# Entropy of a quantum channel

Gilad Gour[1,*] and Mark M. Wilde [2,†]

[1]*Department of Mathematics and Statistics, Institute for Quantum Science and Technology, University of Calgary, Alberta, Canada T2N 1N4*
[2]*Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, and Center for Computation and Technology,*
*Louisiana State University, Baton Rouge, Louisiana 70803, USA*

The von Neumann entropy of a quantum state is a central concept in physics and information theory, having a number of compelling physical interpretations. There is a certain perspective that the most fundamental notion in quantum mechanics is that of a quantum channel, as quantum states, unitary evolutions, measurements, and discarding of quantum systems can each be regarded as certain kinds of quantum channels. Thus, an important goal is to define a consistent and meaningful notion of the entropy of a quantum channel. Motivated by the fact that the entropy of a state $\rho$ can be formulated as the difference of the number of physical qubits and the "relative entropy distance" between $\rho$ and the maximally mixed state, here we define the entropy of a channel $\mathcal{N}$ as the difference of the number of physical qubits of the channel output with the "relative entropy distance" between $\mathcal{N}$ and the completely depolarizing channel. We prove that this definition satisfies all of the axioms, recently put forward by Gour [IEEE Trans. Inf. Theory **65**, 5880 (2019)], required for a channel entropy function. The task of quantum channel merging, in which the goal is for the receiver to merge his share of the channel with the environment's share, gives a compelling operational interpretation of the entropy of a channel. The entropy of a channel can be negative for certain channels, but this negativity has an operational interpretation in terms of the channel merging protocol. We define Rényi and min-entropies of a channel and prove that they satisfy the axioms required for a channel entropy function. Among other results, we also prove that a smoothed version of the min-entropy of a channel satisfies the asymptotic equipartition property.

## I. INTRODUCTION

In his foundational work on quantum statistical mechanics, von Neumann extended the classical Gibbs entropy concept to the quantum realm [1]. This extension, known as the von Neumann or quantum entropy, plays a key role in physics and information theory. It is defined by the following formula [1]:

$$H(A)_\rho \equiv -\operatorname{Tr}\{\rho_A \log_2 \rho_A\}, \qquad (1)$$

where $\rho_A$ is the state of a system $A$. The entropy has operational interpretations in terms of quantum data compression [2] and optimal entanglement manipulation rates of pure bipartite quantum states [3], where the choice of base two for the logarithm becomes clear. In recent developments of quantum thermodynamics, it was shown that the free energy, namely, the difference of the energy and the product of the temperature and the von Neumann entropy, can be interpreted as the rate at which work can be extracted from a large number of copies of a quantum system in a thermal bath at fixed temperature, by using only thermal operations [4].

*gour@ucalgary.ca
†mwilde@lsu.edu

By defining the quantum relative entropy of a state $\rho_A$ and a positive semidefinite operator $\sigma_A$ as [5]

$$D(\rho_A \| \sigma_A) \equiv \operatorname{Tr}\{\rho_A[\log_2 \rho_A - \log_2 \sigma_A]\}, \qquad (2)$$

if $\operatorname{supp}(\rho_A) \subseteq \operatorname{supp}(\sigma_A)$ and $D(\rho_A \| \sigma_A) = +\infty$ otherwise, we can rewrite the formula for quantum entropy as follows:

$$H(A)_\rho = \log_2 |A| - D(\rho_A \| \pi_A), \qquad (3)$$

where $|A|$ denotes the dimension of the system $A$ and $\pi_A \equiv I_A/|A|$ denotes the maximally mixed state. In this way, we can think of entropy as quantifying the difference of the number of physical qubits contained in the system $A$ and the "relative entropy distance" of the state $\rho_A$ to the maximally mixed state $\pi_A$. This way of thinking about quantum entropy is relevant in the resource theory of purity [6–11], in which the goal is to distill local pure states from a given state (or vice versa) by allowing local unitary operations for free. Furthermore, the quantum relative entropy $D(\rho_A \| \pi_A)$ has an operational meaning as the optimal rate at which the state $\rho_A$ can be distinguished from the maximally mixed state $\pi_A$ in the Stein setting of quantum hypothesis testing [12,13]. In what follows, we use the formula in Eq. (3) as the basis for defining the entropy of a quantum channel.

For some time now, there has been a growing realization that the fundamental constituents of quantum mechanics are quantum channels. Recall that a quantum channel $\mathcal{N}_{A \to B}$ is a completely positive, trace preserving map that takes a quantum state for system $A$ to one for system $B$ [14]. Indeed,

all the relevant components of the theory, including quantum states, measurements, unitary evolutions, etc., can be written as quantum channels. A quantum state can be understood as a preparation channel, sending a trivial quantum system to a nontrivial one prepared in a given state. A quantum measurement can be understood as a quantum channel that sends a quantum system to a classical one; and of course a unitary evolution is a kind of quantum channel, as well as the discarding of a quantum system. One might even boldly go as far as to say that there is really only a single postulate of quantum mechanics, and it is that "everything is a quantum channel." With this perspective, one could start from this unified postulate and then understand from there particular kinds of channels, i.e., states, measurements, and unitary evolutions.

Due to the fundamental roles of quantum channels and the entropy of a quantum state, as highlighted above, it is thus natural to ask whether there is a meaningful notion of the entropy of a quantum channel, i.e., a quantifier of the uncertainty of a quantum channel. As far as we are aware, this question has not been fully addressed in prior literature (see Remark 2 for further discussion), and it is the aim of the present paper to provide a convincing notion of a quantum channel's entropy. To define such a notion, we look to Eq. (3) for inspiration. As such, we need generalizations of the quantum relative entropy and the maximally mixed state to the setting of quantum channels:

(1) The quantum relative entropy of channels $\mathcal{N}_{A \to B}$ and $\mathcal{M}_{A \to B}$ is defined as [15,16]

$$D(\mathcal{N} \| \mathcal{M}) \equiv \sup_{\rho_{RA}} D(\mathcal{N}_{A \to B}(\rho_{RA}) \| \mathcal{M}_{A \to B}(\rho_{RA})), \quad (4)$$

where the optimization is with respect to bipartite states $\rho_{RA}$ of a reference system $R$ of arbitrary size and the channel input system $A$. Due to state purification, the data-processing inequality [17], and the Schmidt decomposition theorem, it suffices to optimize over states $\rho_{RA}$ that are pure and such that system $R$ is isomorphic to system $A$. This observation significantly reduces the complexity of computing the channel relative entropy.

(2) The channel that serves as a generalization of the maximally mixed state is the channel $\mathcal{R}_{A \to B}$ that completely randomizes or depolarizes the input state as follows:

$$\mathcal{R}_{A \to B}(X_A) = \mathrm{Tr}\{X_A\} \pi_B, \quad (5)$$

where $X_A$ is an arbitrary operator for system $A$. That is, its action is to discard the input and replace with a maximally mixed state $\pi_B$.

With these notions in place, we can now define the entropy of a quantum channel:

*Definition 1 (Entropy of a quantum channel).* Let $\mathcal{N}_{A \to B}$ be a quantum channel. Its entropy is defined as

$$H(\mathcal{N}) \equiv \log_2 |B| - D(\mathcal{N} \| \mathcal{R}), \quad (6)$$

where $D(\mathcal{N} \| \mathcal{R})$ is the channel relative entropy in Eq. (4) and $\mathcal{R}_{A \to B}$ is the completely randomizing channel in Eq. (5).

We remark here that, in analogy to the operational interpretation for $D(\rho_A \| \pi_A)$ mentioned above, it is known that $D(\mathcal{N} \| \mathcal{R})$ is equal to the optimal rate at which the channel $\mathcal{N}_{A \to B}$ can be distinguished from the completely randomizing channel $\mathcal{R}_{A \to B}$, by allowing for any possible quantum strategy

to distinguish the channels [15]. Again, this statement holds in the Stein setting of quantum hypothesis testing (see Ref. [15] for details). We also emphasize here that the entropy of a channel can be negative for some channels, but this negativity has an operational interpretation in terms of the channel merging protocol (see Remark 7 in this context).

The remainder of our paper contains arguments advocating for this definition of a channel's entropy. In the next section, we show that it satisfies the three basic axioms, put forward in Ref. [18], for any function to be called an entropy function for a quantum channel, including nondecrease under the action of a random unitary superchannel, additivity, and normalization. After that, we provide several alternate representations for the entropy of a channel, the most significant of which is the completely bounded entropy of Ref. [19]. Section III delivers an operational interpretation of a channel's entropy in terms of an information-theoretic task that we call quantum channel merging, which is a dynamical counterpart of the well known task of quantum state merging [20,21]. We calculate channel entropies for several example channels in Sec. IV, which include erasure, dephasing, depolarizing, and Werner–Holevo channels. In the same section, we introduce the energy-constrained and unconstrained entropies of a quantum channel and calculate them for thermal, amplifier, and additive-noise bosonic Gaussian channels. In Sec. V, we define the $\alpha$-Rényi entropy of a channel, prove that it satisfies the basic axioms for certain values of the Rényi parameter $\alpha$, and provide alternate representations for it. In Sec. VI, we define the min-entropy of a channel, establish that it satisfies the basic axioms, and provide alternate representations for it. In Sec. VII, we define the smoothed min-entropy of a channel, and then we prove an asymptotic equipartition property, which relates the smoothed min-entropy of a channel to its entropy. In Sec. VIII, we discuss other entropies of a channel, noting that several of them collapse to the (von Neumann) entropy of a channel. We finally conclude in Sec. IX with a summary and some open questions.

*Remark 2.* We note here that "the entropy of a channel" was also defined in Refs. [22,23], but the definition given there does not satisfy "reduction to states" or the basic axiom of normalization. For this reason, it cannot be considered an entropy function according to the approach of Ref. [18].

## II. ENTROPY OF A QUANTUM CHANNEL

Proceeding with Definition 1 for the entropy of a quantum channel, we now establish several of its properties, and then we provide alternate representations for it.

### A. Properties of the entropy of a quantum channel

In Ref. [18], it was advocated that a function of a quantum channel is an entropy function if it satisfies nondecrease under random unitary superchannels, additivity, and normalization. As shown in the next three subsections, the entropy of a channel, as given in Definition 1, satisfies all three axioms, and in fact, it satisfies stronger properties that imply these.

### 1. Nondecrease under the action of a uniformity preserving superchannel

Before addressing the first axiom, let us first briefly review the notion of superchannels [24], which are linear maps that take as input a quantum channel and output a quantum channel. To define them, let $\mathcal{L}(A \to B)$ denote the set of all linear maps from $\mathcal{L}(A)$ to $\mathcal{L}(B)$. Similarly, let $\mathcal{L}(C \to D)$ denote the set of all linear maps from $\mathcal{L}(C)$ to $\mathcal{L}(D)$. Let $\Theta : \mathcal{L}(A \to B) \to \mathcal{L}(C \to D)$ denote a linear supermap, taking $\mathcal{L}(A \to B)$ to $\mathcal{L}(C \to D)$. A quantum channel is a particular kind of linear map, and any linear supermap $\Theta$ that takes as input an arbitrary quantum channel $\Psi_{A \to B} \in \mathcal{L}(A \to B)$ and is required to output a quantum channel $\Phi_{C \to D} \in \mathcal{L}(C \to D)$ should preserve the properties of complete positivity (CP) and trace preservation (TP). That is, the supermap should be CPTP preserving. Furthermore, for the supermap to be physical, the same should be true when it acts on subsystems of bipartite quantum channels, so that the supermap $\mathrm{id} \otimes \Theta$ should be CPTP preserving, where id represents an arbitrary identity supermap. A supermap satisfying this property is said to be completely CPTP preserving and is then called a superchannel. It was proven in Ref. [24] that any superchannel $\Theta : \mathcal{L}(A \to B) \to \mathcal{L}(C \to D)$ can be physically realized as follows. If

$$\Phi_{C \to D} = \Theta[\Psi_{A \to B}] \tag{7}$$

for an arbitrary input channel $\Psi_{A \to B} \in \mathcal{L}(A \to B)$ and some output channel $\Phi_{C \to D} \in \mathcal{L}(C \to D)$, then the physical realization of the superchannel $\Theta$ is as follows:

$$\Phi_{C \to D} = \Omega_{BE \to D} \circ (\Psi_{A \to B} \otimes \mathrm{id}_E) \circ \Lambda_{C \to AE}, \tag{8}$$

where $\Lambda_{C \to AE} : \mathcal{L}(C) \to \mathcal{L}(AE)$ is a preprocessing channel, system $E$ corresponds to some memory or environment system, and $\Omega_{BE \to D} : \mathcal{L}(BE) \to \mathcal{L}(D)$ is a post-processing channel.

A uniformity preserving superchannel $\Theta$ is a superchannel that takes the completely randomizing channel $\mathcal{R}_{A \to B}$ in (5) to another completely randomizing channel $\mathcal{R}_{C \to D}$, such that $|A| = |C|$ and $|B| = |D|$, i.e.,

$$\Theta(\mathcal{R}_{A \to B}) = \mathcal{R}_{C \to D}. \tag{9}$$

For such superchannels, we have the following:

*Proposition 3.* Let $\mathcal{N}_{A \to B}$ be a quantum channel, and let $\Theta$ be a uniformity preserving superchannel as defined above. Then the entropy of a channel does not decrease under the action of such a superchannel:

$$H(\Theta(\mathcal{N})) \geqslant H(\mathcal{N}). \tag{10}$$

*Proof.* This follows from the fact that the channel relative entropy is nonincreasing under the action of an arbitrary superchannel [18,25]. That is, for two channels $\mathcal{N}_{A \to B}$ and $\mathcal{M}_{A \to B}$, and a superchannel $\Xi$, the following inequality holds

$$D(\mathcal{N} \| \mathcal{M}) \geqslant D(\Xi(\mathcal{N}) \| \Xi(\mathcal{M})). \tag{11}$$

Applying this, we find that

$$H(\mathcal{N}) = \log_2 |B| - D(\mathcal{N} \| \mathcal{R}) \tag{12}$$

$$\leqslant \log_2 |B| - D(\Theta(\mathcal{N}) \| \Theta(\mathcal{R})) \tag{13}$$

$$= \log_2 |B| - D(\Theta(\mathcal{N}) \| \mathcal{R}) \tag{14}$$

$$= \log_2 |D| - D(\Theta(\mathcal{N}) \| \mathcal{R}) \tag{15}$$

$$= H(\Theta(\mathcal{N})). \tag{16}$$

The second equality follows by definition from Eq. (9). ∎

In Ref. [18], a superchannel $\Upsilon$ was called a random unitary superchannel if its action on a channel $\mathcal{N}_{A \to B}$ can be written as

$$\Upsilon(\mathcal{N}_{A \to B}) = \sum_x p_X(x) \mathcal{V}^x_{B \to D} \circ \mathcal{N}_{A \to B} \circ \mathcal{U}^x_{C \to A}, \tag{17}$$

where $\mathcal{U}^x_{C \to A}$ and $\mathcal{V}^x_{B \to D}$ are unitary channels and $p_X(x)$ is a probability distribution. In Ref. [18], it was proved that a random unitary superchannel is a special kind of uniformity preserving superchannel. Thus, due to Proposition 3, it follows that the entropy of a channel, as given in Definition 1, satisfies the first axiom from Ref. [18] required for an entropy function.

### 2. Additivity

In this subsection, we prove that the entropy of a channel is additive, which is the second axiom proposed in Ref. [18] for a channel entropy function. The proof is related to many prior additivity results from Refs. [15,19,26–28].

*Proposition 4 (Additivity).* Let $\mathcal{N}$ and $\mathcal{M}$ be quantum channels. Then the channel entropy is additive in the following sense:

$$H(\mathcal{N} \otimes \mathcal{M}) = H(\mathcal{N}) + H(\mathcal{M}). \tag{18}$$

*Proof.* This can be understood as a consequence of the additivity results from Refs. [15,28], which in turn are related to the earlier additivity results from Refs. [19,26,27]. For channels $\mathcal{N}_{A_1 \to B_1}$ and $\mathcal{M}_{A_2 \to B_2}$, and corresponding randomizing channels $\mathcal{R}^{(1)}_{A_1 \to B_1}$ and $\mathcal{R}^{(2)}_{A_2 \to B_2}$, we have by definition that

$$H(\mathcal{N} \otimes \mathcal{M})$$

$$= \log_2(|B_1||B_2|) - D(\mathcal{N} \otimes \mathcal{M} \| \mathcal{R}^{(1)} \otimes \mathcal{R}^{(2)}) \tag{19}$$

$$= \log_2 |B_1| + \log_2 |B_2| - D(\mathcal{N} \otimes \mathcal{M} \| \mathcal{R}^{(1)} \otimes \mathcal{R}^{(2)}), \tag{20}$$

and so the result follows if

$$D(\mathcal{N} \otimes \mathcal{M} \| \mathcal{R}^{(1)} \otimes \mathcal{R}^{(2)}) = D(\mathcal{N} \| \mathcal{R}^{(1)}) + D(\mathcal{M} \| \mathcal{R}^{(2)}). \tag{21}$$

Note that the inequality "$\geqslant$" for Eq. (21) trivially follows, and so it remains to prove the inequality "$\leqslant$" for Eq. (21). To this end, let $\psi_{RA_1A_2}$ be an arbitrary pure state, and define

$$\rho_{R'A_1} \equiv \mathcal{M}_{A_2 \to B_2}(\psi_{RA_1A_2}), \tag{22}$$

$$\sigma_{R'A_1} \equiv \mathcal{R}_{A_2 \to B_2}(\psi_{RA_1A_2}), \tag{23}$$

where system $R' \equiv RB_2$. Then we find that

$$D((\mathcal{N}_{A_1 \to B_1} \otimes \mathcal{M}_{A_2 \to B_2})(\psi_{RA_1A_2}) \| (\mathcal{R}_{A_1 \to B_1} \otimes \mathcal{R}_{A_2 \to B_2})(\psi_{RA_1A_2}))$$

$$= D(\mathcal{N}_{A_1 \to B_1}(\rho_{R'A_1}) \| \mathcal{R}_{A_1 \to B_1}(\sigma_{R'A_1})) \tag{24}$$

$$\leqslant D(\mathcal{N}_{A_1 \to B_1}(\rho_{R'A_1}) \| \mathcal{R}_{A_1 \to B_1}(\rho_{R'A_1})) + D(\rho_{R'A_1} \| \sigma_{R'A_1}) \tag{25}$$

$$= D(\mathcal{N}_{A_1 \to B_1}(\rho_{R'A_1}) \| \mathcal{R}_{A_1 \to B_1}(\rho_{R'A_1})) + D(\mathcal{M}_{A_2 \to B_2}(\psi_{RA_1A_2}) \| \mathcal{R}_{A_2 \to B_2}(\psi_{RA_1A_2})) \tag{26}$$

$$\leqslant \sup_{\rho_{R'A_1}} D(\mathcal{N}_{A_1 \to B_1}(\rho_{R'A_1}) \| \mathcal{R}_{A_1 \to B_1}(\rho_{R'A_1}))$$

$$+ \sup_{\psi_{RA_1A_2}} D(\mathcal{M}_{A_2 \to B_2}(\psi_{RA_1A_2}) \| \mathcal{R}_{A_2 \to B_2}(\psi_{RA_1A_2})) \tag{27}$$

$$= D(\mathcal{N}_{A_1 \to B_1} \| \mathcal{R}_{A_1 \to B_1}) + D(\mathcal{M}_{A_2 \to B_2} \| \mathcal{R}_{A_2 \to B_2}). \tag{28}$$

The first inequality follows from the same steps given in the proof of Ref. [[28], Lemma 38]. This concludes the proof. ∎

Another approach to establishing additivity is to employ the first identity of Proposition 6 (in Sec. II B) and [[26], Eq. (3.28)], the latter of which was independently formulated in Ref. [[19], Sec. 2.3].

### 3. Reduction to states and normalization

We now prove that the entropy of a channel reduces to the entropy of a state if the channel is one that replaces the input with a given state.

*Proposition 5 (Reduction to states).* Let the channel $\mathcal{N}_{A \to B}$ be a replacer channel, defined such that $\mathcal{N}_{A \to B}(\rho_A) = \sigma_B$ for all states $\rho_A$ and some state $\sigma_B$. Then the following equality holds:

$$H(\mathcal{N}) = H(B)_\sigma. \tag{29}$$

*Proof.* For any input $\psi_{RA}$, the output is $\mathcal{N}_{A \to B}(\psi_{RA}) = \psi_R \otimes \sigma_B$, and we find that

$$D(\mathcal{N}_{A \to B}(\psi_{RA}) \| \mathcal{R}_{A \to B}(\psi_{RA})) = D(\psi_R \otimes \sigma_B \| \psi_R \otimes \pi_B)$$

$$= D(\sigma_B \| \pi_B). \tag{30}$$

This implies that

$$H(\mathcal{N}) = \log_2 |B| - D(\mathcal{N} \| \mathcal{R}) \tag{31}$$

$$= \log |B| - D(\sigma_B \| \pi_B) \tag{32}$$

$$= H(B)_\sigma, \tag{33}$$

concluding the proof. ∎

A final axiom (normalization) for a channel entropy function [18] is that it should be equal to zero for any channel that replaces the input with a pure state and it should be equal to the logarithm of the output dimension for any channel that replaces the input with the maximally mixed state. Clearly, Proposition 5 implies the normalization property if the replaced state is maximally mixed or pure.

### B. Alternate representations for the entropy of a channel

The entropy of a quantum channel has at least three alternate representations, in terms of the completely bounded entropy of Ref. [19], the entropy gain of its complementary channel [27], and the maximum output entropy of the channel conditioned on its environment. We recall these various channel functions now.

Recall that the completely bounded entropy of a quantum channel $\mathcal{N}_{A \to B}$ is defined as [19]

$$H_{\text{CB,min}}(\mathcal{N}) \equiv \inf_{\rho_{RA}} H(B|R)_\omega, \tag{34}$$

where $H(B|R)_\omega \equiv H(BR)_\omega - H(R)_\omega$ is the conditional entropy of the state $\omega_{RB} = \mathcal{N}_{A \to B}(\rho_{RA})$ and the system $R$ is unbounded. However, due to data processing, purification, and the Schmidt decomposition theorem, it follows that

$$H_{\text{CB,min}}(\mathcal{N}) = \inf_{\psi_{RA}} H(B|R)_\omega, \tag{35}$$

where $\psi_{RA}$ is a pure bipartite state with system $R$ isomorphic to the channel input system $A$.

Due to the Stinespring representation theorem [29], every channel $\mathcal{N}_{A \to B}$ can be realized by the action of an isometric channel $\mathcal{U}^{\mathcal{N}}_{A \to BE}$ and a partial trace as follows:

$$\mathcal{N}_{A \to B} = \text{Tr}_E \circ \mathcal{U}^{\mathcal{N}}_{A \to BE}. \tag{36}$$

If we instead trace over the channel output $B$, then this realizes a complementary channel of $\mathcal{N}_{A \to B}$:

$$\mathcal{N}^c_{A \to E} \equiv \text{Tr}_B \circ \mathcal{U}^{\mathcal{N}}_{A \to BE}. \tag{37}$$

Using these notions, we can define the entropy gain of a complementary channel of $\mathcal{N}_{A \to B}$ as follows [27]:

$$G(\mathcal{N}^c_{A \to E}) \equiv \inf_{\rho_A} [H(E)_\tau - H(A)_\rho], \tag{38}$$

where $\tau_{BE} \equiv \mathcal{U}^{\mathcal{N}}_{A \to BE}(\rho_A)$. The entropy gain has been investigated for infinite-dimensional quantum systems in Refs. [30–32]. We can also define the maximum output entropy of the channel conditioned on its environment as

$$\sup_{\rho_A} H(B|E)_\tau, \tag{39}$$

where again $\tau_{BE} \equiv \mathcal{U}^{\mathcal{N}}_{A \to BE}(\rho_A)$.

