



# Geometric Rényi Divergence and its Applications in Quantum Channel Capacities

Kun Fang<sup>1,2</sup> , Hamza Fawzi<sup>1</sup> 

<sup>1</sup> Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK.  
E-mail: [fangfred11@gmail.com](mailto:fangfred11@gmail.com); [h.fawzi@damtp.cam.ac.uk](mailto:h.fawzi@damtp.cam.ac.uk)

<sup>2</sup> Institute for Quantum Computing, Baidu Research, Beijing 100193, China.

Received: 4 November 2019 / Accepted: 8 March 2021  
Published online: 24 May 2021 – © The Author(s) 2021

**Abstract:** Having a distance measure between quantum states satisfying the right properties is of fundamental importance in all areas of quantum information. In this work, we present a systematic study of the *geometric Rényi divergence* (GRD), also known as the maximal Rényi divergence, from the point of view of quantum information theory. We show that this divergence, together with its extension to channels, has many appealing structural properties, which are not satisfied by other quantum Rényi divergences. For example we prove a chain rule inequality that immediately implies the “amortization collapse” for the geometric Rényi divergence, addressing an open question by Berta et al. [Letters in Mathematical Physics 110:2277–2336, 2020, Equation (55)] in the area of quantum channel discrimination. As applications, we explore various channel capacity problems and construct new channel information measures based on the geometric Rényi divergence, sharpening the previously best-known bounds based on the max-relative entropy while still keeping the new bounds single-letter and efficiently computable. A plethora of examples are investigated and the improvements are evident for almost all cases.

## Contents

1. Introduction	1616
1.1 Main contributions	1617
2. Preliminaries	1619
3. Geometric Rényi Divergence	1621
3.1 Definitions and key properties	1621
3.2 Detailed proofs	1624
4. Quantum Communication	1628
4.1 Background	1628
4.2 Summary of results	1628
4.3 Unassisted quantum capacity	1629

4.4	Two-way assisted quantum capacity	1632
4.5	Extension to bidirectional channels	1637
4.6	Examples	1639
5.	Private Communication	1643
5.1	Background	1643
5.2	Summary of results	1645
5.3	Unassisted private capacity	1645
5.4	Two-way assisted private capacity	1647
6.	Classical Communication	1649
6.1	Background	1649
6.2	Summary of results	1650
6.3	Unassisted classical capacity	1651
6.4	Examples	1654
7.	Magic State Generation	1654
7.1	Background	1654
7.2	Summary of results	1656
7.3	Preliminaries of the resource theory of magic	1657
7.4	Geometric Rényi Thaumata of a channel	1659
7.5	Magic state generation capacity	1662
7.6	Quantum channel synthesis	1665
8.	Conclusions	1666
A.	Technical Lemmas	1667
B.	A Hierarchy for Constant-Bounded Subchannels	1669
C.	A Detailed Comparison for Generalized Amplitude Damping Channels	1670
D.	Application in Quantum Channel Discrimination	1670

## 1. Introduction

In information theory, an imperfect communication link between a sender and a receiver is modeled as a noisy channel. The *capacity* of such a channel is defined as the maximum rate at which information can be transmitted through the channel reliably. This quantity establishes the ultimate boundary between communication rates that are achievable in principle by a channel coding scheme and those that are not. A remarkable result by Shannon [1] states that the capacity of a classical channel is equal to the mutual information of this channel, thus completely settling this capacity problem by a single-letter formula. Quantum information theory generalizes the classical theory, incorporating quantum phenomena like entanglement that have the potential to enhance communication capabilities. Notably, the theory of quantum channels is much richer but less well-understood than that of its classical counterpart. For example, quantum channels have several distinct capacities, depending on what one is trying to use them for, and what additional resources are brought into play. These mainly include the classical capacity, private capacity and quantum capacity, with or without the resource assistance such as classical communication and prior shared entanglement. The only solved case for general quantum channels is the entanglement-assisted classical capacity, which is given by the quantum mutual information of the channel [2] and is believed as the most natural analog to Shannon's formula. The capacities in other communication scenarios are still under investigation. Some recent works (e.g [3,4]) also extend the use of quantum channels to generate quantum resources such as magic state, a key ingredient

for fault-tolerant quantum computation. The capability of a channel to generate such resource is thus characterized by its corresponding generation capacity.

In general, the difficulty in finding exact expressions for the channel capacities has led to a wide body of works to construct achievable (lower) and converse (upper) bounds. We will defer the detailed discussion of these bounds to the following individual sections. There are several important and highly desirable criteria that one would like from any bound on channel capacities. Specifically, one is generally interested in bounds that are:

- **single-letter**; i.e., the bound depends only on a single use of the channel. Several well-established channel coding theorems state that the quantum channel capacity is equal to its corresponding regularized information measure (e.g. the quantum capacity of a channel is equal to its regularized coherent information [5–7]). However, these regularized formulas are simply impossible to evaluate in general using finite computational resources, thus not informative enough in spite of being able to write down as formal mathematical expressions. A single-letter formula could be more mathematically tractable and provides a possibility of its evaluation in practice.
- **computable**; i.e., the formula can be explicitly computed for a given quantum channel. This is essentially required by the nature of capacity that quantifies the “capability” of a channel to transmit information or generate resource. An (efficiently) computable converse bound can help to assess the performance of a channel coding scheme in practice and can also be used as a benchmark for the succeeding research. Note that a single-letter formula is not sufficient to guarantee its computability. An example can be given by the quantum squashed entanglement, which admits a single-letter formula but whose computational complexity is proved to be NP-hard [8].
- **general**; i.e., the bounds holds for arbitrary quantum channels without requiring any additional assumption on their structure, such as degradability or covariance. There are bounds working well for specific quantum channels with a certain structure or sufficient symmetry. However, the noise in practice can be much more versatile than expected and more importantly does not necessarily possess the symmetry we need. A general bound is definitely preferable for the sake of practical interest.
- **strong converse**; i.e., if the communication rate exceeds this bound, then the success probability or the fidelity of transmission of any channel coding scheme converges to zero as the number of channel uses increases. In contrast, the (weak) converse bound only requires the convergence to a scalar not equal to one. Thus a strong converse bound is conceptually more informative than a weak converse bound, leaving no room for the tradeoff between the communication rate and its success probability or fidelity. We call *strong converse capacity* of a channel the best (smallest) possible strong converse bound. If the strong converse capacity coincides with the capacity of a channel, then we say this channel admits the strong converse property. This property is known to hold for all memoryless channels in the classical information theory [9] while it remains open in the quantum regime in general (except for the entanglement-assisted classical capacity [10]). A strong converse bound may witness the strong converse property of certain quantum channels (e.g. [11, 12]), further sharpening our understanding of the quantum theory.

*1.1. Main contributions.* In this paper we propose new bounds on quantum channel capacities that satisfy all criteria mentioned above and that improve on previously known bounds. The main novelty of this work is that our bounds all rely on the so-called *geometric Rényi divergence*. We establish several remarkable properties for this Rényi

divergence that are particularly useful in quantum information theory and show how they can be used to provide bounds on quantum channel capacities. A key ingredient used throughout the paper is the semidefinite programming formulation of the weighted matrix geometric mean [13, 14].

*Geometric Rényi divergence.* The *geometric Rényi divergence* (GRD), is defined as [15]

$$\widehat{D}_\alpha(\rho\|\sigma) \equiv \frac{1}{\alpha - 1} \log \text{Tr} \left[ \sigma^{\frac{1}{2}} \left( \sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} \right)^\alpha \sigma^{\frac{1}{2}} \right], \quad \alpha \in (1, 2].$$

The quantity  $\widehat{D}_\alpha$  is also known as the maximal-Rényi divergence [15] as it can be shown to be the maximal divergence among all quantum Rényi divergences satisfying the data-processing inequality. Different from the widely studied Petz Rényi divergence [16] or sandwiched Rényi divergence [17, 18], the GRD converges to the Belavkin–Staszewski relative entropy [19] when  $\alpha \rightarrow 1$ . The geometric Rényi divergence of two channels  $\mathcal{N}$  and  $\mathcal{M}$  is defined in the usual way as:

$$\widehat{D}_\alpha(\mathcal{N}\|\mathcal{M}) \equiv \max_{\rho_A \in \mathcal{S}(A)} \widehat{D}_\alpha(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'})\|\mathcal{M}_{A' \rightarrow B}(\phi_{AA'})),$$

where  $\mathcal{S}(A)$  is the set of quantum states and  $\phi_{AA'}$  is a purification of  $\rho_A$ . We establish the following key properties of GRD which hold for any  $\alpha \in (1, 2]$ :

1. It lies between the Umegaki relative entropy and the max-relative entropy,

$$D(\rho\|\sigma) \leq \widehat{D}_\alpha(\rho\|\sigma) \leq D_{\max}(\rho\|\sigma).$$

2. Its channel divergence admits a closed-form expression,

$$\widehat{D}_\alpha(\mathcal{N}_{A \rightarrow B}\|\mathcal{M}_{A \rightarrow B}) = \frac{1}{\alpha - 1} \log \left\| \text{Tr}_B \left[ J_{\mathcal{M}}^{\frac{1}{2}} \left( J_{\mathcal{M}}^{-\frac{1}{2}} J_{\mathcal{N}} J_{\mathcal{M}}^{-\frac{1}{2}} \right)^\alpha J_{\mathcal{M}}^{\frac{1}{2}} \right] \right\|_\infty,$$

where  $J_{\mathcal{N}}$  and  $J_{\mathcal{M}}$  are the corresponding Choi matrices of  $\mathcal{N}$  and  $\mathcal{M}$  respectively.

3. Its channel divergence is additive under tensor product of channels,

$$\widehat{D}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2\|\mathcal{M}_1 \otimes \mathcal{M}_2) = \widehat{D}_\alpha(\mathcal{N}_1\|\mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2\|\mathcal{M}_2).$$

4. Its channel divergence is sub-additive under channel composition,

$$\widehat{D}_\alpha(\mathcal{N}_2 \circ \mathcal{N}_1\|\mathcal{M}_2 \circ \mathcal{M}_1) \leq \widehat{D}_\alpha(\mathcal{N}_1\|\mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2\|\mathcal{M}_2).$$

5. It satisfies the chain rule for any quantum states  $\rho_{RA}$ ,  $\sigma_{RA}$  and quantum channels  $\mathcal{N}$  and  $\mathcal{M}$ ,

$$\widehat{D}_\alpha(\mathcal{N}_{A \rightarrow B}(\rho_{RA})\|\mathcal{M}_{A \rightarrow B}(\sigma_{RA})) \leq \widehat{D}_\alpha(\rho_{RA}\|\sigma_{RA}) + \widehat{D}_\alpha(\mathcal{N}\|\mathcal{M}).$$

These properties set a clear difference of GRD with other Rényi divergences. Of particular importance is the chain rule property, which immediately implies that the “amortization collapse” for the geometric Rényi divergence, addressing an open question from [20, Eq. (55)] in the area of quantum channel discrimination. Moreover, due to the closed-form expression of the channel divergence and the semidefinite representation of the matrix geometric means [14], any optimization  $\min_{\mathcal{M} \in \mathcal{V}} \widehat{D}_\alpha(\mathcal{N}\|\mathcal{M})$  can be computed as a semidefinite program if  $\mathcal{V}$  is a set of channels characterized by semidefinite conditions.

**Table 1.** Quantum information tasks studied in this paper, and new bounds on capacities obtained using the geometric Rényi divergence  $\widehat{D}_\alpha$

Tasks	Capacities	Previous bounds ( $D$ or $D_{\max}$ )	New bounds ( $\widehat{D}_\alpha$ )
Quantum	(1) Unassisted $Q$	$R$ [11], $R_{\max}$ [21]	$\widehat{R}_\alpha$ [Theorem 12]
	(2) Two-way $Q^{\leftrightarrow}$	$R_{\max}$ [22]	$\widehat{R}_{\alpha, \ominus}$ [Theorem 17]
	(3) Two-way $Q^{\text{bi}, \leftrightarrow}$	$R_{\max}^{\text{bi}}$ [23]	$\widehat{R}_{\alpha, \ominus}^{\text{bi}}$ [Theorem 21]
Private	(4) unassisted $P$	$E_R$ [12, 24], $E_{\max}$ [25]	$\widehat{E}_\alpha$ [Theorem 25]
	(5) Two-way $P^{\leftrightarrow}$	$E_{\max}$ [25]	$\widehat{E}_{\alpha, \Sigma}$ [Theorem 29]
Classical	(6) Unassisted $C$	$C_\beta, C_\zeta$ [26]	$\widehat{Y}_\alpha$ [Theorem 34]
Magic	(7) Adaptive $C_\psi$	$\theta_{\max}$ [3]	$\widehat{\theta}_\alpha$ [Theorem 43]

*Applications in quantum channel capacities.* We utilize the geometric Rényi divergence to study several different channel capacity problems, including (1) unassisted quantum capacity, (2) two-way assisted quantum capacity, (3) two-way assisted quantum capacity of bidirectional quantum channels, (4) unassisted private capacity, (5) two-way assisted private capacity, (6) unassisted classical capacity, (7) magic state generation capacity, as listed in Table 1. Most existing capacity bounds are based on the max-relative entropy due to its nice properties, such as triangle inequality or semidefinite representations. However, these bounds are expected to be loose as the max-relative entropy stands at the top among the family of quantum divergences. For the bounds based on the Umegaki’s relative entropy, they are unavoidably difficult to compute in general due to their minimax optimization formula. In this work, we construct new channel information measures based on the geometric Rényi divergence, sharpening the previous bounds based on the max-relative entropy in general while still keeping the new bounds single-letter and efficiently computable.<sup>1</sup> A plethora of examples are analyzed in each individual sections and the improvements are evident for almost all cases.

The significance of this work is at least two-fold. First, from the technical side, we showcase that the geometric Rényi divergence, which has not been exploited so far in the quantum information literature, is actually quite useful for channel capacity problems. We regard our work as an initial step towards other interesting applications and expect that the technical tools established in this work can also be used in, for example, quantum network theory, quantum cryptography, as the max-relative entropy also appears as the key entropy in these topics. We include another explicit application in the task of quantum channel discrimination in Appendix D. Second, our new capacity bounds meet all the aforementioned desirable criteria and improve the previously best-known results in general, making them suitable as new benchmarks for computing the capacities of quantum channels.

## 2. Preliminaries

A quantum system, denoted by capital letters (e.g.,  $A, B$ ), is usually modeled by finite-dimensional Hilbert spaces (e.g.,  $\mathcal{H}_A, \mathcal{H}_B$ ). The set of linear operators and the set of positive semidefinite operators on system  $A$  are denoted as  $\mathcal{L}(A)$  and  $\mathcal{P}(A)$ , respectively. The identity operator on system  $A$  is denoted by  $\mathbb{1}_A$ . The set of quantum state on system  $A$  is denoted as  $\mathcal{S}(A) \equiv \{\rho_A \mid \rho_A \geq 0, \text{Tr } \rho_A = 1\}$ . A sub-normalized state is a positive semidefinite operator with trace no greater than one. For any two Hermitian operators  $X,$

<sup>1</sup> For unassisted and two-way assisted private capacities, the new bounds are efficiently computable for general qubit channels.

$Y$ , we denote  $X \ll Y$  if their supports has the inclusion  $\text{supp}(X) \subseteq \text{supp}(Y)$ . The trace norm of  $X$  is given by  $\|X\|_1 \equiv \text{Tr} \sqrt{X^\dagger X}$ . The operator norm  $\|X\|_\infty$  is defined as the maximum eigenvalue of  $\sqrt{X^\dagger X}$ . The set of completely positive (CP) maps from  $A$  to  $B$  is denoted as  $\text{CP}(A : B)$ . A *quantum channel or quantum operation*  $\mathcal{N}_{A \rightarrow B}$  is a completely positive and trace-preserving linear map from  $\mathcal{L}(A)$  to  $\mathcal{L}(B)$ . A *subchannel or suboperation*  $\mathcal{M}_{A \rightarrow B}$  is a completely positive and trace non-increasing linear map from  $\mathcal{L}(A)$  to  $\mathcal{L}(B)$ . Let  $|\Phi\rangle_{A'A} = \sum_i |i\rangle_{A'} |i\rangle_A$  be the unnormalized maximally entangled state. Then the *Choi matrix* of a linear map  $\mathcal{E}_{A' \rightarrow B}$  is defined as  $J_{AB}^\mathcal{E} = (\mathcal{I}_A \otimes \mathcal{E}_{A' \rightarrow B})(|\Phi\rangle\langle\Phi|_{A'A})$ . We will drop the identity map  $\mathcal{I}$  and identity operator  $\mathbb{1}$  if they are clear from the context. The logarithms in this work are taken in the base two.

*Notation for semidefinite representation.* For the simplicity of presenting a semidefinite program, we will introduce some new notations to denote semidefinite conditions. Denote the positive semidefinite condition  $X \geq 0$  as  $\llbracket X \rrbracket_{\mathbb{P}}$ , the equality condition  $X = 0$  as  $\llbracket X \rrbracket_{\mathbb{E}}$ , the Hermitian condition  $X = X^\dagger$  as  $\llbracket X \rrbracket_{\mathbb{H}}$  and the linear condition  $\llbracket X \rrbracket_{\mathbb{L}}$  if  $X$  is certain linear operator. We also denote the Hermitian part of  $X$  as  $X^{\mathbb{H}} \equiv X + X^\dagger$ .

*Quantum divergences.* A functional  $\mathbf{D} : \mathcal{S} \times \mathcal{P} \rightarrow \mathbb{R}$  is a *generalized divergence* if it satisfies the data-processing inequality

$$\mathbf{D}(\mathcal{N}(\rho) \|\mathcal{N}(\sigma)) \leq \mathbf{D}(\rho \|\sigma). \tag{1}$$

For any  $\alpha \in [1/2, 1) \cup (1, \infty)$ , the *sandwiched Rényi divergence* is defined as [17, 18]

$$\tilde{D}_\alpha(\rho \|\sigma) \equiv \frac{1}{\alpha - 1} \log \text{Tr} \left[ \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right]^\alpha, \tag{2}$$

which is the smallest quantum Rényi divergence that satisfies a data-processing inequality, and has been widely used to prove the strong converse property (e.g. [11, 18]). In particular, the sandwiched Rényi divergence is non-decreasing in terms of  $\alpha$ , interpolating the *Umegaki relative entropy*  $D(\rho \|\sigma) \equiv \text{Tr}[\rho (\log \rho - \log \sigma)]$  [27] and the *max-relative entropy*  $D_{\max}(\rho \|\sigma) \equiv \min\{\log t \mid \rho \leq t\sigma\}$  [28, 29] as its two extreme cases,

$$D(\rho \|\sigma) = \lim_{\alpha \rightarrow 1} \tilde{D}_\alpha(\rho \|\sigma) \leq \tilde{D}_\alpha(\rho \|\sigma) \leq \lim_{\alpha \rightarrow \infty} \tilde{D}_\alpha(\rho \|\sigma) = D_{\max}(\rho \|\sigma). \tag{3}$$

Another commonly used quantum variant is the *Petz Rényi divergence* [16] defined as

$$\bar{D}_\alpha(\rho \|\sigma) \equiv \frac{1}{\alpha - 1} \log \text{Tr} \left[ \rho^\alpha \sigma^{1-\alpha} \right], \tag{4}$$

which attains operational significance in the quantum generalization of Hoeffding’s and Chernoff’s bounds on the success probability in binary hypothesis testing [30, 31]. At the limit of  $\alpha \rightarrow 0$ , the Petz Rényi divergence converges to the min-relative entropy [29],

$$\lim_{\alpha \rightarrow 0} \bar{D}_\alpha(\rho \|\sigma) = -\log \text{Tr} \Pi_\rho \sigma \equiv D_{\min}(\rho \|\sigma), \tag{5}$$

where  $\Pi_\rho$  is the projector on the support of  $\rho$ . Due to the the Lieb-Thirring trace inequality [32], it holds for all  $\alpha \in (1, \infty)$  that

$$\tilde{D}_\alpha(\rho \|\sigma) \leq \bar{D}_\alpha(\rho \|\sigma). \tag{6}$$

Both  $\widetilde{D}_\alpha$  and  $\overline{D}_\alpha$  recover the Umegaki relative entropy  $D$  at the limit of  $\alpha \rightarrow 1$ . But they are not easy to optimize over in general.

For any generalized divergence  $D$ , the generalized channel divergence between quantum channel  $\mathcal{N}_{A' \rightarrow B}$  and subchannel  $\mathcal{M}_{A' \rightarrow B}$  is defined as [33,34]

$$D(\mathcal{N} \parallel \mathcal{M}) \equiv \max_{\rho_A \in \mathcal{S}(A)} D(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) \parallel \mathcal{M}_{A' \rightarrow B}(\phi_{AA'})), \tag{7}$$

where  $\phi_{AA'}$  is a purification of  $\rho_A$ . In particular, the max-relative channel divergence has a simple closed form [20, Lemma 12],

$$D_{\max}(\mathcal{N} \parallel \mathcal{M}) = D_{\max}(J_{\mathcal{N}} \parallel J_{\mathcal{M}}), \tag{8}$$

where  $J_{\mathcal{N}}$  and  $J_{\mathcal{M}}$  are the corresponding Choi matrices of  $\mathcal{N}$  and  $\mathcal{M}$  respectively.

### 3. Geometric Rényi Divergence

In this section, we investigate the geometric Rényi divergence and its corresponding channel divergence. Our main contribution in this section is to prove several crucial properties of these divergences which are summarized in Theorem 3. These properties will be extensively used in the following sections.

#### 3.1. Definitions and key properties.

**Definition 1** ([15]). Let  $\rho$  be a quantum state and  $\sigma$  be a sub-normalized state with  $\rho \ll \sigma$  and  $\alpha \in (1, 2]$ , their geometric Rényi divergence<sup>2</sup> is defined as

$$\widehat{D}_\alpha(\rho \parallel \sigma) \equiv \frac{1}{\alpha - 1} \log \text{Tr } G_{1-\alpha}(\rho, \sigma), \tag{9}$$

where  $G_\alpha(X, Y)$  is the weighted matrix geometric mean defined as

$$G_\alpha(X, Y) \equiv X^{\frac{1}{2}} \left( X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right)^\alpha X^{\frac{1}{2}}. \tag{10}$$

Note that a useful fact of matrix geometric mean is that  $G_\alpha(X, Y) = G_{1-\alpha}(Y, X)$  (see e.g. [36]).

*Remark 1.* The geometric Rényi divergence converges to the *Belavkin–Staszewski relative entropy* [19],

$$\lim_{\alpha \rightarrow 1} \widehat{D}_\alpha(\rho \parallel \sigma) = \widehat{D}(\rho \parallel \sigma) \equiv \text{Tr } \rho \log [\rho^{1/2} \sigma^{-1} \rho^{1/2}]. \tag{11}$$

Note that  $D(\rho \parallel \sigma) \leq \widehat{D}(\rho \parallel \sigma)$  in general and they coincide for commuting  $\rho$  and  $\sigma$  [37]. Some basic properties such as joint-convexity, data-processing inequality and the continuity of the geometric Rényi divergence (or more generally, maximal  $f$ -divergence) of states can be found in [15]. Further studies on its reversibility under quantum operations are given in [38,39]. Moreover, the weighted matrix geometric mean admits a semidefinite representation [14] (see also Lemma 46 in Appendix A).

<sup>2</sup> It is also called the maximal Rényi divergence (see e.g. [35, Section 4.2.3]) as it is the largest possible quantum Rényi divergence satisfying the data-processing inequality. We here use the term “geometric” as its closed-form expression is given by the matrix geometric means, depicting the nature of this quantity.

**Definition 2.** For any quantum channel  $\mathcal{N}_{A' \rightarrow B}$ , subchannel  $\mathcal{M}_{A' \rightarrow B}$ , and  $\alpha \in (1, 2]$ , their geometric Rényi channel divergence is defined as

$$\widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M}) \equiv \max_{\rho_A \in \mathcal{S}(A)} \widehat{D}_\alpha(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) \parallel \mathcal{M}_{A' \rightarrow B}(\phi_{AA'})), \tag{12}$$

where  $\phi_{AA'}$  is a purification of  $\rho_A$ .

The following Theorem summarizes several crucial properties of the geometric Rényi divergence and its channel divergence. We present their detailed proofs in the next section.

**Theorem 3** (Main technical results). *The following properties of the geometric Rényi divergence and its channel divergence hold.*<sup>3</sup>

1. (Comparison with  $D$  and  $D_{\max}$ ): For any quantum state  $\rho$ , sub-normalized quantum state  $\sigma$  with  $\rho \ll \sigma$  and  $\alpha \in (1, 2]$ , it holds

$$D(\rho \parallel \sigma) \leq \widehat{D}_\alpha(\rho \parallel \sigma) \leq D_{\max}(\rho \parallel \sigma). \tag{13}$$

2. (Closed-form expression of the channel divergence): For any quantum channel  $\mathcal{N}_{A' \rightarrow B}$ , subchannel  $\mathcal{M}_{A' \rightarrow B}$  and  $\alpha \in (1, 2]$ , the geometric Rényi channel divergence is given by

$$\widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M}) = \frac{1}{\alpha - 1} \log \left\| \text{Tr}_B G_{1-\alpha}(J_{AB}^{\mathcal{N}}, J_{AB}^{\mathcal{M}}) \right\|_\infty, \tag{14}$$

where  $J_{AB}^{\mathcal{N}}$  and  $J_{AB}^{\mathcal{M}}$  are the corresponding Choi matrices of  $\mathcal{N}$  and  $\mathcal{M}$  respectively. Moreover, for the Belavkin–Staszewski channel divergence, its has the closed-form expression:

$$\widehat{D}(\mathcal{N} \parallel \mathcal{M}) = \left\| \text{Tr}_B \left\{ (J_{AB}^{\mathcal{N}})^{\frac{1}{2}} \log \left[ (J_{AB}^{\mathcal{N}})^{\frac{1}{2}} (J_{AB}^{\mathcal{M}})^{-1} (J_{AB}^{\mathcal{N}})^{\frac{1}{2}} \right] (J_{AB}^{\mathcal{N}})^{\frac{1}{2}} \right\} \right\|_\infty. \tag{15}$$

3. (Additivity under tensor product): Let  $\mathcal{N}_1$  and  $\mathcal{N}_2$  be two quantum channels and let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be two subchannels. Then for any  $\alpha \in (1, 2]$ , it holds that

$$\widehat{D}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2 \parallel \mathcal{M}_1 \otimes \mathcal{M}_2) = \widehat{D}_\alpha(\mathcal{N}_1 \parallel \mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2 \parallel \mathcal{M}_2). \tag{16}$$

4. (Chain rule): Let  $\rho$  be a quantum state on  $\mathcal{H}_{RA}$ ,  $\sigma$  be a subnormalized state on  $\mathcal{H}_{RA}$  and  $\mathcal{N}_{A \rightarrow B}$  be a quantum channel,  $\mathcal{M}_{A \rightarrow B}$  be a subchannel and  $\alpha \in (1, 2]$ . Then

$$\widehat{D}_\alpha(\mathcal{N}_{A \rightarrow B}(\rho_{RA}) \parallel \mathcal{M}_{A \rightarrow B}(\sigma_{RA})) \leq \widehat{D}_\alpha(\rho_{RA} \parallel \sigma_{RA}) + \widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M}). \tag{17}$$

5. (Sub-additivity under channel composition): For any quantum channels  $\mathcal{N}_{A \rightarrow B}^1$ ,  $\mathcal{N}_{B \rightarrow C}^2$ , any subchannels  $\mathcal{M}_{A \rightarrow B}^1$ ,  $\mathcal{M}_{B \rightarrow C}^2$  and  $\alpha \in (1, 2]$ , it holds

$$\widehat{D}_\alpha(\mathcal{N}_2 \circ \mathcal{N}_1 \parallel \mathcal{M}_2 \circ \mathcal{M}_1) \leq \widehat{D}_\alpha(\mathcal{N}_1 \parallel \mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2 \parallel \mathcal{M}_2). \tag{18}$$

---

<sup>3</sup> Note that all of the properties can be trivially extended to general completely positive maps. We only state these properties for quantum channels or subchannels as they are most frequently used in quantum information theory.

6. (Semidefinite representation): Let  $\mathcal{V}$  be a set of subchannels from  $A$  to  $B$  characterized by certain semidefinite conditions. For any quantum channel  $\mathcal{N}_{A \rightarrow B}$  and  $\alpha(\ell) = 1 + 2^{-\ell}$  with  $\ell \in \mathbb{N}$ , the optimization  $\min_{\mathcal{M} \in \mathcal{V}} \widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M})$  can be computed by a semidefinite program:

$$2^\ell \cdot \log \min y \quad \text{s.t.} \quad \llbracket M, \{N_i\}_{i=0}^\ell, J_{\mathcal{M}}, y \rrbracket_{\text{H}},$$

$$\llbracket \begin{matrix} M & J_{\mathcal{N}} \\ J_{\mathcal{N}} & N_\ell \end{matrix} \rrbracket_{\text{P}}, \left\{ \llbracket \begin{matrix} J_{\mathcal{N}} & N_i \\ N_i & N_{i-1} \end{matrix} \rrbracket_{\text{P}} \right\}_{i=1}^\ell, \llbracket N_0 - J_{\mathcal{M}} \rrbracket_{\text{E}}, \llbracket y \mathbb{1}_A - \text{Tr}_B M \rrbracket_{\text{P}}, \mathcal{M} \in \mathcal{V},$$

(19)

where  $J_{\mathcal{N}}$  and  $J_{\mathcal{M}}$  are the corresponding Choi matrices of  $\mathcal{N}$  and  $\mathcal{M}$  respectively. Here the short notation that  $\llbracket X \rrbracket_{\text{P}}$ ,  $\llbracket X \rrbracket_{\text{E}}$  and  $\llbracket X \rrbracket_{\text{H}}$  represent the positive semidefinite condition  $X \geq 0$ , the equality condition  $X = 0$  and the Hermitian condition  $X = X^\dagger$ , respectively.

*Remark 2.* Inequality (13) acts as a starting point of our improvement on the previous capacity bounds built on the max-relative entropy. The closed-form expression of the channel divergence directly leads to the additivity property in Item 3 and the semidefinite representation in Item 6. These properties should be contrasted with the situation for the Petz or sandwiched Rényi divergence for channels, for which it is unclear how they can be calculated efficiently. The chain rule is another fundamental property that sets a difference of the geometric Rényi divergence with other variants. Using the notion of amortized channel divergence [20]

$$\widehat{D}_\alpha^A(\mathcal{N} \parallel \mathcal{M}) \equiv \sup_{\rho_{RA}, \sigma_{RA} \in \mathcal{S}(RA)} \left[ \widehat{D}_\alpha(\mathcal{N}_{A \rightarrow B}(\rho_{RA}) \parallel \mathcal{M}_{A \rightarrow B}(\sigma_{RA})) - \widehat{D}_\alpha(\rho_{RA} \parallel \sigma_{RA}) \right],$$

(20)

the chain rule is equivalent to

$$\widehat{D}_\alpha^A(\mathcal{N} \parallel \mathcal{M}) = \widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M}) \quad \text{for } \alpha \in (1, 2].$$

(21)

That is, the ‘‘amortization collapse’’ for the geometric Rényi divergence. This solves an open question from [20, Eq. (55)] in the area of quantum channel discrimination.

*Remark 3.* Note that the properties in Item 3,4,5 do not hold for the Umegaki relative entropy in general unless we consider the regularized channel divergence [40]. This implies that these properties are *not* satisfied by the Petz or sandwiched channel Rényi divergences for  $\alpha$  in the neighbourhood of 1. As such, defining a measure with such desirable properties requires going away from the Umegaki relative entropy. We also note that it is unclear whether the Petz or sandwiched channel Rényi divergences are efficiently computable, let alone having a simple closed-form expression. These indicate that the results we obtained in this work based on the geometric Rényi divergence cannot be easily extended to the Umegaki relative entropy.

*Remark 4.* Except for the condition  $\mathcal{M} \in \mathcal{V}$ , the semidefinite representation in the above Item 6 with  $\alpha(\ell) = 1 + 2^{-\ell}$  is described by  $\ell + 3$  linear matrix inequalities, each of size no larger than  $2d \times 2d$  with  $d = |A||B|$ . Thus the computational complexity (time-usage) for computing  $\min_{\mathcal{M} \in \mathcal{V}} \widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M})$  is the same as computing  $\min_{\mathcal{M} \in \mathcal{V}} D_{\max}(\mathcal{N} \parallel \mathcal{M})$ . In practice, taking  $\ell = 0$  ( $\alpha = 2$ ) already gives an improved result and choosing  $\ell$  around 8 – 10 will make the separation more significant. Moreover, a slight modification can be done [14] to compute the optimization for any rational  $\alpha \in (1, 2]$ . But we will restrict our attention, without loss of generality, to the discrete values  $\alpha(\ell) = 1 + 2^{-\ell}$  with  $\ell \in \mathbb{N}$ .

