Strong Converse for the Feedback-Assisted Classical Capacity of Entanglement-Breaking Channels

D. Ding\textsuperscript{a,1,*} and M. M. Wilde\textsuperscript{b,2,**}

\textsuperscript{a}Department of Applied Physics, Stanford University, Stanford, USA
\textsuperscript{b}Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Center for Computation and Technology, Louisiana State University, Baton Rouge, USA

e-mail: *dding@stanford.edu, **mwilde@lsu.edu

Received July 7, 2017; in final form, October 20, 2017

Abstract—Quantum entanglement can be used in a communication scheme to establish a correlation between successive channel inputs that is impossible by classical means. It is known that the classical capacity of quantum channels can be enhanced by such entangled encoding schemes, but this is not always the case. In this paper, we prove that a strong converse theorem holds for the classical capacity of an entanglement-breaking channel even when it is assisted by a classical feedback link from the receiver to the transmitter. In doing so, we identify a bound on the strong converse exponent, which determines the exponentially decaying rate at which the success probability tends to zero, for a sequence of codes with communication rate exceeding capacity. Proving a strong converse, along with an achievability theorem, shows that the classical capacity is a sharp boundary between reliable and unreliable communication regimes. One of the main tools in our proof is the sandwiched Rényi relative entropy. The same method of proof is used to derive an exponential bound on the success probability when communicating over an arbitrary quantum channel assisted by classical feedback, provided that the transmitter does not use entangled encoding schemes.

DOI: 10.1134/S0032946018010015

1. INTRODUCTION

The classical communication theory is one of the modern successes of applied mathematics [1, 2]. It is arguably one of the foundations of our current information age and provides new ways of thinking about problems in many other fields of study, such as physics and in particular quantum mechanics. The interaction between these two fields is mutual; while some problems in quantum mechanics can be turned into communication problems, the existence of quantum phenomena strongly suggests that we should rethink many aspects of communication theory. Not only does the notion of a quantum state challenge what we mean by “information,” but the possibilities due to quantum mechanics give rise to new classes of communication protocols. With respect to this latter consideration, some fundamental motivating questions for quantum information theory have traditionally been and still are the following:

1. Is the theory of classical communication affected at a fundamental level by the consideration of quantum mechanical phenomena?
2. Does using quantum states and measurement for classical communication have practical advantages over using classical techniques?

\textsuperscript{1} Supported from a Stanford Graduate Fellowship.
\textsuperscript{2} Supported from startup funds from the Department of Physics and Astronomy at LSU, the NSF under Award No. CCF-1350397, and the DARPA Quiness Program through US Army Research Office award W31P4Q-12-1-0019.
In an attempt to answer these questions, one of the primary goals is to study the ability of a quantum channel to communicate classical information, i.e., bits. Like many communication problems, this ability is quantified by the notion of channel capacity. The classical capacity $C$ of a quantum channel $\mathcal{N}$ is defined to be the maximum rate of communication such that the decoding error probability can tend to zero in the limit of many channel uses. With this definition, one natural question is to determine how to compute the classical capacity. We know that the Holevo–Schumacher–Westmoreland (HSW) theorem \cite{3,4} provides a lower bound $C(\mathcal{N}) \geq \chi(\mathcal{N})$, where $\chi(\mathcal{N}) = \max_{\{p_X(x), \rho_x\}} I(X; B)_{\rho}$, and $I(X; B)_{\rho} \equiv H(X)_\rho + H(B)_\rho - H(XB)_\rho$ is the quantum mutual information of the following classical-quantum state:

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

where $\{|x\rangle\}$ is an orthonormal basis for the classical reference system and $H(G)_\sigma$ is the von Neumann entropy of a quantum state $\sigma_G$ on system $G$. The quantity $\chi(\mathcal{N})$ is called the Holevo information of the channel.

The classical capacity $C(\mathcal{N})$ can actually be formally rewritten in terms of the Holevo information as well, via a procedure known as regularization. Before doing so, note that we obtain the HSW theorem by considering only encoding procedures that do not use entangled inputs.\footnote{The highest possible rate with this restriction, proved by the HSW theorem to be $\chi(\mathcal{N})$, is then a lower bound on the classical capacity.} That is, each quantum system sent through the channel is not entangled with any other system that is sent. To get the classical capacity, it is generally necessary to incorporate entangled inputs into the calculation. The approach given by \cite{3,4} is to multiplex the channel so that one use of this multiplexed channel corresponds to multiple uses of the original channel. A multipartite entangled state describing several inputs across different uses of the original channel can now be simulated by the corresponding single input state to the multiplexed channel. We can therefore express the maximum rate for blocks of size $n$ in terms of the Holevo information:

$$\frac{1}{n} \chi(\mathcal{N}^\otimes n).$$

To obtain the classical capacity, we simply allow for inputs entangled across arbitrarily many channel uses:

$$C(\mathcal{N}) = \lim_{n \to \infty} \frac{1}{n} \chi(\mathcal{N}^\otimes n).$$

This idea put together with a converse theorem establishes the regularized expression in (2) as being equal to the classical capacity.

Unfortunately, computing the classical capacity this way is clearly intractable. This prompts us to look for special cases. We observe that the limit in (2) is equal to $\chi(\mathcal{N})$ if and only if the Holevo information satisfies a tensor-power additivity property (see Appendix A for a brief derivation):

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

Another way to restate the above is that entangled inputs do not increase capacity if and only if (3) is satisfied. However, this is not true for some channels, as was shown in \cite{5}. The next question then appears: What characterizes channels that satisfy (3)?
A sufficient condition for a quantum channel to satisfy tensor-power additivity is for it to be entanglement-breaking (EB) \cite{6}. Let $\mathcal{N}_{A\rightarrow B}$ denote a quantum channel, where the arrow notation indicates that the channel maps a state of the input system $A$ to a state in an output system $B$. An EB channel is defined such that for any bipartite state $\rho_{AA'}$, the output state 

$$(\mathcal{N}_{A\rightarrow B} \otimes \text{id}_{A'})(\rho_{AA'})$$

is separable. That is, the output of an EB channel can be written as a convex sum of product states. Effectively, the channel “breaks” the entanglement between $A$ and $A'$. Previous results have established that the Holevo information of EB channels is additive\footnote{Shor proved that the Holevo information of a tensor product of an EB channel with any other channel is equal to the sum of their respective Holevo informations. This form of additivity is stronger than (3) and is the one usually found in the literature.} \cite{7}. This is intuitive, since any entanglement of the inputs is broken by the channel. Following this line of thinking, one could consider making a stronger statement by allowing additional resources to assist the communication but not to the point that entanglement can be established. Indeed, \cite{8} proves a generalization of tensor-power additivity for EB channels with noiseless classical feedback. Furthermore, their results show that classical feedback does not increase the capacity of EB channels.