We now prove that the entropy of a channel, as given in Definition 1, is equal to the completely bounded entropy, the entropy gain of a complementary channel, and the negation of the maximum output entropy of the channel conditioned on its environment.

*Proposition 6.* Let $\mathcal{N}_{A \to B}$ be a quantum channel, and let $\mathcal{U}^{\mathcal{N}}_{A \to BE}$ be an isometric channel extending it, as in Eq. (36). Then

$$H(\mathcal{N}) = H_{\text{CB,min}}(\mathcal{N}) = G(\mathcal{N}^c_{A \to E}) = - \sup_{\rho_A} H(B|E)_\tau, \tag{40}$$

where $\tau_{BE} \equiv \mathcal{U}^{\mathcal{N}}_{A \to BE}(\rho_A)$. It then follows that

$$|H(\mathcal{N})| \leqslant \log_2 |B|. \tag{41}$$

*Proof.* Using the identity $D(\rho \| c\sigma) = D(\rho \| \sigma) - \log_2 c$, for a constant $c > 0$, and the fact that the conditional entropy $H(B|R)_{\mathcal{N}(\psi)} = -D(\mathcal{N}_{A \to B}(\psi_{RA}) \| \psi_R \otimes I_B)$, we find that

$$H(\mathcal{N}) = \log_2 |B| - D(\mathcal{N} \| \mathcal{R}) \tag{42}$$

$$= \log_2 |B| - \sup_{\psi_{RA}} D(\mathcal{N}_{A \to B}(\psi_{RA}) \| \mathcal{R}_{A \to B}(\psi_{RA})) \tag{43}$$

$$= \log_2 |B| - \sup_{\psi_{RA}} D(\mathcal{N}_{A \to B}(\psi_{RA}) \| \psi_R \otimes \pi_B) \tag{44}$$

$$= -\sup_{\psi_{RA}} D(\mathcal{N}_{A \to B}(\psi_{RA}) \| \psi_R \otimes I_B) \tag{45}$$

$$= \inf_{\psi_{RA}} H(B|R)_{\mathcal{N}(\psi)} \tag{46}$$

$$= H_{\mathrm{CB,min}}(\mathcal{N}). \tag{47}$$

We can then conclude the dimension bound in Eq. (41) from the fact that it holds uniformly for the conditional entropy $|H(B|R)| \leqslant \log_2 |B|$. Defining $\tau_{RBE} = \mathcal{U}^{\mathcal{N}}_{A \to BE}(\psi_{RA})$, from the identity

$$H(B|R)_{\tau} = H(BR)_{\tau} - H(R)_{\tau} = H(E)_{\tau} - H(A)_{\rho}, \tag{48}$$

for $\rho_A = \mathrm{Tr}_R\{\psi_{RA}\}$, and where we used $\tau_R = \psi_R$, we have that

$$H(\mathcal{N}) = G(\mathcal{N}^c_{A \to E}) \equiv \inf_{\rho_A} [H(E)_{\tau} - H(A)_{\rho}]. \tag{49}$$

We finally conclude that

$$H(\mathcal{N}) = -\sup_{\rho_A} H(B|E)_{\tau}, \tag{50}$$

which follows from the identity (duality of conditional entropy)

$$H(B|R)_{\omega} = -H(B|E)_{\mathcal{U}(\psi_A)}. \tag{51}$$

This concludes the proof. ∎

*Remark 7.* We note here, as observed in Ref. [19], that the dimension lower bound $H(\mathcal{N}) \geqslant -\log_2 |B|$ is saturated by the identity channel, while the dimension upper bound $H(\mathcal{N}) \leqslant \log_2 |B|$ is saturated for the completely randomizing (depolarizing) channel, which sends every state to the maximally mixed state. Also, the entropy $H(\mathcal{N})$ is equal to zero for a replacer channel that replaces the input with a pure quantum state. It is also known that the entropy of a channel is nonnegative for all entanglement-breaking channels, as shown in Ref. [19]. This includes all classical channels.

Thus, unlike entropy of a quantum state, the entropy of a quantum channel can be negative. This negativity captures the ability of the channel to distill quantum entanglement, in a sense made precise by the quantum channel merging theorem stated as Theorem 10 in Sec. III. In the previous subsection we saw that for a replacer channel with pure output state, the entropy of a channel is zero. This replacer channel is also entanglement breaking. However, the identity channel is the least noisy channel, and therefore should have the least entropy possible. Indeed, as stated above, for the identity channel, our entropy function equals the negative of the logarithm of the dimension (which is the smallest possible value).

*Corollary 8.* For any quantum channel $\mathcal{N}_{A \to B}$,

$$H(\mathcal{N}) \geqslant -\log |A|, \tag{52}$$

with equality if and only if $\mathcal{N}_{A \to B}$ is an isometry.

*Proof.* The proof that $-\log |A|$ is the smallest possible value follows trivially from the well known bound $D(\mathcal{N}_{A \to B}(\psi_{RA}) \| \mathcal{R}_{A \to B}(\psi_{RA})) \leqslant \log |AB|$, or alternatively from (40) and (49) and dropping the nonnegative term $H(E)_{\tau}$. Now, from the proposition above,

$$H(\mathcal{N}) = \inf_{\psi_{RA}} H(B|R)_{\mathcal{N}(\psi)}. \tag{53}$$

Therefore, the smallest possible value $-\log |A|$ is achieved if and only if $\mathcal{N}_{A \to B}(\psi_{RA})$ is the maximally entangled state (recall $|R| = |A|$). This is only possible if $|B| \geqslant |A|$ and $\mathcal{N}$ is an isometry. ∎

## III. QUANTUM CHANNEL MERGING

Given a bipartite state $\rho_{BE}$, the goal of quantum state merging is for Bob to use forward classical communication to Eve, as well as entanglement, to merge his share of the state with Eve's share [20,21]. The optimal rate of entanglement consumed is equal to the conditional entropy $H(B|E)_{\rho}$. Alternatively, the optimal rate of entanglement *gained* is equal to the conditional entropy $H(B|R)_{\psi}$, where $\psi_{RBE}$ is a purification of $\rho_{BE}$.

In this section, we define a task, called *quantum channel merging*, that can be considered a dynamical counterpart of state merging. Given a quantum channel $\mathcal{N}_{A \to B}$ with isometric extension $\mathcal{U}^{\mathcal{N}}_{A \to BE}$, the goal is for Bob to merge his share of the channel with Eve's share. We find here that the entanglement cost of the protocol is equal to $\sup_{\rho_A} H(B|E)_{\omega}$, where $\omega_{BE} = \mathcal{U}^{\mathcal{N}}_{A \to BE}(\rho_A)$. Equivalently, by employing Eq. (50), the entanglement gain of the protocol is equal to $H(\mathcal{N})$, the entropy of the channel $\mathcal{N}_{A \to B}$. Thus, the main result of this section is a direct operational interpretation of the entropy of a channel as the entanglement gain in quantum channel merging. We note here that the completely bounded entropy of Ref. [19] (i.e., entropy of a channel) was recently interpreted in terms of a cryptographic task in Ref. [33].

We now specify the quantum channel merging information-processing task in detail. Let $\mathcal{N}_{A \to B}$ be a quantum channel, and suppose that $\mathcal{U}^{\mathcal{N}}_{A \to BE}$ is an isometric channel extending it. Here, we think of the isometric channel $\mathcal{U}^{\mathcal{N}}_{A \to BE}$ as a broadcast channel (three-terminal device), which connects a source to the receivers Bob and Eve. Suppose that a source generates an arbitrary state $\psi_{RA^n}$ and then sends the $A$ systems through the isometric channel $(\mathcal{U}^{\mathcal{N}}_{A \to BE})^{\otimes n}$, which transmits the $B$ systems to Bob and the $E$ systems to Eve. The goal is for Bob to use free one-way local operations and classical communication (one-way LOCC) to generate ebits at the maximum rate possible, while also merging his systems with Eve's.

Let $n \in \mathbb{N}$, $M \in \mathbb{Q}$, and $\varepsilon \in [0, 1]$. An $(n, M, \varepsilon)$ protocol for this task consists of a one-way LOCC channel $\mathcal{P}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}^n_E \widetilde{E}^n \overline{B}_1 \overline{E}_1}$ such that

$$\sup_{\psi_{RA^n}} \frac{1}{2} \left\| \left[ \mathrm{id}^{\otimes n}_{BE \to \widetilde{B}_E \widetilde{E}} \circ (\mathcal{U}^{\mathcal{N}}_{A \to BE})^{\otimes n} \right] (\psi_{RA^n}) \otimes \Phi^L_{\overline{B}_1 \overline{E}_1} \right.$$

$$\left. - \mathcal{P}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}^n_E \widetilde{E}^n \overline{B}_1 \overline{E}_1} \left( \left[ (\mathcal{U}^{\mathcal{N}}_{A \to BE})^{\otimes n} (\psi_{RA^n}) \right] \otimes \Phi^K_{\overline{B}_0 \overline{E}_0} \right) \right\|_1$$

$$\leqslant \varepsilon, \tag{54}$$

FIG. 1. The goal of quantum channel merging is for Bob to merge his share of the channel with Eve's. Given a channel $\mathcal{N}_{A\to B}$, let $\mathcal{V}^{\mathcal{N}} \equiv (\mathcal{U}^{\mathcal{N}}_{A\to BE})^{\otimes n}$, where $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ is an isometric channel extending $\mathcal{N}_{A\to B}$. By consuming a maximally entangled state $\Phi^K$ of Schmidt rank $K$ and applying a one-way LOCC protocol $\mathcal{P}$, Bob and Eve can distill a maximally entangled state $\Phi^L$ of Schmidt rank $L$ and transfer Bob's systems $B^n$ to Eve, in such a way that any third party having access to the inputs $A^n$ and the outputs $B^n$ and $E^n$ would not be able to distinguish the difference between the ideal situation on the left and the simulation on the right. Theorem 10 states that the optimal asymptotic rate of entanglement gain is equal to the entropy of the channel $\mathcal{N}$.

where $\Phi^K_{\bar{B}_0\bar{E}_0}$ and $\Phi^L_{\bar{B}_1\bar{E}_1}$ are maximally entangled states of Schmidt rank $K$ and $L$, respectively, and $M = L/K$, so that the number of ebits gained in the protocol is equal to $\log_2 M = \log_2 L - \log_2 K$. Figure 1 depicts the task of quantum channel merging.

*Definition 9 (Q. channel merging capacity).* A rate $R$ is achievable for quantum channel merging if for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large $n$, there exists an $(n, 2^{n[R-\delta]}, \varepsilon)$ protocol of the above form. The quantum channel merging capacity $C_M(\mathcal{N})$ is defined to be the supremum of all achievable rates:

$$C_M(\mathcal{N}) \equiv \sup \{R \mid R \text{ is achievable for channel merging on } \mathcal{N}\}. \tag{55}$$

*Theorem 10.* The quantum channel merging capacity of a channel $\mathcal{N}$ is equal to its entropy:

$$C_M(\mathcal{N}) = H(\mathcal{N}). \tag{56}$$

We provide a detailed proof of Theorem 10 in Appendix A.

## IV. EXAMPLES

In this section, we provide formulas for the entropy of several fundamental channel models, including erasure channels, dephasing channels, depolarizing channels, and Werner–Holevo channels. We also define the energy-constrained and unconstrained entropies of a channel and determine formulas for them for common bosonic channel models, including thermal, amplifier, and additive-noise channels.

### A. Finite-dimensional channels

A first observation to make is that, for any finite-dimensional channel, it is an "easy" optimization task to calculate its entropy. This is a consequence of the identity $H(\mathcal{N}) = -\sup_{\rho_A} H(B|E)_{\mathcal{U}(\rho)}$ from Proposition 6 and the concavity of conditional entropy [34,35] (in this context, see also [[26], Eq. (3.19)]). Thus, one can exploit numerical optimizations to calculate it [36,37].

For channels with symmetry, it can be much easier to evaluate a channel's entropy, following from some observations from, e.g., Ref. [[38], Sec. 6]. Let us begin by recalling the notion of a covariant channel $\mathcal{N}_{A\to B}$ [39]. For a group $G$ with unitary channel representations $\{\mathcal{U}^g_A\}_g$ and $\{\mathcal{V}^g_B\}_g$ acting on the input system $A$ and output system $B$ of the channel $\mathcal{N}_{A\to B}$, the channel $\mathcal{N}_{A\to B}$ is covariant with respect to the

group $G$ if the following equality holds for all $g \in G$:

$$\mathcal{N}_{A\to B} \circ \mathcal{U}^g_A = \mathcal{V}^g_B \circ \mathcal{N}_{A\to B}. \tag{57}$$

If the averaging channel is such that $\frac{1}{|G|}\sum_g \mathcal{U}^g_A(X) = \text{Tr}[X]I/|A|$ (implementing a unitary one-design), then we simply say that the channel $\mathcal{N}_{A\to B}$ is covariant. It turns out that the entropy of a channel is simple to calculate for covariant channels, with the optimal $\psi_{RA}$ in Eq. (35) being the maximally entangled state, or equivalently, the optimal $\rho_A$ in $-\sup_{\rho_A} H(B|E)_{\mathcal{U}(\rho)}$ being the maximally mixed state.

*Proposition 11.* Let $\mathcal{N}_{A\to B}$ be a quantum channel that is covariant with respect to a group $G$, in the sense of Eq. (57), and let $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending it. Then it suffices to perform the optimization for the entropy of a channel over states that respect the symmetry of the channel:

$$H(\mathcal{N}) = -\sup_{\rho_A = \mathcal{S}_A(\rho_A)} H(B|E)_{\mathcal{U}(\rho)}, \tag{58}$$

where the symmetrizing channel $\mathcal{S}_A = \frac{1}{|G|}\sum_{g\in G} \mathcal{U}^g_A$. Thus, if a channel is covariant, then $H(\mathcal{N}) = -H(B|E)_{\mathcal{U}(\pi)}$; i.e., the optimal state $\rho_A$ is the maximally mixed state $\pi_A$.

*Proof.* First recall from Proposition 6 that $H(\mathcal{N}) = -\sup_{\rho_A} H(B|E)_{\mathcal{U}(\rho)}$. Let $\rho_A$ be an arbitrary state. If a channel $\mathcal{N}_{A\to B}$ is covariant as in Eq. (57), then it is known that there exists a unitary channel $\mathcal{W}^g_E$ such that [14,40]

$$\mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{U}^g_A = (\mathcal{V}^g_B \otimes \mathcal{W}^g_E) \circ \mathcal{U}^{\mathcal{N}}_{A\to BE}. \tag{59}$$

See also Ref. [41], Appendix A for a simple proof. Then we find that

$$H(B|E)_{\mathcal{U}(\rho)} = H(B|E)_{(\mathcal{V}^g\otimes\mathcal{W}^g)\mathcal{U}(\rho)} \tag{60}$$

$$= \frac{1}{|G|}\sum_{g\in G} H(B|E)_{(\mathcal{V}^g\otimes\mathcal{W}^g)\mathcal{U}(\rho)} \tag{61}$$

$$= \frac{1}{|G|}\sum_{g\in G} H(B|E)_{(\mathcal{U}\circ\mathcal{U}^g)(\rho)} \tag{62}$$

$$\leqslant H(B|E)_{(\mathcal{U}\circ\mathcal{S})(\rho)}. \tag{63}$$

The first equality follows from invariance of conditional entropy under the action of a local unitary (the equality holds for all $g \in G$). The third equality follows from channel covariance. The inequality follows from concavity of conditional entropy [34,35]. ∎

A simple example of a channel that is covariant is the quantum erasure channel, defined as [42]

$$\mathcal{E}^p(\rho) \equiv (1 - p)\rho + p|e\rangle\langle e|, \tag{64}$$
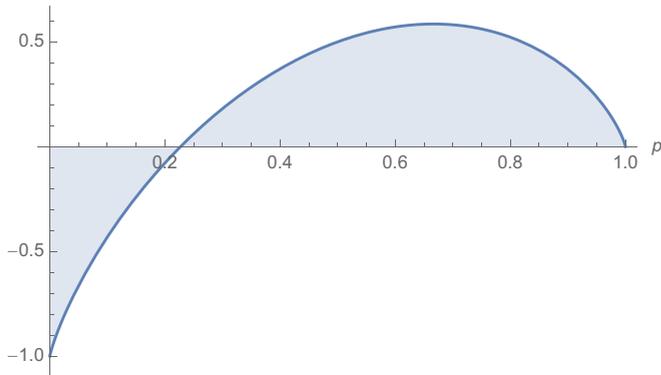
FIG. 2. Entropy of the qubit erasure channel as a function of the erasure probability $p$. When $p = 0$, the erasure channel is the identity qubit channel and thus takes on its smallest value. When $p = 1$, the erasure channel deterministically replaces the input with the pure state $|e\rangle\langle e|$ and thus has entropy equal to zero.

where $\rho$ is a $d$-dimensional input state, $p \in [0, 1]$ is the erasure probability, and $|e\rangle\langle e|$ is a pure erasure state orthogonal to any input state, so that the output state has $d + 1$ dimensions. A $d$-dimensional dephasing channel has the following action:

$$\mathcal{D}^{\mathbf{p}}(\rho) = \sum_{\ell=0}^{d-1} p_\ell Z^\ell \rho Z^{\ell\dagger}, \tag{65}$$

where $\mathbf{p}$ is a vector containing the probabilities $p_\ell$ and $Z$ has the following action on the computational basis $Z|x\rangle = e^{2\pi i x/d}|x\rangle$. This channel is covariant with respect to the Heisenberg–Weyl group of unitaries, which is well known to form a unitary one-design. A particular kind of Werner–Holevo channel performs the following transformation on a $d$-dimensional input state $\rho$ [43]:

$$\mathcal{W}^{(d)}(\rho) \equiv \frac{1}{d-1}[\text{Tr}\{\rho\}I - T(\rho)], \tag{66}$$

where $d \geqslant 2$ and $T$ denotes the transpose map $T(\cdot) = \sum_{i,j} |i\rangle\langle j|(\cdot)|i\rangle\langle j|$. As observed in Ref. [[43], Sec. II], this channel is covariant. The $d$-dimensional depolarizing channel is a common model of noise in quantum information, transmitting the input state with probability $1 - p \in [0, 1]$ and replacing it with the maximally mixed state $\pi \equiv \frac{I}{d}$ with probability $p$:

$$\Delta^p(\rho) = (1 - p)\rho + p\pi. \tag{67}$$

By applying Proposition 11 and evaluating the resulting entropy $-H(B|E)$ for each of the above channels when the maximally mixed state $\pi$ is input, we arrive at the following formulas:

$$H(\mathcal{E}^p) = h_2(p) + (p - 1)\log_2 d, \tag{68}$$

$$H(\mathcal{D}^{\mathbf{p}}) = H(\mathbf{p}) - \log_2 d, \tag{69}$$

$$H(\mathcal{W}^{(d)}) = \log_2[(d - 1)/2], \tag{70}$$

$$H(\Delta^p) = -\left(1 - p + \frac{p}{d^2}\right)\log_2\left(1 - p + \frac{p}{d^2}\right)$$
$$- (d^2 - 1)\frac{p}{d^2}\log_2\frac{p}{d^2} - \log_2 d, \tag{71}$$
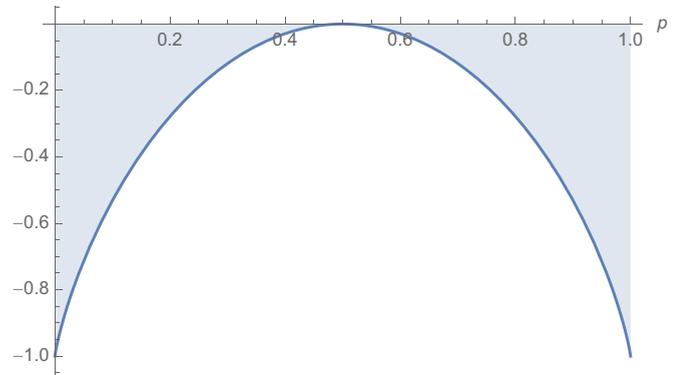
FIG. 3. Entropy of the qubit dephasing channel as a function of the dephasing probability $p$. The optimal input state is the maximally entangled state, so that the channel entropy is evaluated on the Choi state of the channel. When $p = 0$, the dephasing channel is the identity qubit channel and thus takes on its smallest value. When $p = 1/2$, the dephasing channel is a classical channel, so that its Choi state is maximally classically correlated. For such a state, $D(\mathcal{N}\|\mathcal{R}) = 1$ so that the channel entropy is equal to zero.

where $H(\mathbf{p})$ is the Shannon entropy of the probability vector $\mathbf{p}$. These formulas are plotted and interpreted in Figs. 2–4.

### B. Energy-constrained entropy of a channel

We can define the energy-constrained entropy of a channel for infinite-dimensional systems, by employing the identity in Proposition 6 and the definition of conditional entropy from Ref. [44].

To review the definition from Ref. [44], recall that the quantum entropy of a state $\rho$ acting on a separable Hilbert space is defined as

$$H(\rho) \equiv \text{Tr}\{\eta(\rho)\}, \tag{72}$$

where $\eta(x) = -x \log_2 x$ if $x > 0$ and $\eta(0) = 0$. The trace in the above equation can be taken with respect to any countable orthonormal basis of $\mathcal{H}$ [[45], Definition 2]. The quantum entropy is a nonnegative, concave, lower semicontinuous
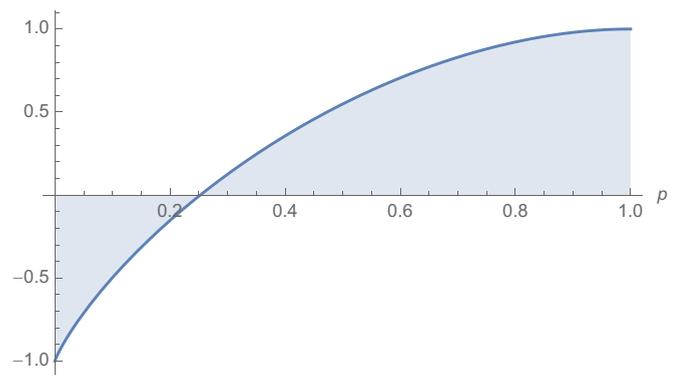


FIG. 4. Entropy of the qubit depolarizing channel as a function of the depolarizing probability $p$. When $p = 0$, the depolarizing channel is the identity qubit channel and thus takes on its smallest value. When $p = 1$, the depolarizing channel replaces the channel input with the maximally mixed state and thus takes on its maximal value.

function on $\mathcal{D}(\mathcal{H})$ [46]. It is also not necessarily finite (see, e.g., Ref. [47]). When $\rho_A$ is assigned to a system $A$, we write $H(A)_\rho \equiv H(\rho_A)$. Recall that the relative entropy of two states $\rho$ and $\sigma$ acting on a separable Hilbert space is given by [48,49]

$$
\begin{aligned}
&D(\rho\|\sigma) \\
&\equiv [\ln 2]^{-1} \sum_{i,j} |\langle\phi_i|\psi_j\rangle|^2 [p(i)\ln\left(\frac{p(i)}{q(j)}\right) + q(j) - p(i)],
\end{aligned}
\tag{73}
$$

where $\rho = \sum_i p(i)|\phi_i\rangle\langle\phi_i|$ and $\sigma = \sum_j q(j)|\psi_j\rangle\langle\psi_j|$ are spectral decompositions of $\rho$ and $\sigma$ with $\{|\phi_i\rangle\}_i$ and $\{|\psi_j\rangle\}_j$ orthonormal bases. The prefactor $[\ln 2]^{-1}$ is there to ensure that the units of the quantum relative entropy are bits. For a bipartite state $\rho_{AB}$, the mutual information is defined as

$$
I(A;B)_\rho \equiv D(\rho_{AB}\|\rho_A \otimes \rho_B). \tag{74}
$$

Finally, for a bipartite state $\rho_{AB}$ such that $H(A)_\rho < \infty$, the conditional entropy is defined as [44]

$$
H(A|B)_\rho \equiv H(A)_\rho - I(A;B)_\rho, \tag{75}
$$

and it is known that $H(A|B)_\rho \in [-H(A)_\rho, H(A)_\rho]$ [44].