3.2. *Detailed proofs.* In the following, we give a detailed proof of each property listed in Theorem 3.

**Lemma 4** (Comparison with  $D$  and  $D_{\max}$ ). *For any quantum state  $\rho$ , sub-normalized quantum state  $\sigma$  with  $\rho \ll \sigma$  and  $\alpha \in (1, 2]$ , the following relation holds*

$$D(\rho \parallel \sigma) \leq \tilde{D}_\alpha(\rho \parallel \sigma) \leq \bar{D}_\alpha(\rho \parallel \sigma) \leq \widehat{D}_\alpha(\rho \parallel \sigma) \leq D_{\max}(\rho \parallel \sigma). \tag{22}$$

*Proof.* The first two inequalities follow from Eqs. (3) and (6). The third inequality follows since the geometric Rényi divergence is the largest Rényi divergence satisfying the data-processing inequality (see [15] or [35, Eq. (4.34)]). It remains to prove the last one. Since the geometric Rényi divergence is monotonically non-decreasing with respect to  $\alpha$ ,<sup>4</sup> it suffices to show that  $\widehat{D}_2(\rho \parallel \sigma) \leq D_{\max}(\rho \parallel \sigma)$ . This has been proved in [28, Remark 5.3.2]. We provide here a different proof by using their semidefinite representations. Recall that  $D_{\max}(\rho \parallel \sigma) = \min\{\log t \mid \rho \leq t\sigma\}$ . Denote the optimal solution as  $t$ , and we have  $D_{\max}(\rho \parallel \sigma) = \log t$  with  $0 \leq \rho \leq t\sigma$ . Note that

$$\begin{aligned} \widehat{D}_2(\rho \parallel \sigma) &= \log \operatorname{Tr} \left[ \rho \sigma^{-1} \rho \right] \\ &= \log \min \left\{ \operatorname{Tr} M \mid \rho \sigma^{-1} \rho \leq M \right\} = \log \min \left\{ \operatorname{Tr} M \mid \begin{bmatrix} M & \rho \\ \rho & \sigma \end{bmatrix} \geq 0 \right\}, \end{aligned} \tag{23}$$

where the last equality follows from the Schur complement characterization of the block positive semidefinite matrix. Take  $M = t\rho$ , and we have

$$\begin{bmatrix} M & \rho \\ \rho & \sigma \end{bmatrix} = \begin{bmatrix} t\rho & \rho \\ \rho & \sigma \end{bmatrix} \geq \begin{bmatrix} t\rho & \rho \\ \rho & t^{-1}\rho \end{bmatrix} = \begin{bmatrix} t & 1 \\ 1 & t^{-1} \end{bmatrix} \otimes \rho \geq 0. \tag{24}$$

Thus  $M = t\rho$  is a feasible solution of optimization (23) which implies  $\widehat{D}_2(\rho \parallel \sigma) \leq \log \operatorname{Tr}[t\rho] = \log t = D_{\max}(\rho \parallel \sigma)$ . This completes the proof.  $\square$

Compared with  $D_{\max}$ , it is clear that  $\widehat{D}_\alpha$  gives a tighter approximation of the Umegaki relative entropy  $D$  from above. We provide a concrete example in Fig. 1 to give an intuitive understanding of the relations between different divergences.

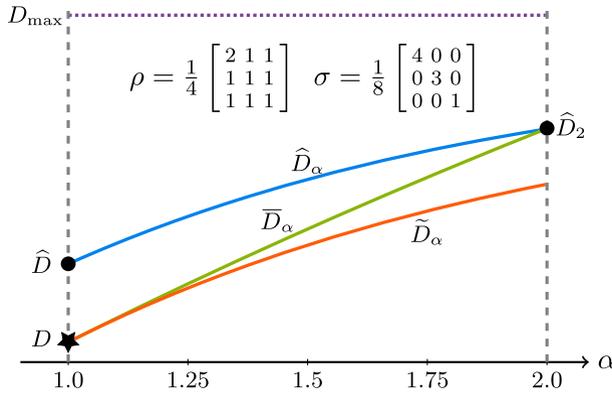
**Lemma 5** (Closed-form expression). *For any quantum channel  $\mathcal{N}_{A' \rightarrow B}$ , subchannel  $\mathcal{M}_{A' \rightarrow B}$  and  $\alpha \in (1, 2]$ , the geometric Rényi channel divergence is given by*

$$\widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M}) = \frac{1}{\alpha - 1} \log \left\| \operatorname{Tr}_B G_{1-\alpha}(J_{AB}^{\mathcal{N}}, J_{AB}^{\mathcal{M}}) \right\|_\infty, \tag{25}$$

where  $J_{AB}^{\mathcal{N}}$  and  $J_{AB}^{\mathcal{M}}$  are the corresponding Choi matrices of  $\mathcal{N}$  and  $\mathcal{M}$  respectively. Moreover, for the Belavkin–Staszewski channel divergence, it has the closed-form expression:

$$\widehat{D}(\mathcal{N} \parallel \mathcal{M}) = \left\| \operatorname{Tr}_B \left\{ (J_{AB}^{\mathcal{N}})^{\frac{1}{2}} \log \left[ (J_{AB}^{\mathcal{N}})^{\frac{1}{2}} (J_{AB}^{\mathcal{M}})^{-1} (J_{AB}^{\mathcal{N}})^{\frac{1}{2}} \right] (J_{AB}^{\mathcal{N}})^{\frac{1}{2}} \right\} \right\|_\infty. \tag{26}$$

<sup>4</sup> This is clear from the minimization formula of  $\widehat{D}_\alpha$  in [15, Eq. (11)] and the monotonicity of classical Rényi divergence.



**Fig. 1.** Relations between Umegaki relative entropy  $D$ , Belavkin–Staszewski relative entropy  $\widehat{D}$ , max-relative entropy  $D_{\max}$ , sandwiched Rényi divergence  $\widetilde{D}_\alpha$ , Petz Rényi divergence  $\overline{D}_\alpha$  and geometric Rényi divergence  $\widehat{D}_\alpha$

*Proof.* Note that for any quantum state  $\rho_A$  and its purification  $\phi_{AA'}$ , we have the relation

$$\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) = \mathcal{N}_{A' \rightarrow B}(\sqrt{\rho_A}|\Phi\rangle\langle\Phi|_{AA'}\sqrt{\rho_A}) = \sqrt{\rho_A}J_{AB}^{\mathcal{N}}\sqrt{\rho_A}. \quad (27)$$

By definition of the geometric Rényi divergence we have

$$\widehat{D}_\alpha(\mathcal{N}\|\mathcal{M}) = \max_{\rho_A \in \mathcal{S}(A)} \widehat{D}_\alpha(\sqrt{\rho_A}J_{AB}^{\mathcal{N}}\sqrt{\rho_A} \parallel \sqrt{\rho_A}J_{AB}^{\mathcal{M}}\sqrt{\rho_A}) \quad (28)$$

$$= \frac{1}{\alpha - 1} \log \max_{\rho_A \in \mathcal{S}(A)} \text{Tr} G_{1-\alpha}(\sqrt{\rho_A}J_{AB}^{\mathcal{N}}\sqrt{\rho_A}, \sqrt{\rho_A}J_{AB}^{\mathcal{M}}\sqrt{\rho_A}) \quad (29)$$

$$= \frac{1}{\alpha - 1} \log \max_{\rho_A \in \mathcal{S}(A)} \text{Tr} \sqrt{\rho_A} G_{1-\alpha}(J_{AB}^{\mathcal{N}}, J_{AB}^{\mathcal{M}})\sqrt{\rho_A} \quad (30)$$

$$= \frac{1}{\alpha - 1} \log \max_{\rho_A \in \mathcal{S}(A)} \text{Tr} G_{1-\alpha}(J_{AB}^{\mathcal{N}}, J_{AB}^{\mathcal{M}})(\rho_A \otimes \mathbb{1}_B) \quad (31)$$

$$= \frac{1}{\alpha - 1} \log \max_{\rho_A \in \mathcal{S}(A)} \text{Tr} \left[ \left[ \text{Tr}_B G_{1-\alpha}(J_{AB}^{\mathcal{N}}, J_{AB}^{\mathcal{M}}) \right] \rho_A \right) \quad (32)$$

$$= \frac{1}{\alpha - 1} \log \left\| \text{Tr}_B G_{1-\alpha}(J_{AB}^{\mathcal{N}}, J_{AB}^{\mathcal{M}}) \right\|_\infty, \quad (33)$$

where the third step follows from the transformer inequality given in Lemma 47 in Appendix A and the fact that we can assume by continuity that  $\rho_A$  has full rank.<sup>5</sup> The last line follows from the semidefinite representation of the infinity norm  $\|X\|_\infty = \max_{\rho \in \mathcal{S}} \text{Tr} X\rho$ .

The expression for  $\widehat{D}$  follows exactly the same steps by using Corollary 48 in Appendix A and replacing the weighted matrix geometric mean with the operator relative entropy.  $\square$

<sup>5</sup> We recall that this follows from a standard continuity argument which works because the geometric Rényi divergence has nice continuity properties, see [41, Page 2] or [15, Eq. (72)].

**Lemma 6** (Additivity). *Let  $\mathcal{N}_1$  and  $\mathcal{N}_2$  be two quantum channels and let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be two subchannels. Then for any  $\alpha \in (1, 2]$  it holds that*

$$\widehat{D}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2 \| \mathcal{M}_1 \otimes \mathcal{M}_2) = \widehat{D}_\alpha(\mathcal{N}_1 \| \mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2 \| \mathcal{M}_2). \tag{34}$$

*Proof.* Due to the closed-form expression in Lemma 5, we have

$$\widehat{D}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2 \| \mathcal{M}_1 \otimes \mathcal{M}_2) = \frac{1}{\alpha - 1} \log \left\| \text{Tr}_{B_1 B_2} G_{1-\alpha}(J_{\mathcal{N}_1} \otimes J_{\mathcal{N}_2}, J_{\mathcal{M}_1} \otimes J_{\mathcal{M}_2}) \right\|_\infty \tag{35}$$

$$= \frac{1}{\alpha - 1} \log \left\| \text{Tr}_{B_1 B_2} G_{1-\alpha}(J_{\mathcal{N}_1}, J_{\mathcal{M}_1}) \otimes G_{1-\alpha}(J_{\mathcal{N}_2}, J_{\mathcal{M}_2}) \right\|_\infty \tag{36}$$

$$= \frac{1}{\alpha - 1} \log \left\| \text{Tr}_{B_1} G_{1-\alpha}(J_{\mathcal{N}_1}, J_{\mathcal{M}_1}) \right\|_\infty \left\| \text{Tr}_{B_2} G_{1-\alpha}(J_{\mathcal{N}_2}, J_{\mathcal{M}_2}) \right\|_\infty \tag{37}$$

$$= \widehat{D}_\alpha(\mathcal{N}_1 \| \mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2 \| \mathcal{M}_2). \tag{38}$$

The first and last lines follow from Lemma 5. The second and third lines follow since the weighted matrix geometric mean and the infinity norm are multiplicative under tensor product.  $\square$

**Lemma 7** (Chain rule). *Let  $\rho$  be a quantum state on  $\mathcal{H}_{RA}$ ,  $\sigma$  be a subnormalized state on  $\mathcal{H}_{RA}$  and  $\mathcal{N}_{A \rightarrow B}$  be a quantum channel,  $\mathcal{M}_{A \rightarrow B}$  be a subchannel and  $\alpha \in (1, 2]$ . Then*

$$\widehat{D}_\alpha(\mathcal{N}_{A \rightarrow B}(\rho_{RA}) \| \mathcal{M}_{A \rightarrow B}(\sigma_{RA})) \leq \widehat{D}_\alpha(\rho_{RA} \| \sigma_{RA}) + \widehat{D}_\alpha(\mathcal{N} \| \mathcal{M}). \tag{39}$$

*Proof.* Let  $|\Phi\rangle_{SA} = \sum_i |i\rangle_S |i\rangle_A$  be the unnormalized maximally entangled state. Denote  $J_{SB}^{\mathcal{N}}$  and  $J_{SB}^{\mathcal{M}}$  as the Choi matrices corresponding to  $\mathcal{N}$  and  $\mathcal{M}$ , respectively. Then we have the identities (see e.g. [22, Eq.(11)])

$$\begin{aligned} \mathcal{N}_{A \rightarrow B}(\rho_{RA}) &= \langle \Phi_{SA} | \rho_{RA} \otimes J_{SB}^{\mathcal{N}} | \Phi_{SA} \rangle \\ \text{and } \mathcal{M}_{A \rightarrow B}(\sigma_{RA}) &= \langle \Phi_{SA} | \sigma_{RA} \otimes J_{SB}^{\mathcal{M}} | \Phi_{SA} \rangle. \end{aligned} \tag{40}$$

For  $y = \left\| \text{Tr}_B G_{1-\alpha}(J_{SB}^{\mathcal{N}}, J_{SB}^{\mathcal{M}}) \right\|_\infty$ , Lemma 5 ensures that

$$\widehat{D}_\alpha(\mathcal{N} \| \mathcal{M}) = \frac{1}{\alpha - 1} \log y \tag{41}$$

and by definition of the infinity norm we find

$$\text{Tr}_B G_{1-\alpha}(J_{SB}^{\mathcal{N}}, J_{SB}^{\mathcal{M}}) \leq y \mathbb{1}_S. \tag{42}$$

By definition of the geometric Rényi divergence and by using (40) we can write

$$\begin{aligned} \widehat{D}_\alpha(\mathcal{N}_{A \rightarrow B}(\rho_{RA}) \| \mathcal{M}_{A \rightarrow B}(\sigma_{RA})) &= \frac{1}{\alpha - 1} \log \text{Tr} G_{1-\alpha}(\langle \Phi_{SA} | \rho_{RA} \otimes J_{SB}^{\mathcal{N}} | \Phi_{SA} \rangle, \langle \Phi_{SA} | \sigma_{RA} \otimes J_{SB}^{\mathcal{M}} | \Phi_{SA} \rangle) \end{aligned} \tag{43}$$

$$\leq \frac{1}{\alpha - 1} \log \text{Tr} \langle \Phi_{SA} | G_{1-\alpha}(\rho_{RA} \otimes J_{SB}^{\mathcal{N}}, \sigma_{RA} \otimes J_{SB}^{\mathcal{M}}) | \Phi_{SA} \rangle \tag{44}$$

$$= \frac{1}{\alpha - 1} \log \text{Tr} \langle \Phi_{SA} | G_{1-\alpha}(\rho_{RA}, \sigma_{RA}) \otimes G_{1-\alpha}(J_{SB}^{\mathcal{N}}, J_{SB}^{\mathcal{M}}) | \Phi_{SA} \rangle \tag{45}$$

$$= \frac{1}{\alpha - 1} \log \text{Tr} \langle \Phi_{SA} | G_{1-\alpha}(\rho_{RA}, \sigma_{RA}) \otimes \text{Tr}_B G_{1-\alpha}(J_{SB}^{\mathcal{N}}, J_{SB}^{\mathcal{M}}) | \Phi_{SA} \rangle \quad (46)$$

$$\leq \frac{1}{\alpha - 1} \log \text{Tr} \langle \Phi_{SA} | G_{1-\alpha}(\rho_{RA}, \sigma_{RA}) \otimes y \mathbb{1}_S | \Phi_{SA} \rangle \quad (47)$$

$$= \frac{1}{\alpha - 1} \log (y \text{Tr} G_{1-\alpha}(\rho_{RA}, \sigma_{RA})) \quad (48)$$

$$= \frac{1}{\alpha - 1} \log y + \frac{1}{\alpha - 1} \log \text{Tr} G_{1-\alpha}(\rho_{RA}, \sigma_{RA}) \quad (49)$$

$$= \widehat{D}_\alpha(\mathcal{N} \| \mathcal{M}) + \widehat{D}_\alpha(\rho_{RA} \| \sigma_{RA}), \quad (50)$$

where the first inequality follows from the transformer inequality given in Lemma 47 in Appendix A. The third line follows from the multiplicativity of weighted matrix geometric mean under tensor product. The second inequality uses (42) and the fact that  $X \mapsto \text{Tr} KX$  is monotone for positive operator  $K$ . Equation (48) follows from the identity  $\langle \Phi_{SA} | Y_{RA} \otimes \mathbb{1}_S | \Phi_{SA} \rangle = \text{Tr}_A Y_{RA}$ .  $\square$

**Lemma 8** (Sub-additivity). *For any quantum channels  $\mathcal{N}_{A \rightarrow B}^1, \mathcal{N}_{B \rightarrow C}^2$ , any subchannels  $\mathcal{M}_{A \rightarrow B}^1, \mathcal{M}_{B \rightarrow C}^2$  and  $\alpha \in (1, 2]$ , it holds*

$$\widehat{D}_\alpha(\mathcal{N}_2 \circ \mathcal{N}_1 \| \mathcal{M}_2 \circ \mathcal{M}_1) \leq \widehat{D}_\alpha(\mathcal{N}_1 \| \mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2 \| \mathcal{M}_2). \quad (51)$$

*Proof.* This is a direct consequence of the chain rule in Lemma 7. For any pure state  $\phi_{AR}$ , we have

$$\widehat{D}_\alpha(\mathcal{N}_2 \circ \mathcal{N}_1(\phi_{AR}) \| \mathcal{M}_2 \circ \mathcal{M}_1(\phi_{AR})) \leq \widehat{D}_\alpha(\mathcal{N}_2 \| \mathcal{M}_2) + \widehat{D}_\alpha(\mathcal{N}_1(\phi_{AR}) \| \mathcal{M}_1(\phi_{AR})) \quad (52)$$

$$\leq \widehat{D}_\alpha(\mathcal{N}_2 \| \mathcal{M}_2) + \widehat{D}_\alpha(\mathcal{N}_1 \| \mathcal{M}_1) + \widehat{D}_\alpha(\phi_{AR} \| \phi_{AR}) \quad (53)$$

$$= \widehat{D}_\alpha(\mathcal{N}_2 \| \mathcal{M}_2) + \widehat{D}_\alpha(\mathcal{N}_1 \| \mathcal{M}_1). \quad (54)$$

Taking a maximization of  $\phi_{AR}$  on the left hand side, we have the desired result.  $\square$

**Lemma 9** (Semidefinite representation). *Let  $\mathcal{V}$  be a set of subchannels from  $A$  to  $B$  characterized by certain semidefinite conditions. For any quantum channel  $\mathcal{N}_{A \rightarrow B}$  and  $\alpha(\ell) = 1 + 2^{-\ell}$  with  $\ell \in \mathbb{N}$ , the optimization  $\min_{\mathcal{M} \in \mathcal{V}} \widehat{D}_\alpha(\mathcal{N} \| \mathcal{M})$  can be computed by a semidefinite program:*

$$2^\ell \cdot \log \min y \quad \text{s.t.} \quad \begin{aligned} & \llbracket M, \{N_i\}_{i=0}^\ell, J_{\mathcal{M}}, y \rrbracket_{\text{H}}, \\ & \left\llbracket \begin{matrix} M & J_{\mathcal{N}} \\ J_{\mathcal{N}} & N_\ell \end{matrix} \right\rrbracket_{\text{P}}, \left\{ \left\llbracket \begin{matrix} J_{\mathcal{N}} & N_i \\ N_i & N_{i-1} \end{matrix} \right\rrbracket_{\text{P}} \right\}_{i=1}^\ell, \left\llbracket N_0 - J_{\mathcal{M}} \right\rrbracket_{\text{E}}, \left\llbracket y \mathbb{1}_A - \text{Tr}_B M \right\rrbracket_{\text{P}}, \mathcal{M} \in \mathcal{V}, \end{aligned} \quad (55)$$

where  $J_{\mathcal{N}}$  and  $J_{\mathcal{M}}$  are the corresponding Choi matrices of  $\mathcal{N}$  and  $\mathcal{M}$  respectively.

*Proof.* This is a direct consequence of the closed-form expression in Lemma 5 and the semidefinite representation of the weighted matrix geometric means in [14] (see also Lemma 46 in Appendix A), as well as the semidefinite representation of the infinity norm of an Hermitian operator  $\|X\|_\infty = \min\{y \mid X \leq y \mathbb{1}\}$ .  $\square$

## 4. Quantum Communication

*4.1. Background.* The *quantum capacity* of a noisy quantum channel is the maximum rate at which it can reliably transmit quantum information over asymptotically many uses of the channel. There are two different quantum capacities of major concern, the (unassisted) quantum capacity  $Q$  and the two-way assisted quantum capacity  $Q^{\leftrightarrow}$ , depending on whether classical communication is allowed between each channel uses.

The well-established quantum capacity theorem shows that the quantum capacity is equal to the regularized channel coherent information [5–7, 42–44],

$$Q(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} I_c(\mathcal{N}^{\otimes n}) = \sup_{n \in \mathbb{N}} \frac{1}{n} I_c(\mathcal{N}^{\otimes n}), \quad (56)$$

where  $I_c(\mathcal{N}) \equiv \max_{\rho \in \mathcal{S}} [H(\mathcal{N}(\rho)) - H(\mathcal{N}^c(\rho))]$  is the channel coherent information,  $H$  is the von Neumann entropy and  $\mathcal{N}^c$  is the complementary channel of  $\mathcal{N}$ . The regularization in (56) is necessary in general since the channel coherent information is non-additive [45, 46] and an unbounded number of channel uses may be required to detect capacity [47]. For this reason, the quantum capacity is notoriously difficult to evaluate, not to mention the scenario with two-way classical communication assistance.

Substantial efforts have been made in providing single-letter lower and upper bounds on  $Q$  and  $Q^{\leftrightarrow}$  (e.g. [48–53]). Most of them require certain symmetries of the channel to be computable or relatively tight. Of particular interest is a strong converse bound given by Tomamichel et al. [11]. Inspired by the Rains bound from entanglement theory [54], they introduced the Rains information ( $R$ ) of a quantum channel and further proved that it is a strong converse rate for quantum communication through the channel. However,  $R$  is not known to be computable in general due to its minimax optimization of the Umegaki relative entropy. For the ease of computability, Wang et al. [21] relaxed the Umegaki relative entropy to the max-relative entropy, obtaining a variant known as the max-Rains information ( $R_{\max}$ ). Leveraging the semidefinite representation of the max-relative entropy, they showed that  $R_{\max}$  is efficiently computable via a semidefinite program. It was later strengthened by Berta and Wilde [22] that  $R_{\max}$  is also a strong converse rate for quantum communication under two-way classical communication assistance. Since then, the max-Rains information  $R_{\max}$  is arguably<sup>6</sup> the best-known computable strong converse bound on both assisted and unassisted quantum capacities in general.

*4.2. Summary of results.* In this part, we aim to improve the bound given by the max-Rains information in both assisted and unassisted scenarios. The structure of this part is organized as follows (see also a schematic diagram in Fig. 2).

In Sect. 4.3 we discuss the unassisted quantum communication. Based on the notion of the generalized Rains information in [11], we exhibit that the generalized Rains information induced by the geometric Rényi divergence ( $\widehat{R}_\alpha$ ) can be computed as a semidefinite program, improving the previously known result of the max-Rains information [21] in general. That is, we show that

$$Q(\mathcal{N}) \leq Q^\dagger(\mathcal{N}) \leq R(\mathcal{N}) \leq \widehat{R}_\alpha(\mathcal{N}) \leq R_{\max}(\mathcal{N}), \quad \text{with } \widehat{R}_\alpha(\mathcal{N}) \text{ SDP computable,}$$

<sup>6</sup> Another known strong converse bound is the entanglement-assisted quantum capacity [10] which can be estimated by an algorithm in [55, 56]. But this bound is usually larger than the max-Rains information since the entanglement assistance is too strong.

where  $Q(\mathcal{N})$  and  $Q^\dagger(\mathcal{N})$  denote the unassisted quantum capacity of channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively.

In Sect. 4.4, we study the quantum communication with PPT assistance, an assistance stronger than the two-way classical communication. We introduce the *generalized Theta-information* which is a new variant of channel information inspired by the channel resource theory (similar to the Upsilon-information in [57]). More precisely, we define the generalized Theta-information as a “channel distance” to the class of subchannels given by the zero set of Holevo–Werner bound ( $Q_\Theta$ ) [48]. Interestingly, we show that the max-Rains information  $R_{\max}$  coincides with the generalized Theta-information induced by the max-relative entropy  $R_{\max, \Theta}$ , i.e.,  $R_{\max} = R_{\max, \Theta}$ , thus providing a completely new perspective of understanding the former quantity. Moreover, we prove that the generalized Theta-information induced by the geometric Rényi divergence ( $\widehat{R}_{\alpha, \Theta}$ ) is a strong converse bound on the PPT-assisted quantum capacity by utilizing an “amortization argument”. Together with its SDP formula, we conclude that  $\widehat{R}_{\alpha, \Theta}$  improves the previous result of the max-Rains information [22] in general. That is, we show that

$$Q^{\text{PPT}, \leftrightarrow}(\mathcal{N}) \leq Q^{\text{PPT}, \leftrightarrow, \dagger}(\mathcal{N}) \leq \widehat{R}_{\alpha, \Theta}(\mathcal{N}) \leq R_{\max}(\mathcal{N}), \quad \text{with } \widehat{R}_{\alpha, \Theta}(\mathcal{N}) \text{ SDP computable,}$$

where  $Q^{\text{PPT}, \leftrightarrow}(\mathcal{N})$  and  $Q^{\text{PPT}, \leftrightarrow, \dagger}(\mathcal{N})$  denote the PPT-assisted quantum capacity of channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively.

In Sect. 4.5, we consider the PPT-assisted quantum communication via bidirectional quantum channels, a more general model than the usual point-to-point channels. We extend the results in Sect. 4.4 to this general model and demonstrate an improvement to the previous result of the bidirectional max-Rains information ( $R_{\max}^{\text{bi}}$ ) [23]. That is, we show that

$$Q^{\text{bi, PPT}, \leftrightarrow}(\mathcal{N}) \leq Q^{\text{bi, PPT}, \leftrightarrow, \dagger}(\mathcal{N}) \leq \widehat{R}_{\alpha, \Theta}^{\text{bi}}(\mathcal{N}) \leq R_{\max}^{\text{bi}}(\mathcal{N}), \quad \text{with } \widehat{R}_{\alpha, \Theta}^{\text{bi}}(\mathcal{N}) \text{ SDP computable,}$$

where  $Q^{\text{bi, PPT}, \leftrightarrow}(\mathcal{N})$  and  $Q^{\text{bi, PPT}, \leftrightarrow, \dagger}(\mathcal{N})$  denote the PPT-assisted quantum capacity of a bidirectional channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively.

Finally in Sect. 4.6 we investigate several fundamental quantum channels, demonstrating the efficiency of our new strong converse bounds. It turns out that our new bounds work exceptionally well and exhibit a significant improvement on the max-Rains information for almost all cases.

**4.3. Unassisted quantum capacity.** In this section, we discuss converse bounds on the unassisted quantum capacity.<sup>7</sup>

**Definition 10** ([11]). For any generalized divergence  $D$ , the generalized Rains bound of a quantum state  $\rho_{AB}$  is defined as

$$R(\rho_{AB}) \equiv \min_{\sigma_{AB} \in \text{PPT}'(A:B)} D(\rho_{AB} \| \sigma_{AB}), \tag{57}$$

where the minimization is taken over the Rains set  $\text{PPT}'(A : B) \equiv \{ \sigma_{AB} \mid \sigma_{AB} \geq 0, \| \sigma_{AB}^{\text{T}_B} \|_1 \leq 1 \}$ .

<sup>7</sup> We refer to the work [11, Section II] for rigorous definitions of the unassisted quantum capacity and its strong converse rate.



*Remark 5.* Note that in the limit of  $\alpha \rightarrow 1$ , the bound  $\widehat{R}_\alpha$  will converge to the Rains information induced by the Belavkin–Staszewski relative entropy due to Eq. (11).

The following result shows how to compute the newly introduced bound  $\widehat{R}_\alpha(\mathcal{N})$  as an SDP.

**Proposition 13** (SDP formula). *For any quantum channel  $\mathcal{N}$  and  $\alpha(\ell) = 1 + 2^{-\ell}$  with  $\ell \in \mathbb{N}$ , it holds*

$$\widehat{R}_\alpha(\mathcal{N}) = \ell \cdot 2^\ell - (2^\ell + 1) \log(2^\ell + 1) + (2^\ell + 1) \log S_\alpha(\mathcal{N}), \tag{62}$$

with  $S_\alpha(\mathcal{N})$  given by the following SDP

$$S_\alpha(\mathcal{N}) = \max \operatorname{Tr} \left[ \left( K^H - \sum_{i=1}^\ell W_i \right) \cdot J_{\mathcal{N}} \right] \quad \text{s.t.} \quad \llbracket K, \{Z_i\}_{i=0}^\ell \rrbracket_{\mathbb{L}}, \llbracket \{W_i\}_{i=1}^\ell, \rho \rrbracket_{\mathbb{H}},$$

$$\left[ \begin{array}{cc} \rho \otimes \mathbb{1} & K \\ K^\dagger & Z_\ell^H \end{array} \right]_{\mathbb{P}}, \left\{ \left[ \begin{array}{cc} W_i & Z_i \\ Z_i^\dagger & Z_{i-1}^H \end{array} \right]_{\mathbb{P}} \right\}_{i=1}^\ell, \left[ \rho \otimes \mathbb{1} \pm \left[ Z_0^H \right]^{\text{T}B} \right]_{\mathbb{P}}, \left[ \operatorname{Tr} \rho - 1 \right]_{\mathbb{E}}, \tag{63}$$

where  $J_{\mathcal{N}}$  is the Choi matrix of  $\mathcal{N}$  and  $X^H \equiv X + X^\dagger$  denotes the Hermitian part of  $X$ .

*Proof.* The proof involves a non-trivial scaling technique for variables replacement, which is important for simplifying the minimax optimization of  $\widehat{R}_\alpha$  to a single SDP. More formally, this proof contains two steps. First we derive a *suitable* SDP formula for  $\widehat{R}_\alpha(\rho_{AB})$  in terms of a maximization problem. Second, we replace  $\rho_{AB}$  as the channel’s output state  $\mathcal{N}_{A' \rightarrow B}(\phi_{AA'})$  and maximize over all the input state  $\rho_A$ . Since the SDP maximization formula for  $\widehat{R}_\alpha(\rho_{AB})$  is not necessarily unique, we need to find a suitable one which is able to give us an overall semidefinite optimization in the second step.