There are also possible stronger statements in another direction. The direct part of the original HSW theorem states that if the communication rate $R$ is less than the Holevo information $\chi$, then there exists a sequence of protocols $\mathcal{P}^n$ such that the probability of error for such a sequence satisfies

$$\lim_{n \to \infty} p_e(n) = 0,$$

where $n$ is the number of channel uses. A result from \cite{9} sharpens this claim by showing that there exists a sequence $\mathcal{P}^n$ such that

$$p_e(n) \leq 2^{-kn},$$

for some $k > 0$ determined by the channel. The converse part of the original HSW theorem can be strengthened in a similar manner. It states that if $R > C(\mathcal{N})$, for any sequence of protocols $\mathcal{P}^n$, then

$$\lim_{n \to \infty} p_e(n) > 0.$$ 

This is known as the weak converse. In contrast, a strong converse is symmetric to the achievability result above and states that regardless of the protocols used, the success probability decreases to zero in the asymptotic limit whenever $R > C(\mathcal{N})$. The strong converse can be sharpened as well whenever there is a constant separation between $R$ and $C(\mathcal{N})$ such that the convergence of the success probability to zero is exponential in $n$.

There are many reasons why we would want to prove a strong converse. First, a strong converse enriches our understanding of the capacity. A strong converse along with an achievability theorem shows that the capacity is a sharp boundary between reliable and unreliable communication regimes. This, amongst other results, indicates that the classical capacity of a quantum channel is a fundamental quantity of interest. Second, a strong converse is more relevant in practice than the weak converse. A realistic quantum communication scheme has a finite blocklength; i.e., the encoding is across a finite number of channel uses. Although the weak converse does provide a lower bound on the probability of error in the nonasymptotic regime, the strong converse improves the bound considerably. It expresses a trade-off between rate, error probability, and blocklength restrictive enough to be easily checked numerically or experimentally.

While a classical version of the strong converse is known for arbitrary discrete memoryless classical channels \cite{10,11}, it is still open whether or not strong converses hold for memoryless...
quantum channels. After some early work [12,13], the strong converse has been proved for special cases, in particular for channels with certain symmetry [14], for EB channels [15], and for a wide class of quantum Gaussian channels [16]. Given the strong converse results for EB channels [15] and the weak converse for EB channels with feedback [8], it is natural to ask if these two statements are true at the same time. We can also ask directly for the strong converse for unentangled inputs when a feedback link is available. These are the main questions that we address in this paper.

2. SUMMARY OF RESULTS

In this paper, we derive an explicit exponential bound on the success probability of a classical communication scheme that uses an entanglement-breaking channel along with classical feedback. The same method of proof can be used to establish a bound for arbitrary quantum channels with classical feedback provided that the inputs are not entangled across multiple uses of the channel. When the communication rate exceeds the classical capacity, these exponential bounds immediately imply strong converse theorems for these settings.

We now provide an outline of the proof:

1. First, taking as a starting point the general approach of [17], relating hypothesis testing to unassisted communication, we bound the success probability of an arbitrary feedback-assisted classical communication protocol by a sandwiched Rényi relative entropy [15,18].

2. Next, one of the main observations from [8] is that the sender and receiver’s systems are separable at all times throughout such a protocol whenever the communication channel is entanglement-breaking. We use this fact and an entropy inequality from [19] to split the relative entropy into two terms. The first term is bounded by an α-information radius, which is a measure of the range of states a channel can output. We then equate this to the sandwiched α-Holevo information [15], which is a Rényi generalization of the Holevo information. The second term is bounded via monotonicity by another sandwiched Rényi relative entropy, to which we recursively apply the same argument.

3. This gives the following bound on the probability of success for any finite blocklength $n$:

$$p_{\text{succ}} \leq 2^{-n \sup_{\alpha > 1} \alpha \chi(R - \chi(N))},$$

where $R$ is the rate of communication and $N$ is the EB channel. It follows from previous arguments [15] that when $R > \chi(N)$, the right hand side of (4) is a decaying exponential, thereby establishing the strong converse. We provide an alternate (arguably simpler) proof of this fact (similar to those in [20,21]) by establishing that the α-Holevo information converges continuously to the Holevo information as $\alpha$ approaches 1.

Appendix B includes a brief review of the argument for the weak converse from [8].

3. PRELIMINARIES

In this section we provide some necessary definitions, concepts, and previous results used in the derivation of (4).

3.1. Quantum States, Measurements, Operator Norms, and Quantum Channels

We start with definitions of relevant mathematical notions from quantum mechanics. Given a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})$ denote the algebra consisting of linear operators acting on $\mathcal{H}$. A relevant measure of a operator $X$ is its Schatten $\alpha$-norm, which is defined as

$$\|X\|_\alpha \equiv \{\text{Tr}[|X|^\alpha]\}^{1/\alpha},$$

where $\alpha \geq 1$ and $|X| \equiv \sqrt{X^\dagger X}$. 
Fig. 1. General protocol for feedback-assisted classical communication. The sender manipulates the systems labeled by $A$ and the receiver those labeled by $B$. Every $X$ system is classical and represents noiseless classical feedback from the receiver to the sender.

The set of quantum states is a convex subset of $\mathcal{B}(\mathcal{H})$ given by

$$S(\mathcal{H}) = \{\rho \in \mathcal{B}(\mathcal{H}) : \rho \geq 0, \quad \text{Tr} \, \rho = 1\},$$

where the notation $\rho \geq 0$ means that $\rho$ is positive semidefinite. For composite states, we consider the tensor product of two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, denoted by $\mathcal{H}_A \otimes \mathcal{H}_B$. We can obtain from the overall density operator $\rho_{AB} \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$ the reduced density operator $\rho_A$ corresponding to the quantum state on only the $A$ system by performing a partial trace:

$$\rho_A \equiv \text{Tr}_B(\rho_{AB}).$$

The partial trace can be defined as

$$\text{Tr}_B(|x_1\rangle\langle x_2|_A \otimes |y_1\rangle\langle y_2|_B) \equiv |x_1\rangle\langle x_2|_A \langle y_2|y_1\rangle$$

for vectors $|x_1\rangle$, $|x_2\rangle$, $|y_1\rangle$, and $|y_2\rangle$, and then extended by linearity. A state $\rho_{AB}$ is separable if it can be written as

$$\rho_{AB} = \sum_x p(x) \rho^x_A \otimes \rho^x_B,$$

where $p(x)$ is a probability distribution and $\{\rho^x_A\}$ and $\{\rho^x_B\}$ are sets of states.

A positive operator-valued measure (POVM) consists of a set $\{\Lambda_m\}$ of positive semidefinite operators indexed by $m$ and corresponding to different measurement results. The set satisfies $\sum_m \Lambda_m = I$, which allows us to interpret the quantity

$$p_m \equiv \text{Tr}(\rho \Lambda_m)$$

as the probability of measuring $m$ given a quantum state $\rho$.

Next consider maps on operators and in particular quantum states. A linear map $\Psi : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$ is called positive if $\forall \tau \in \mathcal{B}(\mathcal{H}_A)$, $\tau \geq 0$ implies $\Psi(\tau) \geq 0$. It is called completely positive if $\text{id}_R \otimes \Psi$ is positive for an arbitrary auxiliary system $R$, where $\text{id}_R$ is the identity map on $R$. A map $\Psi$ is a quantum channel if it is linear, completely positive, and also trace-preserving.