A Gibbs observable is a positive semidefinite operator $G$ acting on a separable Hilbert space such that $\mathrm{Tr}\{e^{-\beta G}\} < \infty$ for all $\beta > 0$ [14,50,51]. This condition for a Gibbs observable means that there is always a well defined thermal state.

Finally, we say that a quantum channel $\mathcal{N}_{A\rightarrow B}$ obeys the finite-output entropy condition [14,50,51] with respect to a Gibbs observable $G$ if for all $P \geqslant 0$, the following inequality holds:

$$
\sup_{\rho_A : \mathrm{Tr}\{G\rho_A\}\leqslant P} H(B)_{\mathcal{N}(\rho)} < \infty. \tag{76}
$$

We now define the energy-constrained and unconstrained channel entropy as follows:

*Definition 12.* Let $\mathcal{N}_{A\rightarrow B}$ be a quantum channel that satisfies the finite-output entropy condition with respect to a Gibbs observable $G$. For $P \geqslant 0$, the energy-constrained entropy of $\mathcal{N}_{A\rightarrow B}$ is defined as

$$
H(\mathcal{N}, G, P) \equiv \inf_{\psi_{RA} : \mathrm{Tr}\{G\psi_A\}\leqslant P} H(B|R)_\omega, \tag{77}
$$

where $\omega_{RB} \equiv \mathcal{N}_{A\rightarrow B}(\psi_{RA})$ and the optimization is with respect to all pure bipartite states with system $R$ isomorphic to system $A$. The unconstrained entropy of $\mathcal{N}_{A\rightarrow B}$ with respect to $G$ is then defined as

$$
H(\mathcal{N}, G) \equiv \inf_{P\geqslant 0} H(\mathcal{N}, G, P). \tag{78}
$$

### C. Bosonic Gaussian channels

In this section, we evaluate the energy-constrained and unconstrained entropy of several important bosonic Gaussian channels [14,52], including the thermal, amplifier, and additive-noise channels. Here we take the Gibbs observable to be the photon number operator $\hat{n}$ [14,52], and we note that each of these channels satisfies the finite-output entropy condition mentioned above. From a practical perspective, we should be most interested in these particular single-mode bosonic Gaussian channels, as these are of the greatest interest in applications, as stressed in Ref. [[14], Sec. 12.6.3] and Ref.

[[53], Sec. 3.5]. Each of these are defined, respectively, by the following Heisenberg input-output relations:

$$
\hat{b} = \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{e}, \tag{79}
$$

$$
\hat{b} = \sqrt{G}\hat{a} + \sqrt{G-1}\hat{e}^\dagger, \tag{80}
$$

$$
\hat{b} = \hat{a} + (x + ip)/\sqrt{2}, \tag{81}
$$

where $\hat{a}$, $\hat{b}$, and $\hat{e}$ are the field-mode annihilation operators for the sender's input, the receiver's output, and the environment's input of these channels, respectively.

The channel in Eq. (79) is a thermalizing channel, in which the environmental mode is prepared in a thermal state $\theta(N_B)$ of mean photon number $N_B \geqslant 0$, defined as

$$
\theta(N_B) \equiv \frac{1}{N_B + 1}\sum_{n=0}^{\infty}\left(\frac{N_B}{N_B+1}\right)^n|n\rangle\langle n|, \tag{82}
$$

where $\{|n\rangle\}_{n=0}^{\infty}$ is the orthonormal, photonic number-state basis. When $N_B = 0$, $\theta(N_B)$ reduces to the vacuum state, in which case the resulting channel in Eq. (79) is called the pure-loss channel. The parameter $\eta \in (0, 1)$ is the transmissivity of the channel, representing the average fraction of photons making it from the input to the output of the channel. Let $\mathcal{L}_{\eta,N_B}$ denote this channel.

The channel in Eq. (80) is an amplifier channel, and the parameter $G > 1$ is its gain. For this channel, the environment is prepared in the thermal state $\theta(N_B)$. If $N_B = 0$, then the amplifier channel is called the pure-amplifier channel. Let $\mathcal{A}_{G,N_B}$ denote this channel.

Finally, the channel in Eq. (81) is an additive-noise channel, representing a quantum generalization of the classical additive white Gaussian noise channel. In Eq. (81), $x$ and $p$ are zero-mean, independent Gaussian random variables each having variance $\xi \geqslant 0$. Let $\mathcal{T}_\xi$ denote this channel. Note that the additive-noise channel arises from the thermal channel in the limit $\eta \rightarrow 1$, $N_B \rightarrow \infty$, but with $(1-\eta)N_B \rightarrow \xi$ [54].

Kraus representations for the channels in Eqs. (79)–(81) are available in Ref. [55], which can be helpful for further understanding their action on input quantum states.

All of the above channels are phase-insensitive or phase-covariant Gaussian channels [14,52]. Let $N_S \geqslant 0$. Since the function $\sup_{\rho : \mathrm{Tr}\{\hat{n}\rho\}\leqslant N_S} H(B|E)_{\mathcal{U}(\rho)}$ we are evaluating is concave in the input and invariant under local unitaries, Ref. [[56], Remark 22] applies, implying that the optimal input state for the entropies of these channels is the bosonic thermal state $\theta(N_S)$. We then find by employing well known entropy formulas from Refs. [57,58] (see also Ref. [59] in this context) that

$$
\begin{aligned}
&H(\mathcal{L}_{\eta,N_B}, \hat{n}, N_S) \\
&= g_2\{[D_1 + (1-\eta)(N_S - N_B) - 1]/2\} \\
&\quad + g_2\{[D_1 - (1-\eta)(N_S - N_B) - 1]/2\} - g_2(N_S),
\end{aligned}
\tag{83}
$$

$$
\begin{aligned}
&H(\mathcal{A}_{G,N_B}, \hat{n}, N_S) \\
&= g_2\{[D_2 + (G-1)(N_S + N_B + 1) - 1]/2\} \\
&\quad + g_2\{[D_2 - (G-1)(N_S + N_B + 1) - 1]/2\} - g_2(N_S),
\end{aligned}
\tag{84}
$$
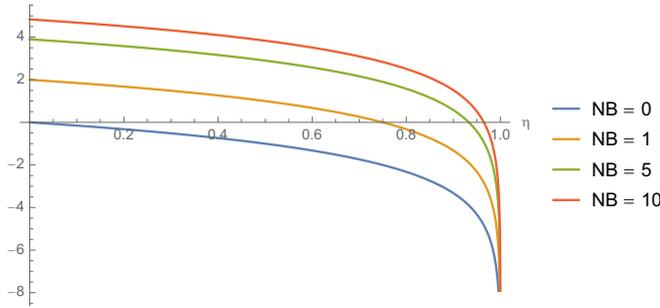
FIG. 5. Entropy of the bosonic thermal channel as a function of the transmissivity $\eta$. When $\eta = 1$, the thermal channel is the identity channel and thus takes on its smallest value of $-\infty$, regardless of the value of $N_B$. When $\eta = 0$ and $N_B = 0$, the thermal channel deterministically replaces the input with the pure vacuum state $|0\rangle\langle 0|$ and thus has entropy equal to zero. As the thermal noise $N_B$ increases, the entropy of the thermal channel increases.

$$H(\mathcal{T}_\xi, \hat{n}, N_S) = g_2\{[D_3 - (\xi + 1)]/2\} + g_2\{[D_3 + \xi - 1]/2\} - g_2(N_S), \quad (85)$$

where $g_2$ is the bosonic entropy function defined in Eq. (210) and

$$D_1 \equiv \sqrt{[(\eta + 1)N_S + (1 - \eta)N_B + 1]^2 - 4\eta N_S(N_S + 1)}, \quad (86)$$

$$D_2$$
$$\equiv \sqrt{[(G + 1)N_S + (G - 1)(N_B + 1) + 1]^2 - 4GN_S(N_S + 1)}, \quad (87)$$

$$D_3 \equiv \sqrt{(\xi + 1)^2 + 4\xi N_S}. \quad (88)$$

Note that we arrived at the formula for $H(\mathcal{T}_\xi, \hat{n}, N_S)$ by considering the limit discussed above. Furthermore, by the same reasoning as given in Ref. [[56], Sec. 6], these functions are decreasing with increasing $N_S$, and so we find that

$$H(\mathcal{L}_{\eta, N_B}, \hat{n}) = \inf_{N_S \geqslant 0} H(\mathcal{L}_{\eta, N_B}, \hat{n}, N_S) \quad (89)$$

$$= \lim_{N_S \to \infty} H(\mathcal{L}_{\eta, N_B}, \hat{n}, N_S), \quad (90)$$

$$H(\mathcal{A}_{G, N_B}, \hat{n}) = \inf_{N_S \geqslant 0} H(\mathcal{A}_{G, N_B}, \hat{n}, N_S) \quad (91)$$

$$= \lim_{N_S \to \infty} H(\mathcal{A}_{G, N_B}, \hat{n}, N_S), \quad (92)$$

$$H(\mathcal{T}_\xi, \hat{n}) = \inf_{N_S \geqslant 0} H(\mathcal{T}_\xi, \hat{n}, N_S) \quad (93)$$

$$= \lim_{N_S \to \infty} H(\mathcal{T}_\xi, \hat{n}, N_S), \quad (94)$$

which leads to the following formulas for the unconstrained entropies of the channels:

$$H(\mathcal{L}_{\eta, N_B}, \hat{n}) = \log_2(1 - \eta) + g_2(N_B), \quad (95)$$

$$H(\mathcal{A}_{G, N_B}, \hat{n}) = \log_2(G - 1) + g_2(N_B), \quad (96)$$

$$H(\mathcal{T}_\xi, \hat{n}) = \log_2(\xi) + \frac{1}{\ln 2}. \quad (97)$$

These formulas are plotted and interpreted in Figs. 5–7. A Mathematica file is available with the arXiv posting of this

paper to automate these calculations, but we note here that the expansion $g_2(x) = \log_2(x) + 1/\ln 2 + O(1/x)$ is helpful for this purpose. We also note that the formulas in Eqs. (95) and (96) were presented in Ref. [[60], Eq. (2)] and the formula in Eq. (97) was presented in Ref. [[57], Sec. V].

## V. RÉNYI ENTROPY OF A QUANTUM CHANNEL

Generalizing the von Neumann entropy of a quantum state, the Rényi entropy finds extensive application in physics and information theory. Given a pure bipartite state, the Rényi entropy of the reduced state is an entanglement measure, which finds application in conformal field theory [61], holography [62], and black holes [63]. The full range of values of the Rényi entropy is known as the entanglement spectrum. The Rényi entropy finds information-theoretic meaning in the expression for the error exponent of entanglement concentration [64] and quantum data compression [65], indicating the exponential rate at which errors in these settings decay to zero. As such, it is worthwhile to understand the Rényi entropy of a channel as a generalization of the Rényi entropy of a state.

In this section, we define the Rényi entropy of a channel, following the same approach discussed in the introduction.



FIG. 6. Entropy of the bosonic amplifier channel as a function of the amplifier gain $G$. When $G = 1$, the amplifier channel is the identity channel and thus takes on its smallest value of $-\infty$, regardless of the value of $N_B$. As the amplifier gain $G$ and the thermal noise $N_B$ increase, the entropy of the thermal channel increases.
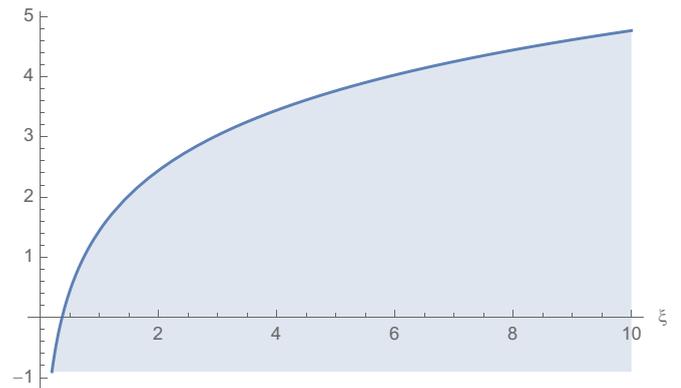


FIG. 7. Entropy of the bosonic additive-noise channel as a function of the noise parameter $\xi$. When $\xi = 0$, the additive-noise channel is the identity channel and thus takes on its smallest value of $-\infty$. As $\xi$ increases, the entropy of the additive-noise channel increases logarithmically.

That is, we first write the Rényi entropy of a state as the difference of the number of physical qubits and the Rényi relative entropy of the state to the maximally mixed state. Then we define the Rényi entropy of a channel in the same way as in Definition 1, but replacing the channel relative entropy with the sandwiched Rényi channel relative entropy from Ref. [15].

The Rényi entropy of a quantum state $\rho_A$ of system $A$ is defined for $\alpha \in (0, 1) \cup (1, \infty)$ as

$$H_\alpha(A)_\rho \equiv \frac{1}{1-\alpha} \log_2 \text{Tr} \left\{ \rho_A^\alpha \right\} \qquad (98)$$

$$= \frac{1}{1-\alpha} \log_2 \|\rho_A\|_\alpha^\alpha, \qquad (99)$$

where $\|X\|_\alpha \equiv [\text{Tr}\{|X|^\alpha\}]^{1/\alpha}$ and $|X| \equiv \sqrt{X^\dagger X}$ for an operator $X$. The Rényi relative entropy of quantum states can be defined in two different ways, known as the Petz–Rényi relative entropy [66,67] and the sandwiched Rényi relative entropy [68,69]. The sandwiched Rényi relative entropy is defined for $\alpha \in (0, 1) \cup (1, \infty)$, a state $\rho$, and a positive semidefinite operator $\sigma$ as

$$D_\alpha(\rho\|\sigma) \equiv \frac{1}{\alpha-1} \log_2 \text{Tr}\{(\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha})^\alpha\}, \quad (100)$$

whenever either $\alpha \in (0, 1)$ or $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ and $\alpha > 1$. Otherwise, it is set to $+\infty$. The sandwiched Rényi relative entropy obeys the data processing inequality for $\rho$ and $\sigma$ as above, a quantum channel $\mathcal{N}$, and $\alpha \in [1/2, 1) \cup (1, \infty)$ [70] (see also Refs. [68,69,71–73]):

$$D_\alpha(\rho\|\sigma) \geqslant D_\alpha[\mathcal{N}(\rho)\|\mathcal{N}(\sigma)]. \qquad (101)$$

It converges to the quantum relative entropy in the limit $\alpha \to 1$ [68,69]:

$$\lim_{\alpha \to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma). \qquad (102)$$

By inspection, the Rényi entropy of a state can be written as

$$H_\alpha(A)_\rho = \log_2 |A| - D_\alpha(\rho_A\|\pi_A). \qquad (103)$$

The sandwiched Rényi channel divergence of channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ is defined for $\alpha \in [1/2, 1) \cup (1, \infty)$ as [15]

$$D_\alpha(\mathcal{N}\|\mathcal{M}) \equiv \sup_{\rho_{RA}} D_\alpha(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})), \quad (104)$$

where the optimization is with respect to bipartite states $\rho_{RA}$ of a reference system $R$ of arbitrary size and the channel input system $A$. Due to state purification, the data-processing inequality in Eq. (101), and the Schmidt decomposition theorem, it suffices to optimize over states $\rho_{RA}$ that are pure and such that system $R$ is isomorphic to system $A$.

We now define the Rényi entropy of a quantum channel as follows:

*Definition 13 (Rényi entropy of a q. channel).* Let $\mathcal{N}_{A\to B}$ be a quantum channel. For $\alpha \in [1/2, 1) \cup (1, \infty)$, the Rényi entropy of the channel $\mathcal{N}$ is defined as

$$H_\alpha(\mathcal{N}) \equiv \log_2 |B| - D_\alpha(\mathcal{N}\|\mathcal{R}), \qquad (105)$$

where $\mathcal{R}_{A\to B}$ is the completely randomizing channel from Eq. (5).

We remark here that $D_\alpha(\mathcal{N}\|\mathcal{R})$, for $\alpha > 1$, has an operational interpretation as the strong converse exponent for discrimination of the channel $\mathcal{N}_{A\to B}$ from the completely randomizing channel $\mathcal{R}_{A\to B}$, when considering any possible channel discrimination strategy [15].

One could alternatively define a different Rényi entropy of a channel according to the above recipe, but in terms of the Petz–Rényi relative entropy. However, it is unclear whether the additivity property is generally satisfied for the resulting Rényi entropy of a channel, and so we do not consider it further here, instead leaving this question open.

### A. Properties of the Rényi entropy of a quantum channel

The Rényi entropy of a channel obeys the three desired axioms from Ref. [18], and in fact, the proofs are essentially the same as the previous ones, but instead using properties of the sandwiched Rényi relative entropy.

*Proposition 14.* Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\Theta$ be a uniformity preserving superchannel as defined above. Then for all $[1/2, 1) \cup (1, \infty)$,

$$H_\alpha(\Theta(\mathcal{N})) \geqslant H_\alpha(\mathcal{N}). \qquad (106)$$

*Proof.* We follow the same steps as in Eqs. (12)–(16), but making the substitutions $H \to H_\alpha$ and $D \to D_\alpha$. Also, we use the fact that, for $[1/2, 1) \cup (1, \infty)$, the sandwiched Rényi channel divergence does not increase under the action of a superchannel, as shown in Ref. [18]. ∎

*Proposition 15 (Additivity).* Let $\mathcal{N}$ and $\mathcal{M}$ be quantum channels. Then the channel Rényi entropy is additive in the following sense for $\alpha \in (1, \infty)$:

$$H_\alpha(\mathcal{N} \otimes \mathcal{M}) = H_\alpha(\mathcal{N}) + H_\alpha(\mathcal{M}). \qquad (107)$$

*Proof.* The proof here follows the same approach given in the proof of Proposition 4, making the substitutions $H \to H_\alpha$ and $D \to D_\alpha$. The steps in Eqs. (24)–(28) follow from the same steps given in the proof of Proposition 41 of Ref. [28], which in turn rely upon the additivity result from Ref. [19]. See also Ref. [15] in this context. ∎

*Proposition 16 (Reduction to states).* Let the channel $\mathcal{N}_{A\to B}$ be a replacer channel, defined such that $\mathcal{N}_{A\to B}(\rho_A) = \sigma_B$ for all states $\rho_A$ and some state $\sigma_B$. Then the following equality holds for all $\alpha \in (0, 1) \cup (1, \infty)$:

$$H_\alpha(\mathcal{N}) = H_\alpha(B)_\sigma. \qquad (108)$$

*Proof.* The proof is essentially the same as the proof of Proposition 5, making the substitutions $H \to H_\alpha$ and $D \to D_\alpha$. ∎

We can then conclude that the Rényi entropy of a channel satisfies the normalization axiom from the fact that $H_\alpha(B)_\sigma = \log |B|$ if $\sigma_B$ is maximally mixed and $H(B)_\sigma = 0$ if $\sigma_B$ is pure.

### B. Alternate representations for the Rényi entropy of a quantum channel

Just as we showed in Sec. II B that there are alternate representations for the entropy of a quantum channel, here we do the same for the Rényi entropy of a channel. We define the conditional Rényi entropy of a bipartite state $\rho_{AB}$ as

$$H_\alpha(A|B)_{\rho|\rho} \equiv -D_\alpha(\rho_{AB}\|I_A \otimes \rho_B), \qquad (109)$$

where $D_\alpha(\rho\|\sigma)$ is the sandwiched Rényi relative entropy from Eq. (100). The conditional Petz–Rényi entropy of a bipartite state $\rho_{AB}$ is defined as

$$\overline{H}_\alpha(A|B)_\rho \equiv -\inf_{\sigma_B} \overline{D}_\alpha(\rho_{AB}\|I_A \otimes \sigma_B), \qquad (110)$$

where the Petz–Rényi relative entropy $\overline{D}_\alpha(\rho\|\sigma)$ is defined for $\alpha \in (0,1) \cup (1,\infty)$ as [66,67]

$$\overline{D}_\alpha(\rho\|\sigma) \equiv \frac{1}{\alpha-1} \log_2 \text{Tr}\{\rho^\alpha \sigma^{1-\alpha}\}, \qquad (111)$$

whenever either $\alpha \in (0,1)$ or $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ and $\alpha > 1$. Otherwise, it is set to $+\infty$. The Petz–Rényi relative entropy obeys the data processing inequality for $\rho$ and $\sigma$ as above, a quantum channel $\mathcal{N}$, and $\alpha \in (0,1) \cup (1,2]$ [66,67]:

$$\overline{D}_\alpha(\rho\|\sigma) \geqslant \overline{D}_\alpha[\mathcal{N}(\rho)\|\mathcal{N}(\sigma)]. \qquad (112)$$

The completely bounded $1 \to \alpha$ norm of a quantum channel is defined for $\alpha \geqslant 1$ as [19]

$$\|\mathcal{N}_{A\to B}\|_{\text{CB},1\to\alpha} \equiv \sup_{\rho_R} \left\|\rho_R^{1/2\alpha} \mathcal{N}_{A\to B}(\Gamma_{RA}) \rho_R^{1/2\alpha}\right\|_\alpha, \qquad (113)$$

where the optimization is with respect to a density operator $\rho_R$ and $\Gamma_{RA} \equiv |\Gamma\rangle\langle\Gamma|_{RA}$ denotes the projection onto the following maximally entangled vector:

$$|\Gamma\rangle_{RA} \equiv \sum_i |i\rangle_R |i\rangle_A, \qquad (114)$$

where $\{|i\rangle_R\}_i$ and $\{|i\rangle_A\}_i$ are orthonormal bases and system $R$ is isomorphic to the channel input system $A$.