*Step One:* Combining the semidefinite representation of the geometric Rényi divergence in Lemma 46 and the semidefinite representation of the Rains set  $\text{PPT}'(A : B) = \{ \sigma_{AB} \geq 0 \mid \sigma_{AB}^{\text{T}B} = X_{AB} - Y_{AB}, \operatorname{Tr}(X_{AB} + Y_{AB}) \leq 1, X_{AB} \geq 0, Y_{AB} \geq 0 \}$ , we have the SDP formula for the geometric Rényi Rains bound as,

$$\widehat{R}_\alpha(\rho_{AB}) = 2^\ell \cdot \log \min \left[ \operatorname{Tr} M \right] \quad \text{s.t.} \quad \llbracket M, \{N_i\}_{i=0}^\ell \rrbracket_{\mathbb{H}}, \llbracket X, Y \rrbracket_{\mathbb{P}},$$

$$\left[ \begin{array}{cc} M & \rho \\ \rho & N_\ell \end{array} \right]_{\mathbb{P}}, \left\{ \left[ \begin{array}{cc} \rho & N_i \\ N_i & N_{i-1} \end{array} \right]_{\mathbb{P}} \right\}_{i=1}^\ell, \left[ N_0^{\text{T}B} - X + Y \right]_{\mathbb{E}}, \left[ 1 - \operatorname{Tr}(X + Y) \right]_{\mathbb{P}}. \tag{64}$$

By the Lagrange multiplier method, the dual SDP is given by

$$\widehat{R}_\alpha(\rho_{AB}) = 2^\ell \cdot \log \max \left[ \operatorname{Tr} \left[ \left( K^H - \sum_{i=1}^\ell W_i \right) \cdot \rho \right] - f \right]$$

$$\text{s.t.} \quad \llbracket K, \{Z_i\}_{i=0}^\ell \rrbracket_{\mathbb{L}}, \llbracket \{W_i\}_{i=1}^\ell, f \rrbracket_{\mathbb{H}},$$

$$\left[ \begin{array}{cc} \mathbb{1} & K \\ K^\dagger & Z_\ell^H \end{array} \right]_{\mathbb{P}}, \left\{ \left[ \begin{array}{cc} W_i & Z_i \\ Z_i^\dagger & Z_{i-1}^H \end{array} \right]_{\mathbb{P}} \right\}_{i=1}^\ell, \left[ f \mathbb{1} \pm \left[ Z_0^H \right]^{\text{T}B} \right]_{\mathbb{P}}. \tag{65}$$

Due to the Slater’s condition, we can easily check that the strong duality holds. Note that both (64) and (65) are already SDPs for any quantum state  $\rho_{AB}$ . However, the last condition in (65) will introduce a non-linear term if we perform the second step of proof at this stage. The following trick will help us get rid of the variable  $f$  which is essential to obtain the final result. Note that the last condition above implies  $f \geq 0$  and together with the rest conditions we necessarily have  $f > 0$ . Replacing the variables as

$$\widetilde{K} = f^{-1/(2^\ell+1)} K, \quad \widetilde{W}_i = f^{-1/(2^\ell+1)} W_i, \quad \widetilde{Z}_i = f^{-(2^{\ell-i}+1)/(2^\ell+1)} Z_i, \tag{66}$$

we obtain an equivalent SDP of  $\widehat{R}_\alpha(\rho_{AB})$  as

$$2^\ell \cdot \log \max \left[ f^{1/(2^\ell+1)} \operatorname{Tr} \left[ \left( \widetilde{K}^H - \sum_{i=1}^\ell \widetilde{W}_i \right) \cdot \rho \right] - f \right] \text{ s.t. } \left[ \widetilde{K}, \{\widetilde{Z}_i\}_{i=0}^\ell \right]_{\mathbb{L}},$$

$$\left[ \{\widetilde{W}_i\}_{i=1}^\ell, f \right]_{\mathbb{H}}, \left[ \begin{array}{c} \mathbb{1} \quad \widetilde{K} \\ \widetilde{K}^\dagger \quad \widetilde{Z}_\ell^H \end{array} \right]_{\mathbb{P}}, \left\{ \left[ \begin{array}{cc} \widetilde{W}_i & \widetilde{Z}_i \\ \widetilde{Z}_i^\dagger & \widetilde{Z}_{i-1}^H \end{array} \right]_{\mathbb{P}} \right\}_{i=1}^\ell, \left[ \mathbb{1} \pm \left[ \widetilde{Z}_0^H \right]^{\mathbb{T}_B} \right]_{\mathbb{P}}. \quad (67)$$

Denote the objective function  $f^{1/(2^\ell+1)} \cdot a - f$  with  $a = \operatorname{Tr} \left[ \left( \widetilde{K}^H - \sum_{i=1}^\ell \widetilde{W}_i \right) \cdot \rho \right] \geq 0$ . For any fixed value  $a$ , the optimal solution is taken at  $f = [a/(2^\ell + 1)]^{1+1/2^\ell}$  with the maximal value  $2^\ell [a/(2^\ell + 1)]^{1+1/2^\ell}$ . Without loss of generality, we can replace the objective function with  $2^\ell [a/(2^\ell + 1)]^{1+1/2^\ell}$  and get rid of the variable  $f$ . Direct calculation gives us

$$\widehat{R}_\alpha(\rho_{AB}) = \ell \cdot 2^\ell - (2^\ell + 1) \log(2^\ell + 1) + (2^\ell + 1) \log S_\alpha(\rho_{AB}) \quad \text{with}$$

$$S_\alpha(\rho_{AB}) = \max \operatorname{Tr} \left[ \left( K^H - \sum_{i=1}^\ell W_i \right) \cdot \rho \right] \text{ s.t. } \left[ K, \{Z_i\}_{i=0}^\ell \right]_{\mathbb{L}}, \left[ \{W_i\}_{i=1}^\ell \right]_{\mathbb{H}},$$

$$\left[ \begin{array}{cc} \mathbb{1} & K \\ K^\dagger & Z_\ell^H \end{array} \right]_{\mathbb{P}}, \left\{ \left[ \begin{array}{cc} W_i & Z_i \\ Z_i^\dagger & Z_{i-1}^H \end{array} \right]_{\mathbb{P}} \right\}_{i=1}^\ell, \left[ \mathbb{1} \pm \left[ Z_0^H \right]^{\mathbb{T}_B} \right]_{\mathbb{P}}. \quad (68)$$

*Step Two:* Note that  $\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) = \sqrt{\rho_A} J_{\mathcal{N}} \sqrt{\rho_A}$  holds for any quantum state  $\rho_A$  with purification  $\phi_{AA'}$ . Thus the final result is straightforward from (68) by replacing the input state  $\rho_{AB}$  as  $\sqrt{\rho_A} J_{\mathcal{N}} \sqrt{\rho_A}$ , replacing  $K, Z_i, W_i$  as  $\rho_A^{-1/2} K \rho_A^{-1/2}, \rho_A^{-1/2} Z_i \rho_A^{-1/2}, \rho_A^{-1/2} W_i \rho_A^{-1/2}$  respectively and maximizing over all input state  $\rho_A$ .  $\square$

**4.4. Two-way assisted quantum capacity.** In this section, we discuss converse bounds on two-way assisted quantum capacity.<sup>8</sup> Recall that the Rains bound in (57) is essentially established as the divergence between the given state and the Rains set—a set of sub-normalized states given by the zero set<sup>9</sup> of the log-negativity  $E_N(\rho_{AB}) \equiv \log \|\rho_{AB}^{\mathbb{T}_B}\|_1$  [58]. With this in mind, we introduce a new variant of the channel’s analog of Rains bound, compatible with the notion of channel resource theory. Specifically, consider the Holevo–Werner bound [48]—a channel’s analog of the log-negativity,

$$Q_\Theta(\mathcal{N}) \equiv \log \|\Theta_B \circ \mathcal{N}_{A \rightarrow B}\|_\diamond, \quad (69)$$

where  $\Theta$  is the transpose map and  $\|\mathcal{F}_{A' \rightarrow B}\|_\diamond \equiv \sup_{X_{AA'} \in \mathcal{L}(AA')} \|\mathcal{F}_{A' \rightarrow B}(X_{AA'})\|_1 / \|X_{AA'}\|_1$  is the diamond norm [59]. In particular, this bound can be represented as the following SDP,

$$Q_\Theta(\mathcal{N}) = \log \min \left\{ y \mid Y_{AB} \pm J_{\mathcal{N}}^{\mathbb{T}_B} \geq 0, \operatorname{Tr}_B Y_{AB} \leq y \mathbb{1}_A \right\}. \quad (70)$$

Inspired by the formulation of the Rains set, we define the set of subchannels given by the zero set of the Holevo–Werner bound  $Q_\Theta$  as

$$\mathcal{V}_\Theta \equiv \left\{ \mathcal{M} \in \operatorname{CP}(A : B) \mid \exists Y_{AB}, \text{ s.t. } Y_{AB} \pm J_{\mathcal{M}}^{\mathbb{T}_B} \geq 0, \operatorname{Tr}_B Y_{AB} \leq \mathbb{1}_A \right\}. \quad (71)$$

<sup>8</sup> We refer to [22, Section 4] for rigorous definitions of the PPT/two-way assisted quantum capacity and its strong converse rate.

<sup>9</sup> It makes no difference by considering  $\|\rho_{AB}^{\mathbb{T}_B}\|_1 = 1$  or  $\|\rho_{AB}^{\mathbb{T}_B}\|_1 \leq 1$ .

**Definition 14** (*Theta-info.*). For any generalized divergence  $\mathbf{D}$ , the generalized Theta-information<sup>10</sup> of a quantum channel  $\mathcal{N}_{A' \rightarrow B}$  is defined as

$$\mathbf{R}_\Theta(\mathcal{N}) \equiv \min_{\mathcal{M} \in \mathcal{V}_\Theta} \mathbf{D}(\mathcal{N} \| \mathcal{M}) = \min_{\mathcal{M} \in \mathcal{V}_\Theta} \max_{\rho_A \in \mathcal{S}(A)} \mathbf{D}(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) \| \mathcal{M}_{A' \rightarrow B}(\phi_{AA'})), \tag{72}$$

where  $\mathcal{V}_\Theta$  is the Theta set in (71) and  $\phi_{AA'}$  is a purification of quantum state  $\rho_A$ .

*Remark 6.* On the r.h.s. of Eq. (72), the objective function is concave in  $\rho_A$  and convex in  $\mathcal{M}$  [57, Proposition 8]. Thus we can swap the min and max by using Sion’s minimax theorem [60].

The following result compares the generalized Theta-information in (72) and the generalized Rains information in (58) presented in the previous section. Interestingly, these two quantities coincide for the max-relative entropy in general.

**Proposition 15.** For any generalized divergence  $\mathbf{D}$  and any quantum channel  $\mathcal{N}$ , it holds

$$\mathbf{R}(\mathcal{N}) \leq \mathbf{R}_\Theta(\mathcal{N}). \tag{73}$$

Moreover, for the max-relative entropy the equality always holds, i.e.,

$$R_{\max}(\mathcal{N}) = R_{\max, \Theta}(\mathcal{N}). \tag{74}$$

*Proof.* We prove the relation (73) first. Note that for any pure state  $\phi_{AA'}$  and  $\mathcal{M}_{A' \rightarrow B} \in \mathcal{V}_\Theta$ , we have

$$\|(\mathcal{M}_{A' \rightarrow B}(\phi_{AA'}))^{\mathsf{T}^B}\|_1 = \|\Theta_B \circ \mathcal{M}_{A' \rightarrow B}(\phi_{AA'})\|_1 \leq \|\Theta_B \circ \mathcal{M}_{A' \rightarrow B}\|_\diamond \leq 1. \tag{75}$$

This implies  $\mathcal{M}_{A' \rightarrow B}(\phi_{AA'}) \in \text{PPT}'(A : B)$ . Then it holds

$$\mathbf{R}(\mathcal{N}) = \max_{\rho_A \in \mathcal{S}(A)} \min_{\sigma_{AB} \in \text{PPT}'(A:B)} \mathbf{D}(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) \| \sigma_{AB}) \tag{76}$$

$$\leq \max_{\rho_A \in \mathcal{S}(A)} \min_{\mathcal{M} \in \mathcal{V}_\Theta} \mathbf{D}(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) \| \mathcal{M}_{A' \rightarrow B}(\phi_{AA'})) \tag{77}$$

$$= \min_{\mathcal{M} \in \mathcal{V}_\Theta} \max_{\rho_A \in \mathcal{S}(A)} \mathbf{D}(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) \| \mathcal{M}_{A' \rightarrow B}(\phi_{AA'})) \tag{78}$$

$$= \mathbf{R}_\Theta(\mathcal{N}). \tag{79}$$

The first and last line follow by definition. The inequality holds since  $\mathcal{M}_{A' \rightarrow B}(\phi_{AA'}) \in \text{PPT}'(A : B)$  and thus the first line is minimizing over a larger set. In the third line, we swap the min and max by the argument in Remark 6.

We next prove the Eq. (74). Recall that the SDP formula of the max-Rains information is given by ([21, Proposition 5] or [61, Eq. (11)])

$$R_{\max}(\mathcal{N}) = \log \min \left\{ \mu \mid (V - Y)^{\mathsf{T}^B} \geq J_{\mathcal{N}}, \text{Tr}_B(V + Y) \leq \mu \mathbb{1}_A, Y, V \geq 0 \right\}. \tag{80}$$

Replace  $V$  and  $Y$  with  $\mu V$  and  $\mu Y$  respectively, and then denote  $N = (V - Y)^{\mathsf{T}^B}$ , we have

$$R_{\max}(\mathcal{N}) = \log \min \left\{ \mu \mid J_{\mathcal{N}} \leq \mu N, N = (V - Y)^{\mathsf{T}^B}, \text{Tr}_B(V + Y) \leq \mathbb{1}_A, Y, V \geq 0 \right\}. \tag{81}$$

<sup>10</sup> The name follows from the Theta set  $\mathcal{V}_\Theta$  where  $\Theta$  was originally used as the transpose map in the Holevo–Werner bound.

Notice that the second to the last conditions define a set of CP maps

$$\mathcal{V} \equiv \left\{ \mathcal{M} \in \text{CP}(A : B) \mid \exists V_{AB}, Y_{AB}, \text{ s.t. } J_{\mathcal{M}}^{\text{T}_B} = V - Y, \text{Tr}_B(V + Y) \leq \mathbb{1}_A, Y, V \geq 0 \right\}. \tag{82}$$

Combining (81) and (82), we obtain  $R_{\max}(\mathcal{N}) = \min_{\mathcal{M} \in \mathcal{V}} D_{\max}(\mathcal{N} \| \mathcal{M})$ . Thus it suffices for us to show the equivalence  $\mathcal{V} = \mathcal{V}_{\ominus}$ . For any  $\mathcal{M} \in \mathcal{V}_{\ominus}$ , take  $V = (R + J_{\mathcal{M}}^{\text{T}_B})/2$  and  $Y = (R - J_{\mathcal{M}}^{\text{T}_B})/2$ . Then  $V \geq 0, Y \geq 0, J_{\mathcal{M}}^{\text{T}_B} = V - Y$  and  $\text{Tr}_B(V + Y) = \text{Tr}_B R \leq \mathbb{1}_A$ , which implies  $\mathcal{M} \in \mathcal{V}$ . On the other hand, for any  $\mathcal{M} \in \mathcal{V}$ , take  $R = V + Y$ . We can check that  $\text{Tr}_B R = \text{Tr}_B(V + Y) \leq \mathbb{1}_A, R + J_{\mathcal{M}}^{\text{T}_B} = (V + Y) + (V - Y) = 2V \geq 0$  and  $R - J_{\mathcal{M}}^{\text{T}_B} = (V + Y) - (V - Y) = 2Y \geq 0$ , which implies  $\mathcal{M} \in \mathcal{V}_{\ominus}$ . Finally we have

$$R_{\max}(\mathcal{N}) = \min_{\mathcal{M} \in \mathcal{V}} D_{\max}(\mathcal{N} \| \mathcal{M}) = \min_{\mathcal{M} \in \mathcal{V}_{\ominus}} D_{\max}(\mathcal{N} \| \mathcal{M}) = R_{\max, \ominus}(\mathcal{N}), \tag{83}$$

which completes the proof.  $\square$

We proceed to consider the geometric Rényi divergence and show its amortization property, a key ingredient to proving the strong converse bound on the assisted quantum capacity in Theorem 17.

Suppose Alice and Bob share a quantum state  $\rho_{A'AB'}$  with the system cut  $A'A : B'$ . Their shared entanglement with respect to the measure  $\widehat{R}_{\alpha}$  is given by  $\widehat{R}_{\alpha}(\rho_{A'AB'})$ . If Alice redistributes part of her system  $A$  through the channel  $\mathcal{N}_{A \rightarrow B}$  and Bob receives the output system  $B$ , then their shared state becomes to  $\omega_{A':BB'} = \mathcal{N}_{A \rightarrow B}(\rho_{A'AB'})$  with the shared entanglement evaluated as  $\widehat{R}_{\alpha}(\omega_{A':BB'})$ . The amortization inequality shows that the amount of entanglement change after the state redistribution is upper bounded by the channel’s information measure  $\widehat{R}_{\alpha, \ominus}(\mathcal{N})$ .

**Proposition 16** (Amortization). *For any quantum state  $\rho_{A'AB'}$ , any quantum channel  $\mathcal{N}_{A \rightarrow B}$  and the parameter  $\alpha \in (1, 2]$ , it holds*

$$\widehat{R}_{\alpha}(\omega_{A':BB'}) \leq \widehat{R}_{\alpha}(\rho_{A'AB'}) + \widehat{R}_{\alpha, \ominus}(\mathcal{N}_{A \rightarrow B}) \quad \text{with} \quad \omega_{A':BB'} = \mathcal{N}_{A \rightarrow B}(\rho_{A'AB'}). \tag{84}$$

*Proof.* This is a direct consequence of the chain rule property of the geometric Rényi divergence in Lemma 7. Suppose the optimal solutions of  $\widehat{R}_{\alpha}(\rho_{A'AB'})$  and  $\widehat{R}_{\alpha, \ominus}(\mathcal{N}_{A \rightarrow B})$  are taken at  $\sigma_{A'AB'} \in \text{PPT}'(A'A : B')$  and  $\mathcal{M} \in \mathcal{V}_{\ominus}$ , respectively. Let  $\gamma_{A'BB'} = \mathcal{M}_{A \rightarrow B}(\sigma_{A'AB'})$ . We have

$$\left\| \gamma_{A'BB'}^{\text{T}_{BB'}} \right\|_1 = \left\| \Theta_B \circ \mathcal{M}_{A \rightarrow B}(\sigma_{A'AB'}^{\text{T}_{B'}}) \right\|_1 \leq \left\| \Theta_B \circ \mathcal{M}_{A \rightarrow B} \right\|_{\diamond} \left\| \sigma_{A'AB'}^{\text{T}_{B'}} \right\|_1 \leq 1, \tag{85}$$

where the first inequality follows from the definition of diamond norm and the second inequality follows from the choice of  $\sigma_{A'AB'}$  and  $\mathcal{M}_{A \rightarrow B}$ . Thus  $\gamma_{A'BB'} \in \text{PPT}'(A' : BB')$  and forms a feasible solution for  $\widehat{R}_{\alpha}(\omega_{A':BB'})$ . Then we have

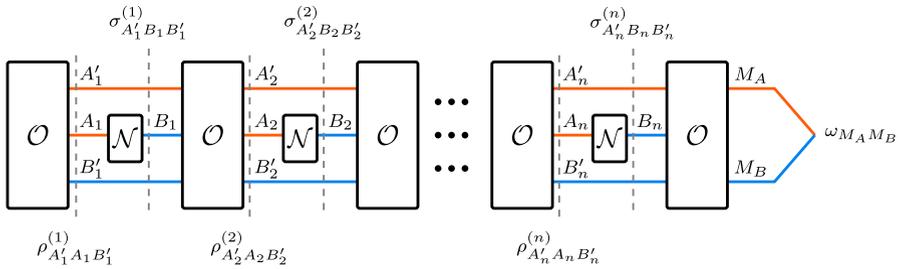
$$\widehat{R}_{\alpha}(\omega_{A':BB'}) \leq \widehat{D}_{\alpha}(\omega_{A':BB'} \| \gamma_{A'BB'}) \tag{86}$$

$$= \widehat{D}_{\alpha}(\mathcal{N}_{A \rightarrow B}(\rho_{A'AB'}) \| \mathcal{M}_{A \rightarrow B}(\sigma_{A'AB'})) \tag{87}$$

$$\leq \widehat{D}_{\alpha}(\mathcal{N} \| \mathcal{M}) + \widehat{D}_{\alpha}(\rho_{A'AB'} \| \sigma_{A'AB'}) \tag{88}$$

$$= \widehat{R}_{\alpha, \ominus}(\mathcal{N}_{A \rightarrow B}) + \widehat{R}_{\alpha}(\rho_{A'AB'}). \tag{89}$$

The second inequality follows from the chain rule of the geometric Rényi divergence in Lemma 7, and the last line follows by the optimality assumption of  $\mathcal{M}$  and  $\sigma$ .  $\square$



**Fig. 3.** A schematic diagram for the protocol of  $\mathcal{O}$ -assisted quantum communication that uses a quantum channel  $\mathcal{N}$  times, where  $\mathcal{O}$  is usually chosen as LOCC or PPT. Every channel use is interleaved by an operation in the class  $\mathcal{O}$ . The goal of such a protocol is to produce an approximate maximally entangled state  $\omega_{M_A M_B}$  between Alice and Bob. The systems in red are held by Alice while the systems in blue are held by Bob

**Theorem 17** (Application 2). *For any quantum channel  $\mathcal{N}$  and  $\alpha \in (1, 2]$ , it holds*

$$Q^{\text{PPT}, \leftrightarrow}(\mathcal{N}) \leq Q^{\text{PPT}, \leftrightarrow, \dagger}(\mathcal{N}) \leq \widehat{R}_{\alpha, \Theta}(\mathcal{N}) \leq R_{\max}(\mathcal{N}), \tag{90}$$

where  $Q^{\text{PPT}, \leftrightarrow}(\mathcal{N})$  and  $Q^{\text{PPT}, \leftrightarrow, \dagger}(\mathcal{N})$  denote the PPT-assisted quantum capacity of channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively.

*Proof.* The first inequality holds by definition. The last inequality holds since we have  $\widehat{R}_{\alpha, \Theta}(\mathcal{N}) \leq R_{\max, \Theta}(\mathcal{N}) = R_{\max}(\mathcal{N})$  by Lemma 4 and Proposition 15, respectively. It remains to prove the second one. Once we have the amortization inequality in Proposition 16, the proof of the second inequality will closely follow the one in [22, Theorem 3]. Consider  $n$  round PPT-assisted quantum communication protocol illustrated in Fig. 3. For the  $i$ -th round, denote the input state of  $\mathcal{N}$  as  $\rho_{A'_i A_i B'_i}^{(i)}$  and the output state as  $\sigma_{A'_i B_i B'_i}^{(i)}$ . The final state after  $n$  rounds communication is denoted as  $\omega_{M_A M_B}$ . Then we have

$$\widehat{R}_{\alpha}(\omega_{M_A M_B}) \leq \widehat{R}_{\alpha}(\sigma_{A':B B'}^{(n)}) \tag{91}$$

$$= \widehat{R}_{\alpha}(\sigma_{A':B B'}^{(n)}) - \widehat{R}_{\alpha}(\rho_{A' A: B'}^{(1)}) \tag{92}$$

$$\leq \widehat{R}_{\alpha}(\sigma_{A':B B'}^{(n)}) + \sum_{i=1}^{n-1} [\widehat{R}_{\alpha}(\sigma_{A':B B'}^{(i)}) - \widehat{R}_{\alpha}(\rho_{A' A: B'}^{(i+1)})] - \widehat{R}_{\alpha}(\rho_{A' A: B'}^{(1)}) \tag{93}$$

$$= \sum_{i=1}^n [\widehat{R}_{\alpha}(\sigma_{A':B B'}^{(i)}) - \widehat{R}_{\alpha}(\rho_{A' A: B'}^{(i)})] \tag{94}$$

$$\leq n \widehat{R}_{\alpha, \Theta}(\mathcal{N}). \tag{95}$$

The first and third lines follow from the monotonicity of the geometric Rényi Rains bound  $\widehat{R}_{\alpha}$  with respect to the PPT operations [11, Eq. (22)]. The second line follows since  $\rho_{A' A: B'}^{(1)}$  is a PPT state and thus  $\widehat{R}_{\alpha}(\rho_{A' A: B'}^{(1)}) = 0$ . The last line follows from Proposition 16.

Note that any communication protocol is characterized by a triplet  $(n, r, \varepsilon)$  with the number of rounds  $n$ , the communication rate  $r$ , and the error tolerance  $\varepsilon$ . Denote  $k \equiv 2^{nr}$  and we have  $\text{Tr } \Phi_k \omega \geq 1 - \varepsilon$  with  $\Phi_k$  being the  $k$ -dimensional maximally entangled state. Moreover, for any  $\sigma \in \text{PPT}'$ , it holds  $\text{Tr } \Phi_k \sigma \leq 1/k$  [54]. Without loss of generality, we can assume that  $\varepsilon \leq 1 - 2^{-nr}$ . Otherwise, any rate above the capacity

would satisfy the strong converse property since  $1 - \varepsilon < 2^{-nr}$ . Thus for any  $\sigma \in \text{PPT}'$  we have

$$1 - \text{Tr } \Phi_k \omega \leq \varepsilon \leq 1 - 2^{-nr} \leq 1 - \text{Tr } \Phi_k \sigma. \tag{96}$$

Let  $\mathcal{N}(\gamma) = (\text{Tr } \Phi_k \gamma)|0\rangle\langle 0| + (\text{Tr}(1 - \Phi_k)\gamma)|1\rangle\langle 1|$ . Due to the data-processing inequality, we have

$$\widehat{D}_\alpha(\omega\|\sigma) \geq \widehat{D}_\alpha(\mathcal{N}(\omega)\|\mathcal{N}(\sigma)) = \delta_\alpha(1 - \text{Tr } \Phi_k \omega\|1 - \text{Tr } \Phi_k \sigma) \geq \delta_\alpha(\varepsilon\|1 - 2^{-nr}), \tag{97}$$

where  $\delta_\alpha(p\|q) \equiv \frac{1}{\alpha-1} \log [p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}]$  is the binary classical Rényi divergence. The last inequality in (97) follows from the monotonicity property that  $\delta_\alpha(p'\|q) \leq \delta_\alpha(p\|q)$  if  $p \leq p' \leq q$  and  $\delta_\alpha(p\|q') \leq \delta_\alpha(p\|q)$  if  $p \leq q' \leq q$  [62]. Then we have

$$\widehat{R}_\alpha(\omega) = \min_{\sigma \in \text{PPT}'} \widehat{D}_\alpha(\omega\|\sigma) \tag{98}$$

$$\geq \delta_\alpha(\varepsilon\|1 - 2^{-nr}) \tag{99}$$

$$\geq \frac{1}{\alpha - 1} \log(1 - \varepsilon)^\alpha (2^{-nr})^{1-\alpha} \tag{100}$$

$$= \frac{\alpha}{\alpha - 1} \log(1 - \varepsilon) + nr. \tag{101}$$

Combining Eqs. (91) and (98), we have

$$\frac{\alpha}{\alpha - 1} \log(1 - \varepsilon) + nr \leq n\widehat{R}_{\alpha, \Theta}(\mathcal{N}), \tag{102}$$

which is equivalent to

$$1 - \varepsilon \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)[r - \widehat{R}_{\alpha, \Theta}(\mathcal{N})]}. \tag{103}$$

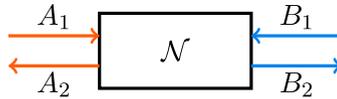
This implies that if the communication rate  $r$  is strictly larger than  $\widehat{R}_{\alpha, \Theta}(\mathcal{N})$ , the fidelity of transmission  $1 - \varepsilon$  decays exponentially fast to zero as the number of channel use  $n$  increases. Or equivalently, we have the strong converse inequality  $Q^{\text{PPT}, \leftrightarrow, \dagger}(\mathcal{N}) \leq \widehat{R}_{\alpha, \Theta}(\mathcal{N})$  and completes the proof.  $\square$

Let  $Q^{\leftrightarrow}$  and  $Q^{\leftrightarrow, \dagger}$  be the two-way assisted quantum capacity and its strong converse capacity respectively. We have the following as a direct consequence of Theorem 17, since PPT assistance is stronger.

**Corollary 18.** *For any quantum channel  $\mathcal{N}$  and  $\alpha \in (1, 2]$ , it holds*

$$Q^{\leftrightarrow}(\mathcal{N}) \leq Q^{\leftrightarrow, \dagger}(\mathcal{N}) \leq \widehat{R}_{\alpha, \Theta}(\mathcal{N}) \leq R_{\max}(\mathcal{N}). \tag{104}$$

Finally, we present how to compute  $\widehat{R}_{\alpha, \Theta}(\mathcal{N})$  as an SDP.



**Fig. 4.** A model of bidirectional quantum channel where  $A_1, A_2$  are held by Alice and  $B_1, B_2$  by Bob

**Proposition 19** (SDP formula). *For any quantum channel  $\mathcal{N}_{A' \rightarrow B}$  and  $\alpha(\ell) = 1 + 2^{-\ell}$  with  $\ell \in \mathbb{N}$ , the geometric Rényi Theta-information can be computed as an SDP:*

$$\begin{aligned} \widehat{R}_{\alpha, \Theta}(\mathcal{N}) &= 2^\ell \cdot \log \min y \quad \text{s.t.} \quad \llbracket M, \{N_i\}_{i=0}^\ell, R, y \rrbracket_{\text{H}}, \\ &\quad \left\{ \left\llbracket \begin{matrix} M & J_{\mathcal{N}} \\ J_{\mathcal{N}} & N_\ell \end{matrix} \right\rrbracket_{\text{P}}, \left\{ \left\llbracket \begin{matrix} J_{\mathcal{N}} & N_i \\ N_i & N_{i-1} \end{matrix} \right\rrbracket_{\text{P}} \right\}_{i=1}^\ell, \left\llbracket R \pm N_0^{\text{T}B} \right\rrbracket_{\text{P}}, \left\llbracket \mathbb{1} - \text{Tr}_B R \right\rrbracket_{\text{P}}, \right. \\ &\quad \left. \left\llbracket y \mathbb{1}_A - \text{Tr}_B M \right\rrbracket_{\text{P}} \right\}, \end{aligned} \tag{105}$$

where  $J_{\mathcal{N}}$  is the Choi matrix of  $\mathcal{N}$ .

*Proof.* This directly follows from Lemma 9 and the definition of the Theta set  $\mathcal{V}_\Theta$  in (71).  $\square$

**4.5. Extension to bidirectional channels.** In this section we showcase that the above results for the PPT/two-way assisted quantum capacity can be extended to a more general scenario where Alice and Bob share a bidirectional quantum channel.

A bipartite quantum channel  $\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}$  is a completely positive trace-preserving map that sends composite system  $A_1 B_1$  to  $A_2 B_2$ . This channel is called *bidirectional channel* if  $A_1 A_2$  are held by Alice and  $B_1 B_2$  are held by Bob. That is, Alice and Bob each input a state to this channel and receive an output [63], as depicted in Fig. 4. This is the most general setting for two-party communications and will reduce to the usual point-to-point channel when the dimensions of Bob’s input and Alice’s output are trivial, i.e.,  $\dim(\mathcal{H}_{B_1}) = \dim(\mathcal{H}_{A_2}) = 1$ .

In [23], the authors introduced the bidirectional version of the max-Rains information as

$$\begin{aligned} R_{\max}^{\text{bi}}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}) &\equiv \log \min \|\text{Tr}_{A_2 B_2}(V_{A_1 B_1 A_2 B_2} + Y_{A_1 B_1 A_2 B_2})\|_\infty \quad \text{s.t.} \quad (106) \\ V_{A_1 B_1 A_2 B_2} &\geq 0, Y_{A_1 B_1 A_2 B_2} \geq 0, (V_{A_1 B_1 A_2 B_2} - Y_{A_1 B_1 A_2 B_2})^{\text{T}B_1 B_2} \geq J_{A_1 B_1 A_2 B_2}^{\mathcal{N}}. \end{aligned}$$

Let  $Q^{\text{bi,PPT},\leftrightarrow}$  and  $Q^{\text{bi,PPT},\leftrightarrow,\dagger}$  be the PPT-assisted quantum capacity of a bidirectional channel and its strong converse capacity respectively.<sup>11</sup> It was proved in [23] that

$$Q^{\text{bi,PPT},\leftrightarrow}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}) \leq Q^{\text{bi,PPT},\leftrightarrow,\dagger}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}) \leq R_{\max}^{\text{bi}}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}). \tag{107}$$

Following a similar approach in Sect. 4.4, we can further strengthen this bound by exploiting the geometric Rényi divergence.

<sup>11</sup> We refer to the work [23, Page 2-3] for rigorous definitions of the PPT/two-way assisted quantum capacity of a bidirectional channel and its strong converse rate.

We start with a bidirectional version of the Werner–Holevo bound<sup>12</sup>

$$Q_{\Theta}^{\text{bi}}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}) \equiv \log \|\Theta_{B_2} \circ \mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2} \circ \Theta_{B_1}\|_{\diamond}, \tag{108}$$

and define its zero set  $\mathcal{V}_{\Theta}^{\text{bi}}$  which admits a semidefinite representation as

$$\mathcal{V}_{\Theta}^{\text{bi}} = \left\{ \mathcal{M} \in \text{CP}(A_1 B_1 : A_2 B_2) \mid \exists R_{A_1 B_1 A_2 B_2}, \text{ s.t. } R \pm J_{\mathcal{M}}^{\top_{B_1 B_2}} \geq 0, \text{Tr}_{A_2 B_2} R \leq \mathbb{1}_{A_1 B_1} \right\}. \tag{109}$$

Using the same idea as the point-to-point scenario, we defined the generalized Theta-information of a bidirectional channel  $\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}$  as the ‘‘channel distance’’<sup>13</sup>

$$R_{\Theta}^{\text{bi}}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}) \equiv \min_{\mathcal{M} \in \mathcal{V}_{\Theta}^{\text{bi}}} D(\mathcal{N} \parallel \mathcal{M}), \tag{110}$$

where  $D$  is a generalized divergence and the channel divergence follows from the usual definition

$$D(\mathcal{N} \parallel \mathcal{M}) \equiv \max_{\phi_{A_1 B_1 A_3 B_3}} D(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}(\phi_{A_1 B_1 A_3 B_3}) \parallel \mathcal{M}_{A_1 B_1 \rightarrow A_2 B_2}(\phi_{A_1 B_1 A_3 B_3})) \tag{111}$$

by maximizing over all the pure states  $\phi_{A_1 B_1 A_3 B_3}$ .