3.2. Classical Feedback-Assisted Classical Communication Protocols

Figure 1 depicts the most general three-round protocol for classical feedback-assisted classical communication. The generalization to $n$ rounds is clear. The protocol begins with Alice preparing
a classical register $M$ with the message to be sent. This is correlated with some system $A'_0$. Bob uses the classical feedback channel to send a classical system $X_0$ correlated with some quantum system $B'_0$ to Alice. The global state is then

$$\rho_{MA'_0X_0B'_0} = \sum_m p_M(m)|m\rangle\langle m|_{M} \otimes \rho_{A'_0}^m \otimes \sum_x p_{X_0}(x)|x\rangle\langle x|_{X_0} \otimes \rho_{B'_0}^x.$$  

We will track the state conditioned on a particular value $m$ of the message register $M$, given by

$$\rho_{A'_0X_0B'_0}^m \equiv \rho_{A'_0}^m \otimes \sum_x p_{X_0}(x)|x\rangle\langle x|_{X_0} \otimes \rho_{B'_0}^x.$$  

Alice performs an encoding $E_{A'_0X_0 \rightarrow A'_1A_1}^1$. The state at this point is

$$\rho_{A'_1A_1B'_0}^m \equiv E_{A'_0X_0 \rightarrow A'_1A_1}^1(\rho_{A'_0X_0B'_0}^m) = \sum_x p_{X_0}(x)E_{A'_0X_0 \rightarrow A'_1A_1}(\rho_{A'_0}^m \otimes |x\rangle\langle x|_{X_0}) \otimes \rho_{B'_0}^x.$$  

We note that the state $\rho_{A'_1A_1B'_0}^m$ is separable with respect to the cut $A'_1A_1 : B'_0$.

Next, Alice uses the EB channel $N_{A_1 \rightarrow B_1}$ for the first time by sending system $A_1$, leading to the state

$$\rho_{A'_1B_1B'_0}^m \equiv N_{A_1 \rightarrow B_1}(\rho_{A'_1A_1B'_0}^m).$$  

Since the channel is entanglement-breaking and the state before the channel was already separable with respect to the $A'_1A_1 : B'_0$ cut, the state after the channel is fully separable, i.e., it is separable across all possible partitions. Now, consider instead that Alice uses an arbitrary channel $N_{A_1 \rightarrow B_1}$ but with an input separable across the $A'_1 : A_1$ cut (such that $\rho_{A'_1A_1B'_0}^m$ is fully separable). Then the state $\rho_{A'_1B_1B'_0}^m$ is again fully separable. We write it in the form

$$\rho_{A'_1B_1B'_0}^m = \sum_y p_Y(y)\rho_{A'_1}^{y,m} \otimes \rho_{B_1}^{y,m} \otimes \rho_{B'_0}^{y,m}.$$  

Bob now applies the decoding map $D_{B_1B'_0 \rightarrow X_1B'_1}^1$, where $X_1$ is the classical system that is sent back to Alice. The state at this point is

$$\rho_{A'_1X_1B'_1}^m \equiv D_{B_1B'_0 \rightarrow X_1B'_1}^1(\rho_{A'_1B_1B'_0}^m) = \sum_{x_1} p_{X_1}(x_1)|x_1\rangle\langle x_1|_{X_1} \otimes \sum_{z_1} p_{Z_1|X_1}(z_1|x_1)\rho_{A'_1}^{x_1,z_1} \otimes \rho_{B'_1}^{x_1,z_1},$$  

which is fully separable. Hence, after Alice applies the second encoder $E^2$ and then sends it through the channel, the state will still be fully separable. The key observation here is that if we use an EB channel or an arbitrary channel with separable inputs, the state is always separable across a cut that divides Alice and Bob’s systems.

The only difference in subsequent rounds is the final measurement. Say there are $n$ rounds in the protocol. At the last round, Bob measures the state $\rho_{B_nB'_n}$ using a POVM given by $\{D^m\}$ with elements corresponding to different possible messages that Alice sent.

### 3.3. Rényi Relative Entropies and Bounds on Success Probability

An important classical information theoretic quantity is the Rényi relative entropy, which can be generalized to the quantum case in a number of ways. In this paper, we use the sandwiched...
quantum Rényi relative entropy [15,18], which is given by

\[
\bar{D}_\alpha(\rho \| \sigma) \equiv \begin{cases} 
\frac{1}{\alpha - 1} \log \left[ \text{Tr}(\left(\sigma^{(1-\alpha)/(2\alpha)} \rho \sigma^{(1-\alpha)/(2\alpha)}\right)^\alpha) \right], \\
\rho \not\perp \sigma \land (\text{supp}(\rho) \subseteq \text{supp}(\sigma) \lor \alpha \in (0,1)), \\
+\infty, \quad \text{otherwise},
\end{cases}
\]

where \( \alpha \in (0,1) \cup (1,\infty) \) and \( \rho \not\perp \sigma \) means \( \rho, \sigma \) are nonorthogonal quantum states. Note that all logarithms in this paper are taken to the base two.

We now recall some properties of the sandwiched Rényi relative entropy. For fixed \( \rho \) and \( \sigma \), the function \( \alpha \mapsto \bar{D}_\alpha(\rho \| \sigma) \) is monotone nondecreasing [18]. It also converges to the quantum relative entropy \( D(\rho \| \sigma) \) [22] in the limit as \( \alpha \to 1 \) [15,18]:

\[
\lim_{\alpha \to 1} \bar{D}_\alpha(\rho \| \sigma) = D(\rho \| \sigma), \tag{6}
\]

where

\[
D(\rho \| \sigma) \equiv \begin{cases} 
\text{Tr}[\rho(\log \rho - \log \sigma)], \\
\in \text{supp}(\rho) \subseteq \text{supp}(\sigma), \\
+\infty, \quad \text{otherwise}.
\end{cases}
\]

Furthermore, it satisfies the data-processing inequality for \( \alpha \in [1/2,1) \cup (1,\infty) \) [23,24]; i.e., for all quantum channels \( \mathcal{N} \),

\[
\bar{D}_\alpha(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) \leq \bar{D}_\alpha(\rho \| \sigma).
\]

In particular, consider the following replacement channel which simply replaces the input with some state \( \omega \):

\[
\mathcal{R}_\omega(\rho) \equiv \text{Tr}(\rho)\omega.
\]

It is easy to see that for all \( \omega \), \( \bar{D}_\alpha(\omega \| \omega) = 0 \), implying that

\[
\bar{D}_\alpha(\rho \| \sigma) \geq \bar{D}_\alpha(\mathcal{R}_\omega(\rho) \| \mathcal{R}_\omega(\sigma)) = 0,
\]

which shows that the sandwiched Rényi relative entropy is nonnegative whenever its arguments are quantum states \( \rho \) and \( \sigma \).