We can now state the alternate representations for the Rényi entropy of a channel:

*Proposition 17.* Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending it, as in Eq. (36). Then for $\alpha \in (0,1) \cup (1,\infty)$,

$$H_\alpha(\mathcal{N}) = \inf_{\psi_{RA}} H_\alpha(B|R)_{\omega|\omega} = -\sup_{\rho_A} \overline{H}_\beta(B|E)_\tau, \qquad (115)$$

where the first optimization is with respect to bipartite pure states with system $R$ isomorphic to system $A$, $\omega_{RB} \equiv \mathcal{N}_{A\to B}(\psi_{RA})$, $\tau_{BE} \equiv \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho_A)$, and $\beta = 1/\alpha$. For $\alpha \in [1/2, 1) \cup (1,\infty)$,

$$|H_\alpha(\mathcal{N})| \leqslant \log_2 |B|. \qquad (116)$$

For $\alpha \in (1,\infty)$, we have that

$$H_\alpha(\mathcal{N}) = \frac{\alpha}{1-\alpha} \log_2 \|\mathcal{N}_{A\to B}\|_{\text{CB},1\to\alpha}. \qquad (117)$$

*Proof.* To establish the equality

$$H_\alpha(\mathcal{N}) = \inf_{\psi_{RA}} H_\alpha(B|R)_{\omega|\omega}, \qquad (118)$$

we follow the same reasoning as in Eqs. (42)–(46), but making the substitutions $H \to H_\alpha$ and $D \to D_\alpha$. To establish the equality

$$\inf_{\psi_{RA}} H_\alpha(B|R)_{\omega|\omega} = -\sup_{\rho_A} \overline{H}_\beta(B|E)_\tau, \qquad (119)$$

we employ the identity [[74], Theorem 2]

$$H_\alpha(B|R)_{\omega|\omega} = -\overline{H}_\beta(B|E)_\tau. \qquad (120)$$

To establish the dimension bounds, consider from data processing that

$$H_\alpha(B|R)_{\omega|\omega} \leqslant H_\alpha(B)_\omega \leqslant \log_2 |B|, \qquad (121)$$

where the second inequality follows from a dimension bound for the Rényi entropy. To establish the other dimension bound, let us employ the identity [[74], Theorem 2] again

$$H_\alpha(B|R)_{\omega|\omega} = -\overline{H}_\beta(B|E)_\tau \qquad (122)$$

$$\geqslant -\inf_{\sigma_E} D_\beta(\tau_{BE}\|I_B \otimes \sigma_E) \qquad (123)$$

$$\geqslant -H_\beta(B)_\tau \qquad (124)$$

$$\geqslant -\log_2 |B|. \qquad (125)$$

The first inequality is stated in Ref. [[74], Corollary 4], and the second follows from data processing of the Petz–Rényi relative entropy under measurements, which holds for $\beta \in (0,1) \cup (1,\infty)$, as shown in Ref. [[75], Sec. 2.2] (note that a measurement in the eigenbasis of $\tau_B$ combined with the partial trace over system $E$ is a particular kind of measurement).

To establish the connection to the completely bounded norm for $\alpha > 1$, we invoke Ref. [[15], Lemma 8] to find that

$$H_\alpha(\mathcal{N}) = \log_2 |B| - D_\alpha(\mathcal{N}\|\mathcal{R}) \qquad (126)$$

$$= \log_2 |B| - \frac{\alpha}{\alpha-1} \log \|\Omega_{\pi_B^{(1-\alpha)/\alpha}} \circ \mathcal{N}_{A\to B}\|_{\text{CB},1\to\alpha} \qquad (127)$$

$$= \frac{\alpha}{1-\alpha} \log \|\mathcal{N}_{A\to B}\|_{\text{CB},1\to\alpha}, \qquad (128)$$

where

$$\Omega_{\pi_B^{(1-\alpha)/\alpha}}(X_B) \equiv \pi_B^{(1-\alpha)/2\alpha} X_B \pi_B^{(1-\alpha)/2\alpha} \qquad (129)$$

$$= |B|^{(\alpha-1)/\alpha} X_B, \qquad (130)$$

concluding the proof. ∎

Again, the dimension lower bound is saturated for the identity channel, while the dimension upper bound is saturated for the completely depolarizing channel.

## VI. MIN-ENTROPY OF A QUANTUM CHANNEL

The min-entropy of a quantum state $\rho_A$ of a system $A$ is defined as [76]

$$H_{\min}(A)_\rho \equiv -\log_2 \|\rho\|_\infty \qquad (131)$$

$$= \lim_{\alpha\to\infty} H_\alpha(A)_\rho. \qquad (132)$$

It has found extensive application in the context of quantum cryptography [76]. The max-relative entropy of a state $\rho$ with a positive semidefinite operator $\sigma$ is defined as [77]

$$D_{\max}(\rho\|\sigma) \equiv \inf\{\lambda : \rho \leqslant 2^\lambda \sigma\} \qquad (133)$$

$$= \log_2 \|\sigma^{-1/2} \rho \sigma^{-1/2}\|_\infty, \qquad (134)$$

whenever $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and otherwise it is set to $+\infty$. The max-relative entropy was recently given an information-theoretic meaning as the distinguishability cost of two quantum states [78]. It is known that [68]

$$D_{\max}(\rho\|\sigma) = \lim_{\alpha\to\infty} D_\alpha(\rho\|\sigma). \qquad (135)$$

Observe that the min-entropy of a quantum state $\rho$ can be written as the difference of the number of physical qubits for the system $A$ and the max-relative entropy of $\rho$ to the maximally mixed state $\pi_A$:

$$H_{\min}(A)_\rho = \log_2 |A| - D_{\max}(\rho_A \| \pi_A). \tag{136}$$

Thus, following the spirit of previous developments, we define the min-entropy of a channel as follows:

*Definition 18 (Min-entropy of a quantum channel).* We define the *min-entropy of a quantum channel* $\mathcal{N}_{A \to B}$ according to the recipe given in the introduction of our paper:

$$H_{\min}(\mathcal{N}) \equiv \log_2 |B| - D_{\max}(\mathcal{N} \| \mathcal{R}), \tag{137}$$

where $D_{\max}(\mathcal{N} \| \mathcal{R})$ is the max-channel divergence [15,16] and $\mathcal{R}_{A \to B}$ is the completely randomizing channel from Eq. (5).

The max-channel divergence is defined for two arbitrary channels $\mathcal{N}_{A \to B}$ and $\mathcal{M}_{A \to B}$ as [15,16]

$$D_{\max}(\mathcal{N} \| \mathcal{M}) \equiv \sup_{\rho_{RA}} D_{\max}(\mathcal{N}_{A \to B}(\rho_{RA}) \| \mathcal{M}_{A \to B}(\rho_{RA})) \tag{138}$$

$$= D_{\max}(\mathcal{N}_{A \to B}(\Phi_{RA}) \| \mathcal{M}_{A \to B}(\Phi_{RA})). \tag{139}$$

The latter equality, that an optimal state is the maximally entangled state $\Phi_{RA}$, was proved in Ref. [[28], Lemma 12] (see also Ref. [[79], Eq. (45)] and Ref. [[28], Remark 13] in this context). In fact, an optimal state is any pure bipartite state with full Schmidt rank (reduced state has full support).

Due to the limit in Eq. (135) and the equality in Eq. (139), it follows that

$$D_{\max}(\mathcal{N} \| \mathcal{M}) = \lim_{\alpha \to \infty} D_\alpha(\mathcal{N} \| \mathcal{M}). \tag{140}$$

As such, we can immediately conclude that the min-entropy of a channel $H_{\min}(\mathcal{N})$ is equal to the following limit:

$$H_{\min}(\mathcal{N}) = \lim_{\alpha \to \infty} H_\alpha(\mathcal{N}), \tag{141}$$

and that it satisfies nondecrease under a uniformity preserving superchannel, additivity, and reduction to states (i.e., for a replacer channel, it reduces to the min-entropy of the replacing state), which, as stated previously, imply the three axioms from Ref. [18].

## A. Alternate representation for the min-entropy of a channel in terms of conditional min-entropies

The conditional min-entropy of a bipartite quantum state $\rho_{AB}$ is defined as [76]

$$H_{\min}(A|B)_\rho \equiv -\inf_{\sigma_B} D_{\max}(\rho_{AB} \| I_A \otimes \sigma_B). \tag{142}$$

We can also define the following related quantity:

$$H_{\min}(A|B)_{\rho|\rho} \equiv -D_{\max}(\rho_{AB} \| I_A \otimes \rho_B), \tag{143}$$

and clearly we have that

$$H_{\min}(A|B)_\rho \geqslant H_{\min}(A|B)_{\rho|\rho}. \tag{144}$$

The identities in Eqs. (35) and (40), as well as the definition of conditional min-entropy, inspire the following quantity:

$$H_{\min}^\uparrow(\mathcal{N}) = \inf_{\psi_{RA}} H_{\min}(B|R)_\omega. \tag{145}$$

In the above, $\omega_{RB} \equiv \mathcal{N}_{A \to B}(\psi_{RA})$ and $\psi_{RA}$ is a pure state with system $R$ isomorphic to the channel input system $A$.

This quantity might seem different from the min-entropy of a channel, but the following proposition states that $H_{\min}^\uparrow(\mathcal{N})$ is actually equal to the min-entropy of the channel $H_{\min}(\mathcal{N})$, thus simplifying the notion of min-entropy of a quantum channel:

*Proposition 19.* Let $\mathcal{N}_{A \to B}$ be a quantum channel. Then

$$H_{\min}(\mathcal{N}) = \inf_{\psi_{RA}} H_{\min}(B|R)_{\omega|\omega} \tag{146}$$

$$= H_{\min}(B|R)_{\Phi^{\mathcal{N}} | \Phi^{\mathcal{N}}} \tag{147}$$

$$= H_{\min}^\uparrow(\mathcal{N}), \tag{148}$$

where $\omega_{RB} \equiv \mathcal{N}_{A \to B}(\psi_{RA})$ and $\psi_{RA}$ is a pure state with system $R$ isomorphic to the channel input system $A$. Also, the state $\Phi_{RB}^{\mathcal{N}} = \mathcal{N}_{A \to B}(\Phi_{RA})$ is the Choi state of the channel.

*Proof.* The first equality follows from the same steps in the proof of Proposition 17 [see the reasoning around Eq. (118)]. The second equality, i.e.,

$$H_{\min}(\mathcal{N}) = H_{\min}(B|R)_{\Phi^{\mathcal{N}} | \Phi^{\mathcal{N}}}, \tag{149}$$

follows by the observation in Eq. (139).

The proof of the equality $H_{\min}(\mathcal{N}) = H_{\min}^\uparrow(\mathcal{N})$ follows from semidefinite programming duality, similar to what was done previously for conditional min-entropy in Ref. [80]. Consider that

$$H_{\min}^\uparrow(\mathcal{N})$$

$$= \inf_{\psi_{RA}} H_{\min}(B|R)_{\mathcal{N}_{A \to B}(\psi_{RA})} \tag{150}$$

$$= \inf_{\psi_{RA}} [-\inf_{\sigma_R} D_{\max}(\mathcal{N}_{A \to B}(\psi_{RA}) \| \sigma_R \otimes I_B)] \tag{151}$$

$$= \inf_{\psi_{RA}} [-\inf_{\sigma_R} \{\log_2 \mathrm{Tr}\{\sigma_R\} : \mathcal{N}_{A \to B}(\psi_{RA}) \leqslant \sigma_R \otimes I_B\}] \tag{152}$$

$$= -\log_2 \sup_{\psi_{RA}} \inf_{\sigma_R} \{\mathrm{Tr}\{\sigma_R\} : \mathcal{N}_{A \to B}(\psi_{RA}) \leqslant \sigma_R \otimes I_B\}. \tag{153}$$

Considering the innermost part of the last line above as the semidefinite program

$$\inf_{\sigma_R} \{\mathrm{Tr}\{\sigma_R\} : \mathcal{N}_{A \to B}(\psi_{RA}) \leqslant \sigma_R \otimes I_B\}, \tag{154}$$

its dual is given by

$$\sup_{X_{RB}} \{\mathrm{Tr}\{X_{RB} \mathcal{N}_{A \to B}(\psi_{RA})\} : X_R \leqslant I_R, \; X_{RB} \geqslant 0\}. \tag{155}$$

Now let us write the pure state $\psi_{RA}$ as $\psi_{RA} = Y_R \Gamma_{RA} Y_R^\dagger$, where $Y_R$ satisfies $\mathrm{Tr}\{Y_R^\dagger Y_R\} = 1$, so that $Y_R^\dagger Y_R$ is a density operator $\rho_R$. Due to the fact that the set of pure states $\psi_{RA}$ with full-rank reduced state $\psi_R$ is dense in the set of all pure states, it suffices to optimize over these. This means that we can rewrite the last line of the first block (without the negative logarithm) as

$$\sup_{Y_R, X_{RB}} \{\mathrm{Tr}\{Y_R^\dagger X_{RB} Y_R \mathcal{N}_{A \to B}(\Gamma_{RA})\} : X_R \leqslant I_R,$$

$$X_{RB} \geqslant 0, \; \mathrm{Tr}\{Y_R^\dagger Y_R\} = 1, \; |Y_R| > 0\}. \tag{156}$$

We now define $X'_{RB} \equiv Y_R^\dagger X_{RB} Y_R$, so that

$$X_{RB} \geqslant 0 \quad \Longleftrightarrow \quad X'_{RB} \geqslant 0 \tag{157}$$

$$X_R \leqslant I_R \quad \Longleftrightarrow \quad X'_R \leqslant Y_R^\dagger Y_R = \rho_R. \tag{158}$$

Then we find that the above is equal to

$$\sup_{\rho_R, X'_{RB}} \{ \mathrm{Tr}\{X'_{RB} \mathcal{N}_{A \to B}(\Gamma_{RA})\} : X'_R \leqslant \rho_R, \, \mathrm{Tr}\{\rho_R\} = 1,$$

$$X'_{RB} \geqslant 0, \, \rho_R \geqslant 0\}$$

$$= \sup_{X'_{RB}} \{ \mathrm{Tr}\{X'_{RB} \mathcal{N}_{A \to B}(\Gamma_{RA})\} : \mathrm{Tr}\{X'_R\} = 1, \, X'_{RB} \geqslant 0\} \tag{159}$$

$$= \sup_{X'_{RB}} \{ \mathrm{Tr}\{X'_{RB} \mathcal{N}_{A \to B}(\Gamma_{RA})\} : \mathrm{Tr}\{X'_{RB}\} = 1, \, X'_{RB} \geqslant 0\} \tag{160}$$

$$= \| \mathcal{N}_{A \to B}(\Gamma_{RA}) \|_\infty. \tag{161}$$

The first equality above follows because we are maximizing over both $\rho_R$ and $X'_{RB}$, and the objective function only increases by taking $X'_R = \rho_R$ and with a maximal value one for the trace. So we conclude that

$$\inf_{\psi_{RA}} H_{\min}(B|R)_{\mathcal{N}_{A \to B}(\psi_{RA})}$$

$$= -\log_2 \| \mathcal{N}_{A \to B}(\Gamma_{RA}) \|_\infty \tag{162}$$

$$= -\log_2 \inf \{\lambda : \mathcal{N}_{A \to B}(\Gamma_{RA}) \leqslant \lambda I_{RB}\} \tag{163}$$

$$= -\log_2 \inf \{\lambda : \mathcal{N}_{A \to B}(\Phi_{RA}) \leqslant \lambda \pi_R \otimes I_B\} \tag{164}$$

$$= H_{\min}(B|R)_{\Phi^{\mathcal{N}} | \Phi^{\mathcal{N}}} \tag{165}$$

$$= H_{\min}(\mathcal{N}), \tag{166}$$

where $\Phi^{\mathcal{N}} = \mathcal{N}_{A \to B}(\Phi_{RA})$ and the last equality follows from Eq. (149). ∎

### B. Relation of min-entropy of a channel to its extended min-entropy

The extended min-entropy of a channel is defined as [18]

$$H_{\min}^{\mathrm{ext}}(\mathcal{N}) \equiv H_{\min}(B|R)_\omega, \tag{167}$$

where $\omega_{RA} = \mathcal{N}_{A \to B}(\Phi_{RA})$, with $\Phi_{RA}$ the maximally entangled state. It is not clear to us whether $H_{\min}^{\mathrm{ext}}(\mathcal{N})$ is generally equal to the min-entropy of a channel $H_{\min}(\mathcal{N})$. However, due to Eq. (149), we conclude that

$$H_{\min}^{\mathrm{ext}}(\mathcal{N}) \geqslant H_{\min}(\mathcal{N}). \tag{168}$$

## VII. ASYMPTOTIC EQUIPARTITION PROPERTY

The smoothed conditional min-entropy of a bipartite state $\rho_{AB}$ is defined for $\varepsilon \in (0, 1)$ as (see, e.g., Ref. [81])

$$H_{\min}^\varepsilon(A|B)_\rho \equiv \sup_{P(\rho_{AB}, \widetilde{\rho}_{AB}) \leqslant \varepsilon} H_{\min}(A|B)_{\widetilde{\rho}}, \tag{169}$$

where the optimization is with respect to all subnormalized states $\widetilde{\rho}_{AB}$ (satisfying $\widetilde{\rho}_{AB} \geqslant 0$, $\mathrm{Tr}\{\widetilde{\rho}_{AB}\} \leqslant 1$, and $\widetilde{\rho}_{AB} \neq 0$) and the sine distance (also called purified distance) of quantum states $\rho$ and $\sigma$ [82–85] is defined in terms of the fidelity

[86] as

$$P(\rho, \sigma) \equiv \sqrt{1 - F(\rho, \sigma)}, \tag{170}$$

$$F(\rho, \sigma) \equiv \| \sqrt{\rho} \sqrt{\sigma} \|_1^2. \tag{171}$$

The definition of fidelity is generalized to subnormalized states $\omega$ and $\tau$ as follows [87]:

$$F(\omega, \tau) \equiv F[\omega \oplus (1 - \mathrm{Tr}\{\omega\}), \tau \oplus (1 - \mathrm{Tr}\{\tau\})], \tag{172}$$

where the right-hand side is the usual fidelity of states (that is, we just add an extra dimension to $\omega$ and $\tau$ and complete them to states). The smoothed conditional min-entropy satisfies the following asymptotic equipartition property [88] (see also Ref. [81]), which is one way that it connects with the conditional entropy of $\rho_{AB}$:

$$\lim_{n \to \infty} \frac{1}{n} H_{\min}^\varepsilon(A^n|B^n)_{\rho^{\otimes n}} = H(A|B)_\rho. \tag{173}$$

The purified channel divergence of two channels $\mathcal{N}_{A \to B}$ and $\mathcal{M}_{A \to B}$ is defined as [16]

$$P(\mathcal{N}, \mathcal{M}) \equiv \sup_{\rho_{RA}} P(\mathcal{N}_{A \to B}(\rho_{RA}), \mathcal{M}_{A \to B}(\rho_{RA})), \tag{174}$$

Again, due to state purification, the data-processing inequality for $P(\rho, \sigma)$, and the Schmidt decomposition theorem, it suffices to optimize over states $\rho_{RA}$ that are pure and such that system $R$ is isomorphic to system $A$. We then use this notion for smoothing the min-entropy of a channel:

*Definition 20 (Smoothed min-entropy of a channel).* The smoothed min-entropy of a channel is defined for $\varepsilon \in (0, 1)$ as

$$H_{\min}^\varepsilon(\mathcal{N}) \equiv \sup_{P(\mathcal{N}, \widetilde{\mathcal{N}}) \leqslant \varepsilon} H_{\min}(\widetilde{\mathcal{N}}), \tag{175}$$

where $P(\mathcal{N}, \widetilde{\mathcal{N}})$ is the purified channel divergence [16].

In the following theorem, we prove that the smoothed min-entropy of a channel satisfies an asymptotic equipartition theorem that generalizes Eq. (173).

*Theorem 21 (Asymptotic equipartition property).* For all $\varepsilon \in (0, 1)$, the following inequality holds

$$\lim_{n \to \infty} \frac{1}{n} H_{\min}^\varepsilon(\mathcal{N}^{\otimes n}) \geqslant H(\mathcal{N}). \tag{176}$$

We also have that

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{\min}^\varepsilon(\mathcal{N}^{\otimes n}) \leqslant H(\mathcal{N}). \tag{177}$$

*Proof.* We first prove the inequality in Eq. (176). Let $\omega_{R^n A^n}$ denote the de Finetti state [89], defined as

$$\omega_{R^n A^n} \equiv \int d(\sigma_{RA}) \, \sigma_{RA}^{\otimes n}, \tag{178}$$

where $\sigma_{RA}$ is a pure state with system $R$ isomorphic to the channel input system $A$, and $d(\sigma_{RA})$ denotes the Haar measure on pure states. This state is the maximally mixed state of the symmetric subspace of the systems $(RA)^n$, and it is permutation invariant [90]. That is, for a unitary channel $\mathcal{W}_{R^n}^\pi \otimes \mathcal{W}_{A^n}^\pi$ corresponding to a permutation $\pi$, we have that $\omega_{R^n A^n} = (\mathcal{W}_{R^n}^\pi \otimes \mathcal{W}_{A^n}^\pi)(\omega_{R^n A^n})$ for all $\pi \in S_n$, with $S_n$ denoting the symmetric group. Let $\omega_{R' R^n A^n}$ denote the purification of the de Finetti state, with the purifying system $R'$ satisfying

the inequality $|R'| \leqslant (n+1)^{|A|^2-1}$ [89]. The reduced state $\omega_{A^n}$ is permutation invariant and has full rank. The latter follows because the set of pure states $\psi_{RA}$ with a full-rank reduced density operator $\psi_R$ is dense in the set of all pure states, and tensor products of full-rank states are full rank. Let $\omega_{R'R^nB^n}^{\widetilde{\mathcal{N}}^n}$ denote the state resulting from the action of the quantum channel $\widetilde{\mathcal{N}}_{A^n \to B^n}^n$ on the input state $\omega_{R'R^nA^n}$, and let $\omega_{R'R^nB^n}^{\mathcal{N}^{\otimes n}}$ denote the state resulting from the action of the quantum channel $\mathcal{N}_{A \to B}^{\otimes n}$ on the input state $\omega_{R'R^nA^n}$. Let CPTP$(A^n \to B^n)$ denote the set of all quantum channels from input system $A^n$ to output system $B^n$. Let Perm$(A^n \to B^n)$ denote the set of all permutation covariant quantum channels from input system $A^n$ to output system $B^n$. Define $\psi_{RB^n}^{\widetilde{\mathcal{N}}^n}$ to be the state resulting from the action of the channel $\widetilde{\mathcal{N}}_{A^n \to B^n}^n$ on the input state $\psi_{RA^n}$. Then consider that

$$H_{\min}^\varepsilon(\mathcal{N}^{\otimes n})$$

$$= \sup_{\substack{\widetilde{\mathcal{N}}^n \in \text{CPTP}(A^n \to B^n): \\ P(\mathcal{N}^{\otimes n}, \widetilde{\mathcal{N}}^n) \leqslant \varepsilon}} \inf_{\psi_{RA^n}} H_{\min}(B^n|R)_{\psi^{\widetilde{\mathcal{N}}^n}|\psi^{\widetilde{\mathcal{N}}^n}}$$

(179)

$$\geqslant \sup_{\substack{\widetilde{\mathcal{N}}^n \in \text{Perm}(A^n \to B^n): \\ P(\mathcal{N}^{\otimes n}, \widetilde{\mathcal{N}}^n) \leqslant \varepsilon}} \inf_{\psi_{RA^n}} H_{\min}(B^n|R)_{\psi^{\widetilde{\mathcal{N}}^n}|\psi^{\widetilde{\mathcal{N}}^n}}$$

(180)

$$= \sup_{\substack{\widetilde{\mathcal{N}}^n \in \text{Perm}(A^n \to B^n): \\ P(\mathcal{N}^{\otimes n}, \widetilde{\mathcal{N}}^n) \leqslant \varepsilon}} H_{\min}(B^n|R^nR')_{\omega^{\widetilde{\mathcal{N}}^n}|\omega^{\widetilde{\mathcal{N}}^n}}$$