Following a similar proof of Proposition 15, we can show that the bidirectional max-Rains information defined in (106) coincides with the bidirectional Theta-information induced by the max-relative entropy. That is,

$$R_{\text{max}}^{\text{bi}}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}) = R_{\text{max}, \Theta}^{\text{bi}}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}). \tag{112}$$

Denote the bidirectional Rains bound as  $\widehat{R}_{\alpha}^{\text{bi}}(\rho) \equiv \min_{\sigma \geq 0, \|\sigma^{\top_{B_1 B_2}}\|_1 \leq 1} \widehat{D}_{\alpha}(\rho \parallel \sigma)$ . A similar proof as Proposition 16 gives us the following amortization inequality.

**Proposition 20** (Amortization). *For any quantum state  $\rho_{A_1 A_3 : B_1 B_3}$ , any bidirectional quantum channel  $\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}$  and  $\alpha \in (1, 2]$ , it holds*

$$\widehat{R}_{\alpha}^{\text{bi}}(\omega_{A_2 A_3 : B_2 B_3}) \leq \widehat{R}_{\alpha}^{\text{bi}}(\rho_{A_1 A_3 : B_1 B_3}) + \widehat{R}_{\alpha, \Theta}^{\text{bi}}(\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}), \tag{113}$$

with the output state  $\omega_{A_2 A_3 : B_2 B_3} = \mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}(\rho_{A_1 A_3 : B_1 B_3})$ .

Using the amortization inequality in Proposition 20 and a standard argument as Theorem 17, we have the analog results of Theorem 17 and Corollary 18 for bidirectional channels as follows:

<sup>12</sup> Note that this quantity was also independently introduced in [64, Eq. (122)] as well as in [65, Eq. (4.38)] when studying the resource theory of bidirectional quantum channels.

<sup>13</sup> Note that the definition of Eq. (110) was independently defined in the recent work [65, Eq. (4.61)] and the relation in Eq. (112) was also independently found in [65, Proposition 4].

**Theorem 21** (Main result 3). *For any bidirectional channel  $\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}$  and  $\alpha \in (1, 2]$ , it holds*

$$Q^{\text{bi,PPT},\leftrightarrow}(\mathcal{N}) \leq Q^{\text{bi,PPT},\leftrightarrow,\dagger}(\mathcal{N}) \leq \widehat{R}_{\alpha,\Theta}^{\text{bi}}(\mathcal{N}) \leq R_{\max}^{\text{bi}}(\mathcal{N}), \tag{114}$$

where  $Q^{\text{bi,PPT},\leftrightarrow}(\mathcal{N})$  and  $Q^{\text{bi,PPT},\leftrightarrow,\dagger}(\mathcal{N})$  denote the PPT-assisted quantum capacity of a bidirectional channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively. As a consequence, it holds

$$Q^{\text{bi},\leftrightarrow}(\mathcal{N}) \leq Q^{\text{bi},\leftrightarrow,\dagger}(\mathcal{N}) \leq \widehat{R}_{\alpha,\Theta}^{\text{bi}}(\mathcal{N}) \leq R_{\max}^{\text{bi}}(\mathcal{N}), \tag{115}$$

where  $Q^{\text{bi},\leftrightarrow}(\mathcal{N})$  and  $Q^{\text{bi},\leftrightarrow,\dagger}(\mathcal{N})$  denote the two-way assisted quantum capacity of a bidirectional channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively.

**Proposition 22** (SDP formula). *For any bidirectional channel  $\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}$  and  $\alpha(\ell) = 1+2^{-\ell}$  with  $\ell \in \mathbb{N}$ , the bidirectional geometric Rényi Theta-information can be computed as an SDP:*

$$\begin{aligned} \widehat{R}_{\alpha,\Theta}^{\text{bi}}(\mathcal{N}) &= 2^\ell \cdot \log \min y \quad \text{s.t.} \quad \llbracket M, \{N_i\}_{i=0}^\ell, R, y \rrbracket_{\text{H}}, \\ &\llbracket \begin{matrix} M & J_{\mathcal{N}} \\ J_{\mathcal{N}} & N_\ell \end{matrix} \rrbracket_{\text{P}}, \left\{ \llbracket \begin{matrix} J_{\mathcal{N}} & N_i \\ N_i & N_{i-1} \end{matrix} \rrbracket_{\text{P}} \right\}_{i=1}^\ell, \llbracket R \pm N_0^{\text{T} B_1 B_2} \rrbracket_{\text{P}}, \llbracket \mathbb{1} - \text{Tr}_{A_2 B_2} R \rrbracket_{\text{P}}, \\ &\llbracket y \mathbb{1} - \text{Tr}_{A_2 B_2} M \rrbracket_{\text{P}}, \end{aligned} \tag{116}$$

where  $J_{\mathcal{N}}$  is the Choi matrix of  $\mathcal{N}$ .

*Proof.* This directly follows from Lemma 9 and the definition of  $\mathcal{V}_{\Theta}^{\text{bi}}$  in (109).  $\square$

**4.6. Examples.** In this section, we investigate several fundamental quantum channels as well as their compositions. We use these toy models to test the performance of our new strong converse bounds, demonstrating the improvement on the previous results. The semidefinite programs are implemented in MATLAB via the CVX package [66,67], by the solver ‘‘Mosek’’ [68] with the best precision.<sup>14</sup>

*Fundamental quantum channels* The quantum depolarizing channel with dimension  $d$  is defined as

$$\mathcal{D}_p(\rho) = (1 - p)\rho + p\mathbb{1}/d, \quad p \in [0, 1]. \tag{117}$$

The quantum erasure channel is defined as

$$\mathcal{E}_p(\rho) = (1 - p)\rho + p|e\rangle\langle e|, \quad p \in [0, 1], \tag{118}$$

where  $|e\rangle$  is an erasure state orthogonal to the input Hilbert space. The quantum dephasing channel is defined as

$$\mathcal{Z}_p(\rho) = (1 - p)\rho + pZ\rho Z, \quad p \in [0, 1], \tag{119}$$

<sup>14</sup> All the data and codes can be found on the GitHub page <https://github.com/fangkunfred>.

where  $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$  is the Pauli- $z$  operator. These three classes of channels are covariant with respect to the whole unitary group. The *generalized amplitude damping* (GAD) channel is defined as

$$\mathcal{A}_{\gamma,N}(\rho) = \sum_{i=1}^4 A_i \rho A_i^\dagger, \quad \gamma, N \in [0, 1] \tag{120}$$

with the Kraus operators

$$A_1 = \sqrt{1-N}(|0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|), \quad A_2 = \sqrt{\gamma(1-N)}|0\rangle\langle 1|, \tag{121}$$

$$A_3 = \sqrt{N}(\sqrt{1-\gamma}|0\rangle\langle 0| + |1\rangle\langle 1|), \quad A_4 = \sqrt{\gamma N}|1\rangle\langle 0|. \tag{122}$$

The GAD channel is one of the realistic sources of noise in superconducting-circuit-based quantum computing [69], which can be viewed as the qubit analogue of the bosonic thermal channel. When  $N = 0$ , it reduces to the conventional *amplitude damping channel* with two Kraus operators  $A_1, A_2$ .

*Comparison for the unassisted quantum capacity* For the unassisted quantum capacity, we compare the qubit depolarizing channel  $\mathcal{D}_p$ , the qubit erasure channel  $\mathcal{E}_p$ , the qubit dephasing channel  $\mathcal{Z}_p$  and the generalized amplitude damping channels  $\mathcal{A}_{p,N}$  with different choices of parameter  $N$ .

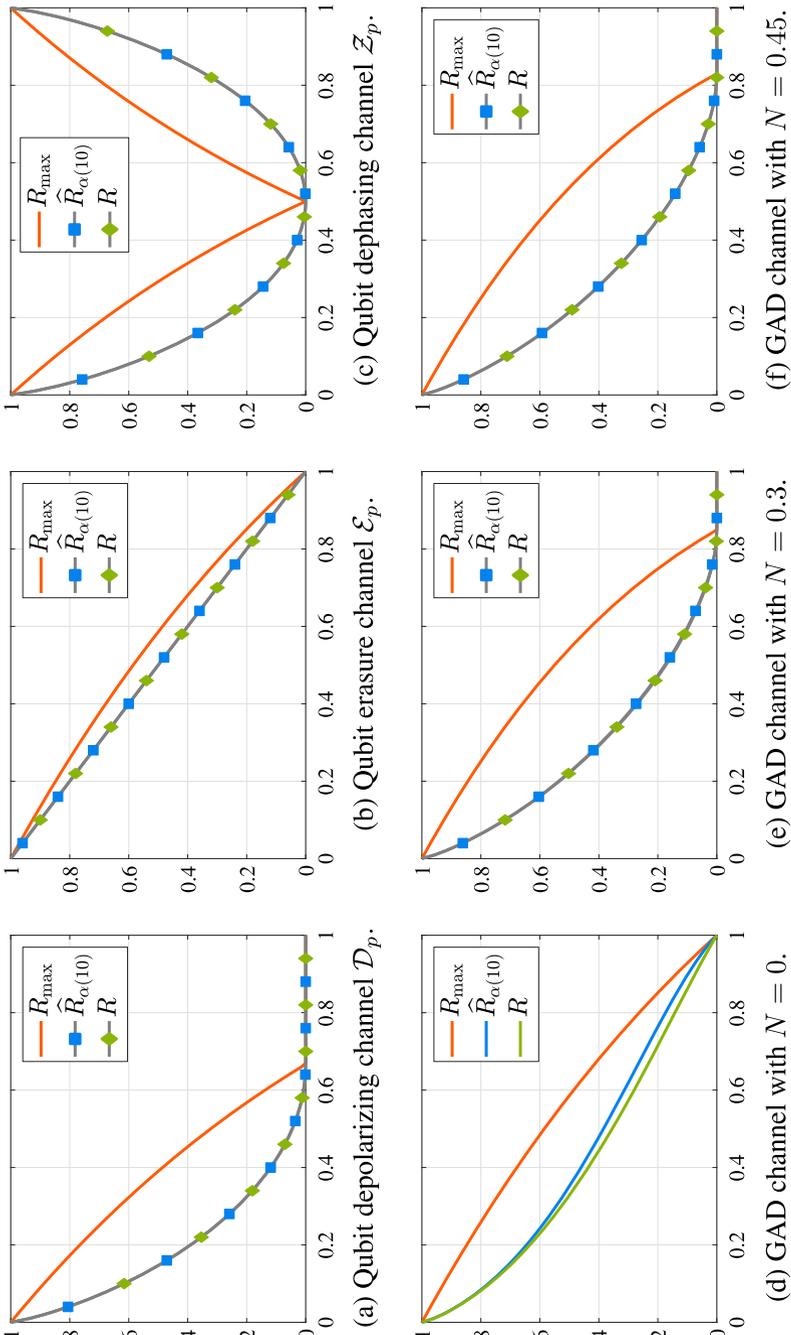
Since  $\mathcal{D}_p, \mathcal{E}_p$  and  $\mathcal{Z}_p$  are covariant with respect to the unitary group, the optimal input state  $\rho_A$  of their Rains information is taken at the maximally mixed state [11, Proposition 2]. Therefore, their Rains information can be computed via the algorithm in [55,56]. Moreover, for any parameters  $\gamma, N \in [0, 1]$ , the GAD channel  $\mathcal{A}_{\gamma,N}$  is covariant with respect to the Pauli- $z$  operator  $Z$ . That is,  $\mathcal{A}_{\gamma,N}(Z\rho Z) = Z\mathcal{A}_{\gamma,N}(\rho)Z$  for all quantum state  $\rho$ . To compute its Rains information, it suffices to perform the maximization over input states with respect to the one-parameter family of states  $\rho_A = (1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|$  [70]. This can be handled, for example, by MATLAB function “fminbnd”.

The comparison results are shown in Fig. 5. It is clear that the geometric Rényi Rains information  $\widehat{R}_{\alpha(10)}$  coincide with the Rains information  $R$  for all these channels except for the particular case  $\mathcal{A}_{p,0}$  in subfigure (d). For all cases,  $\widehat{R}_{\alpha(10)}$  sets a big difference from the max-Rains information  $R_{\max}$ .

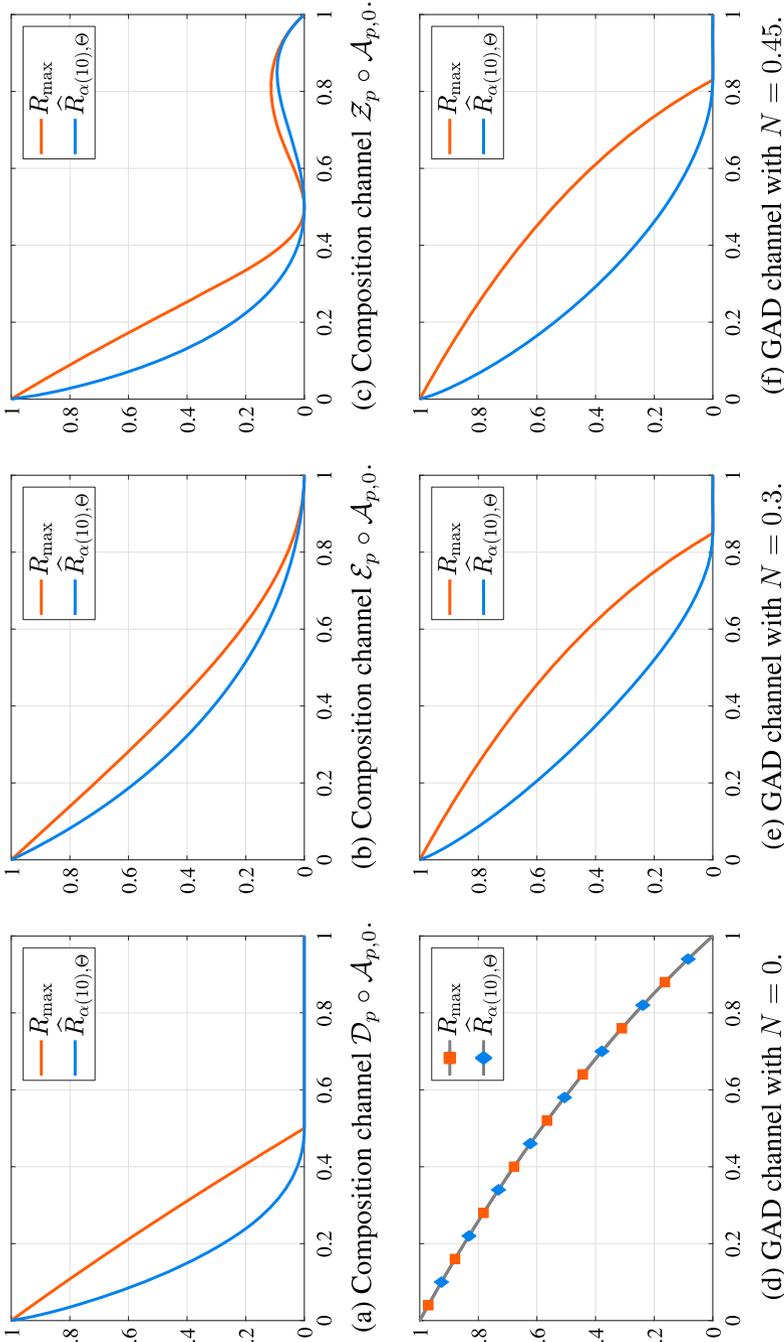
*Comparison for the two-way assisted quantum capacity* For the two-way assisted quantum capacity, we consider the channels  $\mathcal{D}_p, \mathcal{E}_p$  and  $\mathcal{Z}_p$  composed with the amplitude damping channel  $\mathcal{A}_{p,0}$ , and the generalized amplitude damping channel  $\mathcal{A}_{p,N}$  with different choices of parameter  $N$ . Because these channels are not covariant w.r.t. the whole unitary group, their Rains information are not known as valid converse bounds on the two-way assisted quantum capacity.

The comparison result<sup>15</sup> for the two-way assisted quantum capacity is given in Fig. 6. The geometric Rényi Theta-information  $\widehat{R}_{\alpha(10),\Theta}$  demonstrates a significant improvement over the max-Rains information  $R_{\max}$  for all these channels except for one particular case  $\mathcal{A}_{p,0}$  in subfigure (d).

<sup>15</sup> A detailed comparison of the GAD channels with other weak converse bounds in [70] is given in Appendix C.



**Fig. 5.** Comparison of the strong converse bounds on the unassisted quantum capacity of the qubit depolarizing channel  $\mathcal{D}_p$ , the qubit erasure channel  $\mathcal{E}_p$ , the qubit dephasing channel  $\mathcal{Z}_p$  and the generalized amplitude damping channels  $\mathcal{A}_{p,N}$ . The horizontal axis takes value of  $p \in [0, 1]$  and the parameter  $\alpha(10) = 1 + 2^{-10}$



**Fig. 6.** Comparison of the strong converse bounds on the two-way assisted quantum capacity of the channels  $\mathcal{D}_p$ ,  $\mathcal{E}_p$  and  $\mathcal{Z}_p$  composed with the amplitude damping channel  $\mathcal{A}_{p,0}$ , and the generalized amplitude damping channels  $\mathcal{A}_{p,N}$  with different parameters. The horizontal axis takes value of  $p \in [0, 1]$  and the parameter  $\alpha(10) = 1 + 2^{-10}$

*Comparison for the two-way assisted quantum capacity of bidirectional channels* Consider a typical noise in a quantum computer which is modeled as [23]

$$\mathcal{N}_{A_1 B_1 \rightarrow A_2 B_2}(\rho) = p S \rho S^\dagger + (1 - p) U_\phi S \rho S U_\phi^\dagger, \quad p \in [0, 1] \tag{123}$$

where  $S$  is the swap operator and  $U_\phi = |00\rangle\langle 00| + e^{i\phi}|01\rangle\langle 01| + e^{i\phi}|10\rangle\langle 10| + e^{2i\phi}|11\rangle\langle 11|$  is the collective dephasing noise. The comparison result of our new bound  $\widehat{R}_{\alpha(10), \Theta}^{\text{bi}}$  with the previous bound  $R_{\text{max}}^{\text{bi}}$  is given in Fig. 7.

### 5. Private Communication

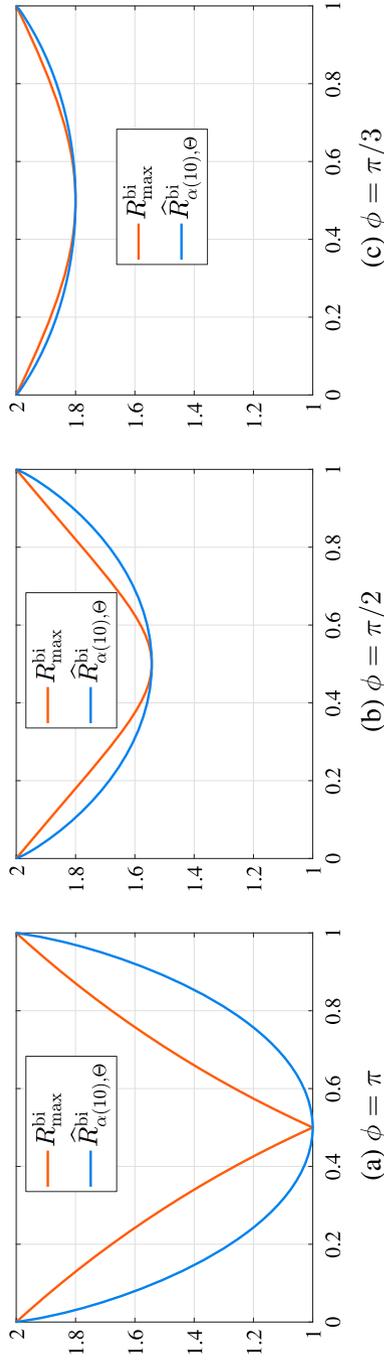
*5.1. Background.* The *private capacity* of a quantum channel is defined as the maximum rate at which classical information can be transmitted privately from the sender (Alice) to the receiver (Bob). By “private”, it means a third party (Eve) who has access to the channel environment cannot learn anything about the information that Alice sends to Bob. There are also two different private capacities of major concern, the (unassisted) private capacity  $P$  and the two-way assisted private capacity  $P^{\leftrightarrow}$ , depending on whether classical communication is allowed between each channel uses.

In the same spirit of the quantum capacity theorem, the private capacity theorem states that the private capacity of a quantum channel is given by its regularized private information [7, 71],

$$P(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} I_p(\mathcal{N}^{\otimes n}) = \sup_{n \in \mathbb{N}} \frac{1}{n} I_p(\mathcal{N}^{\otimes n}), \tag{124}$$

where  $I_p(\mathcal{N}) \equiv \max_{\mathcal{E}} [\chi(\mathcal{E}, \mathcal{N}) - \chi(\mathcal{E}, \mathcal{N}^c)]$  is the private information with the maximization taken over all quantum state ensembles  $\mathcal{E} = \{p_i, \rho_i\}$ ,  $\chi(\mathcal{E}, \mathcal{N}) \equiv H(\sum_i p_i \mathcal{N}(\rho_i)) - \sum_i p_i H(\mathcal{N}(\rho_i))$  is the Holevo information of the ensemble  $\mathcal{E}$ ,  $H$  is the von Neumann entropy and  $\mathcal{N}^c$  is the complementary channel of  $\mathcal{N}$ . The regularization in (124) is necessary in general since the private information is proved to be non-additive [72] and an unbounded number of channel uses may be required to achieve its private capacity [73].

Despite their importance in understanding the fundamental limits of quantum key distributions [74], much less is known about the converse bounds on private capacities, mostly due to their inherently involved settings. The squashed entanglement of a channel was proposed in [75] and proved to be a converse bound on the two-way assisted private capacity. But it remains unknown to be a strong converse rate and the quantity itself is difficult to compute exactly [8]. The entanglement cost of a channel was introduced in [76] and shown to be a strong converse bound on the two-way assisted private capacity [25]. But it was not given by a single-letter formula. A closely related quantity to this part is the relative entropy of entanglement of a channel ( $E_R$ ), which was proved as a (weak) converse bound on the two-way assisted private capacity for channels with “covariant symmetry” [24]. This was later strengthened in [12] as a strong converse bound on the unassisted private capacity for general quantum channels and a strong converse bound on the two-way assisted private capacity for channels with “covariant symmetry”. Moreover, the max-relative entropy of entanglement of a channel ( $E_{\text{max}}$ ) was proved as a strong converse bound on the two-way assisted private capacity in general [25].



**Fig. 7.** Comparison of the strong converse bounds on the two-way assisted quantum capacity of the bidirectional channels in (123) with the dephasing parameter choosing from  $\phi \in \{\pi, \pi/2, \pi/3\}$ . The horizontal axis takes value of  $p \in [0, 1]$  and the parameter  $\alpha(10) = 1 + 2^{-10}$

5.2. *Summary of results.* In this part, we extend the techniques used in the previous sections to the task of private communication and aim to improve the max-relative entropy of entanglement of a channel in both assisted and unassisted scenarios. The structure is organized as follows (see also a schematic diagram in Fig. 8).

In Sect. 5.3 we discuss the unassisted private communication. While the relative entropy of entanglement  $E_R$  established the best known strong converse bound in this case, the difficulties of its evaluation are two-fold: the optimization over the set of separable states and the minimax optimization of the Umegaki relative entropy. The first difficulty will be automatically removed for qubit channels since separability can be completely characterized by the positive partial transpose conditions [77]. The second can be handled by relaxing the Umegaki relative entropy to a semidefinite representable one, such as the max-relative entropy. Based on a notion of the generalized relative entropy of entanglement of a channel, we exhibit that the entanglement of a channel induced by the geometric Rényi divergence ( $\widehat{E}_\alpha$ ) lies between  $E_R$  and  $E_{\max}$ . That is, we show that

$$P(\mathcal{N}) \leq P^\dagger(\mathcal{N}) \leq E_R(\mathcal{N}) \leq \widehat{E}_\alpha(\mathcal{N}) \leq E_{\max}(\mathcal{N}),$$

where  $P(\mathcal{N})$  and  $P^\dagger(\mathcal{N})$  denote the unassisted private capacity of channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively. Moreover,  $\widehat{E}_\alpha(\mathcal{N})$  is given by a conic program in general and reduces to a semidefinite program for all qubit channels.

In Sect. 5.4, we study the private communication with two-way classical communication assistance. We introduce the *generalized Sigma-information* which is a new variant of channel information inspired by the channel resource theory (similar to the Upsilon-information in [57]). More precisely, we define the generalized Sigma-information as a “channel distance” to the class of entanglement breaking subchannels. We show that the max-relative entropy of entanglement  $E_{\max}$  coincides with the generalized Sigma-information induced by the max-relative entropy  $E_{\max, \Sigma}$ , i.e.,  $E_{\max} = E_{\max, \Sigma}$ , thus providing a completely new perspective of understanding the former quantity. Moreover, we prove that the generalized Sigma-information induced by the geometric Rényi divergence ( $\widehat{E}_{\alpha, \Sigma}$ ) is a strong converse bound on the two-way assisted private capacity by utilizing an “amortization argument”, improving the previously best-known result of the max-relative entropy of entanglement [25] in general. That is, we show that

$$P^{\leftrightarrow}(\mathcal{N}) \leq P^{\leftrightarrow, \dagger}(\mathcal{N}) \leq \widehat{E}_{\alpha, \Sigma}(\mathcal{N}) \leq E_{\max}(\mathcal{N}),$$

where  $P^{\leftrightarrow}(\mathcal{N})$  and  $P^{\leftrightarrow, \dagger}(\mathcal{N})$  denote the two-way assisted private capacity of channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively. Moreover,  $\widehat{E}_{\alpha, \Sigma}(\mathcal{N})$  is given by a conic program in general and reduces to a semidefinite program for all qubit channels.

5.3. *Unassisted private capacity.* In this section we discuss converse bounds on the unassisted private capacity.<sup>16</sup>

**Definition 23** ([78]). For any generalized divergence  $D$ , the generalized relative entropy of entanglement of a quantum state  $\rho_{AB}$  is defined as

$$E(\rho_{AB}) \equiv \min_{\sigma_{AB} \in \text{SEP}_{\bullet}(A:B)} D(\rho_{AB} \parallel \sigma_{AB}), \tag{125}$$

<sup>16</sup> We refer to the work [12, Section V.A] for rigorous definitions of the unassisted private capacity and its strong converse rate.



These two quantities are known as strong converse bounds for private communication with and without classical communication assistance, respectively. That is,

$$P^\dagger(\mathcal{N}) \leq E_R(\mathcal{N}) \quad [12] \quad \text{and} \quad P^{\leftrightarrow, \dagger}(\mathcal{N}) \leq E_{\max}(\mathcal{N}) \quad [25]. \quad (129)$$

The computability of  $E_R$  is usually restricted to qubit covariant channels where the input state  $\rho_A$  can be taken as the maximally mixed states and the set of separable states coincides with the set of PPT states [77]. The following result relaxes  $E_R$  to its geometric Rényi version  $\widehat{E}_\alpha$ , which is SDP computable for all qubit channels and is tighter than  $E_{\max}$  in general.

**Theorem 25** (Application 3). *For any quantum channel  $\mathcal{N}_{A' \rightarrow B}$  and  $\alpha \in (1, 2]$ , it holds*

$$P(\mathcal{N}) \leq P^\dagger(\mathcal{N}) \leq E_R(\mathcal{N}) \leq \widehat{E}_\alpha(\mathcal{N}) \leq E_{\max}(\mathcal{N}), \quad (130)$$

where  $P(\mathcal{N})$  and  $P^\dagger(\mathcal{N})$  denote the unassisted private capacity and its corresponding strong converse capacity, respectively. Moreover, the bound  $\widehat{E}_\alpha(\mathcal{N})$  with  $\alpha(\ell) = 1 + 2^{-\ell}$  and  $\ell \in \mathbb{N}$  can be given as

$$\widehat{E}_\alpha(\mathcal{N}) = \ell \cdot 2^\ell - (2^\ell + 1) \log(2^\ell + 1) + (2^\ell + 1) \log T_\alpha(\mathcal{N}), \quad (131)$$

with  $T_\alpha(\mathcal{N})$  given by the following conic program

$$T_\alpha(\mathcal{N}) = \max \operatorname{Tr} \left[ \left( K^H - \sum_{i=1}^\ell W_i \right) \cdot J_{\mathcal{N}} \right] \quad \text{s.t.} \quad \llbracket K, \{Z_i\}_{i=0}^\ell \rrbracket_L, \llbracket \{W_i\}_{i=1}^\ell, \rho \rrbracket_H, \\ \left\| \begin{bmatrix} \rho \otimes \mathbb{1} & K \\ K^\dagger & Z_\ell^H \end{bmatrix} \right\|_{\mathbb{P}}, \left\{ \left\| \begin{bmatrix} W_i & Z_i \\ Z_i^\dagger & Z_{i-1}^H \end{bmatrix} \right\|_{\mathbb{P}} \right\}_{i=1}^\ell, \rho \otimes \mathbb{1} - Z_0^H \in \mathcal{BP}(A : B), \left\| \operatorname{Tr} \rho - 1 \right\|_{\mathbb{E}}, \quad (132)$$

where  $J_{\mathcal{N}}$  is the Choi matrix of  $\mathcal{N}$  and  $\mathcal{BP}(A : B)$  is the set of block-positive operators which reduces to a semidefinite cone if  $|A||B| \leq 6$ .

*Proof.* The first inequality in (130) follows by definition. The second inequality in (130) was proved in [12]. The last two inequalities in (130) are direct consequences of Lemma 4. The derivation of the conic program (132) follows the same steps as Proposition 13. The block positive cone  $\mathcal{BP}(A : B)$  is the dual cone of the set of separable operators. When the channel dimension satisfies  $|A||B| \leq 6$ , this cone admits a semidefinite representation as  $\mathcal{BP}(A : B) = \{X + Y^T B \mid X \geq 0, Y \geq 0\}$  [79, Table 2.2]. Thus the conic program (132) reduces to a semidefinite program.  $\square$

**5.4. Two-way assisted private capacity.** In this section we discuss converse bounds on the two-way assisted private capacity.<sup>17</sup>

A quantum channel  $\mathcal{N}_{A' \rightarrow B}$  is called *entanglement breaking* if its output state  $\mathcal{N}_{A' \rightarrow B}(\rho_{AA'})$  is separable for any input  $\rho_{AA'}$  or equivalently if its Choi matrix is separable [80]. Since every entanglement breaking channel can be simulated by a measurement-preparation scheme [80, 81], any two-way assisted private communication protocol via entanglement breaking channel will end up with a separable state. This indicates that

<sup>17</sup> We refer to [12, Section V.A] for rigorous definitions of the two-way assisted private capacity and its strong converse rate.

these channels are useless for private communication. With this in mind, we consider the set of *entanglement breaking subchannels* as

$$\mathcal{V}_\Sigma \equiv \left\{ \mathcal{M} \in \text{CP}(A : B) \mid J_{\mathcal{M}} \in \text{SEP}(A : B), \text{Tr}_B J_{\mathcal{M}} \leq \mathbb{1}_A \right\}, \quad (133)$$

where  $\text{SEP}(A : B)$  denotes the cone of separable operators.

**Definition 26** (*Sigma-info.*). For any generalized divergence  $\mathbf{D}$ , the generalized Sigma-information of a quantum channel  $\mathcal{N}_{A' \rightarrow B}$  is defined as

$$E_\Sigma(\mathcal{N}) \equiv \min_{\mathcal{M} \in \mathcal{V}_\Sigma} \mathbf{D}(\mathcal{N} \parallel \mathcal{M}) = \min_{\mathcal{M} \in \mathcal{V}_\Sigma} \max_{\rho_A \in \mathcal{S}(A)} \mathbf{D}(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) \parallel \mathcal{M}_{A' \rightarrow B}(\phi_{AA'})), \quad (134)$$

where  $\phi_{AA'}$  is a purification of quantum state  $\rho_A$ .