Using the sandwiched Rényi relative entropy, we can define [15] the sandwiched \( \alpha \)-Holevo information of an ensemble, i.e., a classical probability distribution of quantum states, \( \{p_X(x), \rho_x\} \) as

\[
\bar{\chi}_\alpha(\{p_X(x), \rho_x\}) \equiv \inf_{\sigma_R \in \mathcal{S}(\mathcal{H}_R)} \bar{D}_\alpha(\rho_{XR} \| \rho_X \otimes \sigma_R)
\]

where

\[
\rho_{XR} \equiv \sum_x p_X(x) |x\rangle \langle x| \otimes (\rho_x)_R
\]

and \( \rho_x \) are states of a system \( R \). With this, we define the \( \alpha \)-Holevo information of a quantum channel \( \mathcal{N} \) as

\[
\bar{\chi}_\alpha(\mathcal{N}) \equiv \sup_{\{p_X(x), \rho_x\}} \bar{\chi}_\alpha(\{p_X(x), \mathcal{N}(\rho_x)\}).
\]

We also define the \( \alpha \)-information radius of a channel \( \mathcal{N}: \mathcal{S}(\mathcal{H}_A) \mapsto \mathcal{S}(\mathcal{H}_B) \) as

\[
\bar{K}_\alpha(\mathcal{N}) \equiv \inf_{\sigma \in \mathcal{S}(\mathcal{H}_B)} \sup_{\rho \in \mathcal{S}(\mathcal{H}_A)} \bar{D}_\alpha(\mathcal{N}(\rho) \| \sigma).
\]

We note that since \( \bar{D}_\alpha \) is monotonically nondecreasing in \( \alpha \), so are \( \bar{\chi}_\alpha \) and \( \bar{K}_\alpha \).
4. STRONG CONVERSE

This section is dedicated to proving the main theorem of this paper. We will need the following lemmas. The first is proved using an inequality from [17] in [21, Lemma 5] via monotonicity of $D_\alpha$. The second will be used to take advantage of the separability of the quantum state observed in Section 3.2. The third is an equality between $\alpha$-Holevo information and $\alpha$-information radius. The fourth states that $\alpha$-Holevo information and $\alpha$-information radius, respectively, tend to the conventional Holevo information and information radius in the limit $\alpha \to 1$, which we prove in Appendix C. Note that this establishes Lemma 3 as a generalization of the equality $\chi(N) = K(N)$ [25–27].

**Lemma 1.** Let $\alpha > 1$, $\rho, \sigma \in S(H)$, and $\Lambda$ be such that $0 \leq \Lambda \leq I$. Let

$$p \equiv \text{Tr}[\Lambda \rho], \quad q \equiv \text{Tr}[\Lambda \sigma].$$

Then

$$D_\alpha(\rho \| \sigma) \geq \frac{1}{\alpha - 1} \log[p^\alpha q^{1-\alpha}].$$

**Lemma 2** [19, 28]. Let $P_{AB}$ be a positive semidefinite separable operator. Such an operator can be written in the following form:

$$P_{AB} = \sum_j C_A^j \otimes D_B^j,$$

where $C_A^j, D_B^j \geq 0$ for all $j$. Let $P_B = \text{Tr}_A\{P_{AB}\}$ and let $\mathcal{M}_A$ be a completely positive linear map acting on the $A$ system. Then, for all $\alpha \geq 1$,

$$\|\mathcal{M}_A \otimes \text{id}_B(P_{AB})\|_\alpha \leq \nu_\alpha(\mathcal{M}_A) \cdot \|P_B\|_\alpha,$$

where $\nu_\alpha(\mathcal{M}_A)$ is the $1 \to \alpha$ norm of $\mathcal{M}_A$, defined as

$$\nu_\alpha(\mathcal{M}_A) \equiv \sup_{X \neq 0, X \in B(H_A)} \frac{\|\mathcal{M}_A(X)\|_\alpha}{\|X\|_1}.$$ 

**Lemma 3** [15]. For $\alpha > 1$, the $\alpha$-Holevo information and the $\alpha$-information radius are the same:

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

**Proof.** This statement was essentially proved as [15, Lemma 14], but only for the interval $\alpha \in (1, 2]$. To get the statement for all $\alpha > 1$, we actually need to extend the proof slightly. This follows from the proof given in [15] and the observations that for all $\alpha > 1$, $x^{(1-\alpha)/\alpha}$ is operator convex, $\text{Tr}\{x^\alpha\}$ is convex, and $D_\alpha$ is jointly quasi-convex.

For completeness and convenience, we also give a full proof here, following the steps in the proof of Lemma 14 in [15] closely. We first prove the inequality $\tilde{K}_\alpha(N) \leq \tilde{\chi}_\alpha(N)$ for $\alpha > 1$. Defining $\tilde{Q}_\alpha(\rho \| \sigma) \equiv \text{Tr}((\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha})^\alpha)$, we have

$$\tilde{K}_\alpha(N) = \inf_{\rho} \sup_{\sigma} \tilde{D}_\alpha(N(\rho) \| \sigma)$$

$$= \inf_{\rho} \sup_{\sigma} \frac{1}{\alpha - 1} \log \tilde{Q}_\alpha(N(\rho) \| \sigma) = \frac{1}{\alpha - 1} \log \inf_{\rho} \sup_{\sigma} \tilde{Q}_\alpha(N(\rho) \| \sigma).$$

So now we focus on the $\tilde{Q}_\alpha$ quantity and find that

$$\inf_{\rho} \sup_{\sigma} \tilde{Q}_\alpha(N(\rho) \| \sigma) \leq \inf_{\rho} \sup_{\mu} \int d\mu(\rho) \tilde{Q}_\alpha(N(\rho) \| \sigma) = \sup_{\mu} \int d\mu(\rho) \tilde{Q}_\alpha(N(\rho) \| \sigma)$$

$$= \sup_{\rho_X} \inf_{\mu} \sum_x p_X(\mu) \tilde{Q}_\alpha(N(\rho_x) \| \sigma) = \sup_{\rho_X} \inf_{\sigma_B} \tilde{Q}_\alpha(\rho_X \| \rho \otimes \sigma_B).$$