(181)

$$\geqslant \sup_{\substack{\widetilde{\mathcal{N}}^n \in \text{Perm}(A^n \to B^n): \\ P(\omega^{\mathcal{N}^{\otimes n}}, \omega^{\widetilde{\mathcal{N}}^n}) \leqslant \varepsilon'}} H_{\min}(B^n|R^nR')_{\omega^{\widetilde{\mathcal{N}}^n}|\omega^{\widetilde{\mathcal{N}}^n}}.$$

(182)

The first equality follows from Definition 20. The first inequality follows by restricting the maximization to permutation-covariant channels. The second equality follows because the reduced state $\omega_{A^n}$ has full rank and by applying the remark after Eq. (139), to conclude that

$$\inf_{\psi_{RA^n}} H_{\min}(B^n|R)_{\psi^{\widetilde{\mathcal{N}}^n}|\psi^{\widetilde{\mathcal{N}}^n}}$$

$$= -\sup_{\psi_{RA^n}} D_{\max}(\widetilde{\mathcal{N}}_{A^n \to B^n}^n(\psi_{RA^n}) \| \psi_R \otimes I_{B^n})$$

(183)

$$= -\sup_{\psi_{RA^n}} D_{\max}(\widetilde{\mathcal{N}}_{A^n \to B^n}^n(\psi_{RA^n}) \| \psi_R \otimes \pi_{B^n})$$
$$+ n \log_2 |B|$$

(184)

$$= -\sup_{\psi_{RA^n}} D_{\max}(\widetilde{\mathcal{N}}_{A^n \to B^n}^n(\psi_{RA^n}) \| \mathcal{R}_{A \to B}^{\otimes n}(\psi_{RA^n}))$$
$$+ n \log_2 |B|$$

(185)

$$= -D_{\max}(\widetilde{\mathcal{N}}_{A^n \to B^n}^n(\omega_{R'R^nA^n}) \| \mathcal{R}_{A \to B}^{\otimes n}(\omega_{R'R^nA^n}))$$
$$+ n \log_2 |B|$$

(186)

$$= H_{\min}(B^n|R^nR')_{\omega^{\widetilde{\mathcal{N}}^n}|\omega^{\widetilde{\mathcal{N}}^n}}.$$

(187)

The second inequality follows by applying the post-selection technique [[89], Theorem 1] with

$$\varepsilon' \equiv \varepsilon(n+1)^{-2(|A|^2-1)}.$$

(188)

(See also Proposition D.5 of Ref. [91].) Note that the factor of two in the exponent of Eq. (188) is necessary because we are employing the sine distance as the channel distance measure. To be clear, the statement we are invoking is that if

$$\mathcal{N}^n, \mathcal{M}^n \in \text{Perm}(A^n \to B^n)$$

(189)

satisfy

$$P(\mathcal{N}_{A^n \to B^n}^n(\omega_{R'R^nA^n}), \mathcal{M}_{A^n \to B^n}^n(\omega_{R'R^nA^n})) \leqslant \varepsilon',$$

(190)

then

$$P(\mathcal{N}_{A^n \to B^n}^n, \mathcal{M}_{A^n \to B^n}^n) \leqslant \varepsilon.$$

(191)

Continuing, we have that

Eq. (182)

$$= \sup_{\substack{\widetilde{\mathcal{N}}^n \in \text{CPTP}(A^n \to B^n): \\ P(\omega^{\mathcal{N}^{\otimes n}}, \omega^{\widetilde{\mathcal{N}}^n}) \leqslant \varepsilon'}} H_{\min}(B^n|R^nR')_{\omega^{\widetilde{\mathcal{N}}^n}|\omega^{\widetilde{\mathcal{N}}^n}}$$

(192)

$$= \sup_{\substack{\sigma_{R'R^nB^n}: \\ P(\mathcal{N}^{\otimes n}(\omega_{R'R^nA^n}), \sigma_{R'R^nB^n}) \leqslant \varepsilon', \\ \sigma_{R'R^n} = \omega_{R'R^n}}} H_{\min}(B^n|R^nR')_{\sigma|\sigma}$$

(193)

$$\geqslant \sup_{\substack{\sigma_{R'R^nB^n}: \\ P(\mathcal{N}^{\otimes n}(\omega_{R'R^nA^n}), \sigma_{R'R^nB^n}) \leqslant 2\varepsilon'/3}} H_{\min}(B^n|R^nR')_{\sigma|\sigma}$$
$$- \log_2\left(\frac{8 + [\varepsilon'/3]^2}{[\varepsilon'/3]^2}\right).$$

(194)

The first equality follows from reasoning similar to that given for Lemma 11 in Appendix B of Ref. [92], i.e., that a permutation-covariant channel is optimal among all channels, due to the fact that the original channel $\mathcal{N}^{\otimes n}$ is permutation covariant. In our case, it follows by employing the fact that the channel min-entropy does not decrease under the action of a uniformity preserving superchannel [see the discussion after Eq. (141)], and the superchannel that randomly performs a permutation at the channel input and the inverse permutation at the channel output is one such superchannel. The second equality is a consequence of the fact that the following two sets are equal:

$$\left\{ \widetilde{\mathcal{N}}_{A^n \to B^n}^n(\phi_{RA^n}) : P(\widetilde{\mathcal{N}}_{A^n \to B^n}^n(\phi_{RA^n}), \mathcal{N}_{A \to B}^{\otimes n}(\phi_{RA^n})) \leqslant \varepsilon, \right.$$
$$\left. \widetilde{\mathcal{N}}_{A^n \to B^n}^n \in \text{CPTP} \right\}$$
$$= \left\{ \widetilde{\omega}_{RB^n} \in \mathcal{D}(\mathcal{H}_{RB^n}) : P(\widetilde{\omega}_{RB^n}, \mathcal{N}_{A \to B}^{\otimes n}(\phi_{RA^n})) \leqslant \varepsilon, \right.$$
$$\left. \widetilde{\omega}_R = \phi_R \right\},$$

(195)

which follows from applying Lemma 10 in Appendix B of Ref. [92]. The inequality follows from Theorem 3 of Ref. [93] (while noting that the state $\hat{\rho}_{AB}$ defined therein satisfies $\hat{\rho}_B =$

$\rho_B$, so that the proof Theorem 3 of Ref. [93] applies to our situation).

Continuing, and by applying Ref. [[78], Eq. (L10)] and definitions, we find that

$$\sup_{\substack{\sigma_{R'R^nB^n}: \\ P(\mathcal{N}^{\otimes n}(\omega_{R'R^nA^n}),\sigma_{R'R^nB^n})\leqslant 2\varepsilon'/3}} H_{\min}(B^n|R^nR')_{\sigma|\sigma}$$

$$\geqslant H_\alpha(B^n|R^nR')_{\omega^{\mathcal{N}^{\otimes n}}|\omega^{\mathcal{N}^{\otimes n}}} + f(2\varepsilon'/3,\alpha) \quad (196)$$

$$\geqslant H_\alpha(\mathcal{N}^{\otimes n}) + +f(2\varepsilon'/3,\alpha) \quad (197)$$

$$= nH_\alpha(\mathcal{N}) + f(2\varepsilon'/3,\alpha), \quad (198)$$

where

$$f(\delta,\alpha) \equiv \log_2(1-\delta^2) + \frac{\log_2(\delta^2)}{\alpha-1}. \quad (199)$$

The second inequality follows from the definition of the Rényi entropy of a channel (Definition 13), and the equality follows from the additivity of the Rényi entropy of a channel (Proposition 15). Putting everything above together, we conclude the following bound:

$$\frac{1}{n}H_{\min}^\varepsilon(\mathcal{N}) \geqslant H_\alpha(\mathcal{N}) - \frac{1}{n}\log_2\left(\frac{8+[\varepsilon'/3]^2}{[\varepsilon'/3]^2}\right)$$
$$+ \frac{f(2\varepsilon'/3,\alpha)}{n}. \quad (200)$$

Taking the limit as $n \to \infty$, we conclude that the following inequality holds for all $\alpha > 1$:

$$\lim_{n\to\infty} \frac{1}{n}H_{\min}^\varepsilon(\mathcal{N}) \geqslant H_\alpha(\mathcal{N}). \quad (201)$$

Since this inequality holds for all $\alpha > 1$, we can take the limit as $\alpha \to 1$ to conclude that

$$\lim_{n\to\infty} \frac{1}{n}H_{\min}^\varepsilon(\mathcal{N}) \geqslant H(\mathcal{N}). \quad (202)$$

This concludes the proof of the inequality in Eq. (176).

To arrive at the second inequality in Eq. (177), let $\widetilde{\mathcal{N}}^n$ be a channel such that

$$P(\mathcal{N}^{\otimes n}, \widetilde{\mathcal{N}}^n) \leqslant \varepsilon. \quad (203)$$

Now let $\phi_{RA^n}$ be an arbitrary state. We then have from the definition in Eq. (174) that

$$P\big(\mathcal{N}_{A\to B}^{\otimes n}(\phi_{RA^n}), \widetilde{\mathcal{N}}_{A^n\to B^n}^n(\phi_{RA^n})\big) \leqslant \varepsilon. \quad (204)$$

Defining the states

$$\widetilde{\omega}_{RB^n} \equiv \widetilde{\mathcal{N}}_{A^n\to B^n}^n(\phi_{RA^n}), \quad (205)$$

$$\omega_{RB^n} \equiv \mathcal{N}_{A\to B}^{\otimes n}(\phi_{RA^n}), \quad (206)$$

we find that

$$H_{\min}(\widetilde{\mathcal{N}}^n) \leqslant H_{\min}(B^n|R)_{\widetilde{\omega}|\widetilde{\omega}} \quad (207)$$

$$\leqslant H(B^n|R)_{\widetilde{\omega}} \quad (208)$$

$$\leqslant H(B^n|R)_\omega + \varepsilon 2n\log_2|B| + g_2(\varepsilon), \quad (209)$$

where

$$g_2(\varepsilon) \equiv (\varepsilon+1)\log_2(\varepsilon+1) - \varepsilon\log_2\varepsilon. \quad (210)$$

The second inequality follows from monotonicity of the conditional Rényi entropy with respect to $\alpha$, and the last from the uniform continuity bound in Ref. [[94], Lemma 2]. The above bound holds for any choice of $\phi_{RA^n}$, and so we conclude that

$$H_{\min}(\widetilde{\mathcal{N}}^n) \leqslant H(\mathcal{N}^{\otimes n}) + \varepsilon 2n\log_2|B| + g_2(\varepsilon) \quad (211)$$

$$= nH(\mathcal{N}) + \varepsilon 2n\log_2|B| + g_2(\varepsilon), \quad (212)$$

where the equality follows from the additivity of the entropy of a channel (Proposition 4). Now, the inequality has been shown for all $\widetilde{\mathcal{N}}^n$ satisfying $P(\mathcal{N}^{\otimes n}, \widetilde{\mathcal{N}}^n) \leqslant \varepsilon$, and so we conclude, after dividing by $n$, that

$$\frac{1}{n}H_{\min}^\varepsilon(\mathcal{N}^{\otimes n}) \leqslant H(\mathcal{N}) + 2\varepsilon\log_2|B| + \frac{1}{n}g_2(\varepsilon). \quad (213)$$

Taking the limit as $n \to \infty$, we get that

$$\lim_{n\to\infty} \frac{1}{n}H_{\min}^\varepsilon(\mathcal{N}^{\otimes n}) \leqslant H(\mathcal{N}) + 2\varepsilon\log_2|B|. \quad (214)$$

Now taking the limit as $\varepsilon \to 0$, we arrive at the second inequality in Eq. (177). ∎

In Appendix B, we point out how an approach similar to that in the above proof leads to an alternate proof of the upper bound in Ref. [[92], Theorem 8], regarding an asymptotic equipartition property for the smoothed max-mutual information of a quantum channel.

## VIII. GENERALIZED CHANNEL ENTROPIES FROM GENERALIZED DIVERGENCES

In this section, we discuss other possibilities for defining generalized entropies of a quantum channel. One main concern might be how unique or distinguished our notion of entropy of a channel from Definition 1 is, being based on the channel relative entropy of the channel of interest and the completely randomizing channel. As a consequence of the fact that there are alternate ways of defining channel relative entropies, there could be alternate notions of channel entropies. However, we should recall that one of the main reasons we have chosen the definition in Definition 1 is that the channel relative entropy appearing there has a particularly appealing operational interpretation in the context of channel discrimination [15]. That is, for what one might consider the most natural and general setting of quantum channel discrimination, the optimal rate for distinguishing a channel from the completely randomizing channel is given by the channel relative entropy in Eq. (4) [15]. As we show in what follows, there are further reasons to focus on our definition of the entropy of a channel from Definition 1, as well as our definition of the min-entropy of a channel from Definition 18.

To begin the discussion, let $\mathcal{S}(C)$ denote the set of quantum states for an arbitrary quantum system $C$. Let us recall that a function $\mathbf{D} : \mathcal{S}(C) \times \mathcal{S}(C) \to \mathbb{R} \cup \{+\infty\}$ is a generalized divergence [95,96] if for arbitrary Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, arbitrary states $\rho_A, \sigma_A \in \mathcal{S}(A)$, and an arbitrary channel $\mathcal{N}_{A\to B}$, the following data processing inequality holds

$$\mathbf{D}(\rho_A\|\sigma_A) \geqslant \mathbf{D}(\mathcal{N}_{A\to B}(\rho_A)\|\mathcal{N}_{A\to B}(\sigma_A)). \quad (215)$$

Examples of interest are in particular the quantum relative entropy, the Petz-Rényi divergences, the sandwiched Rényi divergences, as considered in this paper.

Based on generalized divergences, one can define at least two different channel divergences as a measure for the distinguishability of two quantum channels $\mathcal{N}_{A \to B}$ and $\mathcal{M}_{A \to B}$. Here we consider a function of two quantum channels to be a channel divergence if it is monotone under the action of a superchannel.

(1) Generalized channel divergence [16]:

$$\mathbf{D}(\mathcal{N}\|\mathcal{M}) \equiv \sup_{\rho_{RA}} \mathbf{D}(\mathcal{N}_{A \to B}(\rho_{RA})\|\mathcal{M}_{A \to B}(\rho_{RA})). \quad (216)$$

In the above, the optimization can be restricted to pure states of systems $R$ and $A$ with $R$ isomorphic to system $A$. The monotonicity of the generalized channel divergence under the action of a superchannel was proven in Ref. [18].

(2) Amortized channel divergence [28]:

$$\mathbf{D}^{\mathcal{A}}(\mathcal{N}\|\mathcal{M})$$
$$\equiv \sup_{\rho_{RA}, \sigma_{RA}} \mathbf{D}(\mathcal{N}_{A \to B}(\rho_{RA})\|\mathcal{M}_{A \to B}(\sigma_{RA})) - \mathbf{D}(\rho_{RA}\|\sigma_{RA}).$$
$$(217)$$

The monotonicity of the amortized channel divergence under the action of a superchannel was proven in Ref. [28].

We can consider other divergences as follows, but they are not known to be monotone under the action of a general superchannel, and so we do not label them as channel divergences:

(1) Choi divergence:

$$\mathbf{D}^{\Phi}(\mathcal{N}\|\mathcal{M}) \equiv \mathbf{D}(\mathcal{N}_{A \to B}(\Phi_{RA})\|\mathcal{M}_{A \to B}(\Phi_{RA})). \quad (218)$$

As we show in Appendix C, the Choi divergence is monotone under the action of a superchannel consisting of mixtures of a unital preprocessing channel and an arbitrary post-processing channel.

(2) Adversarial divergence:

$$\mathbf{D}^{\text{adv}}(\mathcal{N}\|\mathcal{M}) \equiv \sup_{\rho_{RA}} \inf_{\sigma_{RA}} \mathbf{D}(\mathcal{N}_{A \to B}(\rho_{RA})\|\mathcal{M}_{A \to B}(\sigma_{RA})). \quad (219)$$

In the above, due to state purification, data processing, and the Schmidt decomposition, the maximization can be restricted to pure states $\rho_{RA}$ of systems $R$ and $A$ with $R$ isomorphic to system $A$. The minimization should be taken over mixed states $\sigma_{RA}$. For a proof of this fact, see Appendix D.

(3) Adversarial Choi divergence:

$$\mathbf{D}^{\text{adv}, \Phi}(\mathcal{N}\|\mathcal{M}) \equiv \inf_{\sigma_{RA}} \mathbf{D}(\mathcal{N}_{A \to B}(\Phi_{RA})\|\mathcal{M}_{A \to B}(\sigma_{RA})). \quad (220)$$

(4) "No quantum memory" divergence:

$$\sup_{\rho_A} \mathbf{D}(\mathcal{N}_{A \to B}(\rho_A)\|\mathcal{M}_{A \to B}(\rho_A)). \quad (221)$$

There could certainly even be other divergences to consider. In our context, two effective ways of singling out particular divergences as primary and others as secondary are (1) whether the channel divergence has a compelling operational interpretation for a channel discrimination task and (2) whether the channel divergence leads to an entropy function that satisfies the axioms from Ref. [18].

Based on the recipe given in the Introduction, from a given divergence $\mathbf{D}'(\mathcal{N}\|\mathcal{M})$ (any of the choices above), one could then define a generalized entropy function of a channel $\mathcal{N}_{A \to B}$

as

$$\mathbf{H}(\mathcal{N}) \equiv \log_2 |B| - \mathbf{D}'(\mathcal{N}\|\mathcal{R}), \quad (222)$$

where $\mathcal{R}_{A \to B}$ is the completely randomizing channel from Eq. (5).

Taking the above approach to pruning entropy functions, we can already rule out the last one ("no quantum memory"), as done in Ref. [18], because, after taking $\mathbf{D}$ to be the most prominent case of quantum relative entropy, the resulting entropy function is the minimum output entropy of a channel, which is known to be nonadditive [97]. While an entropy arising from the Choi divergence leads to an entropy function satisfying the axioms desired for an entropy function, the Choi divergence itself does not appear to have a compelling operational interpretation in the sense of being a "channel measure" because it simply reduces a channel discrimination problem to a state discrimination problem (i.e., it does not make use of the most general approach one could take for discriminating arbitrary channels). This point could be debated, and we do return to entropy functions derived from Choi and adversarial Choi divergences in Sec. VIII D below.

### A. Collapse of entropy functions derived from quantum relative entropy

From the list above, by focusing on the operational and axiomatic criteria listed above, this leaves us with the generalized channel divergence and the amortized channel divergence. Here we also consider the adversarial divergence. Interestingly, after taking $\mathbf{D}$ to be the prominent case of quantum relative entropy and the channel $\mathcal{M}$ to be the completely randomizing channel, we find the following collapse of the divergences:

$$D(\mathcal{N}\|\mathcal{R}) = D^{\mathcal{A}}(\mathcal{N}\|\mathcal{R}) = D^{\text{adv}}(\mathcal{N}\|\mathcal{R}). \quad (223)$$

The first equality was shown in Refs. [15,28], and we show the second one now. From the definitions, we have that $\mathbf{D}^{\text{adv}}(\mathcal{N}\|\mathcal{M}) \leqslant \mathbf{D}(\mathcal{N}\|\mathcal{M})$ for any generalized divergence $\mathbf{D}$ and any channel $\mathcal{M}$. So we show the opposite inequality for the special case of $\mathbf{D} = D$ and $\mathcal{M} = \mathcal{R}$. Let $\rho_{RA}$ and $\sigma_{RA}$ be arbitrary states. Then

$$D(\mathcal{N}_{A \to B}(\rho_{RA})\|\mathcal{R}_{A \to B}(\sigma_{RA}))$$
$$= D(\mathcal{N}_{A \to B}(\rho_{RA})\|\sigma_R \otimes \pi_B) \quad (224)$$
$$= -H(\mathcal{N}_{A \to B}(\rho_{RA}))$$
$$\quad - \text{Tr}\{\mathcal{N}_{A \to B}(\rho_{RA}) \log_2(\sigma_R \otimes \pi_B)\} \quad (225)$$
$$= -H(\mathcal{N}_{A \to B}(\rho_{RA})) - \text{Tr}\{\rho_R \log_2 \sigma_R\}$$
$$\quad - \text{Tr}\{\mathcal{N}_{A \to B}(\rho_A) \log_2 \pi_B\} \quad (226)$$
$$= -H(\mathcal{N}_{A \to B}(\rho_{RA})) - \text{Tr}\{\rho_R \log_2 \rho_R\}$$
$$\quad + D(\rho_R\|\sigma_R) - \text{Tr}\{\mathcal{N}_{A \to B}(\rho_A) \log_2 \pi_B\} \quad (227)$$
$$= -H(\mathcal{N}_{A \to B}(\rho_{RA}))$$
$$\quad - \text{Tr}\{\mathcal{N}_{A \to B}(\rho_{RA}) \log_2(\rho_R \otimes \pi_B)\} + D(\rho_R\|\sigma_R)$$
$$(228)$$
$$= D(\mathcal{N}_{A \to B}(\rho_{RA})\|\mathcal{R}_{A \to B}(\rho_{RA})) + D(\rho_R\|\sigma_R). \quad (229)$$

Now taking an infimum over all $\sigma_{RA}$ and invoking the nonnegativity of quantum relative entropy, we conclude that

$$\inf_{\sigma_{RA}} D(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{R}_{A\to B}(\sigma_{RA}))$$

$$= D(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{R}_{A\to B}(\rho_{RA})). \quad (230)$$

By taking a supremum over $\rho_{RA}$, we then conclude that $D^{\mathrm{adv}}(\mathcal{N}\|\mathcal{R}) = D(\mathcal{N}\|\mathcal{R})$.

Thus, the collapse in Eq. (223), as well as the operational interpretation of $D(\mathcal{N}\|\mathcal{R})$ from Ref. [15] and the fact that the resulting entropy function satisfies the axioms from Ref. [18], indicate that our choice of the entropy of a quantum channel in Definition 1 is cogent.

### B. Collapse of entropy functions derived from max-relative entropy

Interestingly, a similar and further collapse occurs when taking **D** to be the max-relative entropy:

$$D_{\max}(\mathcal{N}\|\mathcal{R}) = D_{\max}^{\Phi}(\mathcal{N}\|\mathcal{R}) \quad (231)$$

$$= D_{\max}^{\mathcal{A}}(\mathcal{N}\|\mathcal{R}) \quad (232)$$

$$= D_{\max}^{\mathrm{adv}}(\mathcal{N}\|\mathcal{R}). \quad (233)$$

The first two equalities were shown in Ref. [[28], Proposition 10] for arbitrary channels $\mathcal{N}$ and $\mathcal{M}$. By employing a semidefinite programming approach as in the proof of Proposition 19, we can conclude the last equality. Thus, this collapse, as well as the facts that the max-relative entropy $D_{\max}(\mathcal{N}\|\mathcal{M})$ is an upper bound on the rate at which any two channels can be distinguished in an arbitrary context [[28], Corollary 18] and the resulting entropy function $H_{\min}(\mathcal{N})$ satisfies the axioms from Ref. [18], indicate that our choice of the min-entropy of a quantum channel in Definition 18 is also cogent.