As mentioned in Remark 6, the min and max in the above definition can be swapped.

Analogous to Proposition 15, the following result establishes the relation between the generalized Sigma-information  $E_\Sigma$  and the generalized relative entropy of entanglement of a channel  $\mathbf{E}$ .

**Proposition 27.** For any generalized divergence  $\mathbf{D}$  and any quantum channel  $\mathcal{N}$ , it holds

$$\mathbf{E}(\mathcal{N}) \leq E_\Sigma(\mathcal{N}). \quad (135)$$

Moreover, for the max-relative entropy the equality always holds, i.e.,

$$E_{\max}(\mathcal{N}) = E_{\max, \Sigma}(\mathcal{N}). \quad (136)$$

*Proof.* This first inequality can be proved in a similar manner as Proposition 15 by using the fact that  $\mathcal{M}_{A' \rightarrow B}(\phi_{AA'}) = \sqrt{\rho_A} J_{\mathcal{M}} \sqrt{\rho_A} \in \text{SEP}_\bullet(A : B)$  for any  $\rho_A \in \mathcal{S}(A)$  and  $\mathcal{M} \in \mathcal{V}_\Sigma$ . We now prove the Eq. (136). It has been shown in [22, Lemma 7] that

$$E_{\max}(\mathcal{N}) = \log \min \left\{ \|\text{Tr}_B Y_{AB}\|_\infty \mid J_{\mathcal{N}} \leq Y_{AB}, Y_{AB} \in \text{SEP}(A : B) \right\}. \quad (137)$$

Using the semidefinite representation of infinity norm and replacing  $Y_{AB} = tJ_{\mathcal{M}}$ , we have

$$E_{\max}(\mathcal{N}) = \log \min \left\{ t \mid J_{\mathcal{N}} \leq tJ_{\mathcal{M}}, J_{\mathcal{M}} \in \text{SEP}(A : B), \text{Tr}_B J_{\mathcal{M}} \leq \mathbb{1}_A \right\}. \quad (138)$$

By the definition of  $D_{\max}$  and  $\mathcal{V}_\Sigma$ , we have

$$E_{\max}(\mathcal{N}) = \min_{\mathcal{M} \in \mathcal{V}_\Sigma} D_{\max}(J_{\mathcal{N}} \parallel J_{\mathcal{M}}) = \min_{\mathcal{M} \in \mathcal{V}_\Sigma} D_{\max}(\mathcal{N} \parallel \mathcal{M}) = E_{\max, \Sigma}(\mathcal{N}), \quad (139)$$

where the second equality follows from Eq. (8).  $\square$

*Remark 7.* The idea of considering the set of entanglement breaking channels also appears in [25, Theorem V.2], where an upper bound of  $E_{\max}(\mathcal{N})$  is given as  $E_{\max}(\mathcal{N}) \leq B_{\max}(\mathcal{N})$  with

$$B_{\max}(\mathcal{N}) \equiv \min \left\{ D_{\max}(J_{\mathcal{N}} \parallel J_{\mathcal{M}}) \mid \mathcal{M} \text{ is an entanglement breaking quantum channel} \right\}. \quad (140)$$

However, the key difference here is that  $E_{\max, \Sigma}$  is minimizing over all the entanglement breaking *subchannels* which is a strictly superset of entanglement breaking channels. Such extension is essential to get the equality  $E_{\max}(\mathcal{N}) = E_{\max, \Sigma}(\mathcal{N})$  instead of an upper bound.

We further consider the Sigma-information induced by the geometric Rényi divergence. Following a similar argument as Proposition 16, we can have the amortization property.

**Proposition 28** (Amortization). *For any quantum state  $\rho_{A'AB'}$  and quantum channel  $\mathcal{N}_{A \rightarrow B}$  and the parameter  $\alpha \in (1, 2]$ , it holds*

$$\widehat{E}_\alpha(\omega_{A':BB'}) \leq \widehat{E}_\alpha(\rho_{A'A:B'}) + \widehat{E}_{\alpha,\Sigma}(\mathcal{N}) \quad \text{with } \omega_{A':BB'} = \mathcal{N}_{A \rightarrow B}(\rho_{A'A:B'}). \quad (141)$$

*Proof.* The proof is similar to that of Proposition 16. We only need to show that for any sub-normalized state  $\sigma_{A'AB'} \in \text{SEP}_\bullet(A'A : B')$  and map  $\mathcal{M}_{A \rightarrow B} \in \mathcal{V}_\Sigma$ , it holds  $\gamma_{A'BB'} \equiv \mathcal{M}_{A \rightarrow B}(\sigma_{A'AB'}) \in \text{SEP}_\bullet(A' : BB')$ . This can be checked as follows. First it is clear that  $\text{Tr } \gamma_{A'BB'} \leq 1$  since both  $\mathcal{M}$  and  $\sigma$  are sub-normalized. Denote the tensor product decomposition  $\sigma_{A'AB'} = \sum_{i,j} \sigma_{A'A}^i \otimes \sigma_{B'}^j$  and  $J_{SB}^M = \sum_{k,\ell} J_S^k \otimes J_B^\ell$ . Let  $|\Phi\rangle_{SA}$  be the unnormalized maximally entangled state. Then we have

$$\gamma_{A'BB'} = \langle \Phi_{SA} | J_{SB}^M \otimes \sigma_{A'AB'} | \Phi_{SA} \rangle = \sum_{i,j,k,\ell} \langle \Phi_{SA} | \sigma_{A'A}^i \otimes J_S^k | \Phi_{SA} \rangle \otimes \sigma_{B'}^j \otimes J_B^\ell, \quad (142)$$

where the r.h.s. belongs to  $\mathcal{SEP}(A' : BB')$ . This completes the proof.  $\square$

Combining the amortization inequality and a similar argument in [25, Theorem IV.1.], we can obtain an improved strong converse bound as follows:

**Theorem 29** (Application 4). *For any quantum channel  $\mathcal{N}_{A' \rightarrow B}$  and  $\alpha \in (1, 2]$ , it holds*

$$P^{\leftrightarrow}(\mathcal{N}) \leq P^{\leftrightarrow,\dagger}(\mathcal{N}) \leq \widehat{E}_{\alpha,\Sigma}(\mathcal{N}) \leq E_{\max}(\mathcal{N}), \quad (143)$$

where  $P^{\leftrightarrow}(\mathcal{N})$  and  $P^{\leftrightarrow,\dagger}(\mathcal{N})$  denote the two-way assisted private capacity of channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively. Moreover, the bound  $\widehat{E}_{\alpha,\Sigma}(\mathcal{N})$  with  $\alpha(\ell) = 1 + 2^{-\ell}$  and  $\ell \in \mathbb{N}$  can be given by a conic program

$$\begin{aligned} \widehat{E}_{\alpha,\Sigma}(\mathcal{N}) = 2^\ell \cdot \log \min y \quad \text{s.t.} \quad & \llbracket M, \{N_i\}_{i=0}^\ell, y \rrbracket_{\text{H}}, N_0 \in \mathcal{SEP}(A : B) \\ & \left\llbracket \begin{matrix} M & J_{\mathcal{N}} \\ J_{\mathcal{N}} & N_\ell \end{matrix} \right\rbracket_{\text{P}}, \left\{ \left\llbracket \begin{matrix} J_{\mathcal{N}} & N_i \\ N_i & N_{i-1} \end{matrix} \right\rbracket_{\text{P}} \right\}_{i=1}^\ell, \left\llbracket \mathbb{1}_A - \text{Tr}_B N_0 \right\rbracket_{\text{P}}, \left\llbracket y\mathbb{1} - \text{Tr}_B M \right\rbracket_{\text{P}}, \end{aligned} \quad (144)$$

which reduces to a semidefinite program if the dimension satisfies  $|A||B| \leq 6$ .

## 6. Classical Communication

*6.1. Background.* The classical capacity of a quantum channel is the maximum rate at which it can reliably transmit classical information over asymptotically many uses of the channel. Since classical messages are of major concern here, the communication assistance is usually given by the shared entanglement instead of the two-way classical communication discussed in the quantum/private communication scenarios. The entanglement-assisted classical capacity has been completely solved as the mutual information of the channel [2], which is believed to be a natural counterpart in the classical Shannon theory. In this sense, shared entanglement simplifies the quantum Shannon theory.

When it comes to the unassisted classical capacity, the best-known characterization is given by the classical capacity theorem, which states that the classical capacity of a quantum channel is given by its regularized Holevo information [82,83],

$$C(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}) = \sup_{n \in \mathbb{N}} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}), \tag{145}$$

where  $\chi(\mathcal{N}) \equiv \max_{\mathcal{E}} \chi(\mathcal{E}, \mathcal{N})$  is the Holevo information with the maximization taken over all quantum state ensembles  $\mathcal{E} = \{p_i, \rho_i\}$ ,  $\chi(\mathcal{E}, \mathcal{N}) \equiv H(\sum_i p_i \mathcal{N}(\rho_i)) - \sum_i p_i H(\mathcal{N}(\rho_i))$  is the Holevo information of the ensemble  $\mathcal{E}$ , and  $H$  is the von Neumann entropy. An impressive work by Hastings [84] shows that the Holevo information is not additive in general, indicating the necessity of the regularization in (145). Moreover, as computing  $\chi$  itself is already NP-complete [85], its regularized quantity for a general quantum channel is expected to be more difficult to evaluate. Even for the qubit amplitude damping channel, its unassisted classical capacity is still unknown [26].

Deriving a single-letter expression for the classical capacity of a quantum channel remains a major open problem in quantum information theory. Several general converse bounds are given in [34] by an ‘‘continuity argument’’, extending the idea in [51] from quantum capacity to classical capacity. However, those bounds are not known to be strong converse and typically work well only if the channel possesses certain structures, such as close to entanglement breaking channel or sufficiently covariant.<sup>18</sup> Two best-known strong converse bounds are given by  $C_\beta$  and  $C_\zeta$  in [26], and both bounds are SDP computable. An attempt to improve the bound  $C_\beta$  is discussed in [57] by a notion called Upsilon-information ( $\Upsilon$ ), similar to the Theta-information and Sigma-information in the previous parts. However, a (weak) sub-additivity of the Upsilon-information induced by the sandwiched R enyi divergence is required for showing  $\Upsilon$  as a strong converse bound for general quantum channels. This sub-additivity was only proved in [57] for covariant channels while the general case was left open.

*6.2. Summary of results.* In this part, we aim to push forward the analysis in [57] by considering the geometric R enyi divergence and improve both of the two strong converse bounds  $C_\beta$  and  $C_\zeta$  in general. The structure of this part is organized as follows (see also a schematic diagram in Fig. 9).

In Sect. 6.3, we first study the generalized Upsilon-information induced by the max-relative entropy ( $\Upsilon_{\max}$ ) and prove that it is no greater than  $C_\beta$  and  $C_\zeta$  in general. We then discuss the generalized Upsilon-information induced by the geometric R enyi divergence ( $\widehat{\Upsilon}_\alpha$ ) and show that it is a strong converse bound on classical capacity by proving its sub-additivity. Due to the relation that  $\widehat{D}_\alpha \leq D_{\max}$ , we have  $\widehat{\Upsilon}_\alpha \leq \Upsilon_{\max}$ . Then we have an improved strong converse bound  $\widehat{\Upsilon}_\alpha$  satisfying

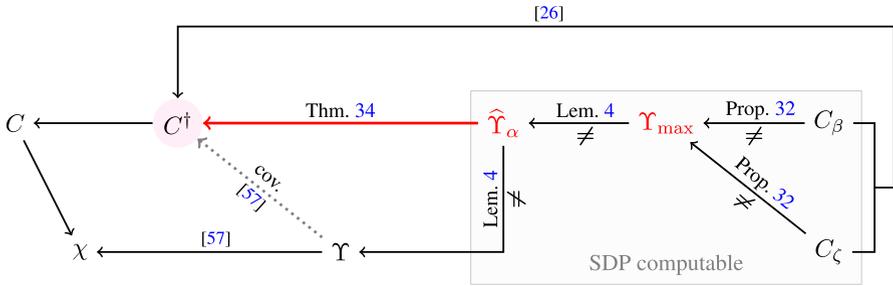
$$C(\mathcal{N}) \leq C^\dagger(\mathcal{N}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}) \leq \min \{C_\beta(\mathcal{N}), C_\zeta(\mathcal{N})\} \quad \text{with } \widehat{\Upsilon}_\alpha(\mathcal{N}) \text{ SDP computable,}$$

where  $C(\mathcal{N})$  and  $C^\dagger(\mathcal{N})$  denote the unassisted classical capacity of channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively.

In Sect. 6.4, we investigate several fundamental quantum channels, demonstrating the efficiency of our new strong converse bounds. It turns out that our new bounds work exceptionally well and exhibit a significant improvement on previous results for almost all cases.

---

<sup>18</sup> These are two main limitations of converse bounds established by using the continuity of the channel capacities.



**Fig. 9.** Relations between different converse bounds for classical communication.  $C$  and  $C^\dagger$  are the classical capacity and the strong converse capacity, respectively.  $\Upsilon$ ,  $\hat{\Upsilon}_\alpha$  and  $\Upsilon_{\max}$  are the generalized Upsilon-information induced by different quantum divergences.  $C_\beta$  and  $C_\zeta$  are the SDP strong converse bounds in [26].  $\chi$  is the Holevo information. The circled quantities is the one of particular interest in quantum information theory. The key quantities and the main contributions in this section are marked in red. The quantity at the start point of an arrow is no smaller than the one at the end point. The inequality sign represents two quantities are not the same in general. The dotted arrow represents that the relation holds under certain restrictions, where “cov.” stands for “covariant”. The parameter  $\alpha$  is taken in the interval  $(1, 2]$ . The quantities in the shaded area are SDP computable in general

**6.3. Unassisted classical capacity.** In this section we discuss converse bounds on the unassisted classical capacity of a quantum channel.<sup>19</sup>

A quantum channel is called *constant channel* or *replacer channel* if it always output a fixed quantum state, i.e., there exists  $\sigma_B \in \mathcal{S}(B)$  such that  $\mathcal{N}_{A \rightarrow B}(\rho_A) = \sigma_B$  for all  $\rho_A \in \mathcal{S}(A)$ . Unlike quantum or private communication where the sets of useless channels are not completely determined yet, the useless set for classical communication is fully characterized by the set of constant channels. That is,  $C(\mathcal{N}) = 0$  if and only if  $\mathcal{N}$  is a constant channel.<sup>20</sup> As a natural extension, the work [57] proposed to consider the set of *constant-bounded subchannels*,

$$\mathcal{V}_{cb} \equiv \{ \mathcal{M} \in \text{CP}(A : B) \mid \exists \sigma_B \in \mathcal{S}(B) \text{ s.t. } \mathcal{M}_{A \rightarrow B}(\rho_A) \leq \sigma_B, \forall \rho_A \in \mathcal{S}(A) \}. \quad (146)$$

It seems not easy to find a semidefinite representation for the set  $\mathcal{V}_{cb}$  directly. Thus a restriction of  $\mathcal{V}_{cb}$  was given in [57] as

$$\mathcal{V}_\beta \equiv \{ \mathcal{M} \in \text{CP}(A : B) \mid \beta(J_{\mathcal{M}}) \leq 1 \} \text{ with } \beta(J_{\mathcal{M}}) \equiv \min \left\{ \text{Tr } S_B \mid R_{AB} \pm J_{\mathcal{M}}^{T_B} \geq 0, \mathbb{1}_A \otimes S_B \pm R_{AB}^{T_B} \geq 0 \right\}. \quad (147)$$

This subset can be seen as the zero set<sup>21</sup> of the strong converse bound  $C_\beta(\mathcal{N}) \equiv \log \beta(J_{\mathcal{N}})$  [26]. As discussed in Appendix B, we will see that  $\mathcal{V}_{cb}$  can be approximated by a complete semidefinite hierarchy, where the subset  $\mathcal{V}_\beta$  can be considered as a symmetrized version of its first level. In the following, we proceed our analysis, without loss of generality, over the set  $\mathcal{V}_\beta$ . A more detailed discussion of  $\mathcal{V}_{cb}$  can be found in Appendix B.

<sup>19</sup> We refer to [26, Section IV.A] for rigorous definitions of the unassisted classical capacity and its strong converse rate.

<sup>20</sup> This can be easily seen from the radius characterization of the Holevo capacity  $\chi(\mathcal{N}) = \min_\sigma \max_\rho D(\mathcal{N}(\rho) \parallel \sigma)$  [86].

<sup>21</sup> It makes no difference by considering  $\beta(J_{\mathcal{M}}) \leq 1$  or  $\beta(J_{\mathcal{M}}) = 1$ .

**Definition 30** (*Upsilon-info.* [57]). For any generalized divergence  $\mathbf{D}$ , the generalized Upsilon-information of a quantum channel  $\mathcal{N}_{A' \rightarrow B}$  with respect to the set  $\mathcal{V}_\beta$  is defined as

$$\Upsilon(\mathcal{N}) \equiv \min_{\mathcal{M} \in \mathcal{V}_\beta} \mathbf{D}(\mathcal{N} \| \mathcal{M}) = \min_{\mathcal{M} \in \mathcal{V}_\beta} \max_{\rho_A \in \mathcal{S}(A)} \mathbf{D}(\mathcal{N}_{A' \rightarrow B}(\phi_{AA'}) \| \mathcal{M}_{A' \rightarrow B}(\phi_{AA'})), \tag{148}$$

where  $\phi_{AA'}$  is a purification of quantum state  $\rho_A$ .

As mentioned in Remark 6, the min and max in the above definition can be swapped.

Let us first consider the generalized Upsilon-information induced by the max-relative entropy  $\Upsilon_{\max}$ .

**Proposition 31.** *For any quantum channel  $\mathcal{N}$ , the generalized Upsilon-information induced by the max-relative entropy  $\Upsilon_{\max}(\mathcal{N})$  is given as an SDP,*

$$\Upsilon_{\max}(\mathcal{N}) = \log \min \left\{ \text{Tr } S_B \mid J_{\mathcal{N}} \leq K_{AB}, R_{AB} \pm K_{AB}^{\text{T}_B} \geq 0, \mathbb{1}_A \otimes S_B \pm R_{AB}^{\text{T}_B} \geq 0 \right\}. \tag{149}$$

*Proof.* By definition we have  $\Upsilon_{\max}(\mathcal{N}) \equiv \min_{\mathcal{M} \in \mathcal{V}_\beta} D_{\max}(\mathcal{N} \| \mathcal{M}) = \min_{\mathcal{M} \in \mathcal{V}_\beta} D_{\max}(J_{\mathcal{N}} \| J_{\mathcal{M}})$ , where the second equality follows from Eq. (8). Then it is clear that

$$\Upsilon_{\max}(\mathcal{N}) = \log \min \left\{ t \mid J_{\mathcal{N}} \leq t J_{\mathcal{M}}, \text{Tr } G \leq 1, W \pm J_{\mathcal{M}}^{\text{T}_B} \geq 0, \mathbb{1} \otimes G \pm W^{\text{T}_B} \geq 0 \right\}. \tag{150}$$

Replacing  $K = t J_{\mathcal{M}}, S = t G$  and  $R = t W$ , we have the desired result.  $\square$

Besides the bound  $C_\beta$ , there is another SDP strong converse bound given in [26] as

$$C_\zeta(\mathcal{N}) \equiv \log \min \left\{ \text{Tr } S_B \mid J_{\mathcal{N}} \leq K_{AB}, \mathbb{1}_A \otimes S_B \pm K_{AB}^{\text{T}_B} \geq 0 \right\}. \tag{151}$$

We can show that  $\Upsilon_{\max}$  is no greater than both of these quantities in general.

**Proposition 32.** *For any quantum channel  $\mathcal{N}$ , it holds  $\Upsilon_{\max}(\mathcal{N}) \leq \min \{C_\beta(\mathcal{N}), C_\zeta(\mathcal{N})\}$ .*

*Proof.* The result is clear by comparing their SDP formulas. Specifically, by restricting  $K_{AB} = J_{\mathcal{N}}$  in  $\Upsilon_{\max}$ , we can retrieve  $C_\beta$ . By restricting  $R_{AB} = \mathbb{1}_A \otimes S_B$  in  $\Upsilon_{\max}$ , we can retrieve  $C_\zeta$ .  $\square$

We further consider the generalized Upsilon-information induced by the geometric Rényi divergence  $\widehat{\Upsilon}_\alpha$ . The following sub-additivity is a key ingredient to proving that  $\widehat{\Upsilon}_\alpha$  is a strong converse bound in Theorem 34.

**Proposition 33** (Sub-additivity). *For any quantum channels  $\mathcal{N}_1, \mathcal{N}_2$  and  $\alpha \in (1, 2]$ , it holds*

$$\widehat{\Upsilon}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_1) + \widehat{\Upsilon}_\alpha(\mathcal{N}_2). \tag{152}$$

*Proof.* This is a direct consequence of the additivity of the geometric Rényi channel divergence in Lemma 6 and the sub-additivity of the quantity  $\beta(\cdot)$  in (147). More specifically, suppose the optimal solution of  $\widehat{\Upsilon}_\alpha(\mathcal{N}_1)$  and  $\widehat{\Upsilon}_\alpha(\mathcal{N}_2)$  are taken at  $\mathcal{M}_1 \in \mathcal{V}_\beta^1$  and  $\mathcal{M}_2 \in \mathcal{V}_\beta^2$  respectively. Then we can check that  $\mathcal{M}_1 \otimes \mathcal{M}_2 \in \mathcal{V}_\beta^{12}$  which is a feasible solution for  $\widehat{\Upsilon}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2)$ . Thus we have

$$\begin{aligned} \widehat{\Upsilon}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) &\leq \widehat{D}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2 \| \mathcal{M}_1 \otimes \mathcal{M}_2) = \widehat{D}_\alpha(\mathcal{N}_1 \| \mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2 \| \mathcal{M}_2) \\ &= \widehat{\Upsilon}_\alpha(\mathcal{N}_1) + \widehat{\Upsilon}_\alpha(\mathcal{N}_2), \end{aligned}$$

where the last inequality follows by the optimality assumption of  $\mathcal{M}_1$  and  $\mathcal{M}_2$ .  $\square$

Based on the sub-additivity, we are ready to show our improved strong converse bound.

**Theorem 34** (Application 5). *For any quantum channel  $\mathcal{N}$  and  $\alpha \in (1, 2]$ , it holds*

$$C(\mathcal{N}) \leq C^\dagger(\mathcal{N}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}) \leq \Upsilon_{\max}(\mathcal{N}) \leq \min \{C_\beta(\mathcal{N}), C_\zeta(\mathcal{N})\}, \quad (153)$$

where  $C(\mathcal{N})$  and  $C^\dagger(\mathcal{N})$  denote the unassisted classical capacity of channel  $\mathcal{N}$  and its corresponding strong converse capacity, respectively.

*Proof.* The first inequality holds by definition. The third inequality follows from Lemma 4. The last inequality was proved in Proposition 32. It remains to prove the second inequality  $C^\dagger(\mathcal{N}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N})$ . For any classical communication protocol with a triplet  $(r, n, \varepsilon)$ , it holds by a standard argument [57, Proposition 20] that

$$1 - \varepsilon \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)} \left[ r^{-\frac{1}{n}} \widetilde{\Upsilon}_\alpha(\mathcal{N}^{\otimes n}) \right], \quad (154)$$

where  $\widetilde{\Upsilon}_\alpha$  is the Upsilon information induced by the sandwiched Rényi divergence  $\widetilde{D}_\alpha$ . Due to the sub-additivity of  $\widehat{\Upsilon}_\alpha$  in Proposition 33 and the inequality in Lemma 4, we have

$$n \widehat{\Upsilon}_\alpha(\mathcal{N}) \geq \widehat{\Upsilon}_\alpha(\mathcal{N}^{\otimes n}) \geq \widetilde{\Upsilon}_\alpha(\mathcal{N}^{\otimes n}). \quad (155)$$

Combining (154) and (155), we have

$$1 - \varepsilon \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)} \left[ r^{-\frac{1}{n}} \widetilde{\Upsilon}_\alpha(\mathcal{N}^{\otimes n}) \right] \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)} [r - \widehat{\Upsilon}_\alpha(\mathcal{N})]. \quad (156)$$

This implies that if the communication rate  $r$  is strictly larger than  $\widehat{\Upsilon}_\alpha(\mathcal{N})$ , the success probability of the transmission  $1 - \varepsilon$  decays exponentially fast to zero as the number of channel use  $n$  increases. Or equivalently, we have the strong converse inequality  $C^\dagger(\mathcal{N}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N})$  and completes the proof.  $\square$

Finally, we present how to compute the geometric Rényi Upsilon information.

**Proposition 35.** (SDP formula) *For any quantum channel  $\mathcal{N}_{A' \rightarrow B}$  and  $\alpha(\ell) = 1 + 2^{-\ell}$  with  $\ell \in \mathbb{N}$ , the geometric Rényi Upsilon information can be computed by the following SDP:*

$$\begin{aligned} \widehat{\Upsilon}_\alpha(\mathcal{N}) &= 2^\ell \cdot \log \min y \quad \text{s.t.} \quad \llbracket M, \{N_i\}_{i=0}^\ell, R, S, y \rrbracket_{\text{H}}, \llbracket y \mathbb{1} - \text{Tr}_B M \rrbracket_{\text{P}} \\ &\quad \llbracket \begin{matrix} M & J_{\mathcal{N}} \\ J_{\mathcal{N}} & N_\ell \end{matrix} \rrbracket_{\text{P}}, \left\{ \llbracket \begin{matrix} J_{\mathcal{N}} & N_i \\ N_i & N_{i-1} \end{matrix} \rrbracket_{\text{P}} \right\}_{i=1}^\ell, \llbracket R \pm N_0^{\text{T}B} \rrbracket_{\text{P}}, \llbracket \mathbb{1} \otimes S \pm R^{\text{T}B} \rrbracket_{\text{P}}, \llbracket 1 - \text{Tr} S \rrbracket_{\text{P}}. \end{aligned} \quad (157)$$

*Proof.* This directly follows from Lemma 9 and the definition of the set  $\mathcal{V}_\beta$  in (147).  $\square$

**6.4. Examples.** In this section, we study several fundamental quantum channels as well as their compositions. We use these toy models to test the performance of our new strong converse bounds, demonstrating the improvement on the previously known results.

Consider the depolarizing channel  $\mathcal{D}_p$  defined in (117), the erasure channel  $\mathcal{E}_p$  defined in (118) and the dephrasing channel  $\mathcal{N}_{p,q}(\rho) \equiv (1-q)[(1-p)\rho + pZ\rho Z] + q \text{Tr}(\rho)|e\rangle\langle e|$ , where  $|e\rangle$  is an erasure flag orthogonal to the input Hilbert space. Since these channels are covariant with respect to the unitary group, their Upsilon informations are known as strong converse bounds [57, Proposition 20] and can be computed via the algorithm in [55, 56]. As for the generalized amplitude damping (GAD) channel  $\mathcal{A}_{\gamma,N}$  defined in (120), its Upsilon information is not known as a valid converse bound.

The comparison results<sup>22</sup> are shown in Fig. 10. It is clear that  $\widehat{\Upsilon}_{\alpha(10)}$  demonstrates significant improvements over  $C_\beta$  and  $C_\zeta$  for all these channels except for one particular case  $\mathcal{A}_{p,0}$  in subfigure (d) where all bounds coincide. It is interesting to note that an analytical expression of the bounds  $C_\beta(\mathcal{A}_{\gamma,N}) = C_\zeta(\mathcal{A}_{\gamma,N}) = \log(1 + \sqrt{1 - \gamma})$  is given in [70, Proposition 6], which is independent on the parameter  $N$ . However, this is clearly not the case for our new bound  $\widehat{\Upsilon}_\alpha$ . For covariant channels  $\mathcal{D}_p$ ,  $\mathcal{E}_p$  and  $\mathcal{N}_{p,p^2}$ , the bound  $\widehat{\Upsilon}_{\alpha(10)}$  also coincides with the Upsilon information  $\Upsilon$  in subfigures (a–c). In particular,  $\widehat{\Upsilon}_{\alpha(10)}$  is given by  $1 - p$  in subfigure (b), witnessing again the strong converse property of the qubit erasure channel  $C(\mathcal{E}_p) = C^\dagger(\mathcal{E}_p) = 1 - p$  [87]. Such tightness can also be observed here for the dephrasing channel  $\mathcal{N}_{p,q}$  and it would be easy to show that  $\chi(\mathcal{N}_{p,q}) = C(\mathcal{N}_{p,q}) = C^\dagger(\mathcal{N}_{p,q}) = 1 - q$  which is independent of the dephasing noise parameter  $p$ .

From Fig. (10e),  $\widehat{\Upsilon}_\alpha$  does not give improvement for the amplitude damping channel  $\mathcal{A}_{\gamma,0}$ . However, when considering the composition channel  $\mathcal{M}_p \equiv \mathcal{A}_{p,0} \circ \mathcal{Z}_p$ , which was studied by Aliferis et al. [88] in the context of fault-tolerant quantum computation, the strong converse bound  $\widehat{\Upsilon}_\alpha$  works unexpectedly well, as shown in Fig. 11.

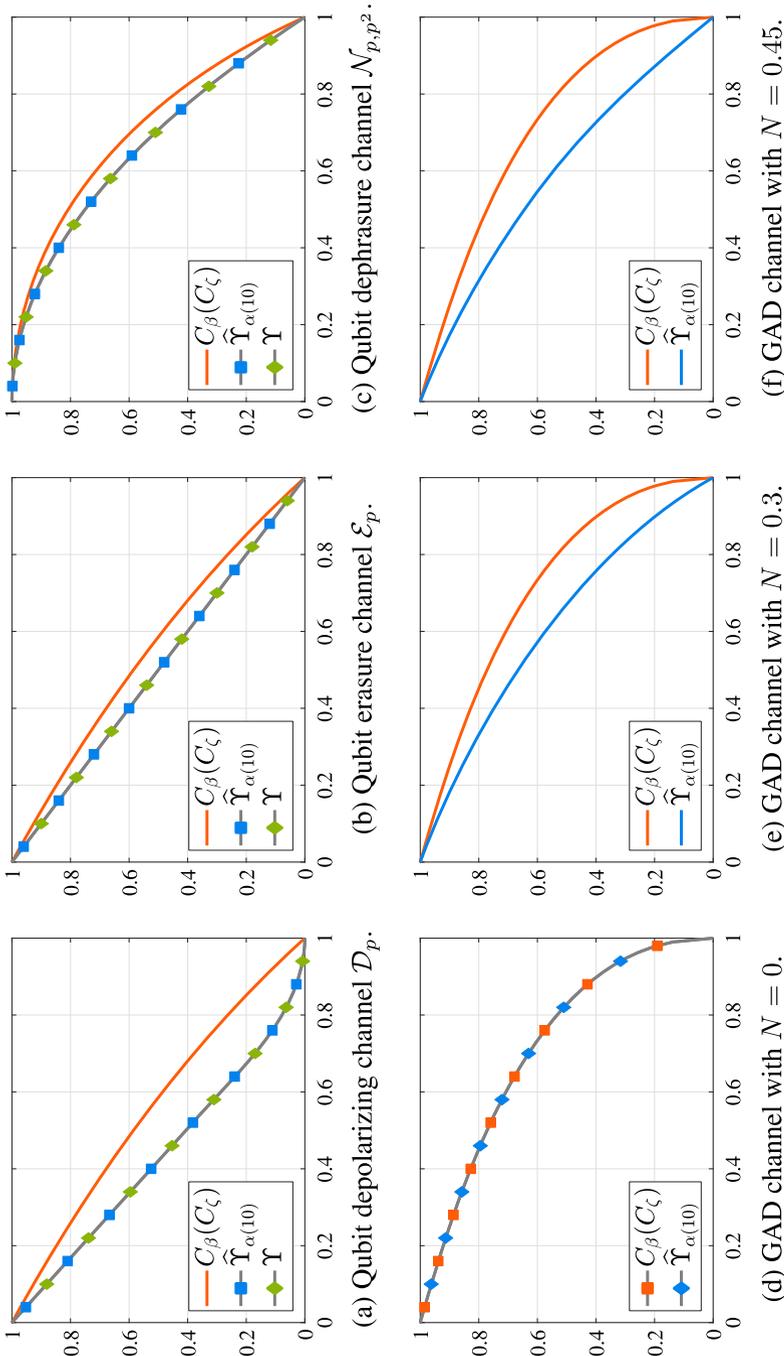
Since  $\mathcal{M}_p$  is an entanglement-breaking channel at  $p = 1/2$ , it is expected that  $\mathcal{M}_p$  is approximately entanglement-breaking around this point. Therefore, a converse bound  $C_{\text{EB}}$  was established in [34, Corollary III.7] by using a continuity argument of the classical capacity. It has been shown in [34, Figure 5] that this continuity bound  $C_{\text{EB}}$  gives certain improvement on  $C_\beta$  for the interval around  $p = 1/2$ . However, Fig. 11 shows that the new strong converse bound  $\widehat{\Upsilon}_{\alpha(10)}$  is much tighter than both  $C_{\text{EB}}$  and  $C_\beta$  for all  $\mathcal{M}_p$  with  $p \in [0, 0.75]$ . The Holevo information  $\chi$  is also numerically computed by utilizing the algorithm<sup>23</sup> in [89]. We observe that the upper bound  $\widehat{\Upsilon}_{\alpha(10)}$  and the lower bound  $\chi$  are very close, leading to a good estimation to the classical capacity of  $\mathcal{M}_p$ .