(7)
The first inequality follows by taking a supremum over all probability measures $\mu$ on the set of all states $\rho$. The first equality is a result of applying the Sion minimax theorem [29]—we can do so because the function $\int d\mu(\rho) \tilde{Q}_\alpha(\mathcal{N}(\rho)\|\sigma)$ is linear in the probability measure $\mu$ and convex in states $\sigma$. Convexity of $\tilde{Q}_\alpha(\mathcal{N}(\rho)\|\sigma)$ in $\sigma$ follows because by cyclicity of trace
\[
\tilde{Q}_\alpha(\mathcal{N}(\rho)\|\sigma) = \text{Tr}\{([\mathcal{N}(\rho)]^{1/2}\sigma^{(1-\alpha)/\alpha}[\mathcal{N}(\rho)]^{1/2})^\alpha\},
\]
x$^{(1-\alpha)/\alpha}$ is operator convex for $\alpha > 1$, and Tr$(x^\alpha)$ is convex for $\alpha > 1$. The second equality follows by applying the Fenchel–Eggleston–Caratheodory theorem (see [2], for example): the function $\tilde{Q}_\alpha(\mathcal{N}(\rho)\|\sigma)$ is continuous in $\rho$, which is a density operator acting on a $d$-dimensional Hilbert space, so that for each $\mu$ there exists a probability distribution $p_X(x)$ on no more than $d^2$ letters such that
\[
\int d\mu(\rho) \tilde{Q}_\alpha(\mathcal{N}(\rho)\|\sigma) = \sum_x p_X(x)\tilde{Q}_\alpha(\mathcal{N}(\rho_x)\|\sigma).
\]
The last equality in (7) follows from the properties of $\tilde{Q}_\alpha$ and by defining
\[
\rho_{XB} = \sum_x p_X(x)|x\rangle\langle x| \otimes [\mathcal{N}(\rho_x)]_B.
\]
Thus, we can then conclude that
\[
\tilde{K}_\alpha(\mathcal{N}) \leq \frac{1}{\alpha - 1} \log \sup_{\{p_X(x), \rho_x\}} \inf_{\sigma_B} \tilde{Q}_\alpha(\rho_{XB}\|\rho_X \otimes \sigma_B) = \sup_{\{p_X(x), \rho_x\}} \inf_{\sigma_B} \frac{1}{\alpha - 1} \log \tilde{Q}_\alpha(\rho_{XB}\|\rho_X \otimes \sigma_B) = \tilde{\chi}_\alpha(\mathcal{N}).
\]
The proof of the other inequality $\tilde{K}_\alpha(\mathcal{N}) \geq \tilde{\chi}_\alpha(\mathcal{N})$ is simpler. We have
\[
\tilde{\chi}_\alpha(\mathcal{N}) = \sup_{\{p_X(x), \rho_x\}} \inf_{\sigma_B} \tilde{D}_\alpha(\rho_{XB}\|\rho_X \otimes \sigma_B) \leq \sup_{\{p_X(x), \rho_x\}} \tilde{D}_\alpha(\rho_{XB}\|\rho_X \otimes \sigma) \\
\leq \sup_{\{p_X(x), \rho_x\}} \sup_x \tilde{D}_\alpha(|x\rangle\langle x| \otimes \mathcal{N}(\rho_x)\||x\rangle\langle x| \otimes \sigma) \\
= \sup_{\{p_X(x), \rho_x\}} \sup_x \tilde{D}_\alpha(\mathcal{N}(\rho_x)\|\sigma) \leq \sup_\rho \tilde{D}_\alpha(\mathcal{N}(\rho)\|\sigma). \quad (8)
\]
The second inequality follows from joint quasi-convexity of $\tilde{D}_\alpha$ for $\alpha > 1$ [23,24], where a function $f$ is jointly quasi-convex if
\[
f(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2) \leq \max\{f(x_1, y_1), f(x_2, y_2)\}.
\]
Since the above inequality (8) holds for all states $\sigma$, we can conclude that $\tilde{K}_\alpha(\mathcal{N}) \geq \tilde{\chi}_\alpha(\mathcal{N})$. \triangle

**Lemma 4.** For a quantum channel $\mathcal{N}$, the following limits hold:
\[
\lim_{\alpha \to 1} \tilde{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N})
\]
and
\[
\lim_{\alpha \to 1} \tilde{K}_\alpha(\mathcal{N}) = K(\mathcal{N}),
\]
where
\[
K(\mathcal{N}) \equiv \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\rho \in \mathcal{S}(\mathcal{H})} D(\mathcal{N}(\rho)\|\sigma)
\]
is the information radius.
Fig. 2. Protocol for feedback-assisted classical communication when using the replacement channel. Note that the communication line between $A_1$ and $\hat{M}$ is broken.

We now state the theorem.

**Theorem 1.** Given any $n$-round protocol for classical feedback-assisted classical communication over an entanglement-breaking channel $\mathcal{N}$ with rate $R$, the average probability of success is bounded from above by an exponential in $n$:

$$p_{\text{succ}} \leq 2^{-n \sup_{\alpha > 1} \frac{1}{\alpha} (R - \chi_\alpha(\mathcal{N}))},$$

where $\chi_\alpha$ is the $\alpha$-Holevo information. The same bound holds for an arbitrary channel $\mathcal{N}$ given that the encoder does not entangle inputs across different uses of the channel.

**Proof.** We take as a starting point the approach of Nagaoka [17], connecting hypothesis testing with data processing of a Rényi information quantity. Let $\mathcal{P}_n$ be such a protocol. We are bounding the average probability of success, so we assume that Alice chooses her messages uniformly at random. Using the notation of Section 3.2, the state we have at the final round of the protocol $\mathcal{P}_n$ is

$$\rho_{MB_nB'_{n-1}} = \frac{1}{L} \sum_{m=1}^L |m\rangle\langle m|_M \otimes \rho_{B_nB'_{n-1}}^m,$$

where $L$ is the number of possible messages.

Following the argument in [21], we can write the success probability as

$$p_{\text{succ}} = \frac{1}{L} \sum_m \text{Tr}[D_{B_nB'_{n-1}}^m \rho_{B_nB'_{n-1}}^m] = \text{Tr}[T_{MB_nB'_{n-1}} \rho_{MB_nB'_{n-1}}],$$

where

$$T_{MB_nB'_{n-1}} \equiv \sum_m |m\rangle\langle m|_M \otimes D_{B_nB'_{n-1}}^m.$$

Note that $0 \leq T \leq I$, so that $\{T, I - T\}$ is a POVM.

We now consider the state $\tau_{MB_nB'_{n-1}}$ defined to be the final state if we had implemented $\mathcal{P}_n$ using a replacement channel $\mathcal{R}_\sigma$ instead of the original channel (see Fig. 2).

The overall state in this alternate scenario has a simple expression, since Bob’s states are now independent of $m$:

$$\tau_{MB_nB'_{n-1}} = \tau_M \otimes \sigma_{B_n} \otimes \tau_{B'_{n-1}},$$

where $\tau_M = I_M/L$. In the above, we recall that $B_n$ is a label for the output system of the $n$th replacement channel $\mathcal{R}_\sigma$, and the state of this system is equal to $\sigma$. We can then compute the
following:
\[
\text{Tr}ig[ T_{MBnB'_n} \tau_{MBnB'_n} \big] = \frac{1}{L} \sum_m \text{Tr}ig[ D_{BnB'_n}^{m} \sigma_{Bn} \otimes \tau_{B'_n} \big] = \frac{1}{L},
\]
where the last equality follows because \( \sum_m D_{BnB'_n}^{m} = I \) and \( \text{Tr}[\sigma_{Bn} \otimes \tau_{B'_n}] = 1 \). This equality is intuitive, since for a replacement channel, the receiver Bob cannot do any better than to guess the input message \( m \) at random. Letting \( T \) be the operator \( \Lambda \) in Lemma 1, we conclude that for \( \alpha > 1 \),
\[
\tilde{D}_{\alpha} (\rho_{MBnB'_n} \parallel \tau_{MBnB'_n}) \geq \frac{1}{\alpha - 1} \log \left( \frac{1}{\text{Tr}[\sigma_{Bn} \otimes \tau_{B'_n}]} \right),
\]
which can be rewritten as follows:
\[
\frac{\alpha}{\alpha - 1} \log(p_{\text{succ}}) + \log L \leq \tilde{D}_{\alpha} (\rho_{MBnB'_n} \parallel \tau_{MBnB'_n}).
\]
We note that if the support of the state \( \sigma \) does not contain the support of the image of the channel \( N_{A \rightarrow B} \), then the \( \tilde{D}_{\alpha} \) upper bound is trivially equal to \( +\infty \). Thus, in what follows, we choose \( \sigma \) such that its support contains the support of the image of the channel in order to guarantee that all quantities involved in the forthcoming proof are finite.