### C. Entropy functions derived from Rényi relative entropies

In Sec. V, we defined the Rényi entropy of a channel as in Definition 13, in terms of the sandwiched Rényi relative entropy. The following collapse is known for the sandwiched Rényi relative entropy for $\alpha \in (1, \infty)$ [15,28]:

$$D_\alpha(\mathcal{N}\|\mathcal{R}) = D_\alpha^{\mathcal{A}}(\mathcal{N}\|\mathcal{R}). \quad (234)$$

However, it is not known whether these quantities are equal for $\alpha \in (0, 1)$ or whether they are equal to the adversarial divergence $D_\alpha^{\mathrm{adv}}(\mathcal{N}\|\mathcal{R})$ for any $\alpha \in (0, 1) \cup (1, \infty)$. At the same time, one of the most compelling reasons to fix the definition of channel Rényi entropy as we have done is that the channel divergence $D_\alpha(\mathcal{N}\|\mathcal{R})$ has both a convincing operational interpretation in channel discrimination as the optimal strong converse exponent and the entropy function satisfies all of the desired axioms for an entropy function. Furthermore, the entropy function $H_\alpha(\mathcal{N})$ represents a useful bridge between the entropy and min-entropy of a quantum channel, due to the facts that $\lim_{\alpha\to 1} H_\alpha(\mathcal{N}) = H(\mathcal{N})$, $\lim_{\alpha\to\infty} H_\alpha(\mathcal{N}) = H_{\min}(\mathcal{N})$, and $H_\alpha(\mathcal{N}) \leqslant H_\beta(\mathcal{N})$ for $\alpha \geqslant \beta \geqslant 1$.

One should notice that we did not define the Rényi entropy of a channel in terms of the Petz–Rényi relative entropy and the resulting channel divergence, amortized channel divergence, or adversarial divergence. One of the main reasons for this is that it is not known whether the resulting entropy func-

tions are additive. Furthermore, operational interpretations for these divergences have not been established, having been open since the paper [15] appeared. As such, it very well could be the case that one could derive cogent notions of channel entropy from the Petz–Rényi relative entropy, but this remains the topic of future work.

### D. Entropy functions derived from Choi and adversarial Choi divergences

In this subsection, we discuss various entropy functions derived from Choi and adversarial Choi divergences. As emphasized previously, we note again here that the operational interpretations for these divergences are really about state discrimination tasks rather than channel discrimination tasks. Nevertheless, the resulting entropy functions satisfy the axioms put forward in Ref. [18].

By picking the divergence **D** to be the quantum relative entropy $D$, we find that the Choi and adversarial Choi divergences are equal when discriminating an arbitrary channel $\mathcal{N}_{A\to B}$ from the completely randomizing channel $\mathcal{R}_{A\to B}$:

$$D^{\Phi}(\mathcal{N}\|\mathcal{R}) = D^{\mathrm{adv},\Phi}(\mathcal{N}\|\mathcal{R}). \quad (235)$$

The proof of this statement follows along the lines of Eqs. (224)–(229). There is a simple operational interpretation for $D^{\Phi}(\mathcal{N}\|\mathcal{R})$ in terms of state discrimination [12,13], while an operational interpretation for $D^{\mathrm{adv},\Phi}(\mathcal{N}\|\mathcal{R})$ in terms of state discrimination was given recently in Ref. [98].

We could also pick the divergence **D** to be Petz–Rényi relative entropy $\overline{D}_\alpha$ or the sandwiched Rényi relative entropy $D_\alpha$. The resulting Choi and adversarial Choi divergences are then generally not equal when discriminating an arbitrary channel $\mathcal{N}_{A\to B}$ from the completely randomizing channel $\mathcal{R}_{A\to B}$. There is an operational interpretation for $\overline{D}_\alpha^{\Phi}(\mathcal{N}\|\mathcal{R})$ for $\alpha \in (0, 1)$ in terms of state discrimination [99,100] (error exponent problem), and there is an operational interpretation for $D_\alpha^{\Phi}(\mathcal{N}\|\mathcal{R})$ for $\alpha \in (1, \infty)$ in terms of state discrimination [72] (strong converse exponent problem). Interestingly, Ref. [98] has given a meaningful operational interpretation for the adversarial Choi divergences $\overline{D}_\alpha^{\mathrm{adv},\Phi}(\mathcal{N}\|\mathcal{R})$ for $\alpha \in (0, 1)$ and $D_\alpha^{\mathrm{adv},\Phi}(\mathcal{N}\|\mathcal{R})$ for $\alpha \in (1, \infty)$ in terms of error exponent and strong converse exponent state discrimination problems, respectively.

For $\mathcal{N}_{A\to B}$ a quantum channel and $\Phi_{RB}^{\mathcal{N}} \equiv \mathcal{N}_{A\to B}(\Phi_{RA})$ the Choi state, the resulting channel entropy functions are then as follows:

$$H^{\Phi}(\mathcal{N}) \equiv H(B|R)_{\Phi^{\mathcal{N}}} = H^{\mathrm{adv},\Phi}(\mathcal{N}), \quad (236)$$

$$H_\alpha^{\Phi}(\mathcal{N}) \equiv H_\alpha(B|R)_{\Phi^{\mathcal{N}}|\Phi^{\mathcal{N}}}, \quad (237)$$

$$\overline{H}_\alpha^{\Phi}(\mathcal{N}) \equiv \overline{H}_\alpha(B|R)_{\Phi^{\mathcal{N}}|\Phi^{\mathcal{N}}}, \quad (238)$$

$$H_\alpha^{\mathrm{adv},\Phi}(\mathcal{N}) \equiv H_\alpha(B|R)_{\Phi^{\mathcal{N}}}, \quad (239)$$

$$\overline{H}_\alpha^{\mathrm{adv},\Phi}(\mathcal{N}) \equiv \overline{H}_\alpha(B|R)_{\Phi^{\mathcal{N}}}. \quad (240)$$

It then follows that all of the above entropy functions are additive for $\alpha \in (0, 1) \cup (1, \infty)$ (with the exception of additivity holding for $H_\alpha^{\mathrm{adv},\Phi}(\mathcal{N})$ for $\alpha \in [1/2, 1) \cup (1, \infty)$), due to the facts that the Choi state of a tensor-product channel is equal to the tensor product of the Choi states of the individual channels, as well as the additivity of the underlying

conditional entropies, for $H_\alpha^{\mathrm{adv},\Phi}(\mathcal{N})$ shown in Ref. [81] and for $\overline{H}_\alpha^{\mathrm{adv},\Phi}(\mathcal{N})$ following from the quantum Sibson identity [[96], Lemma 7] (see also Ref. [[74], Lemma 1]). Normalization and reduction to states (as in Proposition 5) follows for all of the above quantities. What remains is monotonicity under random unitary superchannels, and what we can show is something stronger: monotonicity under doubly stochastic superchannels, the latter defined in Ref. [18] as superchannels $\Theta$ such that their adjoint $\Theta^\dagger$ is also a superchannel, where the adjoint is defined with respect to the inner product for supermaps considered in Ref. [18].

*Theorem 22.* Let $\Theta$ be a doubly stochastic superchannel given by

$$\Theta[\mathcal{N}_{A\to B}] \equiv \Omega_{BE\to D} \circ \mathcal{N}_{A\to B} \circ \Lambda_{C\to AE}, \quad (241)$$

with $\Omega_{BE\to D}$ and $\Lambda_{C\to AE}$ quantum channels, $E$ a quantum memory system, $|A| = |C|$, and $|B| = |D|$. Then, for **H** any of the entropy functions in Eqs. (236)–(240), the following inequality holds:

$$\mathbf{H}(\Theta(\mathcal{N}_{A\to B})) \geqslant \mathbf{H}(\mathcal{N}_{A\to B}). \quad (242)$$

The inequality above holds for $\alpha \in [1/2, 1) \cup (1, \infty)$ for the functions in Eqs. (237) and (239) and for $\alpha \in (0, 1) \cup (1, 2]$ for the functions in Eqs. (238) and (240).

*Proof.* Recall from Ref. [18] that, since $\Theta$ is doubly stochastic, we have that

$$\mathrm{Tr}_E\{\Lambda_{C\to AE}(I_C)\} = I_A, \quad (243)$$

$$\Omega_{BE\to D}(I_B \otimes \rho_{ER}) = I_D \otimes \rho_R. \quad (244)$$

Let $\Theta$ be as above, and let us begin by considering the adversarial quantities for the ranges of $\alpha$ for which data processing holds. Let $\omega_R$ be an arbitrary state. Let $\xi_{AER} \equiv \Lambda_{C\to AE}(\Phi_{CR})$, and note that the marginal $\xi_A$ is the maximally mixed state due to Eq. (243) and the dimension constraint $|A| = |C|$. Therefore, there exists a quantum channel $\mathcal{E}_{R\to ER}$ such that

$$\xi_{AER} = \mathcal{E}_{R\to ER}(\Phi_{AR}). \quad (245)$$

Let $\sigma_{ER} \equiv \mathcal{E}_{R\to ER}(\omega_R)$. With these notations set, and working with the specific entropy function in Eq. (239), we find that

$$H_\alpha^{\mathrm{adv},\Phi}(\Theta(\mathcal{N}_{A\to B})) \geqslant -D_\alpha\{\Theta(\mathcal{N}_{A\to B})(\Phi_{CR})\|I_D \otimes \sigma_R\} \quad (246)$$

$$= -D_\alpha(\Omega_{BE\to D} \circ \mathcal{N}_{A\to B} \circ \Lambda_{C\to AE}(\Phi_{CR})\|I_D \otimes \sigma_R) \quad (247)$$

$$= -D_\alpha(\Omega_{BE\to D} \circ \mathcal{N}_{A\to B} \circ \Lambda_{C\to AE}(\Phi_{CR})\|\Omega_{BE\to D}(I_B \otimes \sigma_{ER})) \quad (248)$$

$$\geqslant -D_\alpha(\mathcal{N}_{A\to B} \circ \Lambda_{C\to AE}(\Phi_{CR})\|I_B \otimes \sigma_{ER}) \quad (249)$$

$$= -D_\alpha(\mathcal{N}_{A\to B}(\xi_{AER})\|I_B \otimes \sigma_{ER}) \quad (250)$$

$$= -D_\alpha(\mathcal{E}_{R\to ER} \circ \mathcal{N}_{A\to B}(\Phi_{AR})\|\mathcal{E}_{R\to ER}(I_B \otimes \omega_R)) \quad (251)$$

$$\geqslant -D_\alpha(\mathcal{N}_{A\to B}(\Phi_{AR})\|I_B \otimes \omega_R). \quad (252)$$

Since the inequality holds for an arbitrary state $\omega_R$, we conclude that

$$H_\alpha^{\mathrm{adv},\Phi}(\Theta(\mathcal{N}_{A\to B})) \geqslant H_\alpha^{\mathrm{adv},\Phi}(\mathcal{N}_{A\to B}), \quad (253)$$

which is the inequality in Eq. (242) for the adversarial Choi Rényi entropy $H_\alpha^{\mathrm{adv},\Phi}(\mathcal{N})$. The proof for the entropy functions in Eqs. (236) and (240) goes the same way, since the above proof only relied upon the data processing inequality.

To arrive at the inequality in Eq. (242) for the entropy functions in Eqs. (237)–(238), we exploit the same proof, but we choose $\omega_R$ to be the maximally mixed state. By tracing over systems $AE$ in Eq. (245), we find that

$$\pi_R = \mathrm{Tr}_{AE}\{\xi_{AER}\} = \mathrm{Tr}_{AE}\{\mathcal{E}_{R\to ER}(\Phi_{AR})\} \quad (254)$$

$$= (\mathrm{Tr}_E \circ \mathcal{E}_{R\to ER})(\pi_A), \quad (255)$$

and so we conclude that the reduced channel $\mathrm{Tr}_E \circ \mathcal{E}_{R\to ER}$ is unital. This means that, by choosing $\sigma_{ER} \equiv \mathcal{E}_{R\to ER}(\omega_R)$ again, we can conclude that $\sigma_R = \pi_R$. By applying the same steps as above, we then find that

$$H_\alpha^\Phi(\Theta(\mathcal{N}_{A\to B})) = -D_\alpha\{\Theta(\mathcal{N}_{A\to B})(\Phi_{CR})\|I_D \otimes \pi_R\} \quad (256)$$

$$\geqslant -D_\alpha(\mathcal{N}_{A\to B}(\Phi_{AR})\|I_B \otimes \pi_R) \quad (257)$$

$$= H_\alpha^\Phi(\mathcal{N}_{A\to B}), \quad (258)$$

which is the inequality in Eq. (242) for the entropy function in Eq. (237). The proof for the entropy function in Eq. (238) then goes the same way. ∎

As a final remark to conclude this section, we note that the following limit holds

$$\lim_{\alpha\to\infty} H_\alpha^{\mathrm{adv},\Phi}(\mathcal{N}_{A\to B}) = H_{\min}^{\mathrm{ext}}(\mathcal{N}_{A\to B}), \quad (259)$$

as a consequence of Eq. (135), and so the proof given above represents a different way, from that given in Ref. [18], for arriving at the conclusion that the extended min-entropy of a channel is nondecreasing under the action of a doubly stochastic superchannel.

## IX. CONCLUSION AND OUTLOOK

In this paper, we have introduced a definition for the entropy of a quantum channel, based on the channel relative entropy between the channel of interest and the completely randomizing channel. Building on this approach, we defined the Rényi and min-entropy of a channel. We proved that these channel entropies satisfy the axioms for entropy functions, recently put forward in Ref. [18]. We also proved that the entropy of a channel is equal to the completely bounded entropy of Ref. [19], and the Rényi entropy of a channel is related to the completely bounded $1 \to p$ norm considered in

Ref. [19]. The smoothed min-entropy of a channel satisfies an asymptotic equipartition property that generalizes the same property for smoothed min-entropy of quantum states [88]. We showed that the entropy of a channel has an operational interpretation in terms of a task called quantum channel merging, in which the goal is for the receiver to merge his share of the channel with the environment's share, and this task is a dynamical counterpart of the known task of quantum state merging [20,21]. We evaluated the entropy of a channel for several common channel models. Finally, we considered other generalized entropies of a quantum channel and gave further evidence that Definition 1 is a cogent approach for defining entropy of a quantum channel.

Going forward from here, one of the most interesting open questions is to determine if there is a set of axioms that uniquely identifies the entropy of a quantum channel, similar to how there is a set of axioms that uniquely characterizes Shannon entropy [101]. We wonder the same for the Rényi entropy of a channel, given that the Rényi entropies were originally identified [102] by removing one of the axioms that uniquely characterizes Shannon entropy. On a different front, one could alternatively define the entropy of $n$ uses of a quantum channel in terms of an optimization over quantum costrategies [103,104] or quantum combs [105], and for analyzing the asymptotic equipartition property in this scenario, one could alternatively smooth with respect to the strategy norm of Refs. [105,106]. The results of Ref. [15] suggest that the asymptotic equipartition property might still hold in this more complex scenario, but further analysis is certainly required. Note that a related scenario has been considered recently in Ref. [107]. Finally, if the Petz–Rényi channel divergence between an arbitrary channel and the completely depolarizing channel is additive, then a Rényi channel entropy defined from it would be convincing. This question about Petz–Rényi channel divergence has been open since Ref. [15].

*Note added.* After completing the results in our related preprint [108], we noticed Ref. [[25], Eq. (6)], in which Yuan proposed to define the entropy of a quantum channel in the same way as we have proposed in Definition 1. Yuan's work is now published as Ref. [25].

## ACKNOWLEDGMENTS

## APPENDIX A: QUANTUM CHANNEL MERGING CAPACITY PROOF

This Appendix details the proof of Theorem 10.

### 1. Converse bound

Let us begin by considering the converse part, following the approach given in Ref. [21] for quantum state merging.

*Proposition 23.* Fix $n, L, K \in \mathbb{N}$ and $\varepsilon \in [0, 1]$. Let $\mathcal{N}_{A \to B}$ be a quantum channel. Then an $(n, L/K, \varepsilon)$ quantum channel merging protocol for $\mathcal{N}_{A \to B}$ satisfies the following bound:

$$\frac{1}{n}[(1 - \sqrt{\varepsilon}) \log_2 L - \log_2 K]$$
$$\leqslant H(\mathcal{N}) + \sqrt{\varepsilon} \log_2 |A| + g_2(\sqrt{\varepsilon}). \quad \text{(A1)}$$

*Proof.* We closely follow the approach given in Ref. [[21], Sec. IV-B], which established the converse part of the quantum state merging theorem. Consider an arbitrary $(n, L/K, \varepsilon)$ quantum channel merging protocol of the form described above. To prove the converse, we can really employ any entanglement measure that reduces to the entropy of entanglement for pure states and is asymptotically continuous. So let us choose the entanglement of formation [109], which is defined for a bipartite state $\rho_{AB}$ as

$$E_F(A; B)_\rho$$
$$\equiv \inf \left\{ \sum_x p_X(x) H(A)_{\psi^x} : \rho_{AB} = \sum_x p_X(x) \psi^x_{AB} \right\}, \quad \text{(A2)}$$

where the infimum is with respect to all convex decompositions of $\rho_{AB}$ into pure states $\psi^x_{AB}$. The entanglement of formation does not increase under the action of an LOCC channel [109]. For the purposes of the converse, as in Ref. [[21], Sec. IV-B], we imagine that the reference party $R$ is working together with Bob $B$, and they are spatially separated from Eve $E$. Let $\omega_{R\widetilde{B}^n_E \widetilde{E}^n \overline{B}_1 \overline{E}_1}$ and $\widetilde{\omega}_{R\widetilde{B}^n_E \widetilde{E}^n \overline{B}_1 \overline{E}_1}$ denote the following respective states:

$$\omega_{R\widetilde{B}^n_E \widetilde{E}^n \overline{B}_1 \overline{E}_1}$$
$$\equiv \left[ \text{id}^{\otimes n}_{BE \to \widetilde{B}_E \widetilde{E}} \circ \left( \mathcal{U}^{\mathcal{N}}_{A \to BE} \right)^{\otimes n} \right] (\psi_{RA^n}) \otimes \Phi^L_{\overline{B}_1 \overline{E}_1}, \quad \text{(A3)}$$

$$\widetilde{\omega}_{R\widetilde{B}^n_E \widetilde{E}^n \overline{B}_1 \overline{E}_1}$$
$$\equiv \mathcal{P}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}^n_E \widetilde{E}^n \overline{B}_1 \overline{E}_1} \left( \left[ \left( \mathcal{U}^{\mathcal{N}}_{A \to BE} \right)^{\otimes n} (\psi_{RA^n}) \right] \otimes \Phi^K_{\overline{B}_0 \overline{E}_0} \right). \quad \text{(A4)}$$

Define

$$f(n, \varepsilon, |A|, L) \equiv \sqrt{\varepsilon} n \log_2 |A| + \sqrt{\varepsilon} \log_2 L + g_2(\sqrt{\varepsilon}). \quad \text{(A5)}$$

We then have that

$$\log_2 L + H(R)_\omega$$
$$= H(\overline{B}_1)_\omega + H(R)_\omega \quad \text{(A6)}$$
$$= H(R\overline{B}_1)_\omega \quad \text{(A7)}$$
$$= E_F(R\overline{B}_1; \widetilde{B}^n_E \widetilde{E}^n \overline{E}_1)_\omega \quad \text{(A8)}$$
$$\leqslant E_F(R\overline{B}_1; \widetilde{B}^n_E \widetilde{E}^n \overline{E}_1)_{\widetilde{\omega}} + f(n, \varepsilon, |A|, L). \quad \text{(A9)}$$

The first equality follows because $\log_2 L = H(\overline{B}_1)_\omega$ for a maximally entangled state of Schmidt rank $L$. The second equality follows because quantum entropy is additive with respect to product states. The third equality follows because the entanglement of formation reduces to entropy of entanglement for pure states. The inequality is a consequence of the uniform continuity bound from Ref. [[94], Corollary 4]. Continuing,

we have that

$$
E_F\big(R\overline{B}_1; \widetilde{B}_E^n \widetilde{E}^n \overline{E}_1\big)_{\widetilde{\omega}}
$$

$$
\leqslant E_F(RB^n\overline{B}_0; E^n\overline{E}_0)_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\psi)\otimes\Phi^K} \tag{A10}
$$

$$
= H(RB^n\overline{B}_0)_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\psi)\otimes\Phi^K} \tag{A11}
$$

$$
= H(RB^n)_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\psi)} + H(\overline{B}_0)_{\Phi^K} \tag{A12}
$$

$$
= H(RB^n)_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\psi)} + \log_2 K. \tag{A13}
$$

The first inequality follows from LOCC monotonicity of the entanglement of formation under the action of the one-way LOCC channel $\mathcal{P}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}_E^n \widetilde{E}^n \overline{B}_1 \overline{E}_1}$. The last three equalities follow for reasons similar to what have been given above. Putting everything together, we find that

$$
\log_2 M = \log_2 L - \log_2 K \tag{A14}
$$

$$
\leqslant H(B^n|R)_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\psi)} + f(n, \varepsilon, |A|, L). \tag{A15}
$$

Since the protocol is required to work for every possible input state $\psi_{RA^n}$, we conclude the following bound:

$$
\log_2 M \leqslant \inf_{\psi_{RA^n}} H(B^n|R)_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\psi)} + f(n, \varepsilon, |A|, L) \tag{A16}
$$

$$
= nH(\mathcal{N}) + f(n, \varepsilon, |A|, L), \tag{A17}
$$

with the equality following from the additivity of the entropy of a channel [19] (recalled here as Proposition 4). The inequality in the statement of the proposition follows by dividing by $n$ and rearranging. ∎

### 2. Achievability bound

Now let us consider the achievability part.

*Proposition 24.* Fix $n, L, K \in \mathbb{N}$ and $\varepsilon \in (0, 1)$. Let $\mathcal{N}_{A\to B}$ be a quantum channel. Then there exists an $(n, L/K, \varepsilon)$ channel merging protocol for $\mathcal{N}_{A\to B}$ such that its entanglement gain satisfies the following inequality for all $\alpha > 1$:

$$
\frac{1}{n}[\log_2 L - \log_2 K] \geqslant H_\alpha(\mathcal{N})
$$

$$
- \frac{\alpha}{n(\alpha-1)}[4\log_2(1/\varepsilon) + 4(|A|^2 - 1)\log_2(n+1)]
$$

$$
- \frac{\alpha}{n(\alpha-1)}[1/\alpha + 2\log_2 13]. \tag{A18}
$$

*Proof.* For the achievability part, we employ ideas used in the theory of quantum channel simulation [91,110–112]. In particular, the main challenge of quantum channel merging over quantum state merging is that it is necessary for the protocol to work for every possible state $\psi_{RA^n}$ that could be input, not merely for a fixed state input. In prior work on quantum channel simulation [91,110–112], this challenge has been met by appealing to the post-selection technique [[89], Theorem 1]. Here, we use the same approach. In the context of the post-selection technique, it is helpful to consult the unpublished note [113] for further details.