## 7. Magic State Generation

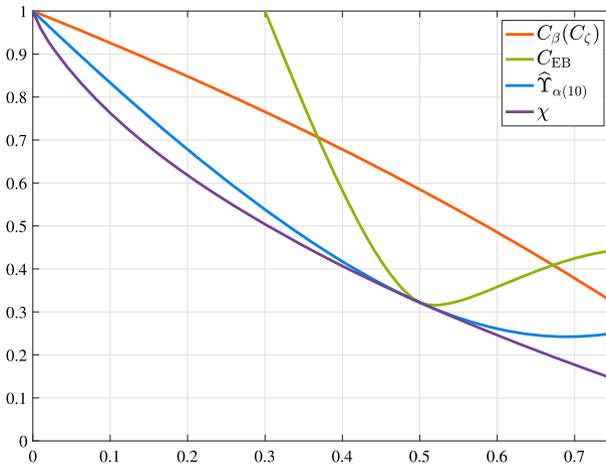
**7.1. Background.** The idea of fault-tolerant quantum computation proposes a reliable framework to implement practical quantum computation against noise and decoherence (e.g. [90–93]). Due to the Gottesman-Knill theorem [94, 95], quantum circuits constructed by stabilizer operations can be efficiently simulated by a classical computer. Therefore, to fully power the universal quantum computation the stabilizer operations

<sup>22</sup> A detailed comparison of the GAD channels with other weak converse bounds in [70] is given in Appendix C.

<sup>23</sup> The MATLAB codes we use are given from [34].



**Fig. 10.** Comparison of the strong converse bounds on the classical capacity for the qubit depolarizing channel  $\mathcal{D}_p$ , the qubit erasure channel  $\mathcal{E}_p$ , the qubit dephasing channel  $\mathcal{N}_{p,q}$  with  $q = p^2$ , and the generalized amplitude damping channels  $\mathcal{A}_{p,N}$  with different parameters. Note that  $C_\beta$  and  $C_\zeta$  coincide for all these examples. The horizontal axis takes value of  $p \in [0, 1]$  and the parameter  $\alpha(10) = 1 + 2^{-10}$



**Fig. 11.** Upper and lower bounds on the classical capacity of the composition channel  $\mathcal{M}_p = \mathcal{A}_{p,0} \circ \mathcal{Z}_p$  with the amplitude damping channel  $\mathcal{A}_{p,0}$  and the dephasing channel  $\mathcal{Z}_p$ . All of  $C_\beta$ ,  $C_\zeta$  and  $\hat{T}_{\alpha(10)}$  are strong converse upper bounds.  $C_\beta$  coincides with  $C_\zeta$  in this case.  $C_{EB}$  is known as a weak converse bound given in [34, Corollary III.7]. The Holevo information  $\chi$  is a lower bound. The horizontal axis takes value of  $p \in [0, 0.75]$  and the parameter  $\alpha(10) = 1 + 2^{-10}$

must be supplemented with some other fault-tolerant non-stabilizer resource. A celebrated scheme for this is given by the state injection technique that allows us to implement non-stabilizer operations by mixing the stabilizer operations with a key ingredient called “magic states” [96,97]. These are non-stabilizer states that must be prepared using the experimentally costly process of magic state distillation (e.g. [98,99]). While extensive efforts have been devoted to construct efficient distillation codes (e.g. [94,99–101]), recent study in [3] as well as [4] initiate the investigation of magic state generation via a general quantum channel, aiming to exploit the power and the limitations of a noisy quantum channel in the scenario of fault-tolerant quantum computation.

Of particular interest is the work [3] which identifies a larger class of operations (*completely positive Wigner preserving* (CPWP) operations), that can be efficiently simulated via classical algorithms. Based on this notion of free operations, the authors established a complete resource theory framework and introduced two efficiently computable magic measures for quantum channels, named *mana* ( $\mathbb{M}$ ) and *max-Thauma* ( $\theta_{\max}$ ) respectively. They proved several desirable properties of these two measures, and further showcased that these channel measures provided strong converse bounds for the task of magic state generation as well as lower bounds for the task of quantum channel synthesis.

**7.2. Summary of results.** In this part, we aim to push forward the study in [3] by considering the Thauma measure induced by the geometric Rényi divergence. Our results can be summarized as follows.

In Sect. 7.4, we prove that the *geometric Rényi Thauma* of a channel ( $\hat{\theta}_\alpha$ ) possesses all the nice properties that are held by the mana and max-Thauma, including the reduction to states, monotonicity under CPWP superchannels, faithfulness, amortization inequality, subadditivity under channel composition, additivity under tensor product as well as a semidefinite representation.

In Sect. 7.5, we exhibit that the geometric Rényi Thaumata of a channel is a strong converse bound on the magic state generation capacity (the maximum number of magic state that can be produced per channel use of  $\mathcal{N}$  via adaptive protocols), improving the max-Thaumata in general. More precisely, we show that

$$C_\psi(\mathcal{N}) \leq C_\psi^\dagger(\mathcal{N}) \leq \frac{\widehat{\theta}_\alpha(\mathcal{N})}{\theta_{\min}(\psi)} \leq \frac{\theta_{\max}(\mathcal{N})}{\theta_{\min}(\psi)}, \quad \text{with } \widehat{\theta}_\alpha(\mathcal{N}) \text{ SDP computable,}$$

where  $C_\psi(\mathcal{N})$  and  $C_\psi^\dagger(\mathcal{N})$  denote the capacity of a channel  $\mathcal{N}$  to generate magic state  $\psi$  and its corresponding strong converse capacity, respectively, and  $\theta_{\min}(\psi)$  is a constant coefficient for given  $\psi$ .

In Sect. 7.6, we show that the geometric Rényi Thaumata can also provide lower bounds for the task of quantum channel synthesis. That is, we prove that the number of uses of a channel  $\mathcal{N}'$  required to implement another channel  $\mathcal{N}$  is bounded from below by  $\widehat{\theta}_\alpha(\mathcal{N})/\widehat{\theta}_\alpha(\mathcal{N}')$  for all  $\alpha \in (1, 2]$ . This new bound complements the previous results by mana and max-Thaumata.

*7.3. Preliminaries of the resource theory of magic.* We first review some basic formalism of the resource theory of magic. Throughout this part, a Hilbert space implicitly has an odd dimension, and if the dimension is not prime, it should be understood to be a tensor product of Hilbert spaces each having odd prime dimension. Let  $\{|j\rangle\}_{j=0}^{d-1}$  be the standard computational basis. For a prime number  $d$ , the generalized Pauli operator (or sometimes called the shift and boost operators)  $X, Z$  are respectively defined as

$$X|j\rangle = |j \oplus 1\rangle, \quad Z|j\rangle = \omega^j|j\rangle, \quad \text{with } \omega = e^{2\pi i/d}, \quad (158)$$

where  $\oplus$  denotes the addition modulo  $d$ . The Heisenberg–Weyl operators are defined as<sup>24</sup>

$$T_{\mathbf{u}} = \tau^{-a_1 a_2} Z^{a_1} X^{a_2}, \quad \text{with } \tau = e^{(d+1)\pi i/d}, \quad \mathbf{u} = (a_1, a_2) \in \mathbb{Z}_d \times \mathbb{Z}_d. \quad (159)$$

For a system with composite Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}_B$ , the Heisenberg–Weyl operators are the tensor product of the Heisenberg–Weyl operators on subsystems  $T_{\mathbf{u}_A \otimes \mathbf{u}_B} = T_{\mathbf{u}_A} \otimes T_{\mathbf{u}_B}$ . For each point  $\mathbf{u} \in \mathbb{Z}_d \times \mathbb{Z}_d$  in the discrete phase space, there is a corresponding operator

$$\mathbb{A}_{\mathbf{u}} \equiv T_{\mathbf{u}} \mathbb{A}_0 T_{\mathbf{u}}^\dagger \quad \text{with } \mathbb{A}_0 \equiv \frac{1}{d} \sum_{\mathbf{u}} T_{\mathbf{u}} \quad (160)$$

The value of the discrete Wigner representation of a quantum state  $\rho$  at  $\mathbb{A}_{\mathbf{u}}$  is given by

$$W_\rho(\mathbf{u}) \equiv \frac{1}{d} \text{Tr}[\mathbb{A}_{\mathbf{u}} \rho]. \quad (161)$$

The Wigner trace and Wigner spectral norm of an Hermitian operator  $V$  are defined as

$$\|V\|_{W,1} \equiv \sum_{\mathbf{u}} |W_V(\mathbf{u})|, \quad \text{and } \|V\|_{W,\infty} \equiv d \max_{\mathbf{u}} |W_V(\mathbf{u})|, \quad (162)$$

<sup>24</sup> The definition here is slightly different from some literatures. We adopt the same notion as in [102].

respectively. For any Hermiticity-preserving map  $\mathcal{N}$ , its discrete Wigner function is defined as

$$W_{\mathcal{N}}(\mathbf{v}|\mathbf{u}) \equiv \frac{1}{d_B} \text{Tr}[\mathbb{A}_{\mathbf{v}_B} \mathcal{N}(\mathbb{A}_{\mathbf{u}_A})] = \frac{1}{d_B} \text{Tr}[J_{\mathcal{N}}(\mathbb{A}_{\mathbf{u}_A} \otimes \mathbb{A}_{\mathbf{v}_B})], \quad (163)$$

with  $J_{\mathcal{N}}$  being the Choi matrix of  $\mathcal{N}$ . The set of quantum states with a non-negative Wigner function is denoted as

$$\mathcal{W}_+ \equiv \{\rho \mid \rho \geq 0, \text{Tr} \rho = 1, W_{\rho}(\mathbf{u}) \geq 0, \forall \mathbf{u}\}. \quad (164)$$

A quantum operation  $\mathcal{E}$  is CPWP if the following holds for any system  $R$  with odd dimension [3]

$$\text{id}_R \otimes \mathcal{E}_{A \rightarrow B}(\rho_{RA}) \in \mathcal{W}_+ \quad \forall \rho_{RA} \in \mathcal{W}_+. \quad (165)$$

**Definition 36** (*Mana*). The mana of a quantum state  $\rho$  is defined as [103]

$$\mathbb{M}(\rho) \equiv \log \|\rho\|_{W,1}. \quad (166)$$

The mana of a quantum channel  $\mathcal{N}_{A \rightarrow B}$  is defined as<sup>25</sup> [3]

$$\mathbb{M}(\mathcal{N}_{A \rightarrow B}) \equiv \log \max_{\mathbf{u}_A} \|\mathcal{N}_{A \rightarrow B}(\mathbb{A}_{\mathbf{u}_A})\|_{W,1} = \log \max_{\mathbf{u}_A} \sum_{\mathbf{v}_B} \frac{1}{d_B} |\text{Tr} J_{\mathcal{N}}(\mathbb{A}_{\mathbf{u}_A} \otimes \mathbb{A}_{\mathbf{v}_B})|. \quad (167)$$

**Definition 37** (*Thauma*). Let  $\mathbf{D}$  be a generalized quantum divergence. The generalized Thauma of a quantum state  $\rho$  is defined as [102]

$$\theta(\rho) \equiv \min_{\sigma \in \mathcal{W}} \mathbf{D}(\rho \|\sigma), \quad (168)$$

where  $\mathcal{W} \equiv \{\sigma \mid \mathbb{M}(\sigma) \leq 0, \sigma \geq 0\}$  is the set of sub-normalized states with non-positive mana. The generalized Thauma of a quantum channel  $\mathcal{N}_{A \rightarrow B}$  is defined as [3]

$$\theta(\mathcal{N}) \equiv \min_{\mathcal{E} \in \mathcal{V}_{\mathbb{M}}} \mathbf{D}(\mathcal{N} \|\mathcal{E}), \quad (169)$$

where  $\mathcal{V}_{\mathbb{M}} \equiv \{\mathcal{E} \in \text{CP}(A : B) \mid \mathbb{M}(\mathcal{E}) \leq 0\}$  is the set of subchannels with non-positive mana.

In particular, the max-Thauma of a channel is induced by the max-relative entropy [3]

$$\theta_{\max}(\mathcal{N}) \equiv \min_{\mathcal{E} \in \mathcal{V}_{\mathbb{M}}} D_{\max}(\mathcal{N} \|\mathcal{E}). \quad (170)$$

**Table 2.** Comparison of properties for the geometric Rényi Thaumata, max-Thaumata and mana of quantum channels

Property	Quantifier		
	Geometric Rényi Thaumata ( $\widehat{\theta}_\alpha$ )	Max-Thaumata ( $\theta_{\max}$ )	Mana ( $\mathbb{M}$ )
Reduction to states	✓ Lemma 38	✓	✓
Monotonicity under CPWP	✓ Lemma 38	✓	✓
Faithfulness	✓ Lemma 38	✓	✓
Amortization	✓ Lemma 39	✓	✓
Subadditivity under $\circ$	✓ Lemma 40	✓	✓
SDP computable	✓ Lemma 41	✓	✓
Additivity under $\otimes$	✓ Lemma 42 ( $\alpha = 2$ )	✓	✓

7.4. *Geometric Rényi Thaumata of a channel.* In this section, we investigate the generalized Thaumata induced by the geometric Rényi divergence:

$$\widehat{\theta}_\alpha(\rho) \equiv \min_{\sigma \in \mathcal{W}} \widehat{D}_\alpha(\rho \parallel \sigma) \quad \text{and} \quad \widehat{\theta}_\alpha(\mathcal{N}) \equiv \min_{\mathcal{E} \in \mathcal{V}_M} \widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{E}). \tag{171}$$

The authors in [3] proved that the mana and max-Thaumata of a channel possess several nice properties, as listed in Table 2. Here we aim to show that all the desirable properties are also held by the geometric Rényi Thaumata as well. These basic properties will be utilized in the next two sections for improving the converse bound on magic state generation capacity and the efficiency of quantum channel synthesis.

**Lemma 38.** *The following properties hold for the geometric Rényi Thaumata of a channel when  $\alpha \in (1, 2]$ :*

- (Reduction to states): Let  $\mathcal{N}(\rho) = \text{Tr}[\rho]\sigma$  be a replacer channel with fixed  $\sigma$  for any  $\rho$ . Then

$$\widehat{\theta}_\alpha(\mathcal{N}) = \widehat{\theta}_\alpha(\sigma). \tag{172}$$

- (Monotonicity): Let  $\mathcal{N}$  be a quantum channel and  $\Gamma$  be a CPWP superchannel. Then

$$\widehat{\theta}_\alpha(\Gamma(\mathcal{N})) \leq \widehat{\theta}_\alpha(\mathcal{N}). \tag{173}$$

- (Faithfulness):  $\widehat{\theta}_\alpha(\mathcal{N})$  is nonnegative for any quantum channel  $\mathcal{N}$  and

$$\widehat{\theta}_\alpha(\mathcal{N}) = 0 \quad \text{if and only if} \quad \mathcal{N} \in \text{CPWP}. \tag{174}$$

*Proof.* The first two properties directly follow from the argument for the generalized Thaumata in [3, Proposition 9 and 10]. The third property follows from the argument in [3, Proposition 11] and the fact that the geometric Rényi divergence is continuous and strongly faithful (i.e.,  $\widehat{D}_\alpha(\rho \parallel \sigma) \geq 0$  in general and  $\widehat{D}_\alpha(\rho \parallel \sigma) = 0$  if and only if  $\rho = \sigma$ ).

□

**Lemma 39** (Amortization). *For any quantum state  $\rho_{RA}$ , any quantum channel  $\mathcal{N}_{A \rightarrow B}$  and the parameter  $\alpha \in (1, 2]$ , it holds*

$$\widehat{\theta}_\alpha(\mathcal{N}_{A \rightarrow B}(\rho_{RA})) - \widehat{\theta}_\alpha(\rho_{RA}) \leq \widehat{\theta}_\alpha(\mathcal{N}_{A \rightarrow B}). \tag{175}$$

<sup>25</sup> This can be seen as an analog of Holevo–Werner bound for quantum capacity of a channel or log-negativity of a quantum state.

*Proof.* The proof follows the similar steps as Proposition 16. We only need to show that for any sub-normalized state  $\sigma_{RA} \in \mathcal{W}$  and any subchannel  $\mathcal{E} \in \mathcal{V}_{\mathbb{M}}$ , it holds  $\gamma_{RB} \equiv \mathcal{E}_{A \rightarrow B}(\sigma_{RA}) \in \mathcal{W}$ . This can be checked as follows:

$$\|\gamma_{RB}\|_{W,1} = \sum_{\mathbf{u}_R, \mathbf{v}_B} |W_{\gamma_{RB}}(\mathbf{u}_R, \mathbf{v}_B)| \tag{176}$$

$$= \sum_{\mathbf{u}_R, \mathbf{v}_B} \left| \sum_{\mathbf{y}_A} W_{\mathcal{E}}(\mathbf{v}_B | \mathbf{y}_A) W_{\sigma_{RA}}(\mathbf{y}_A, \mathbf{u}_R) \right| \tag{177}$$

$$\leq \sum_{\mathbf{u}_R, \mathbf{v}_B, \mathbf{y}_A} \left| W_{\mathcal{E}}(\mathbf{v}_B | \mathbf{y}_A) \right| \left| W_{\sigma_{RA}}(\mathbf{y}_A, \mathbf{u}_R) \right| \tag{178}$$

$$= \sum_{\mathbf{u}_R, \mathbf{y}_A} \left[ \sum_{\mathbf{v}_B} \left| W_{\mathcal{E}}(\mathbf{v}_B | \mathbf{y}_A) \right| \right] \left| W_{\sigma_{RA}}(\mathbf{y}_A, \mathbf{u}_R) \right| \tag{179}$$

$$\leq \sum_{\mathbf{u}_R, \mathbf{y}_A} \left| W_{\sigma_{RA}}(\mathbf{y}_A, \mathbf{u}_R) \right| \tag{180}$$

$$\leq 1. \tag{181}$$

The first line is the definition of the Wigner trace norm in (162). The second line is a chain relation in [3, Lemma 1]. The third line follows from the triangle inequality of the absolute value function. The fourth line follows by grouping the components with respect to index  $\mathbf{v}_B$ . The fifth line follows since  $\mathcal{E} \in \mathcal{V}_{\mathbb{M}}$  and thus  $\sum_{\mathbf{v}_B} |W_{\mathcal{E}}(\mathbf{v}_B | \mathbf{y}_A)| \leq \max_{\mathbf{y}_A} \sum_{\mathbf{v}_B} |W_{\mathcal{E}}(\mathbf{v}_B | \mathbf{y}_A)| \leq 1$ . The last line follows since  $\sigma_{RA} \in \mathcal{W}$ . Thus we can conclude that  $\gamma_{RB} \in \mathcal{W}$ . This completes the proof.  $\square$

**Lemma 40** (Sub-additivity). *For any two quantum channels  $\mathcal{N}_1, \mathcal{N}_2$  and  $\alpha \in (1, 2]$ , it holds*

$$\widehat{\theta}_{\alpha}(\mathcal{N}_2 \circ \mathcal{N}_2) \leq \widehat{\theta}_{\alpha}(\mathcal{N}_1) + \widehat{\theta}_{\alpha}(\mathcal{N}_2). \tag{182}$$

*Proof.* Suppose the optimal solution of  $\widehat{\theta}_{\alpha}(\mathcal{N}_1)$  and  $\widehat{\theta}_{\alpha}(\mathcal{N}_2)$  are taken at  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively. By the subadditivity of the mana under composition, we have  $\mathbb{M}(\mathcal{E}_2 \circ \mathcal{E}_1) \leq 0$  (see [3, Proposition 5]). Thus  $\mathcal{E}_2 \circ \mathcal{E}_1$  is a feasible solution for  $\widehat{\theta}_{\alpha}(\mathcal{N}_2 \circ \mathcal{N}_2)$  and we have

$$\widehat{\theta}_{\alpha}(\mathcal{N}_2 \circ \mathcal{N}_2) \leq \widehat{D}_{\alpha}(\mathcal{N}_2 \circ \mathcal{N}_2 \| \mathcal{E}_2 \circ \mathcal{E}_1) \leq \widehat{D}_{\alpha}(\mathcal{N}_1 \| \mathcal{E}_1) + \widehat{D}_{\alpha}(\mathcal{N}_2 \| \mathcal{E}_2) = \widehat{\theta}_{\alpha}(\mathcal{N}_1) + \widehat{\theta}_{\alpha}(\mathcal{N}_2), \tag{183}$$

where the second inequality follows from Lemma 8, the last equality follows from the optimality assumption of  $\mathcal{E}_1$  and  $\mathcal{E}_2$ .  $\square$

**Lemma 41** (SDP formula). *For any quantum channel  $\mathcal{N}$  and  $\alpha(\ell) = 1 + 2^{-\ell}$  with  $\ell \in \mathbb{N}$ , it holds*

$$\begin{aligned} \widehat{\theta}_{\alpha}(\mathcal{N}) = 2^{\ell} \cdot \log \min y \quad \text{s.t.} \quad & \llbracket M, \{N_i\}_{i=0}^{\ell} \rrbracket_{\mathbb{H}}, \llbracket y \mathbb{1} - \text{Tr}_B M \rrbracket_{\mathbb{P}} \\ & \left\{ \left\llbracket \begin{matrix} J_{\mathcal{N}} & N_i \\ J_{\mathcal{N}} & N_{i-1} \end{matrix} \right\rrbracket_{\mathbb{P}}, \left\{ \left\llbracket \begin{matrix} J_{\mathcal{N}} & N_i \\ N_i & N_{i-1} \end{matrix} \right\rrbracket_{\mathbb{P}} \right\}_{i=1}^{\ell}, \left[ \left[ 1 - \frac{1}{d_B} \sum_{\mathbf{v}_B} |\text{Tr} N_0(\mathbb{A}_{\mathbf{u}_A} \otimes \mathbb{A}_{\mathbf{v}_B})| \right]_{\mathbb{P}}, \forall \mathbf{u}_A. \right. \end{aligned} \tag{184}$$

*Proof.* This directly follows from Lemma 9 and the definition of mana in (167). Note that the absolute value conditions can be written as semidefinite conditions by introducing slack variables.  $\square$

**Lemma 42** (Additivity). *The geometric Rényi Thaumata at  $\alpha = 2$  is additive under tensor product. That is, for any two quantum channels  $\mathcal{N}_1, \mathcal{N}_2$ , it holds*

$$\widehat{\theta}_2(\mathcal{N}_1 \otimes \mathcal{N}_2) = \widehat{\theta}_2(\mathcal{N}_1) + \widehat{\theta}_2(\mathcal{N}_2). \tag{185}$$

*Proof.* We first prove the sub-additivity. Suppose the optimal solution of  $\widehat{\theta}_2(\mathcal{N}_1)$  and  $\widehat{\theta}_2(\mathcal{N}_2)$  are taken at  $\mathcal{E}_1$  and  $\mathcal{E}_2$  respectively. Then we have  $\mathbb{M}(\mathcal{E}_1 \otimes \mathcal{E}_2) = \mathcal{M}(\mathcal{E}_1) + \mathcal{M}(\mathcal{E}_2) \leq 0$  since mana is additive under tensor product [3, Proposition 4]. This implies that  $\mathcal{E}_1 \otimes \mathcal{E}_2$  is a feasible solution for  $\widehat{\theta}_2(\mathcal{N}_1 \otimes \mathcal{N}_2)$ . Thus we have

$$\begin{aligned} \widehat{\theta}_2(\mathcal{N}_1 \otimes \mathcal{N}_2) &\leq \widehat{D}_2(\mathcal{N}_1 \otimes \mathcal{N}_2 \| \mathcal{E}_1 \otimes \mathcal{E}_2) = \widehat{D}_2(\mathcal{N}_1 \| \mathcal{E}_1) + \widehat{D}_2(\mathcal{N}_2 \| \mathcal{E}_2) \\ &= \widehat{\theta}_2(\mathcal{N}_1) + \widehat{\theta}_2(\mathcal{N}_2), \end{aligned} \tag{186}$$

where the first equality follows from Lemma 6 and the second equality follows from the optimality assumption of  $\mathcal{E}_1$  and  $\mathcal{E}_2$ .

We now show the super-additivity by utilizing the dual formula of (184). According to the Lagrangian method, we have the dual problem as

$$\begin{aligned} \widehat{\theta}_2(\mathcal{N}) &= \log \max \text{Tr} \left[ J_{\mathcal{N}} \left( K + K^\dagger \right) \right] - \sum_u f_u \quad \text{s.t.} \\ \begin{bmatrix} \rho_A \otimes \mathbb{1}_B & K \\ K^\dagger & Z \end{bmatrix} &\geq 0, \quad |\text{Tr} Z(\mathbb{A}_u \otimes \mathbb{A}_v)/d_A| \leq f_u, \quad \forall u, v, \quad \text{Tr} \rho = 1. \end{aligned} \tag{187}$$

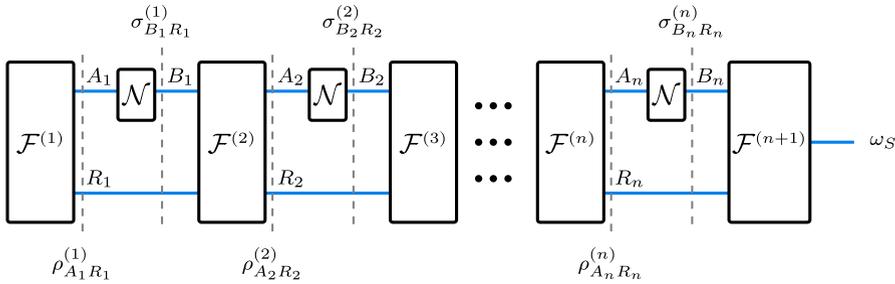
It is easy to check that the strong duality holds. Note that if we replace  $K$  as  $xK$  with  $|x| = 1$ , the optimization is unchanged. Thus we can choose scalar  $x = \text{Tr}(J_{\mathcal{N}}K)^\dagger / |\text{Tr} J_{\mathcal{N}}K|$  to make the term  $\text{Tr} J_{\mathcal{N}}(xK) = |\text{Tr} J_{\mathcal{N}}K|$  to a real scalar. Thus optimization (187) is equivalent to

$$\begin{aligned} \widehat{\theta}_2(\mathcal{N}) &= \log \max 2|\text{Tr} J_{\mathcal{N}}K| - \sum_u f_u \quad \text{s.t.} \\ \begin{bmatrix} \rho_A \otimes \mathbb{1}_B & K \\ K^\dagger & Z \end{bmatrix} &\geq 0, \quad |\text{Tr} Z(\mathbb{A}_u \otimes \mathbb{A}_v)/d_A| \leq f_u, \quad \forall u, v, \quad \text{Tr} \rho = 1. \end{aligned} \tag{188}$$

Again, by replacing  $\widetilde{K} = K/w$ ,  $\widetilde{Z} = Z/w^2$  and  $\widetilde{f}_u = f_u/w^2$ , we have

$$\begin{aligned} \widehat{\theta}_2(\mathcal{N}) &= \log \max 2w|\text{Tr} J_{\mathcal{N}}\widetilde{K}| - w^2 \sum_u \widetilde{f}_u \quad \text{s.t.} \\ \begin{bmatrix} \rho_A \otimes \mathbb{1}_B & \widetilde{K} \\ \widetilde{K}^\dagger & \widetilde{Z} \end{bmatrix} &\geq 0, \quad |\text{Tr} \widetilde{Z}(\mathbb{A}_u \otimes \mathbb{A}_v)/d_A| \leq \widetilde{f}_u, \quad \forall u, v, \quad \text{Tr} \rho = 1. \end{aligned} \tag{189}$$

For any fixed  $|\text{Tr} J_{\mathcal{N}}\widetilde{K}|$  and  $\sum_u \widetilde{f}_u$ , we can quickly check that the optimal solution of the objective function is always taken at  $w = |\text{Tr} J_{\mathcal{N}}\widetilde{K}| / (\sum_u \widetilde{f}_u)$  with the optimal value  $|\text{Tr} J_{\mathcal{N}}\widetilde{K}|^2 / (\sum_u \widetilde{f}_u)$ . Thus the optimization (189) is equivalent to



**Fig. 12.** A schematic diagram for the magic state generation protocol that uses a quantum channel  $n$  times. Every channel use is interleaved by a free CPWP operation  $\mathcal{F}^{(i)}$ . The goal of such a protocol is to produce an approximate magic state  $\omega_S$  in the end

$$\widehat{\theta}_2(\mathcal{N}) = \log \max |\text{Tr } J_{\mathcal{N}} K|^2 / \left( \sum_{\mathbf{u}} f_{\mathbf{u}} \right) \quad \text{s.t.}$$

$$\begin{bmatrix} \rho_A \otimes \mathbb{1}_B & K \\ K^\dagger & Z \end{bmatrix} \geq 0, |\text{Tr } Z(\mathbb{A}_{\mathbf{u}} \otimes \mathbb{A}_{\mathbf{v}}) / d_A| \leq f_{\mathbf{u}}, \forall \mathbf{u}, \mathbf{v}, \text{Tr } \rho = 1. \quad (190)$$

Suppose the optimal solution of  $\widehat{\theta}_2(\mathcal{N}_1)$  and  $\widehat{\theta}_2(\mathcal{N}_2)$  are taken at  $\{K_1, Z_1, f_{\mathbf{u}}^1, \rho_1\}$  and  $\{K_2, Z_2, f_{\mathbf{v}}^2, \rho_2\}$  respectively. We can check that their tensor product  $\{K_1 \otimes K_2, Z_1 \otimes Z_2, f_{\mathbf{u}}^1 f_{\mathbf{v}}^2, \rho_1 \otimes \rho_2\}$  forms a feasible solution for  $\widehat{\theta}_2(\mathcal{N}_1 \otimes \mathcal{N}_2)$  in (190). Thus we have

$$\widehat{\theta}_2(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \log \frac{|\text{Tr}(J_{\mathcal{N}_1} \otimes J_{\mathcal{N}_2})(K_1 \otimes K_2)|^2}{\sum_{\mathbf{u}, \mathbf{v}} f_{\mathbf{u}}^1 f_{\mathbf{v}}^2} = \widehat{\theta}_2(\mathcal{N}_1) + \widehat{\theta}_2(\mathcal{N}_2), \quad (191)$$

which completes the proof.  $\square$

*Remark 8.* Based on numerical observations, we expect that the additivity of the geometric Rényi Thaumata holds for general  $\alpha \in (1, 2]$ . However, the current proof seems to only work for  $\alpha = 2$ .

**7.5. Magic state generation capacity.** In [3], the authors study an information task which uses a quantum channel to produce magic states and quantifies the “magic of channel” by the amount of magic state generated per channel use. Here, we simply dub it as the *magic state generation capacity*<sup>26</sup> as it characterizes the capability of a channel to generate magic states. The most general protocol to produce a magic state can be proceeded as follows (see Fig. 12).

First, we start from preparing a quantum state  $\rho_{R_1 A_1}^{(1)}$  via a free CPWP operation  $\mathcal{F}_{\emptyset \rightarrow R_1 A_1}^{(1)}$ . Then we apply the given channel  $\mathcal{N}$  on system  $A_1$  and obtain a quantum state  $\sigma_{R_1 B_1}^{(1)} = \mathcal{N}_{A_1 \rightarrow B_1}(\rho_{R_1 A_1}^{(1)})$ . After this, we perform another free CPWP operation  $\mathcal{F}_{R_1 B_1 \rightarrow R_2 A_2}^{(2)}$  and then apply the channel  $\mathcal{N}$  again. These processes can be conducted iteratively  $n$  times, and we obtain a quantum state  $\sigma_{R_n B_n}^{(n)}$ . At the end of such a protocol, a final free CPWP operation  $\mathcal{F}_{R_n B_n \rightarrow S}^{(n+1)}$  is performed, producing a quantum state  $\omega_S$ .

