degree (Chancellor's List for Best Thesis) in quantum information from the University of Technology Sydney in 2018. He was a Hartree Post-Doctoral Fellow with the Joint Center for Quantum Information and Computer Science, University of Maryland, College Park. He is currently a Staff Researcher with the Institute for Quantum Computing, Baidu Research. His research investigates a broad range of perspectives of quantum computing and quantum information, including quantum communication, entanglement theory, fault-tolerant quantum computing, near-term quantum algorithms, and quantum machine learning.

Mark M. Wilde (M'99–SM'13–F'22) was born in Metairie, Louisiana, USA. He received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, California, in 2008. He is an Associate Professor in the Department of Physics and Astronomy and the Center for Computation and Technology at Louisiana State University, and, in July 2022, he will begin as Associate Professor of Electrical and Computer Engineering at Cornell University. His current research interests are in quantum Shannon theory, quantum computation, quantum optical communication, quantum computational complexity theory, and quantum error correction.