Let $\zeta_{A^n\hat{A}^n}$ denote the maximally mixed state of the symmetric subspace of the $A^n\hat{A}^n$ systems [90], where $\hat{A}$ is isomorphic to the channel input system $A$. Note that this state can be written as [[90], Proposition 6]

$$
\zeta_{A^n\hat{A}^n} = \int d\psi_{A\hat{A}} \, \psi_{A\hat{A}}^{\otimes n}, \tag{A19}
$$

where $\psi_{A\hat{A}}$ denotes a pure state and $d\psi_{A\hat{A}}$ is the Haar measure over the pure states. This state is permutation invariant; i.e., for a unitary channel $\mathcal{W}_{A^n}^\pi \otimes \mathcal{W}_{\hat{A}^n}^\pi$ corresponding to a permutation $\pi$, we have that $\zeta_{A^n\hat{A}^n} = (\mathcal{W}_{A^n}^\pi \otimes \mathcal{W}_{\hat{A}^n}^\pi)(\zeta_{A^n\hat{A}^n})$ for all $\pi \in S_n$, with $S_n$ denoting the symmetric group. Let $\zeta_{R'\hat{A}^n A^n}$ be a purification of $\zeta_{A^n\hat{A}^n}$, and note that it can be chosen such that [113]

$$
\zeta_{R'\hat{A}^n A^n} = \big(\mathcal{W}_{R'}^\pi \otimes \mathcal{W}_{\hat{A}^n}^\pi \otimes \mathcal{W}_{A^n}^\pi\big)(\zeta_{R'\hat{A}^n A^n}), \tag{A20}
$$

where $\mathcal{W}_{R'}^\pi$ is some unitary, which implies that

$$
\big(\mathcal{W}_{R'}^{\pi^{-1}} \otimes \mathcal{W}_{\hat{A}^n}^{\pi^{-1}}\big)(\zeta_{R'\hat{A}^n A^n}) = \mathcal{W}_{A^n}^\pi(\zeta_{R'\hat{A}^n A^n}). \tag{A21}
$$

The first goal is to show the existence of a state merging protocol for the state $(\mathcal{U}_{A\to BE}^{\mathcal{N}})^{\otimes n}(\zeta_{R'\hat{A}^n A^n})$. As shown in Ref. [[114], Theorem 5.2] (see also the earlier Ref. [[115], Proposition 4.7] in this context), there exists a state merging protocol with error $\sqrt{13\varepsilon'}$, with the entanglement gain satisfying

$$
\log_2 L - \log_2 K \geqslant H_{\min}^{\varepsilon'}(B^n|\hat{A}^n R')_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\zeta)}
$$

$$
- 2\log_2\left(\frac{1}{\varepsilon'}\right). \tag{A22}
$$

(To arrive at the inequality in Eq. (A22), one needs to use the fact that $P(\rho, \sigma) \geqslant \frac{1}{2}\|\rho - \sigma\|_1$ for any two states.) That is, there exists a one-way LOCC channel $\mathcal{P}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}_E^n \widetilde{E}^n \overline{B}_1 \overline{E}_1}$

such that the following inequality holds:

$$
\frac{1}{2}\left\|\big[\mathrm{id}_{BE\to\widetilde{B}_E\widetilde{E}}^{\otimes n} \circ (\mathcal{U}_{A\to BE}^{\mathcal{N}})^{\otimes n}\big](\zeta_{R'\hat{A}^n A^n}) \otimes \Phi_{\overline{B}_1\overline{E}_1}^L - \mathcal{P}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}_E^n \widetilde{E}^n \overline{B}_1 \overline{E}_1}\big(\big[(\mathcal{U}_{A\to BE}^{\mathcal{N}})^{\otimes n}(\zeta_{\hat{R}'A^n A^n})\big] \otimes \Phi_{\overline{B}_0\overline{E}_0}^K\big)\right\|_1 \leqslant \sqrt{13\varepsilon'}. \tag{A23}
$$

Now our goal is for Eq. (54) to be satisfied for all possible states $\psi_{RA^n}$. As a first step toward this goal, note that we can symmetrize the protocol $\mathcal{P}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}_E^n \widetilde{E}^n \overline{B}_1 \overline{E}_1}$ as follows:

$$
\overline{\mathcal{P}}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}_E^n \widetilde{E}^n \overline{B}_1 \overline{E}_1} \equiv \frac{1}{n!} \sum_{\pi \in S_n} \big(\mathcal{W}_{\widetilde{B}_E^n}^{\pi^{-1}} \otimes \mathcal{W}_{\widetilde{E}^n}^{\pi^{-1}}\big) \circ \mathcal{P}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}_E^n \widetilde{E}^n \overline{B}_1 \overline{E}_1} \circ \big(\mathcal{W}_{B^n}^\pi \otimes \mathcal{W}_{E^n}^\pi\big), \tag{A24}
$$

and the inequality in Eq. (A23) is still satisfied, i.e.,

$$
\frac{1}{2}\left\|\big[\mathrm{id}_{BE\to\widetilde{B}_E\widetilde{E}}^{\otimes n} \circ (\mathcal{U}_{A\to BE}^{\mathcal{N}})^{\otimes n}\big](\zeta_{R'\hat{A}^n A^n}) \otimes \Phi_{\overline{B}_1\overline{E}_1}^L - \overline{\mathcal{P}}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}_E^n \widetilde{E}^n \overline{B}_1 \overline{E}_1}\big(\big[(\mathcal{U}_{A\to BE}^{\mathcal{N}})^{\otimes n}(\zeta_{R'\hat{A}^n A^n})\big] \otimes \Phi_{\overline{B}_0\overline{E}_0}^K\big)\right\|_1 \leqslant \sqrt{13\varepsilon'}. \tag{A25}
$$

This follows from the unitary invariance and convexity of the trace norm, the permutation covariance of the maps $[\mathrm{id}_{BE \to \widetilde{B}_E \widetilde{E}}^{\otimes n} \circ (\mathcal{U}_{A \to BE}^{\mathcal{N}})^{\otimes n}]$ and $(\mathcal{U}_{A \to BE}^{\mathcal{N}})^{\otimes n}$:

$$\forall \pi \in S_n : \left( \mathcal{W}_{\widetilde{B}_E^n}^{\pi^{-1}} \otimes \mathcal{W}_{\widetilde{E}^n}^{\pi^{-1}} \right) \circ \mathrm{id}_{BE \to \widetilde{B}_E \widetilde{E}}^{\otimes n} \circ \left( \mathcal{U}_{A \to BE}^{\mathcal{N}} \right)^{\otimes n} \circ \mathcal{W}_{A^n}^{\pi} = \mathrm{id}_{BE \to \widetilde{B}_E \widetilde{E}}^{\otimes n} \circ \left( \mathcal{U}_{A \to BE}^{\mathcal{N}} \right)^{\otimes n}, \tag{A26}$$

$$\forall \pi \in S_n : \left( \mathcal{W}_{B^n}^{\pi^{-1}} \otimes \mathcal{W}_{E^n}^{\pi^{-1}} \right) \circ \left( \mathcal{U}_{A \to BE}^{\mathcal{N}} \right)^{\otimes n} \circ \mathcal{W}_{A^n}^{\pi} = \left( \mathcal{U}_{A \to BE}^{\mathcal{N}} \right)^{\otimes n}, \tag{A27}$$

and the equality in Eq. (A21). Furthermore, the symmetrization can be accomplished by one-way LOCC (Bob randomly picks $\pi$, applies $\mathcal{W}_{B^n}^{\pi}$, communicates the value to Eve, who applies $\mathcal{W}_{E^n}^{\pi}$ at the input and $\mathcal{W}_{\widetilde{B}_E^n}^{\pi^{-1}} \otimes \mathcal{W}_{\widetilde{E}^n}^{\pi^{-1}}$ at the output), and is thus free in our model. Since the symmetrized protocol, the target channel $[\mathrm{id}_{BE \to \widetilde{B}_E \widetilde{E}}^{\otimes n} \circ (\mathcal{U}_{A \to BE}^{\mathcal{N}})^{\otimes n}]$, and the channel $(\mathcal{U}_{A \to BE}^{\mathcal{N}})^{\otimes n}$ are permutation covariant, we can now invoke the post-selection technique [[89], Theorem 1] to conclude that as long as we choose $\varepsilon' = \varepsilon(n+1)^{-2(|A|^2-1)}$, then it is guaranteed that

$$\sup_{\psi_{RA^n}} \frac{1}{2} \left\| \left[ \mathrm{id}_{BE \to \widetilde{B}_E \widetilde{E}}^{\otimes n} \circ \left( \mathcal{U}_{A \to BE}^{\mathcal{N}} \right)^{\otimes n} \right] (\psi_{RA^n}) \otimes \Phi_{\overline{B}_1 \overline{E}_1}^L - \overline{\mathcal{P}}_{B^n E^n \overline{B}_0 \overline{E}_0 \to \widetilde{B}_E^n \widetilde{E}^n \overline{B}_1 \overline{E}_1} \left( \left[ \left( \mathcal{U}_{A \to BE}^{\mathcal{N}} \right)^{\otimes n} (\psi_{RA^n}) \right] \otimes \Phi_{\overline{B}_0 \overline{E}_0}^K \right) \right\|_1 \leqslant \sqrt{13\varepsilon}. \tag{A28}$$

Propagating this choice of $\varepsilon'$ to the quantity in Eq. (A22), this means that we require

$$\log_2 L - \log_2 K \geqslant H_{\min}^{\varepsilon(n+1)^{-2(|A|^2-1)}} (B^n | \hat{A}^n R')_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\zeta)} - 2\log_2 \left( \frac{1}{\varepsilon} \right) - 4(|A|^2 - 1) \log_2 (n+1). \tag{A29}$$

At this point, we invoke Ref. [[81], Eq. (6.92)], as well as the inequality $1 - \sqrt{1 - \delta^2} \geqslant \delta^2/2$ holding for all $\delta \in [0, 1]$, to conclude the following bound for $\alpha > 1$:

$$H_{\min}^{\varepsilon(n+1)^{-2(|A|^2-1)}} (B^n | \hat{A}^n R')_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\zeta)}$$

$$\geqslant H_\alpha(B^n | \hat{A}^n R')_{(\mathcal{U}^{\mathcal{N}})^{\otimes n}(\zeta) | (\mathcal{U}^{\mathcal{N}})^{\otimes n}(\zeta)} + \frac{2\log_2 \left( \varepsilon(n+1)^{-2(|A|^2-1)} \right) - 1}{\alpha - 1} \tag{A30}$$

$$\geqslant \inf_{\phi_{RA^n}} H_\alpha(B^n | R)_{\omega | \omega} - \frac{2\log_2(1/\varepsilon) + 4(|A|^2 - 1)\log_2 (n+1) - 1}{\alpha - 1} \tag{A31}$$

$$= H_\alpha(\mathcal{N}^{\otimes n}) - \frac{2\log_2(1/\varepsilon) + 4(|A|^2 - 1)\log_2 (n+1) - 1}{\alpha - 1} \tag{A32}$$

$$= n H_\alpha(\mathcal{N}) - \frac{2\log_2(1/\varepsilon) + 4(|A|^2 - 1)\log_2 (n+1) - 1}{\alpha - 1}, \tag{A33}$$

where the first equality follows from Proposition 17, with $\omega_{RB^n} \equiv \mathcal{N}_{A \to B}^{\otimes n}(\phi_{RA^n})$, and the last equality critically relies upon the additivity $H_\alpha(\mathcal{N}^{\otimes n}) = n H_\alpha(\mathcal{N})$ from Proposition 15, which in turn directly follows from the main result of Ref. [19]. Putting everything together, we conclude that for $\varepsilon \in (0, 1/13)$, there exists an $(n, L/K, \sqrt{13\varepsilon})$ channel merging protocol for $\mathcal{N}_{A \to B}$ such that its entanglement gain satisfies the following inequality for all $\alpha > 1$:

$$\frac{1}{n}[\log_2 L - \log_2 K] \geqslant H_\alpha(\mathcal{N})$$
$$- \frac{\alpha}{n(\alpha - 1)} \left[ 2\log_2 \left( \frac{1}{\varepsilon} \right) + 4(|A|^2 - 1)\log_2(n+1) + \frac{1}{\alpha} \right]. \tag{A34}$$

We arrive at the statement of the proposition by a final substitution $\varepsilon'' = \sqrt{13\varepsilon} \in (0, 1)$, which implies that $\varepsilon = (\varepsilon'')^2/13$ and $2\log_2(1/\varepsilon) = 4\log_2(1/\varepsilon'') + 2\log_2 13$. ∎

### 3. Quantum channel merging capacity is equal to the entropy of a channel

We can now put together the previous two propositions to conclude the following theorem:

*Proof of Theorem* 10. By applying the limits $n \to \infty$ and $\varepsilon \to 0$, the following bound is a consequence of Proposition 23:

$$C_M(\mathcal{N}) \leqslant H(\mathcal{N}). \tag{A35}$$

For an arbitrary $\alpha > 1$, $\varepsilon \in (0, 1)$, and $\delta > 0$, we can conclude from Proposition 24 that there exists an $(n, 2^{n[H_\alpha(\mathcal{N}) - \delta]}, \varepsilon)$ channel merging protocol by taking $n$ sufficiently large. This implies that $H_\alpha(\mathcal{N})$ is an achievable rate for all $\alpha > 1$. However, since this statement is true for all $\alpha > 1$, we can conclude that the rate $\sup_{\alpha > 1} H_\alpha(\mathcal{N}) = H(\mathcal{N})$ is achievable also. This establishes that $C_M(\mathcal{N}) \geqslant H(\mathcal{N})$. ∎

### APPENDIX B: MAX-MUTUAL INFORMATION OF A CHANNEL AND THE ASYMPTOTIC EQUIPARTITION PROPERTY

In this Appendix, we point out how the max-mutual information of a quantum channel is a limit of the sandwiched Rényi mutual information of a channel, the latter having been defined in Ref. [116]. We then show how to arrive at an alternate proof of the asymptotic equipartition property in Ref. [[92], Theorem 8] by making use of this connection.

First recall that the sandwiched Rényi mutual information of a channel is defined for $\alpha \in (0, 1) \cup (1, \infty)$ as

[[116], Eq. (3.5)]

$$I_\alpha(\mathcal{N}) \equiv \max_{\psi_{RA}} I_\alpha(R;B)_\omega, \tag{B1}$$

where

$$\omega_{RB} \equiv \mathcal{N}_{A \to B}(\psi_{RA}), \tag{B2}$$

$$I_\alpha(R;B)_\omega \equiv \min_{\sigma_B} D_\alpha(\omega_{RB} \| \omega_R \otimes \sigma_B), \tag{B3}$$

where $D_\alpha$ is the sandwiched Rényi relative entropy from Eq. (100). It was subsequently used in Ref. [15]. The max-mutual information of a channel is equal to [[92], Definition 4]

$$I_{\max}(\mathcal{N}) \equiv \max_{\psi_{RA}} I_{\max}(R;B)_\omega, \tag{B4}$$

$$I_{\max}(R;B)_\omega \equiv \min_{\sigma_B} D_{\max}(\omega_{RB} \| \omega_R \otimes \sigma_B). \tag{B5}$$

*Proposition 25.* For a quantum channel $\mathcal{N}_{A \to B}$, the following limit holds:

$$I_{\max}(\mathcal{N}) = \lim_{\alpha \to \infty} I_\alpha(\mathcal{N}). \tag{B6}$$

*Proof.* To see this, consider that

$$\lim_{\alpha \to \infty} I_\alpha(\mathcal{N})$$

$$= \lim_{\alpha \to \infty} \max_{\psi_{RA}} \min_{\sigma_B} D_\alpha(\mathcal{N}_{A \to B}(\psi_{RA}) \| \psi_R \otimes \sigma_B) \tag{B7}$$

$$= \sup_{\alpha > 1} \max_{\psi_{RA}} \min_{\sigma_B} D_\alpha(\mathcal{N}_{A \to B}(\psi_{RA}) \| \psi_R \otimes \sigma_B) \tag{B8}$$

$$\leqslant \max_{\psi_{RA}} \min_{\sigma_B} \sup_{\alpha > 1} D_\alpha(\mathcal{N}_{A \to B}(\psi_{RA}) \| \psi_R \otimes \sigma_B) \tag{B9}$$

$$= \max_{\psi_{RA}} \min_{\sigma_B} D_{\max}(\mathcal{N}_{A \to B}(\psi_{RA}) \| \psi_R \otimes \sigma_B) \tag{B10}$$

$$= I_{\max}(\mathcal{N}). \tag{B11}$$

Now consider that

$$\lim_{\alpha \to \infty} I_\alpha(\mathcal{N})$$

$$= \sup_{\alpha > 1} \max_{\psi_{RA}} \min_{\sigma_B} D_\alpha(\mathcal{N}_{A \to B}(\psi_{RA}) \| \psi_R \otimes \sigma_B) \tag{B12}$$

$$\geqslant \sup_{\alpha > 1} \min_{\sigma_B} D_\alpha(\mathcal{N}_{A \to B}(\Phi_{RA}) \| \Phi_R \otimes \sigma_B) \tag{B13}$$

$$= \min_{\sigma_B} \sup_{\alpha > 1} D_\alpha(\mathcal{N}_{A \to B}(\Phi_{RA}) \| \Phi_R \otimes \sigma_B) \tag{B14}$$

$$= \min_{\sigma_B} D_{\max}(\mathcal{N}_{A \to B}(\Phi_{RA}) \| \Phi_R \otimes \sigma_B) \tag{B15}$$

$$= I_{\max}(\mathcal{N}), \tag{B16}$$

with the exchange of min and sup in the last line following from Ref. [[117], Corollary A.2]. The last equality follows from the remark after Ref. [[92], Definition 4]. ∎

The smoothed max-mutual information of a quantum channel $\mathcal{N}_{A \to B}$ is then defined for $\varepsilon \in (0, 1)$ as [[92], Definition 5]

$$I_{\max}^\varepsilon(\mathcal{N}) \equiv \inf_{\widetilde{\mathcal{N}} \, : \, P(\mathcal{N}, \widetilde{\mathcal{N}}) \leqslant \varepsilon} I_{\max}(\widetilde{\mathcal{N}}). \tag{B17}$$

(Here we smooth with respect to purified distance for convenience.) We then have that [[92], Theorem 8]

$$\lim_{n \to \infty} \frac{1}{n} I_{\max}^\varepsilon(\mathcal{N}^{\otimes n}) \leqslant I(\mathcal{N}), \tag{B18}$$

where $I(\mathcal{N})$ is the mutual information of a channel [26], defined as

$$I(\mathcal{N}) = \lim_{\alpha \to 1} \widetilde{I}_\alpha(\mathcal{N}) = \sup_{\psi_{RA}} I(R;B)_\omega, \tag{B19}$$

where $\omega_{RB} \equiv \mathcal{N}_{A \to B}(\psi_{RA})$.

To arrive at an alternate proof of the upper bound in Ref. [[92], Theorem 8], consider that an application of Ref. [[81], Eq. (6.92)], definitions, and arguments similar to those in the first part of Theorem 21 imply the following inequality for all $\alpha > 1$ and $\varepsilon \in (0, 1)$:

$$I_{\max}^\varepsilon(\mathcal{N}^{\otimes n}) \leqslant I_\alpha(\mathcal{N}^{\otimes n}) + f(\varepsilon, \alpha) \tag{B20}$$

$$= nI_\alpha(\mathcal{N}) + f(\varepsilon, \alpha), \tag{B21}$$

where the equality follows from Ref. [[116], Lemma 6] and $f(\varepsilon, \alpha)$ is a function of $\varepsilon$ and $\alpha$ that vanishes when dividing by $n$ and taking the large $n$ limit. Dividing by $n$ and taking the limit $n \to \infty$, we find that

$$\lim_{n \to \infty} \frac{1}{n} I_{\max}^\varepsilon(\mathcal{N}^{\otimes n}) \leqslant I_\alpha(\mathcal{N}). \tag{B22}$$

Since the inequality holds for all $\alpha > 1$, we can take the limit $\alpha \to 1$, apply Eq. (B19), and conclude the bound in Eq. (B18).

## APPENDIX C: DATA PROCESSING OF THE CHOI DIVERGENCE UNDER PARTICULAR SUPERCHANNELS

*Proposition 26.* Let $\Theta$ be a superchannel of the following form:

$$\Theta(\mathcal{N}_{A \to B}) = \sum_x p(x) \, \Omega_{BE \to D}^x \circ \mathcal{N}_{A \to B} \circ \Lambda_{C \to AE}^x, \tag{C1}$$

where $p(x)$ is a probability distribution, and for each $x$ the map $\Omega_{BE \to D}^x$ is an arbitrary quantum channel, and $\Lambda_{C \to AE}^x$ is a unital quantum channel (hence $|C| = |A||E|$). Then the Choi divergence is monotone under such superchannels:

$$\mathbf{D}^\Phi(\mathcal{N} \| \mathcal{M}) \geqslant \mathbf{D}^\Phi[\Theta(\mathcal{N}) \| \Theta(\mathcal{M})]. \tag{C2}$$

*Proof.* To prove it, we first prove the monotonicity under any superchannel of the form

$$\Upsilon(\mathcal{N}_{A \to B}) = \Omega_{BE \to D} \circ \mathcal{N}_{A \to B} \circ \Lambda_{C \to AE}, \tag{C3}$$

with $\Omega_{BE \to D}$ an arbitrary quantum channel, and $\Lambda_{C \to AE}$ a unital quantum channel. Indeed, denoting by $\Lambda_{AE \to C}^t$ the quantum channel obtained from $\Lambda_{C \to AE}$ by taking the transpose on each of its Kraus operators, and denoting by $\tilde{A}$, $\tilde{C}$, and $\tilde{E}$, replicas of systems $A$, $C$, and $E$,

we find that

$$\mathbf{D}^{\Phi}(\Upsilon(\mathcal{N})\|\Upsilon(\mathcal{M})) = \mathbf{D}(\Upsilon(\mathcal{N}_{A\to B})(\Phi_{\tilde{C}C})\|\Upsilon(\mathcal{M}_{A\to B})(\Phi_{\tilde{C}C})) \tag{C4}$$

$$= \mathbf{D}((\Omega_{BE\to D}\circ\mathcal{N}_{A\to B}\circ\Lambda_{C\to AE})(\Phi_{\tilde{C}C})\|(\Omega_{BE\to D}\circ\mathcal{M}_{A\to B}\circ\Lambda_{C\to AE})(\Phi_{\tilde{C}C})) \tag{C5}$$

$$\leqslant \mathbf{D}((\mathcal{N}_{A\to B}\circ\Lambda_{C\to AE})(\Phi_{\tilde{C}C})\|(\mathcal{M}_{A\to B}\circ\Lambda_{C\to AE})(\Phi_{\tilde{C}C})) \tag{C6}$$

$$= \mathbf{D}\big((\Lambda_{\tilde{A}\tilde{E}\to\tilde{C}}^{t}\circ\mathcal{N}_{A\to B})(\Phi_{\tilde{A}A}\otimes\Phi_{\tilde{E}E})\|(\Lambda_{\tilde{A}\tilde{E}\to\tilde{C}}^{t}\circ\mathcal{M}_{A\to B})(\Phi_{\tilde{A}A}\otimes\Phi_{\tilde{E}E})\big) \tag{C7}$$

$$\leqslant \mathbf{D}(\mathcal{N}_{A\to B}(\Phi_{\tilde{A}A}\otimes\Phi_{\tilde{E}E})\|\mathcal{M}_{A\to B}(\Phi_{\tilde{A}A}\otimes\Phi_{\tilde{E}E})) \tag{C8}$$

$$= \mathbf{D}(\mathcal{N}_{A\to B}(\Phi_{\tilde{A}A})\|\mathcal{M}_{A\to B}(\Phi_{\tilde{A}A})) \tag{C9}$$

$$= \mathbf{D}^{\Phi}(\mathcal{N}\|\mathcal{M}), \tag{C10}$$

where, in both inequalities, we used data processing of the divergence **D**, for the second equality we used the relation $\Lambda_{C\to AE}(\Phi_{\tilde{C}C}) = \Lambda_{\tilde{A}\tilde{E}\to\tilde{C}}^{t}(\Phi_{\tilde{A}A}\otimes\Phi_{\tilde{E}E})$, and for the third equality we used the property $\mathbf{D}(\rho\otimes\omega\|\sigma\otimes\omega) = \mathbf{D}(\rho\|\sigma)$ [69]. Now, to prove the monotonicity under $\Theta$ as in Eq. (C1), we write $\Theta = \sum_{x}p(x)\Upsilon_{x}$, where each $\Upsilon_{x}$ has the form Eq. (C3). With this notation, we find that

$$\mathbf{D}^{\Phi}(\Theta(\mathcal{N})\|\Theta(\mathcal{M})) = \mathbf{D}\left(\sum_{x}p(x)\Upsilon_{x}(\mathcal{N}_{A\to B})(\Phi_{\tilde{C}C})\right\|\left.\sum_{x}p(x)\Upsilon_{x}(\mathcal{M}_{A\to B})(\Phi_{\tilde{C}C})\right) \tag{C11}$$

$$= \mathbf{D}\left(\mathrm{Tr}_{X}\left[\sum_{x}p(x)\Upsilon_{x}(\mathcal{N}_{A\to B})(\Phi_{\tilde{C}C})\otimes|x\rangle\langle x|_{X}\right]\right\|\left.\mathrm{Tr}_{X}\left[\sum_{x}p(x)\Upsilon_{x}(\mathcal{M}_{A\to B})(\Phi_{\tilde{C}C})\otimes|x\rangle\langle x|_{X}\right]\right) \tag{C12}$$

$$\leqslant \mathbf{D}\left(\sum_{x}p(x)\Upsilon_{x}(\mathcal{N}_{A\to B})(\Phi_{\tilde{C}C})\otimes|x\rangle\langle x|_{X}\right\|\left.\sum_{x}p(x)\Upsilon_{x}(\mathcal{M}_{A\to B})(\Phi_{\tilde{C}C})\otimes|x\rangle\langle x|_{X}\right) \tag{C13}$$

$$\leqslant \mathbf{D}\left(\sum_{x}p(x)\mathcal{N}_{A\to B}(\Phi_{\tilde{A}A})\otimes|x\rangle\langle x|_{X}\right\|\left.\sum_{x}p(x)\mathcal{M}_{A\to B}(\Phi_{\tilde{A}A})\otimes|x\rangle\langle x|_{X}\right) \tag{C14}$$

$$= \mathbf{D}(\mathcal{N}_{A\to B}(\Phi_{\tilde{A}A})\|\mathcal{M}_{A\to B}(\Phi_{\tilde{A}A})) \tag{C15}$$

$$= \mathbf{D}^{\Phi}(\mathcal{N}\|\mathcal{M}), \tag{C16}$$

where, in the first inequality, we used the monotonicity of the divergence under data processing, and for the second inequality, we used the monotonicity under maps of the form in Eq. (C3). ∎

## APPENDIX D: OPTIMIZING THE ADVERSARIAL CHANNEL DIVERGENCE

By definition, we always have that

$$\mathbf{D}^{\mathrm{adv}}(\mathcal{N}\|\mathcal{M}) \geqslant \sup_{\psi_{RA}}\inf_{\sigma_{RA}}\mathbf{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{M}_{A\to B}(\sigma_{RA})), \tag{D1}$$

where $\psi_{RA}$ is pure with system $R$ isomorphic to system $A$.