<sup>26</sup> It is named as “distillable magic of a channel” in [3]. We here call it magic state generation capacity as it is consistent with the name of entanglement/coherence generation capacity in the existing literature.

For any error tolerance  $\varepsilon \in [0, 1]$ , the above procedure defines an  $(n, k, \varepsilon)$   $\psi$ -magic state generation protocol, if the final state  $\omega$  has a sufficiently high fidelity with  $k$  copies of the target magic state  $\psi$ ,

$$\text{Tr } \omega_S |\psi\rangle\langle\psi|^{\otimes k} \geq 1 - \varepsilon. \tag{192}$$

A rate  $r$  is achievable if for all  $\varepsilon \in (0, 1]$  and  $\delta > 0$  and sufficiently large  $n$ , there exists an  $(n, n(r - \delta), \varepsilon)$   $\psi$ -magic state generation protocol as depicted above. Then the  $\psi$ -magic state generation capacity of the channel  $\mathcal{N}$  is defined as the supremum of all achievable rates and is denoted as  $C_\psi(\mathcal{N})$ . On the other hand,  $r_0$  is called a strong converse rate if for every  $r > r_0$ , the fidelity  $1 - \varepsilon$  of any generation protocol will decay to zero as the number of rounds  $n$  increases. The strong converse capacity, denoted as  $C_\psi^\dagger(\mathcal{N})$  is the infimum of all strong converse rates. By definition, we have  $C_\psi(\mathcal{N}) \leq C_\psi^\dagger(\mathcal{N})$  in general.

Based on the amortization inequality in Lemma 39, a similar argument as [3, Proposition 20] gives us the following improved bound on the magic state generation capacity:

**Theorem 43** (Application 6). *For any quantum channel  $\mathcal{N}$  and  $\alpha \in (1, 2]$ , it holds*

$$C_\psi(\mathcal{N}) \leq C_\psi^\dagger(\mathcal{N}) \leq \frac{\widehat{\theta}_\alpha(\mathcal{N})}{\theta_{\min}(\psi)} \leq \frac{\theta_{\max}(\mathcal{N})}{\theta_{\min}(\psi)}, \tag{193}$$

where  $\theta_{\min}(\psi) = \min_{\sigma \in \mathcal{W}} D_{\min}(\psi \| \sigma)$  is the min-Thauma of the magic state  $\psi$ .

*Proof.* The first inequality holds by definition. The last inequality is a direct consequence of the relation  $\widehat{D}_\alpha(\rho \| \sigma) \leq D_{\max}(\rho \| \sigma)$  proved in Lemma 4. It remains to show the second inequality. The main ingredient to prove this is the amortization property of the geometric Rényi Thauma in Lemma 39. Consider  $n$  round magic state generation protocol as shown in Fig. 12. For each round, denote the input state of the channel  $\mathcal{N}$  as  $\rho_{R_i A_i}^{(i)}$  and the output state as  $\sigma_{R_i B_i}^{(i)}$ . The final state after  $n$  round operations is denoted as  $\omega_S$ . Thus we have

$$\widehat{\theta}_\alpha(\omega_S) \leq \widehat{\theta}_\alpha(\sigma_{R_n B_n}^{(n)}) \tag{194}$$

$$= \widehat{\theta}_\alpha(\sigma_{R_n B_n}^{(n)}) - \widehat{\theta}_\alpha(\rho_{R_1 A_1}^{(1)}) \tag{195}$$

$$\leq \widehat{\theta}_\alpha(\sigma_{R_n B_n}^{(n)}) + \sum_{i=1}^{n-1} \left[ \widehat{\theta}_\alpha(\sigma_{R_i B_i}^{(i)}) - \widehat{\theta}_\alpha(\rho_{R_i A_i}^{(i+1)}) \right] - \widehat{\theta}_\alpha(\rho_{R_1 A_1}^{(1)}) \tag{196}$$

$$= \sum_{i=1}^n \left[ \widehat{\theta}_\alpha(\sigma_{R_i B_i}^{(i)}) - \widehat{\theta}_\alpha(\rho_{R_i A_i}^{(i)}) \right] \tag{197}$$

$$\leq n \widehat{\theta}_\alpha(\mathcal{N}). \tag{198}$$

The first and third lines follow from the monotonicity of the geometric Rényi Thauma of a quantum state under CPWP operations. The second line follows since  $\widehat{\theta}_\alpha(\rho_{R_1 A_1}^{(1)}) = 0$ . The last line follows from Lemma 39.

Note that any  $\psi$ -magic state generation protocol is characterized by a triplet  $(n, k, \varepsilon)$  with the number of rounds  $n$ , the number of target magic states  $k$  and the error tolerance  $\varepsilon$ . Denote  $r = k/n$ . By definition, we have  $\text{Tr} [\omega_S |\psi\rangle\langle\psi|^{\otimes k}] \geq 1 - \varepsilon$ . Moreover, for any  $\sigma_S \in \mathcal{W}$ , it holds  $\text{Tr} [\sigma_S |\psi\rangle\langle\psi|^{\otimes k}] \leq 2^{-nr\theta_{\min}(\psi)}$  [102]. Without loss of generality, we can assume that  $\varepsilon \leq 1 - 2^{-nr\theta_{\min}(\psi)}$ . Otherwise, any value above the capacity would

already be a strong converse rate since  $1 - \varepsilon < 2^{-nr\theta_{\min}(\psi)}$ . Thus for any  $\sigma_S \in \mathcal{W}$ , we have the inequalities

$$1 - \text{Tr} \left[ \omega_S |\psi\rangle\langle\psi|^{\otimes k} \right] \leq \varepsilon \leq 1 - 2^{-nr\theta_{\min}(\psi)} \leq 1 - \text{Tr} \left[ \sigma_S |\psi\rangle\langle\psi|^{\otimes k} \right]. \tag{199}$$

Consider a quantum channel  $\mathcal{N}(\gamma) = [\text{Tr} |\psi\rangle\langle\psi|^{\otimes k} \gamma] |0\rangle\langle 0| + [1 - \text{Tr} |\psi\rangle\langle\psi|^{\otimes k} \gamma] |1\rangle\langle 1|$ . Due to the data-processing inequality, we have

$$\begin{aligned} \widehat{D}_\alpha(\omega \| \sigma) &\geq \widehat{D}_\alpha(\mathcal{N}(\omega) \| \mathcal{N}(\sigma)) \\ &= \delta_\alpha \left( 1 - \text{Tr} |\psi\rangle\langle\psi|^{\otimes k} \omega \left\| 1 - \text{Tr} |\psi\rangle\langle\psi|^{\otimes k} \sigma \right. \right) \geq \delta_\alpha \left( \varepsilon \left\| 1 - 2^{-nr\theta_{\min}(\psi)} \right. \right), \end{aligned} \tag{200}$$

where  $\delta_\alpha(p \| q) \equiv \frac{1}{\alpha-1} \log [p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}]$ . The last inequality follows from the monotonicity property that  $\delta_\alpha(p' \| q) \leq \delta_\alpha(p \| q)$  if  $p \leq p' \leq q$  and  $\delta_\alpha(p \| q') \leq \delta_\alpha(p \| q)$  if  $p \leq q' \leq q$  [62]. Then we have

$$\begin{aligned} \widehat{\theta}_\alpha(\omega_S) &= \min_{\sigma \in \mathcal{W}} \widehat{D}_\alpha(\omega \| \sigma) \geq \delta_\alpha \left( \varepsilon \left\| 1 - 2^{-nr\theta_{\min}(\psi)} \right. \right) \\ &\geq \frac{1}{\alpha-1} \log(1-\varepsilon)^\alpha \left( 2^{-nr\theta_{\min}(\psi)} \right)^{1-\alpha} = \frac{\alpha}{\alpha-1} \log(1-\varepsilon) + nr\theta_{\min}(\psi). \end{aligned} \tag{201}$$

Combining Eqs. (198) and (201), we have

$$\frac{\alpha}{\alpha-1} \log(1-\varepsilon) + nr\theta_{\min}(\psi) \leq n\widehat{\theta}_\alpha(\mathcal{N}), \tag{202}$$

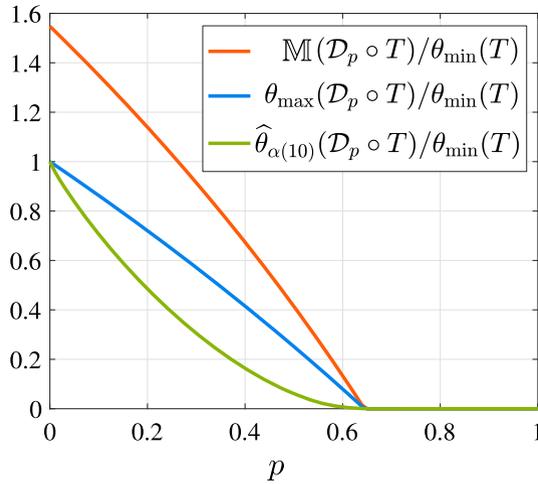
which is equivalent to

$$1 - \varepsilon \leq 2^{-n\theta_{\min}(\psi) \left( \frac{\alpha-1}{\alpha} \right) [r - \widehat{\theta}_\alpha(\mathcal{N}) / \theta_{\min}(\psi)]}. \tag{203}$$

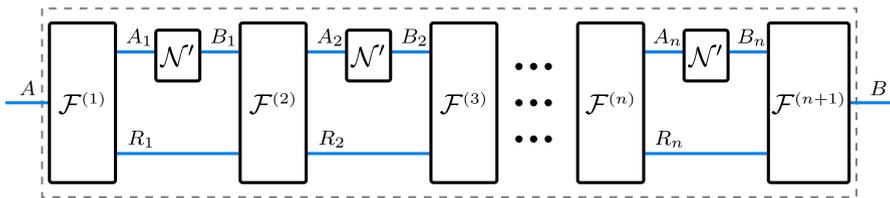
This implies that if the generation rate  $r$  is strictly larger than  $\widehat{\theta}_\alpha(\mathcal{N}) / \theta_{\min}(\psi)$ , the fidelity of the generation protocol  $1 - \varepsilon$  decays exponentially fast to zero as the number of rounds  $n$  increases. Or equivalently, we have  $C_\psi^\dagger(\mathcal{N}) \leq \widehat{\theta}_\alpha(\mathcal{N}) / \theta_{\min}(\psi)$  and completes the proof.  $\square$

*Remark 9.* If the target magic state is  $T$  state  $|T\rangle := (\xi|0\rangle + |1\rangle + \xi^{-1}|2\rangle) / \sqrt{3}$  with  $\xi = e^{2\pi i/9}$  or  $H_+$  state (the eigenstate of a qutrit Hadamard gate with corresponding eigenvalue  $+1$ ), we have  $\theta_{\min}(T) = \log(1 + 2 \sin(\pi/18))$  and  $\theta_{\min}(H_+) = \log(3 - \sqrt{3})$ , respectively [102, Proposition 2].

Consider a qutrit quantum channel  $\mathcal{D}_p \circ T$  composed by a  $T$ -gate with a qutrit depolarizing noise  $\mathcal{D}_p$ . The above Fig. 13 compares different converse bounds on the  $T$ -magic state generation capacity of the channel  $\mathcal{D}_p \circ T$ . It is clear that our new bound based on the geometric Rényi divergence is significantly tighter than the others.



**Fig. 13.** Comparison of the strong converse bounds on the  $T$ -magic state generation capacity of the qutrit quantum channel  $\mathcal{D}_p \circ T$ , where the depolarizing noise parameter  $p \in [0, 1]$  and  $\alpha(10) = 1 + 2^{-10}$



**Fig. 14.** A schematic diagram for the quantum channel synthesis protocol that uses  $n$  times of the channel  $\mathcal{N}'$ . Every channel use is interleaved by a free CPWP operation  $\mathcal{F}^{(i)}$ . The goal of such a protocol is to make the effective channel in the dashed box simulate a target channel  $\mathcal{N}$

**7.6. Quantum channel synthesis.** Another fundamental question in the resource theory of magic asks how many instances of a given quantum channel  $\mathcal{N}'$  are required to simulate another quantum channel  $\mathcal{N}$ , when supplemented with free CPWP operations. Such a general scheme is illustrated in Fig. 14. Denote  $S(\mathcal{N}' \rightarrow \mathcal{N})$  as the smallest number of  $\mathcal{N}'$  channels required to implement the target channel  $\mathcal{N}$  exactly.

The following result establishes new fundamental limits on the quantum channel synthesis problem by employing the geometric Rényi Thaum of the resource and target channels, respectively.

**Theorem 44.** *Let  $\mathcal{N}'$  and  $\mathcal{N}$  be two qudit quantum channels. Then the number of uses of a channel  $\mathcal{N}'$  required to implement  $\mathcal{N}$  is bounded from below as*

$$S(\mathcal{N}' \rightarrow \mathcal{N}) \geq \hat{\theta}_\alpha(\mathcal{N}) / \hat{\theta}_\alpha(\mathcal{N}'), \quad \forall \alpha \in (1, 2]. \tag{204}$$

*Proof.* Suppose the optimal simulation protocol requires to use the resource channel  $n = S(\mathcal{N}' \rightarrow \mathcal{N})$  times and the protocol is given by

$$\mathcal{N} = \mathcal{F}^{(n+1)} \circ \mathcal{N}' \circ \mathcal{F}^{(n)} \circ \dots \circ \mathcal{F}^{(2)} \circ \mathcal{N}' \circ \mathcal{F}^{(1)}, \tag{205}$$

with  $\mathcal{F}^{(i)}$  being CPWP operations. Using the subadditivity of the geometric Rényi Thaumata in Lemma 40 iteratively, we have

$$\widehat{\theta}_\alpha(\mathcal{N}) \leq n\widehat{\theta}_\alpha(\mathcal{N}') + \sum_{i=1}^{n+1} \widehat{\theta}_\alpha(\mathcal{F}^{(i)}) = n\widehat{\theta}_\alpha(\mathcal{N}'), \tag{206}$$

where the equality follows from the faithfulness of the geometric Rényi Thaumata in Lemma 38. Therefore, we have  $S(\mathcal{N}' \rightarrow \mathcal{N}) = n \geq \widehat{\theta}_\alpha(\mathcal{N})/\widehat{\theta}_\alpha(\mathcal{N}')$ , which concludes the proof.  $\square$

Together with the previous result in [3, Proposition 23], we have

**Corollary 45.** *Let  $\mathcal{N}'$  and  $\mathcal{N}$  be two qudit quantum channels. Then the number of uses of a channel  $\mathcal{N}'$  required to implement  $\mathcal{N}$  is bounded from below as*

$$S(\mathcal{N}' \rightarrow \mathcal{N}) \geq \max \left\{ \frac{\mathbb{M}(\mathcal{N})}{\mathbb{M}(\mathcal{N}')} , \frac{\theta_{\max}(\mathcal{N})}{\theta_{\max}(\mathcal{N}')} , \frac{\widehat{\theta}_\alpha(\mathcal{N})}{\widehat{\theta}_\alpha(\mathcal{N}')} \right\}, \quad \forall \alpha \in (1, 2]. \tag{207}$$

*Remark 10.* Note that each lower bound is given by a quotient of two quantities. It is thus not known which one is tighter in general, despite that  $\widehat{\theta}_\alpha(\mathcal{N}) \leq \theta_{\max}(\mathcal{N}) \leq \mathbb{M}(\mathcal{N})$ .

### 8. Conclusions

We have established several fundamental properties of the geometric Rényi divergence as well as its channel divergence. We further demonstrated the usefulness of these properties in the study of quantum channel capacity problems, strengthening the previously best-known results based on the max-relative entropy in general. We expect that the technical tools established in this work can find a diverse range of applications in other research areas, such as quantum network theory and quantum cryptography. For example, we illustrate one more application of the geometric Rényi divergence in the task of quantum channel discrimination in Appendix D.

There are also some interesting problems left for future investigation. The Umegaki relative entropy is the most commonly studied quantum divergence because of its operational interpretation as an optimal error exponent in the hypothesis testing problem (known as the quantum Stein’s lemma) [37, 104]. One open question is to know whether the geometric Rényi divergence as well as the Belavkin–Staszewski relative entropy have any operational interpretation.

*Acknowledgements.* We would like to thank Omar Fawzi for bringing to our attention the open question by Berta et al. [20, Eq. (55)] and for suggesting the chain rule for the geometric Rényi divergence, which helped us simplify and unify the proofs of Proposition 16, 20, 28, 39. We also thank David Sutter for encouraging us to make some proof steps more precise. We thank Sergey N. Filippov, Stefano Pirandola and Mark Wilde for their helpful feedback to this manuscript. KF and HF acknowledge the support of the University of Cambridge Isaac Newton Trust Early Career grant RG74916.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**A. Technical Lemmas**

In this section, we present several technical lemmas that are used in the main text.

**Lemma 46** ([14]). *For any positive semidefinite operators  $X$  and  $Y$  with  $X \ll Y$ , Hermitian operator  $M$  and  $\alpha(\ell) = 1 + 2^{-\ell}$  with  $\ell \in \mathbb{N}$ , the matrix inequality  $G_{1-\alpha}(X, Y) \leq M$  holds if and only if*

$$\exists \llbracket N_0, N_1, \dots, N_\ell \rrbracket_{\mathbb{H}}, \text{ s.t. } \begin{bmatrix} M & X \\ X & N_\ell \end{bmatrix}_{\mathbb{P}}, \left\{ \begin{bmatrix} X & N_i \\ N_i & N_{i-1} \end{bmatrix}_{\mathbb{P}} \right\}_{i=1}^{\ell}, \llbracket N_0 - Y \rrbracket_{\mathbb{E}}. \tag{208}$$

When  $\ell = 0$ , the conditions in the loop are taken as trivial. Here the short notation that  $\llbracket X \rrbracket_{\mathbb{P}}$ ,  $\llbracket X \rrbracket_{\mathbb{E}}$  and  $\llbracket X \rrbracket_{\mathbb{H}}$  represent the positive semidefinite condition  $X \geq 0$ , the equality condition  $X = 0$  and the Hermitian condition  $X = X^\dagger$ , respectively.

The following lemma proves a transformer inequality of the weighted geometric matrix means. Here we require this result to hold for a specific range of the weighting parameter that to the best of our knowledge has not been stated properly before.

**Lemma 47** (Transformer inequality). *Let  $X$  and  $Y$  be two positive operators,  $K$  be any linear operator, and  $\alpha \in (1, 2]$ . Then it holds*

$$G_{1-\alpha}(K X K^\dagger, K Y K^\dagger) \leq K G_{1-\alpha}(X, Y) K^\dagger. \tag{209}$$

Furthermore, if  $K$  is invertible the statement above holds with equality.

*Proof.* Before proving the assertion of the lemma we need to collect some basic properties. We start by recalling the known result [41] that for  $\beta \in [0, 1]$ , we have

$$G_\beta(K X K^\dagger, K Y K^\dagger) \geq K G_\beta(X, Y) K^\dagger. \tag{210}$$

As a next preparatory fact we show that the desired statement is correct for  $\alpha = 2$ , i.e.,

$$G_{-1}(K X K^\dagger, K Y K^\dagger) \leq K G_{-1}(X, Y) K^\dagger. \tag{211}$$

To see this we recall that by Schur’s complement [105, Theorem 1.3.3] we have

$$G_{-1}(X, Y) = X Y^{-1} X \leq M \iff \begin{pmatrix} M & X \\ X & Y \end{pmatrix} \geq 0. \tag{212}$$

Choosing  $M = G_{-1}(X, Y)$  thus gives

$$\begin{pmatrix} G_{-1}(X, Y) & X \\ X & Y \end{pmatrix} \geq 0, \tag{213}$$

which then implies

$$\begin{pmatrix} K G_{-1}(X, Y) K^\dagger & K X K^\dagger \\ K X K^\dagger & K Y K^\dagger \end{pmatrix} \geq 0, \tag{214}$$

because  $Z \mapsto K Z K^\dagger$  is a positive map [105, Exercise 3.2.2]. Using (212) again then implies (211). Because the maps  $t \mapsto t^{-1}$  is operator anti-monotone [106, Table 2.2] we have

$$Y \geq \omega \implies G_{-1}(X, Y) \leq G_{-1}(X, \omega). \tag{215}$$

As a final property we recall a fact from [56, Equation 19] stating that

$$G_s(X, G_t(X, Y)) = G_{st}(X, Y). \tag{216}$$

Now we are ready to prove the assertion of the lemma. For any  $\beta \in [-1, 0)$ , using (216) we have

$$G_\beta(KXK^\dagger, KYK^\dagger) = G_{-1}(KXK^\dagger, G_{-\beta}(KXK^\dagger, KYK^\dagger)) \tag{217}$$

$$\leq G_{-1}(KXK^\dagger, KG_{-\beta}(X, Y)K^\dagger) \tag{218}$$

$$\leq KG_{-1}(X, G_{-\beta}(X, Y))K^\dagger \tag{219}$$

$$= KG_\beta(X, Y)K^\dagger, \tag{220}$$

where the first inequality step follows from (210) and (215). The second inequality is implied by (211). The final step uses (216) again.

The fact that the transformer inequality holds with equality in case  $K$  is invertible follows by applying the inequality twice as

$$G_{1-\alpha}(KXK^\dagger, KYK^\dagger) \leq KG_{1-\alpha}(X, Y)K^\dagger \tag{221}$$

$$= KG_{1-\alpha}(K^{-1}KXK^\dagger(K^\dagger)^{-1}, K^{-1}KYK^\dagger(K^\dagger)^{-1})K^\dagger \tag{222}$$

$$\leq G_{1-\alpha}(KXK^\dagger, KYK^\dagger), \tag{223}$$

which proves that the two inequalities above actually hold with equality.  $\square$

**Corollary 48.** *Let  $X$  and  $Y$  be two positive operators,  $K$  be any linear operator. Let  $D_{op}(X, Y) = X^{\frac{1}{2}} \log(X^{\frac{1}{2}}Y^{-1}X^{\frac{1}{2}})X^{\frac{1}{2}}$  be the operator relative entropy. Then the  $D_{op}$  satisfies the transformer inequality:*

$$D_{op}(KXK^\dagger, KYK^\dagger) \leq KD_{op}(X, Y)K^\dagger. \tag{224}$$

Furthermore, if  $K$  is invertible the statement above holds with equality.

*Proof.* Due to the fact that  $\lim_{\gamma \rightarrow 0} -\frac{1}{\gamma}(x^\gamma - 1) = \log(x)$ , we have the limit identity

$$\lim_{\gamma \rightarrow 0} -\frac{1}{\gamma}(G_\gamma(X, Y) - X) = D_{op}(X, Y). \tag{225}$$

Then we have

$$D_{op}(KXK^\dagger, KYK^\dagger) = \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \left[ G_{1-\alpha}(KXK^\dagger, KYK^\dagger) - KXK^\dagger \right] \tag{226}$$

$$\leq \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \left[ KG_{1-\alpha}(X, Y)K^\dagger - KXK^\dagger \right] \tag{227}$$

$$= K \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \left[ G_{1-\alpha}(X, Y) - X \right] K^\dagger \tag{228}$$

$$= KD_{op}(X, Y)K^\dagger, \tag{229}$$

where the first and last equalities follow from Eq. (225), the inequality follows from Lemma 47.  $\square$

### B. A Hierarchy for Constant-Bounded Subchannels

In this section we discuss the set of constant-bounded subchannels

$$\mathcal{V}_{cb} \equiv \{ \mathcal{M} \in \text{CP}(A : B) \mid \exists \sigma_B \in \mathcal{S}(B) \text{ s.t. } \mathcal{M}_{A \rightarrow B}(\rho_A) \leq \sigma_B, \forall \rho_A \in \mathcal{S}(A) \}. \quad (230)$$

Denote  $\mathcal{N}_\sigma$  as the constant map induced by the state  $\sigma$ . For any  $\mathcal{M} \in \mathcal{V}_{cb}$  the condition  $\mathcal{M}(\rho) \leq \sigma$  for all  $\rho$  is equivalent that  $\mathcal{N}_\sigma - \mathcal{M}$  is a positive map. In terms of their Choi matrices, we have  $\mathbb{1}_A \otimes \sigma_B - J_{\mathcal{M}} \in \mathcal{BP}(A : B)$  where  $\mathcal{BP}(A : B)$  is the cone of block positive operators. Thus we have

$$\mathcal{V}_{cb} = \{ \mathcal{M} \in \text{CP}(A : B) \mid \exists \sigma_B \in \mathcal{S}(B) \text{ s.t. } \mathbb{1}_A \otimes \sigma_B - J_{\mathcal{M}} \in \mathcal{BP}(A : B) \}. \quad (231)$$

Due to the difficulty of finding a semidefinite representation for  $\mathcal{BP}$  [107], we do not expect that there is a semidefinite representation for the set  $\mathcal{V}_{cb}$ . Nevertheless, the cone  $\mathcal{BP}$  can be approximated by a complete hierarchy from the inside as

$$\mathcal{DPS}_1^* \subseteq \mathcal{DPS}_2^* \subseteq \dots \subseteq \mathcal{DPS}_k^* \subseteq \dots \subseteq \mathcal{BP}, \quad (232)$$

where  $\mathcal{DPS}_k^*$  is the dual cone of the well-known DPS hierarchy [108, 109] and is given by the semidefinite representation [110]

$$\mathcal{DPS}_k^* = \left\{ M_{AB_1} \mid M_{AB_1} \otimes \mathbb{1}_{B_{[2:k]}} = (Y_{AB_{[k]}} - \Pi_k Y_{AB_{[k]}} \Pi_k) + \sum_{s=0}^k W_{s, AB_{[k]}}^{\text{T}_{B_{[s]}}} \right. \\ \left. \text{where } Y_{AB_{[k]}} \in \text{Herm}, W_{s, AB_{[k]}} \geq 0, \forall s \in [0 : k] \right\}, \quad (233)$$

where the index  $[s_1 : s_2] \equiv \{s_1, s_1 + 1, \dots, s_2\}$ ,  $[s] \equiv [1 : s]$  and  $\Pi_k$  is the projector on the symmetry subspace of  $\mathcal{H}_B^{\otimes k}$ . Then we can construct a complete semidefinite hierarchy for the set  $\mathcal{V}_{bc}$  as

$$\mathcal{V}_{cb}^1 \subseteq \mathcal{V}_{cb}^2 \subseteq \dots \subseteq \mathcal{V}_{cb}^k \subseteq \dots \subseteq \mathcal{V}_{cb}, \quad (234)$$

with each level given by

$$\mathcal{V}_{cb}^k = \{ \mathcal{M} \in \text{CP}(A : B) \mid \exists \sigma_B \in \mathcal{S}(B) \text{ s.t. } \mathbb{1}_A \otimes \sigma_B - J_{\mathcal{M}} \in \mathcal{DPS}_k^* \}, \quad (235)$$

Consider the first level of the hierarchy

$$\mathcal{V}_{cb}^1 = \left\{ \mathcal{M} \in \text{CP}(A : B) \mid \exists \sigma_B \in \mathcal{S}(B), W_0, W_1 \geq 0 \text{ s.t. } \mathbb{1}_A \otimes \sigma_B - J_{\mathcal{M}} = W_0 + W_1^{\text{T}_B} \right\} \quad (236)$$

Denote  $R = \mathbb{1}_A \otimes \sigma_B - W_0$ , we obtain

$$\mathcal{V}_{cb}^1 = \left\{ \mathcal{M} \in \text{CP}(A : B) \mid \exists \sigma_B \in \mathcal{S}(B), \text{ s.t. } \mathbb{1}_A \otimes \sigma_B - R \geq 0, R^{\text{T}_B} - J_{\mathcal{M}}^{\text{T}_B} \geq 0 \right\} \quad (237)$$

By symmetrizing the conditions  $X - Y \geq 0$  to  $X \pm Y \geq 0$ , we will retrieve the set  $\mathcal{V}_\beta$ . Similarly, by using a different way of variable replacement  $V = \mathbb{1}_A \otimes \sigma - W_1$ , we have

$$\mathcal{V}_{cb}^1 = \left\{ \mathcal{M} \in \text{CP}(A : B) \mid \exists \sigma_B \in \mathcal{S}(B), \text{ s.t. } \mathbb{1}_A \otimes \sigma_B - V^{\text{T}_B} \geq 0, V - J_{\mathcal{M}} \geq 0 \right\}. \quad (238)$$

Then symmetrizing the conditions, we will obtain a set of subchannels  $\mathcal{V}_\zeta$  which is exactly the zero set of the strong converse bound  $C_\zeta$ .

The restriction to symmetric conditions ensures that the set is closed under tensor product, which is a key ingredient to proving the sub-additivity of the geometric Rényi Upsilon-information in Proposition 33. Thus a further improvement of the result in the main text can be made by considering a symmetric restriction of a higher level set  $\mathcal{V}_{cb}^k$ . But we also note that the size of the SDP will exponentially increase in the level of  $k$ .

### C. A Detailed Comparison for Generalized Amplitude Damping Channels

In this section we give a detailed comparison of our new strong converse bounds with previously known results for the generalized amplitude damping (GAD) channels. This class of channels has been systematically investigated in [70], with several converse bounds being established therein based on the data-processing inequality and the continuity of channel capacities as well as a few different techniques. The following plots of previous weak converse bounds are generated via the MATLAB codes provided in the arXiv submission of [70].

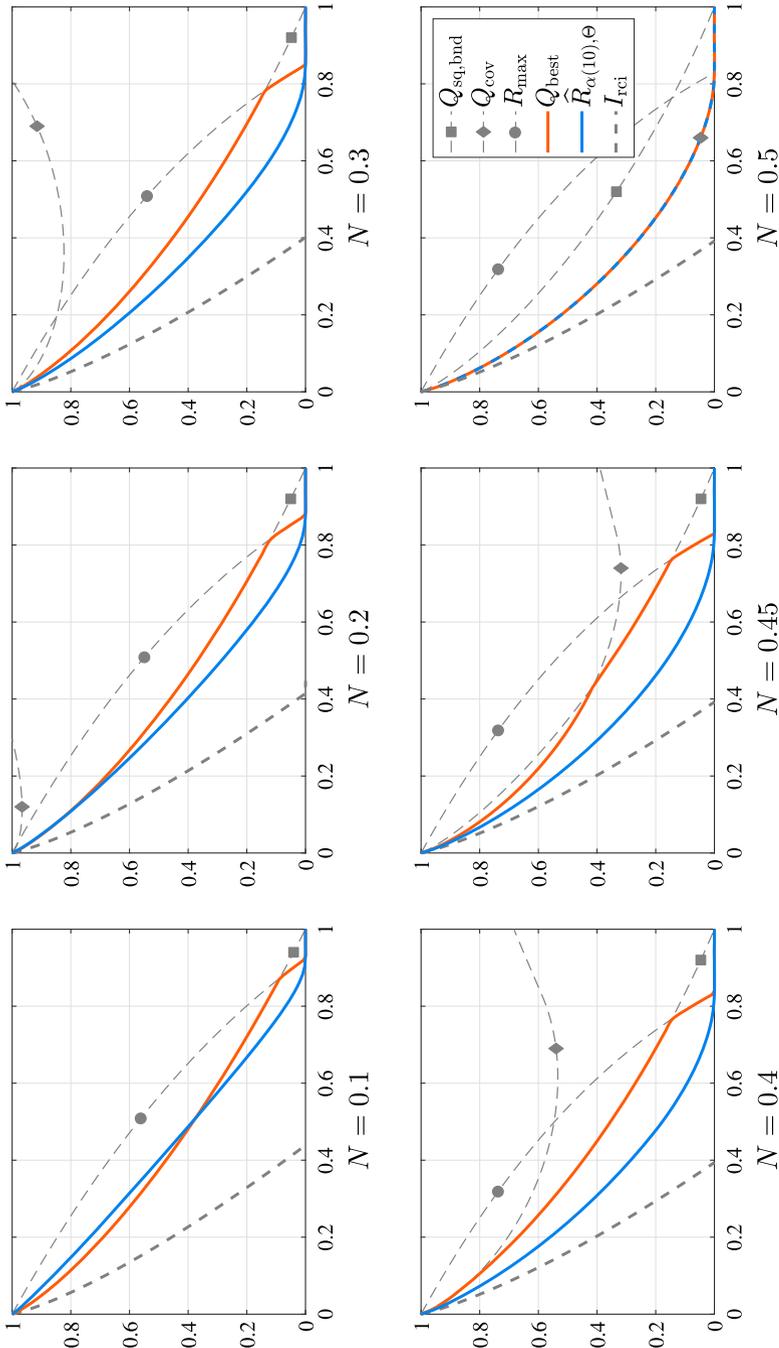
Due to the covariance symmetry of the GAD channels under the Pauli- $z$  operator, the quantities introduced in this work do not provide advantage over the Rains information  $R$  (resp. the relative entropy of entanglement  $E_R$ ) in terms of the unassisted quantum (resp. private) capacity. In the following, we will focus our comparison for the two-way assisted scenarios where both  $R$  and  $E_R$  are not known as valid converse bounds. The comparison result for the two-way assisted quantum capacity is given in Fig. 15. The red solid line is the previously best-known converse bound composed by several different quantities. It is clear that our new strong converse bound  $\widehat{R}_{\alpha(10),\Theta}$  can be much tighter in most cases. Note that for the two-way assisted private capacity, we will obtain exactly the same result as Fig. 15 by replacing  $R_{\max}$  with  $E_{\max}$  and  $\widehat{R}_{\alpha(10),\Theta}$  with  $\widehat{E}_{\alpha(10),\Sigma}$  respectively.