We would therefore like to bound \( \tilde{D}_{\alpha} (\rho_{MBnB'_n} \parallel \tau_{MBnB'_n}) \). To do so, consider that
\[
\tilde{D}_{\alpha} (\rho_{MBnB'_n} \parallel \tau_{MBnB'_n}) = \tilde{D}_{\alpha} (N_{A \rightarrow B_n} (\rho_{MA_nB'_n}) \parallel \tau_{MBnB'_n} \otimes \sigma_{Bn})
= \frac{\alpha}{\alpha - 1} \log \left( \left\| \Theta_{\sigma^{(1-\alpha)/\alpha} \circ N_{A \rightarrow B_n}} \left( \tau_{MBnB'_n} \otimes \rho_{MA_nB'_n} \right) \right\|_{\alpha} \right),
\]
where \( \circ \) denotes function composition and in the last equality we define \( \Theta \) by
\[
\Theta_{\sigma}(\rho) \equiv \sigma^{1/2} \rho \sigma^{1/2},
\]
and the identity operation on the other systems is implied. We now use the key observation from Section 3.2 (and used in [8]) that if \( \mathcal{N} \) is EB or if Alice uses separable inputs, throughout \( P_n \), Alice and Bob’s systems are always separable. Furthermore, the \( M \) system is classical, so \( \rho_{MA_nB'_n} \) is separable with respect to the \( A_n : MBnB'_n \) cut. This implies that
\[
\tau_{MBnB'_n} \otimes \rho_{MA_nB'_n} \tau_{MBnB'_n}^{(1-\alpha)/(2\alpha)}
\]
is a positive semidefinite separable operator:
\[
\tau_{MBnB'_n} \otimes \rho_{MA_nB'_n} \tau_{MBnB'_n}^{(1-\alpha)/(2\alpha)} = \tau_{MBnB'_n}^{(1-\alpha)/(2\alpha)} \left( \sum_j p(j) \rho_{A_n}^j \otimes \rho_{MBnB'_n}^j \right) \tau_{MBnB'_n}^{(1-\alpha)/(2\alpha)}
= \sum_j p(j) \rho_{A_n}^j \otimes \tau_{MBnB'_n}^{(1-\alpha)/(2\alpha)} \rho_{MBnB'_n}^j \tau_{MBnB'_n}^{(1-\alpha)/(2\alpha)}.
\]
Since conjugation by a positive semidefinite operator is clearly a completely positive map, we can apply Lemma 2 to conclude that
\[
\left\| \Theta_{\sigma^{(1-\alpha)/\alpha} \circ N_{A \rightarrow B_n}} \left( \tau_{MBnB'_n} \otimes \rho_{MA_nB'_n} \tau_{MBnB'_n}^{(1-\alpha)/(2\alpha)} \right) \right\|_{\alpha}
\leq \nu_{\alpha} \left( \Theta_{\sigma^{(1-\alpha)/\alpha} \circ N_{A \rightarrow B_n}} \right) \left\| \tau_{MBnB'_n} \otimes \rho_{MA_nB'_n} \tau_{MBnB'_n}^{(1-\alpha)/(2\alpha)} \right\|_{\alpha}.
\]

PROBLEMS OF INFORMATION TRANSMISSION Vol. 54 No. 1 2018
We then have the following chain of (in)equalities:

\[
\tilde{D}_\alpha (\rho_{MBn}B_{n-1} B_{n-1}/\tau_{MBn}B_{n-1} B_{n-1}) \leq \frac{\alpha}{\alpha - 1} \log \nu_\alpha (\Theta_{\sigma^{(1-\alpha)/\alpha}} \circ \mathcal{N}) + \tilde{D}_\alpha (\rho_{MBn} B_{n-1} B_{n-1} \parallel \tau_{MBn} B_{n-1} B_{n-1}) \\
\leq \frac{\alpha}{\alpha - 1} \log \nu_\alpha (\Theta_{\sigma^{(1-\alpha)/\alpha}} \circ \mathcal{N}) + \tilde{D}_\alpha (\rho_{MBn} \parallel \tau_{MBn}) \\
\leq n \frac{\alpha}{\alpha - 1} \log \nu_\alpha (\Theta_{\sigma^{(1-\alpha)/\alpha}} \circ \mathcal{N}) + \tilde{D}_\alpha (\rho_{MBn} \parallel \tau_{MBn}) \\
= n \frac{\alpha}{\alpha - 1} \log \nu_\alpha (\Theta_{\sigma^{(1-\alpha)/\alpha}} \circ \mathcal{N}) = n \frac{\alpha}{\alpha - 1} \log \sup_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \| \left( \Theta_{\sigma^{(1-\alpha)/\alpha}} \circ \mathcal{N}_{A \rightarrow B} \right) (\rho_A) \|_\alpha.
\]

The first inequality follows by combining (10) and (11). The second inequality follows from monotonicity of the sandwiched Rényi relative entropy under the partial trace channel. The third inequality follows by recognizing that \( \tilde{D}_\alpha (\rho_{MBn} B_{n-1} B_{n-1} \parallel \tau_{MBn} B_{n-1} B_{n-1}) \) is the relative entropy at round \( n - 1 \) of \( \mathcal{P}_n \), which allows us to apply the above argument inductively. This is a crucial step of the argument and proves a form of additivity similar to that of (3). The first equality is a consequence of the fact that \( \rho_{MBn} = \tau_{MBn} \), since no channels have been applied at that point in the protocol. The last equality follows from the main result of [30], which allows us to take the supremum over quantum states instead of all operators (furthermore, from the fact that quantum states have trace equal to one). Hence, we have

\[
\frac{\alpha}{\alpha - 1} \log p_{\text{succ}} (\mathcal{P}_n) + \log L \leq n \frac{\alpha}{\alpha - 1} \log \sup_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \| \left( \Theta_{\sigma^{(1-\alpha)/\alpha}} \circ \mathcal{N}_{A \rightarrow B} \right) (\rho_A) \|_\alpha.
\]

Hence, we have proved that the upper bound holds for all states \( \sigma_B \), so we can conclude that

\[
\frac{\alpha}{\alpha - 1} \log p_{\text{succ}} (\mathcal{P}_n) + \log L \\
\leq n \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \sup_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \frac{\alpha}{\alpha - 1} \log \left( \Theta_{\sigma^{(1-\alpha)/\alpha}} \circ \mathcal{N}_{A \rightarrow B} \right) (\rho_A) \|_\alpha = n \tilde{\kappa}_\alpha (\mathcal{N}) = n \tilde{\chi}_\alpha (\mathcal{N}).
\]

The first equality comes from the definition of the \( \alpha \)-information radius, and the second equality is due to Lemma 3.

Now, the protocol uses the channel \( n \) times, and the rate \( R \) of \( \mathcal{P}_n \) is defined to be the number of bits per channel use, so that \( R \leq \frac{\log L}{n} \). This allows us to introduce \( R \) into the inequality:

\[
\frac{1}{n} \log p_{\text{succ}} (\mathcal{P}_n) \leq - \frac{\alpha - 1}{\alpha} (R - \tilde{\chi}_\alpha (\mathcal{N})).
\]

Since this is true for all \( \alpha > 1 \), we can take a supremum over \( \alpha > 1 \) and arrive at the bound stated in (9). \( \triangle \)

The strong converse itself follows now from Theorem 1 and Lemma 4.