To see the claim after Eq. (219), let $\rho_{RA}$ be an arbitrary state with purification $\phi_{R'RA}$. It thus holds that $\phi_{R'RA}$ is a purification of $\rho_{A}$, with $R'R$ acting as the purifying systems. By taking a "canonical" purification of $\rho_{A}$ that is in direct correspondence with its eigendecomposition, there exists a purification $\varphi_{SA}$ of $\rho_{A}$ with system $S$ isomorphic to system $A$. Since the purification $\phi_{R'RA}$ is related by an isometric channel $\mathcal{U}_{S\to R'R}$ to the purification $\varphi_{SA}$ as $\phi_{R'RA} = \mathcal{U}_{S\to R'R}(\varphi_{SA})$ and applying the isometric invariance of generalized divergences [118], we conclude for an arbitrary state $\omega_{SA}$ that

$$\mathbf{D}(\mathcal{N}_{A\to B}(\varphi_{SA})\|\mathcal{M}_{A\to B}(\omega_{SA}))$$
$$= \mathbf{D}((\mathcal{U}_{S\to R'R}\circ\mathcal{N}_{A\to B})(\varphi_{SA})\|(\mathcal{U}_{S\to R'R}\circ\mathcal{M}_{A\to B})(\omega_{SA})) \tag{D2}$$

$$= \mathbf{D}\{\mathcal{N}_{A\to B}(\phi_{R'RA})\|\mathcal{M}_{A\to B}(\mathcal{U}_{S\to R'R}(\omega_{SA}))\} \tag{D3}$$

$$\geqslant \mathbf{D}\{\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}((\mathrm{Tr}_{R'}\circ\mathcal{U}_{S\to R'R})(\omega_{SA}))\} \tag{D4}$$

$$\geqslant \inf_{\sigma_{RA}}\mathbf{D}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\sigma_{RA})). \tag{D5}$$

The first inequality is from data processing under the partial trace over $R'$. Since the inequality holds for arbitrary $\omega_{SA}$, we conclude that

$$\inf_{\sigma_{RA}}\mathbf{D}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\sigma_{RA}))$$
$$\leqslant \inf_{\omega_{SA}}\mathbf{D}(\mathcal{N}_{A\to B}(\varphi_{SA})\|\mathcal{M}_{A\to B}(\omega_{SA})). \tag{D6}$$

We can then take a supremum to conclude that

$$\inf_{\sigma_{RA}}\mathbf{D}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\sigma_{RA}))$$
$$\leqslant \sup_{\varphi_{SA}}\inf_{\omega_{SA}}\mathbf{D}(\mathcal{N}_{A\to B}(\varphi_{SA})\|\mathcal{M}_{A\to B}(\omega_{SA})). \tag{D7}$$

Since the inequality holds for an arbitrary choice of $\rho_{RA}$, we conclude that

$$\mathbf{D}^{\mathrm{adv}}(\mathcal{N}\|\mathcal{M}) \leqslant \sup_{\varphi_{SA}}\inf_{\omega_{SA}}\mathbf{D}(\mathcal{N}_{A\to B}(\varphi_{SA})\|\mathcal{M}_{A\to B}(\omega_{SA})). \tag{D8}$$

This concludes the proof.

[1] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932).

[2] B. Schumacher, Quantum coding, Phys. Rev. A **51**, 2738 (1995).

[3] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Concentrating partial entanglement by local operations, Phys. Rev. A **53**, 2046 (1996).

[4] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Resource Theory of Quantum States out of Thermal Equilibrium, Phys. Rev. Lett. **111**, 250404 (2013).

[5] H. Umegaki, Conditional expectations in an operator algebra IV (entropy and information), Kodai Math. Sem. Rep. **14**, 59 (1962).

[6] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, Thermodynamical Approach to Quantifying Quantum Correlations, Phys. Rev. Lett. **89**, 180402 (2002).

[7] M. Horodecki, K. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen(De), and U. Sen, Local Information as a Resource in Distributed Quantum Systems, Phys. Rev. Lett. **90**, 100402 (2003).

[8] M. Horodecki, P. Horodecki, and J. Oppenheim, Reversible transformations from pure to mixed states and the unique measure of information, Phys. Rev. A **67**, 062104 (2003).

[9] J. Oppenheim, K. Horodecki, M. Horodecki, P. Horodecki, and R. Horodecki, Mutually exclusive aspects of information carried by physical systems: Complementary between local and nonlocal information, Phys. Rev. A **68**, 022307 (2003).

[10] I. Devetak, Distillation of local purity from quantum states, Phys. Rev. A **71**, 062303 (2005).

[11] H. Krovi and I. Devetak, Local purity distillation with bounded classical communication, Phys. Rev. A **76**, 012321 (2007).

[12] F. Hiai and D. Petz, The proper formula for relative entropy and its asymptotics in quantum probability, Commun. Math. Phys. **143**, 99 (1991).

[13] T. Ogawa and H. Nagaoka, Strong converse and Stein's lemma in quantum hypothesis testing, IEEE Trans. Inf. Theory **46**, 2428 (2000).

[14] A. S. Holevo, *Quantum Systems, Channels, Information, de Gruyter Studies in Mathematical Physics*, Book 16 (de Gruyter, Berlin, 2012).

[15] T. Cooney, M. Mosonyi, and M. M. Wilde, Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication, Commun. Math. Phys. **344**, 797 (2016).

[16] F. Leditzky, E. Kaur, N. Datta, and M. M. Wilde, Approaches for approximate additivity of the Holevo information of quantum channels, Phys. Rev. A **97**, 012332 (2018).

[17] G. Lindblad, Completely positive maps and entropy inequalities, Commun. Math. Phys. **40**, 147 (1975).

[18] G. Gour, Comparison of quantum channels with superchannels, IEEE Trans. Inf. Theory **65**, 5880 (2019).

[19] I. Devetak, C. King, M. Junge, and M. B. Ruskai, Multiplicativity of completely bounded *p*-norms implies a new additivity result, Commun. Math. Phys. **266**, 37 (2006).

[20] M. Horodecki, J. Oppenheim, and A. Winter, Partial quantum information, Nature **436**, 673 (2005).

[21] M. Horodecki, J. Oppenheim, and A. Winter, Quantum state merging and negative information, Commun. Math. Phys. **269**, 107 (2007).

[22] W. Roga, K. Zyczkowski, and M. Fannes, Entropic characterization of quantum operations, Int. J. Quantum. Inform. **9**, 1031 (2011).

[23] W. Roga, Entropy of quantum channel in the theory of quantum information, Ph.D. thesis, Uniwersytet Jagiellonski, 2011.

[24] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Transforming quantum operations: Quantum supermaps, Europhys. Lett. **83**, 30004 (2008).

[25] X. Yuan, Hypothesis testing and entropies of quantum channels, Phys. Rev. A **99**, 032317 (2019).

[26] C. Adami and N. J. Cerf, von Neumann capacity of noisy quantum channels, Phys. Rev. A **56**, 3470 (1997).

[27] R. Alicki, Isotropic quantum spin channels and additivity questions, arXiv:quant-ph/0402080 (2004).

[28] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, Amortized channel divergence for asymptotic quantum channel discrimination, Lett Math Phys **110**, 2277 (2020).

[29] W. F. Stinespring, Positive functions on C*-algebras, Proc. Am. Math. Soc. **6**, 211 (1955).

[30] A. S. Holevo, The entropy gain of infinite-dimensional quantum evolutions, Doklady Mathematics **82**, 730 (2010).

[31] A. S. Holevo, Entropy gain and the Choi-Jamiolkowski correspondence for infinite-dimensional quantum evolutions, Theor. Math. Phys. **166**, 123 (2011).

[32] A. S. Holevo, The entropy gain of quantum channels, in *Proceedings of the IEEE International Symposium on Information Theory Proceedings, St. Petersburg, Russia* (IEEE, Piscataway, NJ, 2011), pp. 289–292.

[33] D. Yang, K. Horodecki, and A. Winter, Distributed Private Randomness Distillation, Phys. Rev. Lett. **123**, 170501 (2019).

[34] E. H. Lieb and M. B. Ruskai, Proof of the strong subadditivity of quantum-mechanical entropy, J. Math. Phys. **14**, 1938 (1973).

[35] E. H. Lieb and M. B. Ruskai, A fundamental property of quantum-mechanical entropy, Phys. Rev. Lett. **30**, 434 (1973).

[36] H. Fawzi, J. Saunderson, and P. A. Parrilo, Semidefinite approximations of the matrix logarithm, Found. Comput. Math. **19**, 259 (2019).

[37] H. Fawzi and O. Fawzi, Efficient optimization of the quantum relative entropy, J. Phys. A: Math. Theor. **51**, 154003 (2018).

[38] E. Kaur and M. M. Wilde, Amortized entanglement of a quantum channel and approximately teleportation-simulable channels, J. Phys. A **51**, 035303 (2018).

[39] A. S. Holevo, Remarks on the classical capacity of quantum channel, arXiv:quant-ph/0212025 (2002).

[40] A. S. Holevo, Multiplicativity of *p*-norms of completely positive maps and the additivity problem in quantum information theory, Russian Math. Surveys **61**, 301 (2006).

[41] S. Das, S. Bäuml, and M. M. Wilde, Entanglement and secret-key-agreement capacities of bipartite quantum interactions and read-only memory devices, Phys. Rev. A **101**, 012344 (2020).

[42] M. Grassl, T. Beth, and T. Pellizzari, Codes for the quantum erasure channel, Phys. Rev. A **56**, 33 (1997).

[43] R. F. Werner and A. S. Holevo, Counterexample to an additivity conjecture for output purity of quantum channels, J. Math. Phys. **43**, 4353 (2002).

[44] A. A. Kuznetsova, Conditional entropy for infinite-dimensional quantum systems, Theory Probab. Appl. **55**, 709 (2011).

[45] H. Araki and E. H. Lieb, Entropy inequalities, Commun. Math. Phys. **18**, 160 (1970).

[46] A. Wehrl, Three theorems about entropy and convergence of density matrices, Rep. Math. Phys. **10**, 159 (1976).

[47] V. Baccetti and M. Visser, Infinite Shannon entropy, J. Stat. Mech.: Theory Exp. (2013) P04010.

[48] H. Falk, Inequalities of J. W. Gibbs, Am. J. Phys. **38**, 858 (1970).

[49] G. Lindblad, Entropy, information and quantum measurements, Commun. Math. Phys. **33**, 305 (1973).

[50] A. S. Holevo, Entanglement-assisted capacity of constrained channels, *Proc. SPIE 5128, First International Symposium on Quantum Informatics, Lipki, Russian Federation* (SPIE, 2003), pp. 62–69.

[51] A. S. Holevo, Entanglement-assisted capacities of constrained quantum channels, Theory Probab. Appl. **48**, 243 (2004).

[52] A. Serafini, *Quantum Continuous Variables* (CRC Press, Boca Raton, FL, 2017).

[53] A. S. Holevo and V. Giovannetti, Quantum channels and their entropic characteristics, Rep. Prog. Phys. **75**, 046001 (2012).

[54] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and J. H. Shapiro, Minimum output entropy of bosonic channels: A conjecture, Phys. Rev. A **70**, 032315 (2004).

[55] J. S. Ivan, K. K. Sabapathy, and R. Simon, Operator-sum representation for bosonic Gaussian channels, Phys. Rev. A **84**, 042311 (2011).

[56] K. Sharma, M. M. Wilde, S. Adhikari, and M. Takeoka, Bounding the energy-constrained quantum and private capacities of phase-insensitive Gaussian channels, New J. Phys. **20**, 063025 (2018).

[57] A. S. Holevo and R. F. Werner, Evaluating capacities of bosonic Gaussian channels, Phys. Rev. A **63**, 032312 (2001).

[58] V. Giovannetti, S. Lloyd, L. Maccone, and P. W. Shor, Broadband channel capacities, Phys. Rev. A **68**, 062323 (2003).

[59] M. M. Wilde, P. Hayden, and S. Guha, Quantum trade-off coding for bosonic communication, Phys. Rev. A **86**, 062306 (2012).

[60] S. Pirandola, R. García-Patrón, S. L. Braunstein, and S. Lloyd, Direct and Reverse Secret-Key Capacities of a Quantum Channel, Phys. Rev. Lett. **102**, 050503 (2009).

[61] P. Calabrese and J. Cardy, Entanglement entropy and conformal field theory, J. Phys. A: Math. Theor. **42**, 504005 (2009).

[62] T. Nishioka, S. Ryu, and T. Takayanagi, Holographic entanglement entropy: An overview, J. Phys. A: Math. Theor. **42**, 504008 (2009).

[63] S. N. Solodukhin, Entanglement entropy of black holes, Living Rev. Relativ. **14**, 8 (2011).

[64] M. Hayashi, M. Koashi, K. Matsumoto, F. Morikoshi, and A. Winter, Error exponents for entanglement concentration, J. Phys. A: Math. Gen. **36**, 527 (2003).

[65] M. Hayashi, Exponents of quantum fixed-length pure-state source coding, Phys. Rev. A **66**, 032321 (2002).

[66] D. Petz, Quasi-entropies for states of a von neumann algebra, Publ. Res. Inst. Math. Sci. **21**, 787 (1985).

[67] D. Petz, Quasi-entropies for finite quantum systems, Rep. Math. Phys. **23**, 57 (1986).

[68] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, J. Math. Phys. **54**, 122203 (2013).

[69] M. M. Wilde, A. Winter, and D. Yang, Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy, Commun. Math. Phys. **331**, 593 (2014).

[70] R. L. Frank and E. H. Lieb, Monotonicity of a relative Rényi entropy, J. Math. Phys. **54**, 122201 (2013).

[71] S. Beigi, Sandwiched Rényi divergence satisfies data processing inequality, J. Math. Phys. **54**, 122202 (2013).

[72] M. Mosonyi and T. Ogawa, Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies, Commun. Math. Phys. **334**, 1617 (2015).

[73] M. M. Wilde, Optimized quantum f-divergences and data processing, J. Phys. A **51**, 374002 (2018).

[74] M. Tomamichel, M. Berta, and M. Hayashi, Relating different quantum generalizations of the conditional Rényi entropy, J. Math. Phys. **55**, 082206 (2014).

[75] M. Hayashi, *Quantum Information: An Introduction* (Springer Verlag, Berlin, 2006).

[76] R. Renner, Security of quantum key distribution, Ph.D. thesis, ETH Zurich, 2005.

[77] N. Datta, Min- and max-relative entropies and a new entanglement monotone, IEEE Trans. Inf. Theory **55**, 2816 (2009).

[78] X. Wang and M. M. Wilde, Resource theory of asymmetric distinguishability, Phys. Rev. Research **1**, 033170 (2019).

[79] M. García Díaz, K. Fang, X. Wang, M. Rosati, M. Skotiniotis, J. Calsamiglia, and A. Winter, Using and reusing coherence to realize quantum processes, Quantum **2**, 100 (2018).

[80] R. Koenig, R. Renner, and C. Schaffner, The operational meaning of min- and max-entropy, IEEE Trans. Inf. Theory **55**, 4337 (2009).

[81] M. Tomamichel, *Quantum Information Processing with Finite Resources: Mathematical Foundations*, Springer Briefs in Mathematical Physics (Springer, Berlin, 2015), Vol. 5.

[82] A. E. Rastegin, Relative error of state-dependent cloning, Phys. Rev. A **66**, 042304 (2002).

[83] A. E. Rastegin, A lower bound on the relative error of mixed-state cloning and related operations, J. Opt. B: Quantum Semiclassical Opt. **5**, S647 (2003).

[84] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, Phys. Rev. A **71**, 062310 (2005).

[85] A. E. Rastegin, Sine distance for quantum states, arXiv:quant-ph/0602112 (2006).

[86] A. Uhlmann, The "transition probability" in the state space of a *-algebra, Rep. Math. Phys. **9**, 273 (1976).

[87] M. Tomamichel, R. Colbeck, and R. Renner, Duality between smooth min- and max-entropies, IEEE Trans. Inf. Theory **56**, 4674 (2010).

[88] M. Tomamichel, R. Colbeck, and R. Renner, A fully quantum asymptotic equipartition property, IEEE Trans. Inf. Theory **55**, 5840 (2009).

[89] M. Christandl, R. König, and R. Renner, Post-Selection Technique for Quantum Channels with Applications to Quantum Cryptography, Phys. Rev. Lett. **102**, 020504 (2009).

[90] A. Harrow, The church of the symmetric subspace, arXiv:1308.6595 (2013).

[91] M. Berta, M. Christandl, and R. Renner, The quantum reverse Shannon theorem based on one-shot information theory, Commun. Math. Phys. **306**, 579 (2011).

[92] K. Fang, X. Wang, M. Tomamichel, and M. Berta, Quantum channel simulation and the channel's smooth max-information, IEEE Trans. Inf. Theory **66**, 2129 (2019).

[93] A. Anshu, M. Berta, R. Jain, and M. Tomamichel, A minimax approach to one-shot entropy inequalities, J. Math. Phys. **60**, 122201 (2019).

[94] A. Winter, Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints, Commun. Math. Phys. **347**, 291 (2016).

[95] Y. Polyanskiy and S. Verdú, Arimoto channel coding converse and Rényi divergence, in *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computation, Monticello, IL* (IEEE, Pisataway, NJ, 2010), pp. 1327–1333.

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

[97] M. B. Hastings, Superadditivity of communication capacity using entangled inputs, Nat. Phys. **5**, 255 (2009).

[98] M. Hayashi and M. Tomamichel, Correlation detection and an operational interpretation of the Rényi mutual information, J. Math. Phys. **57**, 102201 (2016).

[99] M. Hayashi, Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding, Phys. Rev. A **76**, 062301 (2007).

[100] H. Nagaoka, The converse part of the theorem for quantum Hoeffding bound, arXiv:quant-ph/0611289 (2006).

[101] I. Csiszár, Axiomatic characterizations of information measures, Entropy **10**, 261 (2008).

[102] A. Rényi, On measures of entropy and information, *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, held at the Statistical Laboratory, University of California, 1960*, edited by J. Neyman (University of California Press, Berkeley, 1961), pp. 547–561.

[103] G. Gutoski and J. Watrous, Toward a general theory of quantum games, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing* (ACM, 2007), pp. 565–574.

[104] G. Gutoski, Quantum strategies and local operations, Ph.D. thesis, University of Waterloo, 2009.

[105] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Memory Effects in Quantum Channel Discrimination, Phys. Rev. Lett. **101**, 180501 (2008).

[106] G. Gutoski, On a measure of distance for quantum strategies, J. Math. Phys. **53**, 032202 (2012).

[107] G. Chiribella and D. Ebler, Optimal quantum networks and one-shot entropies, New J. Phys. **18**, 093053 (2016).

[108] G. Gour and M. M. Wilde, Entropy of a quantum channel, arXiv:1808.06980v1 (2018).

[109] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, Phys. Rev. A **54**, 3824 (1996).

[110] M. Berta, F. G. S. L. Brandão, M. Christandl, and S. Wehner, Entanglement cost of quantum channels, IEEE Trans. Inf. Theory **59**, 6779 (2013).

[111] M. Berta, J. M. Renes, and M. M. Wilde, Identifying the information gain of a quantum measurement, IEEE Trans. Inf. Theory **60**, 7987 (2014).

[112] M. Berta, Quantum side information: uncertainty relations, extractors, channel simulations, Ph.D. thesis, ETH Zurich, 2013.

[113] W. Matthews, The post-selection method: Bounding the diamond distance between covariant maps, Available at https://www.math.uwaterloo.ca/~wcleung/CO781-QMI-2010-lec20.pdf and https://www.scribd.com/document/386027566/ (2010).

[114] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, One-shot decoupling, Commun. Math. Phys. **328**, 251 (2014).

[115] M. Berta, Single-shot quantum state merging, Diploma thesis, ETH Zurich, 2009.

[116] M. Gupta and M. M. Wilde, Multiplicativity of completely bounded *p*-norms implies a strong converse for entanglement-assisted capacity, Commun. Math. Phys. **334**, 867 (2015).

[117] M. Mosonyi and F. Hiai, On the quantum Rényi relative entropies and related capacity formulas, IEEE Trans. Inf. Theory **57**, 2474 (2011).

[118] M. Tomamichel, M. M. Wilde, and A. Winter, Strong converse rates for quantum communication, IEEE Trans. Inf. Theory **63**, 715 (2017).