The comparison result for the classical capacity is given in Fig. 16. The red solid line is the previously best-known converse bound composed by several different quantities. It is clear that our new strong converse bound  $\widehat{\Upsilon}_{\alpha(10)}$  can make further improvement at some parameter range, particularly for low to medium amplitude damping noise. In the range of high noise, the GAD channel becomes entanglement-breaking. Thus the  $\varepsilon$ -entanglement breaking upper bound  $C_{EB}$  will be the tightest one, as expected. We do not show the plot for  $N = 0.5$ , because the channel becomes a qubit unital channel and thus its Holevo information is already tight [112]. We do not explicitly depict the weak converse bound in [113] since it is not tight to the cases we present.

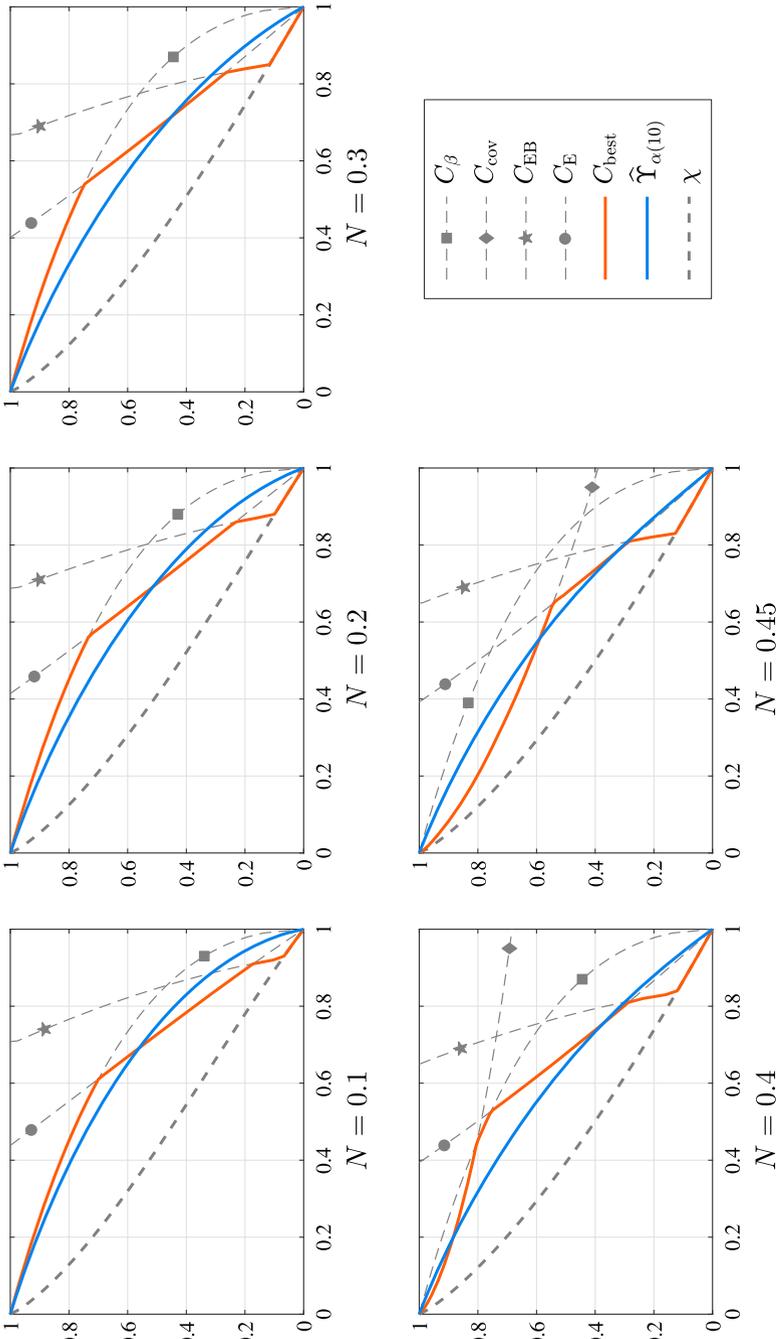
### D. Application in Quantum Channel Discrimination

A fundamental problem in quantum information theory is to distinguish between two quantum channels  $\mathcal{N}$  and  $\mathcal{M}$ . In the asymmetric hypothesis testing setting (Stein's setting), we aim to minimize the type II error probability, under the condition that the type I error probability does not exceed a constant  $\varepsilon \in (0, 1)$ . More precisely, for any given two quantum channels  $\mathcal{N}$  and  $\mathcal{M}$ , denote the corresponding type I and type II error of the adaptive protocol  $\{Q, \mathcal{A}\}$  as  $\alpha_n(\{Q, \mathcal{A}\})$  and  $\beta_n(\{Q, \mathcal{A}\})$ . Then the asymmetric distinguishability is defined as

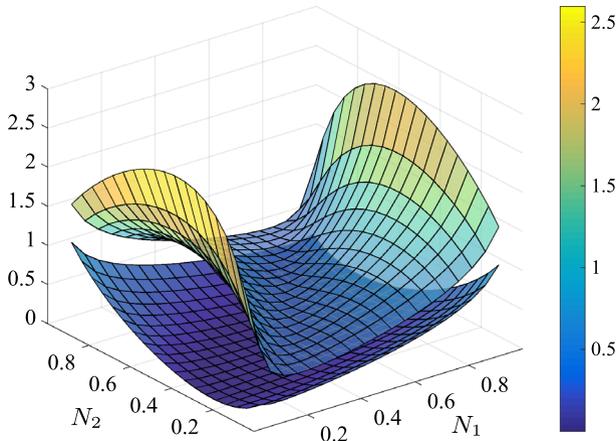
$$\zeta_n(\varepsilon, \mathcal{N}, \mathcal{M}) := \sup_{\{Q, \mathcal{A}\}} \left\{ -\frac{1}{n} \log \beta_n(\{Q, \mathcal{A}\}) \mid \alpha_n(\{Q, \mathcal{A}\}) \leq \varepsilon \right\}. \quad (239)$$



**Fig. 15.** Comparison of the converse bounds on two-way assisted quantum capacities of the GAD channels. The quantity  $Q_{sq,bnd}$  is the squashed entanglement upper bound in [70, Eq. (176)]. The quantity  $Q_{cov}$  is the approximate covariance upper bound in [70, Eq. (205)]. The quantity  $R_{max}$  is the max-Rains information in [21, 61] which is later proved to be a strong converse bound for two-way assisted quantum capacity in [22]. The red solid line  $Q_{best}$  is the previously tightest upper bound composed by  $Q_{sq,bnd}$ ,  $Q_{cov}$  and  $R_{max}$ , i.e.,  $Q_{best} = \min\{Q_{sq,bnd}, Q_{cov}, R_{max}\}$ . The blue solid line  $R_{\alpha(10),\theta}$  is our new strong converse bound in [this work, Eq. (105)] with level  $\ell = 10$ . The quantity  $I_{rci}$  is the reverse coherent information in [70, Eq. (188)] which is a lower bound on the two-way assisted quantum capacity [111]. The parameter  $\alpha(10) = 1 + 2^{-10}$



**Fig. 16.** Comparison of the converse bounds on the classical capacity of the GAD channels. The quantity  $C_\beta$  is the strong converse bound in [26]. The quantity  $C_{\text{cov}}$  is the  $\varepsilon$ -covariance upper bound in [70, Eq. (113)]. The quantity  $C_{\text{EB}}$  is the  $\varepsilon$ -entanglement breaking upper bound in [70, Eq. (112)]. The quantity  $C_E$  is the entanglement-assisted classical capacity in [114]. The red solid line  $C_{\text{best}}$  is the previously tightest upper bound composed by  $C_\beta, C_{\text{cov}}, C_{\text{EB}}$  and  $C_E$ , i.e.  $C_{\text{best}} = \min\{C_\beta, C_{\text{cov}}, C_{\text{EB}}, C_E\}$ . The blue solid line  $\hat{Y}_{\alpha(10)}$  is our new strong converse bound in [this work, Eq. (157)] with level  $\ell = 10$ . The quantity  $\chi$  is the Holevo information given in [115] which is a lower bound on the classical capacity. The parameter  $\alpha(10) = 1 + 2^{-10}$



**Fig. 17.** This figure displays the difference between the upper and lower bounds in the Stein setting for the GAD channels  $\mathcal{A}_{0.8,N_1}$  and  $\mathcal{A}_{0.7,N_2}$ . We vary the parameter  $N_1, N_2 \in [0, 1]$ . The upper surface is the difference  $D_{\max}(\mathcal{A}_{0.8,N_1} \parallel \mathcal{A}_{0.7,N_2}) - D(\mathcal{A}_{0.8,N_1} \parallel \mathcal{A}_{0.7,N_2})$ . The lower surface is the difference  $\widehat{D}_{\alpha(10)}(\mathcal{A}_{0.8,N_1} \parallel \mathcal{A}_{0.7,N_2}) - D(\mathcal{A}_{0.8,N_1} \parallel \mathcal{A}_{0.7,N_2})$

Its asymptotic quantities are defined as

$$\underline{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M}) := \liminf_{n \rightarrow \infty} \zeta_n(\varepsilon, \mathcal{N}, \mathcal{M}), \quad \bar{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M}) := \limsup_{n \rightarrow \infty} \zeta_n(\varepsilon, \mathcal{N}, \mathcal{M}) \quad (240)$$

The best-known single-letter strong converse bound on  $\bar{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M})$  is given by the channel’s max-relative entropy  $D_{\max}(\mathcal{N} \parallel \mathcal{M})$  [20, Corollary 18], i.e.,

$$D(\mathcal{N} \parallel \mathcal{M}) \leq \underline{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M}) \leq \bar{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M}) \leq D_{\max}(\mathcal{N} \parallel \mathcal{M}). \quad (241)$$

In the following, we sharpen this upper bound by the geometric Rényi channel divergence in general. This gives a more accurate estimation of the fundamental limits of channel discrimination under adaptive strategies.

**Theorem 49.** *Let  $\mathcal{N}$  and  $\mathcal{M}$  be two quantum channels and  $\varepsilon \in (0, 1)$ ,  $\alpha \in (1, 2]$ . It holds*

$$D(\mathcal{N} \parallel \mathcal{M}) \leq \underline{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M}) \leq \bar{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M}) \leq \widehat{D}_{\alpha}(\mathcal{N} \parallel \mathcal{M}) \leq D_{\max}(\mathcal{N} \parallel \mathcal{M}). \quad (242)$$

Moreover,  $\widehat{D}_{\alpha}(\mathcal{N} \parallel \mathcal{M})$  is also a strong converse bound.

*Proof.* The first two inequalities follow from (241). The last inequality follows from the relation that  $\widehat{D}_{\alpha}(\rho \parallel \sigma) \leq D_{\max}(\rho \parallel \sigma)$  in Lemma 4. In the following, we show that  $\bar{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M}) \leq \widehat{D}_{\alpha}(\mathcal{N} \parallel \mathcal{M})$ . This follows a similar step as [20, Proposition 17].

Let  $\{Q, \mathcal{A}\}$  be an arbitrary adaptive protocol for discriminating  $\mathcal{N}$  and  $\mathcal{M}$ . Let  $p$  and  $q$  denote the final decision probabilities. As argued in [20, Proposition 17], we can take  $\alpha_n(\{Q, \mathcal{A}\}) = \varepsilon$ . Then

$$\widehat{D}_{\alpha}(p \parallel q) \geq \frac{1}{\alpha - 1} \log p^{\alpha} q^{1-\alpha} = \frac{1}{\alpha - 1} \log(1 - \varepsilon)^{\alpha} q^{1-\alpha} = \frac{\alpha}{\alpha - 1} \log(1 - \varepsilon) - \log q. \quad (243)$$

By applying the meta-converse in [20, Lemma 14] as well as the chain rule of the geometric Rényi divergence, we have

$$-\frac{1}{n} \log q \leq \widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M}) + \frac{1}{n} \frac{\alpha}{\alpha - 1} \log \frac{1}{1 - \varepsilon}. \quad (244)$$

Since Eq. (244) holds for any channel discrimination protocol, we have  $\bar{\zeta}(\varepsilon, \mathcal{N}, \mathcal{M}) \leq \widehat{D}_\alpha(\mathcal{N} \parallel \mathcal{M})$ .  $\square$

Note that our new strong converse bound is also single-letter and efficient computable via semidefinite program (it even admits a closed-form expression as shown in Lemma 5). The following example of the GAD channels  $\mathcal{A}_{0.8, N_1}$  and  $\mathcal{A}_{0.7, N_2}$  demonstrates that  $\widehat{D}_\alpha$  is much tighter than  $D_{\max}$ .

## References

- Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948)
- Bennett, C.H., Shor, P.W., Smolin, J.A., Thapliyal, A.V.: Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Trans. Inf. Theory* **48**(10), 2637–2655 (2002)
- Wang, X., Wilde, M.M., Su, Y.: Quantifying the magic of quantum channels (2019). [arXiv:1903.04483](https://arxiv.org/abs/1903.04483)
- Seddon, J.R., Campbell, E.: Quantifying magic for multi-qubit operations (2019). [arXiv:1901.03322](https://arxiv.org/abs/1901.03322)
- Lloyd, S.: Capacity of the noisy quantum channel. *Phys. Rev. A* **55**(3), 1613–1622 (1997)
- Shor, P.W.: The quantum channel capacity and coherent information. In: *Lecture Notes, MSRI Workshop on Quantum Computation* (2002)
- Devetak, I.: The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory* **51**(1), 44–55 (2005)
- Huang, Y.: Computing quantum discord is NP-complete. *New J. Phys.* **16**(3), 33027 (2014)
- Wolfowitz, J.: Coding theorems of information theory. *Mathematics of Computation* (1978)
- Bennett, C.H., Devetak, I., Harrow, A.W., Shor, P.W., Winter, A.: The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Trans. Inf. Theory* **60**(5), 2926–2959 (2014)
- Tomamichel, M., Wilde, M.M., Winter, A.: Strong converse rates for quantum communication. *IEEE Trans. Inf. Theory* **63**(1), 715–727 (2017)
- Wilde, M.M.: Tomamichel, marco, berta, mario: converse bounds for private communication over quantum channels. *IEEE Trans. Inf. Theory* **63**(3), 1792–1817 (2017)
- Sagnol, G.: On the semidefinite representation of real functions applied to symmetric matrices. *Linear Algebra Appl.* **439**(10), 2829–2843 (2013)
- Fawzi, H., Saunderson, J.: Lieb’s concavity theorem, matrix geometric means, and semidefinite optimization. *Linear Algebra Appl.* **513**, 240–263 (2017)
- Matsumoto, K.: A new quantum version of f-divergence. In: *Nagoya Winter Workshop: Reality and Measurement in Algebraic Quantum Theory*, pp. 229–273. Springer (2015)
- Petz, D.: Quasi-entropies for finite quantum systems. *Rep. Math. Phys.* **23**(1), 57–65 (1986)
- Müller-Lennert, M., Dupuis, F., Szechr, O., Fehr, S., Tomamichel, M.: On quantum Rényi entropies: a new generalization and some properties. *J. Math. Phys.* **54**(12), 122203 (2013)
- Wilde, M.M., Winter, A., Yang, D.: Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Commun. Math. Phys.* **331**(2), 593–622 (2014)
- Belavkin, V.P., Staszewski, P.: C\*-algebraic generalization of relative entropy and entropy. *Annales de l’IHP Physique théorique* **37**, 51–58 (1982)
- Berta, M., Hirche, C., Kaur, E., Wilde, M.M.: Amortized channel divergence for asymptotic quantum channel discrimination (2018). [arXiv:1808.01498](https://arxiv.org/abs/1808.01498)
- Wang, X., Fang, K., Duan, R.: Semidefinite programming converse bounds for quantum communication. *IEEE Trans. Inf. Theory* **65**(4), 2583–2592 (2018)
- Berta, M., Wilde, M.M.: Amortization does not enhance the max-Rains information of a quantum channel. *New J. Phys.* **20**(5), 053044 (2018)
- Bäumli, S., Das, S., Wilde, M.M.: Fundamental limits on the capacities of bipartite quantum interactions. *Phys. Rev. Lett.* **121**(25), 250504 (2018)
- Pirandola, S., Laurenza, R., Ottaviani, C., Banchi, L.: Fundamental limits of repeaterless quantum communications. *Nat. Commun.* **8**, 15043 (2017)

25. Christandl, M., Alexander, A.: Relative entropy bounds on quantum, private and repeater capacities. *Commun. Math. Phys.* **353**(2), 821–852 (2017)
26. Wang, X., Xie, W., Duan, R.: Semidefinite programming strong converse bounds for classical capacity. *IEEE Trans. Inf. Theory* **64**(1), 640–653 (2018)
27. Umegaki, H.: Conditional expectation in an operator algebra. IV. Entropy and information. *Kodai Math. Semin. Rep.* **14**(2), 59–85 (1962)
28. Renner, R.: Security of quantum key distribution. Ph.D. thesis, ETH Zurich **6**(01), 1–127 (2005)
29. Datta, N.: Min- and max-relative entropies and a new entanglement monotone. *IEEE Trans. Inf. Theory* **55**(6), 2816–2826 (2009)
30. Nussbaum, M., Szkoła, A.: The chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Stat.* **37**(2), 1040–1057 (2009)
31. Audenaert, K.M.R., Calsamiglia, J., Muñoz-Tapia, R., Bagan, E., Masanes, Ll., Acín, A., Verstraete, F.: Discriminating States: The Quantum Chernoff Bound (2007)
32. Lieb, E.H., Thirring, W.E.: Inequalities for the Moments of the Eigenvalues of the Schrodinger Hamiltonian and Their Relation to Sobolev Inequalities, pp. 135–169. Springer, Heidelberg (1991)
33. Cooney, T., Mosonyi, M., Wilde, M.M.: Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication. *Commun. Math. Phys.* **344**(3), 797–829 (2016)
34. Leditzky, F., Kaur, E., Datta, N., Wilde, M.M.: Approaches for approximate additivity of the Holevo information of quantum channels. *Phys. Rev. A* **97**(1), 012332 (2018)
35. Tomamichel, M.: Quantum information processing with finite resources. Springer Briefs in Mathematical Physics, vol. 5. Springer, Cham (2016)
36. Lawson, J.: Weighted means and Karcher equations of positive operators. *Proc. Natl. Acad. Sci.* **110**(39), 15626–15632 (2013)
37. Hiai, F., Petz, D.: The proper formula for relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.* **143**(1), 99–114 (1991)
38. Hiai, F., Mosonyi, M.: Different quantum f-divergences and the reversibility of quantum operations. *Rev. Math. Phys.* **29**(07), 1750023 (2017)
39. Bluhm, A., Capel, A.: A strengthened data processing inequality for the belavkin-staszewski relative entropy (2019). [arXiv:1904.10768](https://arxiv.org/abs/1904.10768)
40. Fang, K., Fawzi, O., Renner, R., Sutter, D.: A chain rule for the quantum relative entropy (2019). to appear on the arXiv
41. Kubo, F., Ando, T.: Means of positive linear operators. *Math. Ann.* **246**(3), 205–224 (1980)
42. Schumacher, B., Nielsen, M.A.: Quantum data processing and error correction. *Phys. Rev. A* **54**(4), 2629–2635 (1996)
43. Barnum, H., Knill, E., Nielsen, M.A.: On quantum fidelities and channel capacities. *IEEE Trans. Inf. Theory* **46**(4), 1317–1329 (2000)
44. Barnum, H., Nielsen, M.A., Schumacher, B.: Information transmission through a noisy quantum channel. *Phys. Rev. A* **57**(6), 4153–4175 (1998)
45. DiVincenzo, D.P., Shor, P.W., Smolin, J.A.: Quantum-channel capacity of very noisy channels. *Phys. Rev. A* **57**(2), 830–839 (1998)
46. Leditzky, F., Leung, D., Smith, G.: Dephasure channel and superadditivity of coherent information. *Phys. Rev. Lett.* **121**(16), 160501 (2018)
47. Cubitt, T., Elkouss, D., Matthews, W., Ozols, M., Perez-Garcia, D., Strelchuk, S.: Unbounded number of channel uses may be required to detect quantum capacity. *Nat. Commun.* **6**(1), 6739 (2015)
48. Holevo, A., Werner, R.: Evaluating capacities of bosonic Gaussian channels. *Phys. Rev. A* **63**(3), 032312 (2001)
49. Müller-Hermes, A., Reeb, D., Wolf, M.M.: Positivity of linear maps under tensor powers. *J. Math. Phys.* **57**(1), 1–25 (2016)
50. Smith, G., Smolin, J.A., Winter, A.: The quantum capacity with symmetric side channels. *IEEE Trans. Inf. Theory* **54**(9), 4208–4217 (2008)
51. Sutter, D., Scholz, V.B., Winter, A., Renner, R.: Approximate degradable quantum channels. *IEEE Trans. Inf. Theory* **63**(12), 7832–7844 (2017)
52. Gao, L., Junge, M., LaRacuente, N.: Capacity bounds via operator space methods. *J. Math. Phys.* **59**(12), 122202 (2018)
53. Smith, G., Smolin, J.A.: Additive extensions of a quantum channel. In: Proceedings of IEEE Information Theory Workshop (ITW), pp. 368–372. IEEE (2008)
54. Rains, E.M.: A semidefinite program for distillable entanglement. *IEEE Trans. Inf. Theory* **47**(7), 2921–2933 (2001)
55. Fawzi, H., Fawzi, O.: Efficient optimization of the quantum relative entropy. *J. Phys. A: Math. Theor.* **51**(15), 154003 (2018)

56. Fawzi, H., Saunderson, J., Parrilo, P.A.: Semidefinite approximations of the matrix logarithm. *Found. Comput. Math.* **19**, 259–296 (2018)
57. Wang, X., Fang, K., Tomamichel, M.: On converse bounds for classical communication over quantum channels. *IEEE Trans. Inf. Theory* **65**(7), 4609–4619 (2019)
58. Plenio, M.B.: Logarithmic negativity: a full entanglement monotone that is not convex. *Phys. Rev. Lett.* **95**(9), 090503 (2005)
59. Yu Kitaev, A.: Quantum computations: algorithms and error correction. *Russ. Math. Surv.* **52**, 1191–1249 (1997)
60. Sion, M.: On general minimax theorems. *Pac. J. Math.* **8**(1), 171–176 (1958)
61. Wang, X., Duan, R.: A semidefinite programming upper bound of quantum capacity. In: 2016 IEEE International Symposium on Information Theory (ISIT), volume 2016-Augus, pp. 1690–1694. IEEE (2016)
62. Polyanskiy, Y., Verdu, S.: Arimoto channel coding converse and Renyi divergence. In: 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), vol. 0, pp. 1327–1333. IEEE (2010)
63. Bennett, C.H., Harrow, A.W., Leung, D.W., Smolin, J.A.: On the capacities of bipartite hamiltonians and unitary gates. *IEEE Trans. Inf. Theory* **49**(8), 1895–1911 (2003)
64. Gour, G., Scandolo, C.M.: The Entanglement of a Bipartite Channel (2019). [arXiv:1907.02552](https://arxiv.org/abs/1907.02552)
65. Bäuml, S., Das, S., Wang, X., Wilde, M.M.: Resource theory of entanglement for bipartite quantum channels (2019). [arXiv:1907.04181](https://arxiv.org/abs/1907.04181)
66. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 2.1 (2014). <http://cvxr.com/cvx>
67. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (eds.) *Recent Advances in Learning and Control. Lecture Notes in Control and Information Sciences*, pp. 95–110. Springer, Berlin (2008)
68. MOSEK ApS: The MOSEK optimization toolbox for MATLAB manual. Version 9.0. (2019)
69. Chirolli, L., Burkard, G.: Decoherence in solid-state qubits. *Adv. Phys.* **57**(3), 225–285 (2008)
70. Khatri, S., Sharma, K., Wilde, M.M.: Information-theoretic aspects of the generalized amplitude damping channel (2019). [arXiv:1903.07747](https://arxiv.org/abs/1903.07747)
71. Cai, N., Winter, A., Yeung, R.W.: Quantum privacy and quantum wiretap channels. *Probl. Inf. Transm.* **40**(4), 318–336 (2004)
72. Smith, G., Renes, J.M., Smolin, J.A.: Structured codes improve the Bennett-Brassard-84 quantum key rate. *Phys. Rev. Lett.* **100**(17), 1–4 (2008)
73. Elkouss, D., Strelchuk, S.: Superadditivity of private information for any number of uses of the channel. *Phys. Rev. Lett.* **115**(4), 040501 (2015)
74. Bennett, C.H., Brassard, G.: Quantum cryptography: public key distribution and coin tossing. In: *International Conference on Computers, Systems & Signal Processing*, Bangalore, India, Dec 9-12, 1984, pp. 175–179 (1984)
75. Takeoka, M., Guha, S., Wilde, M.M.: The squashed entanglement of a quantum channel. *IEEE Trans. Inf. Theory* **60**(8), 4987–4998 (2014)
76. Berta, M., Brandao, F.G.S.L., Christandl, M., Wehner, S.: Entanglement cost of quantum channels. *IEEE Trans. Inf. Theory* **59**(10), 6779–6795 (2013)
77. Horodecki, M., Horodecki, P., Horodecki, R.: Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A* **223**(1–2), 1–8 (1996)
78. Vedral, V., Plenio, M.B.: Entanglement measures and purification procedures. *Phys. Rev. A* **57**(3), 1619–1633 (1998)
79. Aubrun, G., Szarek, S.J.: *Alice and Bob meet Banach*, vol. 223. American Mathematical Soc (2017)
80. Horodecki, M., Shor, P.W., Ruskai, M.B.: Entanglement breaking channels. *Rev. Math. Phys.* **15**(06), 629–641 (2003)
81. Holevo, A.S.: Quantum coding theorems. *Russ. Math. Surv.* **53**(6), 1295–1331 (1998)
82. Holevo, A.S.: The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory* **44**(1), 269–273 (1998)
83. Schumacher, B., Westmoreland, M.D.: Sending classical information via noisy quantum channels. *Phys. Rev. A* **56**(1), 131–138 (1997)
84. Hastings, M.B.: Superadditivity of communication capacity using entangled inputs. *Nat. Phys.* **5**(4), 255–257 (2009)
85. Beigi, S., Shor, P.W.: On the complexity of computing zero-error and Holevo capacity of quantum channels (2007). [arXiv:0709.2090](https://arxiv.org/abs/0709.2090)
86. Ohya, M., Petz, D., Watanabe, N.: On capacities of quantum channels. *Probab. Math. Statist.* **17**, 179–196 (1997)
87. Wilde, M.M., Winter, A.: Strong converse for the quantum capacity of the erasure channel for almost all codes (2014). [arXiv:1402.3626](https://arxiv.org/abs/1402.3626)

88. Aliferis, P., Brito, F., DiVincenzo, D.P., Preskill, J., Steffen, M., Terhal, B.M.: Fault-tolerant computing with biased-noise superconducting qubits: a case study. *New J. Phys.* **11**(1), 013061 (2009)
89. Sutter, D., Sutter, T., Esfahani, P.M., Renner, R.: Efficient approximation of quantum channel capacities. *IEEE Trans. Inf. Theory* **62**(1), 578–598 (2016)
90. Shor, P.W.: Fault-tolerant quantum computation. In: *Proceedings of 37th Conference on Foundations of Computer Science*, pp. 56–65. IEEE Comput. Soc. Press (1996)
91. Yu Kitaev, A.: Fault-tolerant quantum computation by anyons. *Ann. Phys.* **303**(1), 2–30 (2003)
92. Knill, E.: Fault-tolerant postselected quantum computation: Threshold analysis (2004). [arXiv:quant-ph/0404104](https://arxiv.org/abs/quant-ph/0404104)
93. Raussendorf, R., Harrington, J.: Fault-tolerant quantum computation with high threshold in two dimensions. *Phys. Rev. Lett.* **98**(19), 190504 (2007)
94. Gottesman, D.: Stabilizer codes and quantum error correction (1997). [arXiv:quant-ph/9705052](https://arxiv.org/abs/quant-ph/9705052)
95. Aaronson, S.: Improved simulation of stabilizer circuits. *Phys. Rev. A* **70**(5), 052328 (2004)
96. Gottesman, D., Chuang, I.L.: Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**(6760), 390 (1999)
97. Zhou, X., Leung, D.W., Chuang, I.L.: Methodology for quantum logic gate construction. *Phys. Rev. A* **62**(5), 052316 (2000)
98. Bravyi, S., Haah, J.: Magic-state distillation with low overhead. *Phys. Rev. A* **86**(5), 052329 (2012)
99. Hastings, M.B., Haah, J.: Distillation with sublogarithmic overhead. *Phys. Rev. Lett.* **120**(5), 050504 (2018)
100. Gottesman, D.: Theory of fault-tolerant quantum computation. *Phys. Rev. A* **57**, 127–137 (1998)
101. Aharonov, D., Ben-Or, M.: Fault-tolerant quantum computation with constant error rate. *SIAM J. Comput.* **38**(4), 1207–1282 (2008)
102. Wang, X., Wilde, M.M., Su, Y.: Efficiently computable bounds for magic state distillation (2018). [arXiv:1812.10145](https://arxiv.org/abs/1812.10145)
103. Veitch, V., Hamed Mousavian, S.A., Gottesman, D., Emerson, J.: The resource theory of stabilizer quantum computation. *New J. Phys.* **16**(1), 013009 (2014)
104. Nagaoka, H., Ogawa, T.: Strong converse and Stein’s lemma in quantum hypothesis testing. *IEEE Trans. Inf. Theory* **46**(7), 2428–2433 (2000)
105. Bhatia, R.: *Positive Definite Matrices*. Princeton University Press (2007)
106. Sutter, D.: *Approximate Quantum Markov Chains*. Springer Briefs in Mathematical Physics, vol. 28. Springer, Cham (2018)
107. Fawzi, H.: The set of separable states has no finite semidefinite representation except in dimension 3 by 2 (2019). [arXiv:1905.02575](https://arxiv.org/abs/1905.02575)
108. Doherty, A.C., Parrilo, P.A., Spedalieri, F.M.: Distinguishing separable and entangled states. *Phys. Rev. Lett.* **88**(18), 187904 (2002)
109. Doherty, A.C., Parrilo, P.A., Spedalieri, F.M.: Complete family of separability criteria. *Phys. Rev. A* **69**(2), 22308 (2004)
110. Fang, K., Fawzi, H.: The sum-of-squares hierarchy on the sphere, and applications in quantum information theory (2019). [arXiv:1908.05155](https://arxiv.org/abs/1908.05155)
111. Garcia-Patrón, R., Pirandola, S., Lloyd, S., Shapiro, J.H.: Reverse coherent information. *Phys. Rev. Lett.* **102**(21), 210501 (2009)
112. King, C.: Additivity for unital qubit channels. *J. Math. Phys.* **43**(10), 4641–4653 (2002)
113. Filippov, S.N.: Lower and upper bounds on nonunital qubit channel capacities. *Rep. Math. Phys.* **82**(2), 149–159 (2018)
114. Li-Zhen, H., Mao-Fa, F.: Entanglement-assisted classical capacity of a generalized amplitude damping channel. *Chin. Phys. Lett.* **24**(9), 2482 (2007)
115. Li-Zhen, H., Mao-Fa, F.: The holevo capacity of a generalized amplitude-damping channel. *Chin. Phys.* **16**(7), 1843 (2007)

Communicated by M. Christandl