**Corollary 1** (strong converse). The probability of success of any sequence of protocols which use an entanglement-breaking channel with classical feedback at a rate greater than the classical capacity is bounded from above by a decaying exponential. The same is true for arbitrary channels with separable inputs.

**Proof.** Recall that since \( \tilde{D}_\alpha \) is monotonically increasing in \( \alpha \), so is \( \tilde{\chi}_\alpha \). This along with Lemma 4 implies

\[
\inf_{\alpha > 1} \tilde{\chi}_\alpha (\mathcal{N}) = \lim_{\alpha \searrow 1} \tilde{\chi}_\alpha (\mathcal{N}) = \chi (\mathcal{N}).
\]

Hence, if \( R > \chi (\mathcal{N}) \), by the continuity of \( \tilde{\chi}_\alpha (\mathcal{N}) \) as \( \alpha \searrow 1 \), there exists a value of \( \alpha > 1 \) such that \( \frac{\alpha - 1}{\alpha} (R - \tilde{\chi}_\alpha (\mathcal{N})) > 0 \). Then the bound in (9) implies an exponential decay of the success probability. \( \triangle \)
5. CONCLUSION

By studying the classical capacity of a quantum channel, quantum information theorists found that an implication of quantum mechanics for classical communication is the possibility of superior encoding schemes that use entanglement, a uniquely quantum phenomenon [5]. However, entanglement does not give an advantage for every channel. The results in this and previous papers show that for entanglement-breaking channels, no communication protocols, even with classical feedback, can take advantage of this extra resource. They cannot use it to increase their capacity, nor, as is proved in this paper, to even lift the exponentially decaying ceiling on their success probabilities. We perhaps expect this, since by definition EB channels destroy entanglement, and classical feedback channels cannot create entanglement. This is the guiding principle behind our proof, in particular the key fact that the transmitter and receiver states are separable throughout the protocol.

More generally, it is known that classical feedback can give a large boost to the classical capacity of channels which are not entanglement-breaking in at least two different ways: first, there exist channels for which the Holevo information is small but the classical capacity with feedback included can be quite large [31]. Similarly, there exist channels for which the classical capacity is small but becomes large when classical feedback is available [32]. Thus, showing that the Holevo information is a strong-converse bound suggests that for feedback not to help, we require special channels, such as entanglement-breaking ones.

Going in the opposite direction, it is known that EB channels form a proper superset of classical channels [33]. We thus obtain the result of [34] as a direct corollary. Furthermore, since the original posting of our paper as arXiv:1506.02228, an open question that we posed has now been answered: the strong converse exponent from Theorem 1 is tight, due to the results in [35]. Hence, we obtain as a special case from this and our result the classical results of [36–38] as well.

A possible direction for future research is to ask the same question for Hadamard channels, which are defined to be complements of EB channels. To define the complementary channel, we first explain the interpretation of channel as a model for open quantum dynamics. That is, a channel from system $A$ to system $B$ can be interpreted as the restriction of a unitary interaction with larger system $BR$, where $R$ is referred to as the “environment.” The complementary channel is the map from $A$ to $BR$ tracing out the $B$ system instead and is itself a channel from $A$ to $R$. Hence, Hadamard channels break any entanglement with the environment system that is traced out and is thus related to our guiding principle. The strong converse has already been proved for Hadamard channels [15], but with the addition of classical feedback, even a weak converse has yet to be proved. Examples of Hadamard channels include generalized dephasing channels, cloning channels, and the Unruh channel [39].

Finally, we remark here that it should be possible to use the methods given here and in [21] in order to characterize a particular adaptive hypothesis testing scenario. Assume that the goal is to distinguish an entanglement-breaking channel from a replacement channel by means of adaptive, separability-preserving channels. Then the optimal strong converse exponent should be given in terms of a quantity similar to that in (9). However, we leave the details for future work.

APPENDIX A

Additivity. Here we justify the statement in (3). The backward implication is trivial. As for the forward implication, we first note that given $n$, for all $k \in \mathbb{Z}^+$,

$$\frac{1}{n} \chi(N^{\otimes n}) \leq \frac{1}{kn} \chi(N^{\otimes kn}),$$

PROBLEMS OF INFORMATION TRANSMISSION Vol. 54 No. 1 2018
because we can split the \( kn \) channels into \( k \) blocks of \( n \) channels and encode each block independently. In particular, \( \frac{1}{n} \chi(N^{\otimes n}) \geq \chi(N) \). To show the opposite inequality, we assume the contrapositive: \( \frac{1}{n} \chi(N^{\otimes n}) > \chi(N) \) for some \( n \). However, this means

\[
\frac{1}{kn} \chi(N^{\otimes kn}) - \chi(N) \geq \frac{1}{n} \chi(N^{\otimes n}) - \chi(N) > 0,
\]

for all \( k \in \mathbb{Z}^+ \), and thus the limit in (2) does not converge to \( \chi(N) \).

**APPENDIX B**

**Weak converse and finite bounds.** We recall for motivation and the reader’s convenience the argument for the weak converse from [8].

**Theorem 2.** Let \( \mathcal{P}_n \) be a protocol for classical feedback-assisted classical communication over an entanglement-breaking channel \( N \) such that it uses the channel \( n \) times, has communication rate \( R \), and has average probability of decoding error \( \varepsilon \). Then it satisfies the following inequality:

\[
R \leq \chi(N) + g(n, \varepsilon),
\]

where \( \chi(N) \) is the Holevo information of the channel and \( g(n, \varepsilon) \) is a real-valued function such that

\[
\lim_{\varepsilon \searrow 0} \lim_{n \to \infty} g(n, \varepsilon) = 0.
\]

The same is true for a protocol for communication over an arbitrary channel given that the encoder does not entangle inputs across different uses of the channel.

**Proof.** We use the notation in Section 3.2 and take a general approach that is, for instance, presented in [40, Section 21.5.1]. Let \( \overline{\Phi} \) denote the following “shared randomness” state:

\[
\overline{\Phi}_{M\hat{M}} = \frac{1}{|\mathcal{M}|} \sum_m |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{\hat{M}},
\]

where \( |\mathcal{M}| \) is the size of the message set \( \mathcal{M} \). Now, assume that the information processing task is common randomness generation instead of classical communication. If we require the state \( \rho \) at the end of the protocol to be \( \varepsilon' \)-close to \( \overline{\Phi} \),

\[
\|\overline{\Phi} - \rho\|_1 \leq \varepsilon',
\]

the Fannes–Audenaert inequality for mutual information [41,42] gives

\[
nR = I(M; \hat{M})_{\Phi} \leq I(M; \hat{M})_{\rho} + f(n, \varepsilon'),
\]

where \( f(n, \varepsilon') \) is some continuous function of \( n \) and \( \varepsilon' \) with the property \( \lim_{\varepsilon' \searrow 0} \lim_{n \to \infty} \frac{1}{n} f(n, \varepsilon') = 0 \). Continuing, we have

\[
I(M; \hat{M})_{\rho} \leq I(M; B_n B'_{n-1})_{\rho} = I(M; B'_{n-1})_{\rho} + I(M; B_n | B'_{n-1})_{\rho},\]

where we applied the data processing inequality and the chain rule for conditional quantum mutual information. We use the chain rule again to get

\[
I(M; \hat{M})_{\rho} \leq I(M | B'_{n-1}; B_n)_{\rho}.
\]
From Section 3.2, if \( \mathcal{N} \) is entanglement-breaking or if Alice uses separable inputs, Alice and Bob’s systems are always separable. Hence, the global state before the \( n \)th channel use can be written as

\[
\rho_{MA'_nB'_n} = \sum_m p_M(m) |m\rangle_M \otimes \rho^{m}_{A'_nA_nB'_n} = \sum_m p_M(m) |m\rangle_M \otimes \sum_y p_Y |y\rangle_M \rho^{m,y}_{A'_nA_n} \otimes \rho^{m,y}_{B'_n}.
\]

Hence, the state after the channel is given by

\[
\rho_{MA'_nB'_n} = \sum_m p_M(m) |m\rangle_M \otimes \sum_y p_Y |y\rangle_M \rho^{m,y}_{A'_nA_n} \otimes \rho^{m,y}_{B'_n}.
\]

We introduce an auxiliary system \( Y \) that labels the separable sum over the index \( y \). That is, \( Y \) is chosen such that the global state is

\[
\rho_{MYA'_nB'_n} = \sum_{m,y} p_M(m) p_Y |m\rangle_M \langle m| \otimes |y\rangle_Y \rho^{m,y}_{A'_nA_n} \otimes \rho^{m,y}_{B'_n}.
\]

We trace over \( A'_n \) to get

\[
\rho_{MYB'_n} = \sum_{m,y} p_M(m) p_Y |m\rangle_M \langle m| \otimes |y\rangle_Y \rho^{m,y}_{B'_n},
\]

where

\[
\rho^{m,y}_{B'_n} = \text{Tr}_{A'_n} (\mathcal{N}_{A_n} (\rho^{m,y}_{A'_nA_n})) = \mathcal{N}_{A_n} (\rho^{m,y}_{A_n}). \tag{13}
\]

This allows us to argue

\[
I(MB'_n;B_n) \leq I(MYB'_n-1;B_n) = I(MY;B_n) + I(B'_n-1;B_n|MY) = I(MY;B_n) \leq \chi(\mathcal{N}),
\]

where the first inequality is from the data processing inequality, the first equality from the chain rule, and the second equality because the \( B'_n-1B_n \) system is in a product state when conditioning on \( M \) and \( Y \). Finally, given equation (13), the state \( \rho_{MYB_n} \) is a classical-quantum state of the form

\[
\rho_{MYB_n} = \text{Tr}_{B'_n} (\rho_{MYB'_n}) = \sum_{m,y} p_M(m) p_Y |m\rangle_M \langle m| \otimes |y\rangle_Y \otimes \mathcal{N}_{A_n} (\rho^{m,y}_{A_n}).
\]

Hence, the definition of the Holevo information of the channel \( \mathcal{N} \) gives us final inequality. Putting things together, we find that

\[
I(M;\hat{M}) \leq \chi(\mathcal{N}) + I(M;B'_n) \leq \chi(\mathcal{N}) + I(M;B_n-1B'_n-2),
\]

where the last inequality follows from the data processing inequality. But now, we recognize the quantity \( I(M;B'_n) \) is of the same form as \( I(M;B_nB'_n-1) \) in (12), so that we can iterate through the same sequence of arguments to get

\[
I(M;\hat{M}) \leq 2\chi(\mathcal{N}) + I(M;B_nB'_n-2).
\]

Continuing all the way back to the first channel use, we find

\[
I(M;\hat{M}) \leq n\chi(\mathcal{N}),
\]

since \( I(M;B'_0) = 0 \) (see (5)). Thus, we conclude

\[
R \leq \chi(\mathcal{N}) + \frac{1}{n} f(n,\varepsilon').
\]

Now, the rate for classical communication with \( n \) channel uses and error \( \varepsilon \) is at most that of randomness generation with \( n \) uses and error \( \varepsilon' \), where \( \varepsilon' \) is some function of \( \varepsilon \) such that \( \lim_{\varepsilon \to 0} \varepsilon' = 0 \). Hence, \( g(n,\varepsilon) \equiv \frac{1}{n} f(n,\varepsilon') \) is what we need. \( \triangle \)
**Corollary 2** (weak converse). The classical capacity of entanglement-breaking channels with classical feedback is given by the Holevo information. The same is true for arbitrary channels with separable inputs.

The proof is immediate.

**APPENDIX C**

**Limit value of the α-Holevo information.** The arguments here are essentially the same as those in [20,21], along with an additional insight from [43, Appendix A]. We first recall a minimax

classical feedback is given by the Holevo information. The same is true for arbitrary channels

monotone nondecreasing. We can therefore invoke Lemma 5 with

where

Furthermore, for fixed

\( \forall y \in Y, f(\cdot, y) \) is lower semicontinuous;

2. \( \forall x \in X, f(x, \cdot) \) is monotonic.

Then

\[
\inf_{x \in X} \sup_{y \in Y} f(x, y) = \sup_{y \in Y} \inf_{x \in X} f(x, y).
\]

If the first condition were instead

1. \( \forall y \in Y, f(\cdot, y) \) is upper semicontinuous,

then

\[
\inf_{y \in Y} \sup_{x \in X} f(x, y) = \sup_{x \in X} \inf_{y \in Y} f(x, y).
\]

We now prove Lemma 4.

**Proof of Lemma 4.** For fixed \( \rho_{XR'} \) and \( \alpha, \sigma_{R'} \mapsto \tilde{D}_\alpha(\rho_{XR'}, \rho_X \otimes \sigma_{R'}) \) is lower semicontinuous (see, e.g., [21, Appendix A]). Furthermore, for fixed \( \rho_{XR'} \) and \( \sigma_{R'} \), \( \alpha \mapsto \tilde{D}_\alpha(\rho_{XR'}, \rho_X \otimes \sigma_{R'}) \) is monotone nondecreasing. We can therefore invoke Lemma 5 with \( X = \mathcal{S}(\mathcal{H}_{R'}) \) and \( Y = (0,1) \).

We use this, the definition of the Holevo information, and properties of the quantum relative entropy to find the following one-sided limit of the α-Holevo information of a quantum channel \( \mathcal{N} : \mathcal{B}(\mathcal{H}_{R'}) \rightarrow \mathcal{B}(\mathcal{H}_{R'}) \):

\[
\lim_{\alpha \searrow 1} \chi_\alpha(\mathcal{N}) = \sup_{\alpha \in (0,1)} \chi_\alpha(\mathcal{N}) = \sup_{\alpha \in (0,1)} \sup_{\{p_X(x), \rho_x\}} \chi_\alpha\{p_X(x), \mathcal{N}(\rho_x)\}
\]

\[
= \sup_{\{p_X(x), \rho_x\}} \sup_{\alpha \in (0,1)} \sup_{\sigma_{R'} \in \mathcal{S}(\mathcal{H}_{R'})} \tilde{D}_\alpha(\rho_{XR'} \| \rho_X \otimes \sigma_{R'})
\]

\[
= \sup_{\{p_X(x), \rho_x\}} \sup_{\sigma_{R'} \in \mathcal{S}(\mathcal{H}_{R'})} \inf_{\alpha \in (0,1)} \tilde{D}_\alpha(\rho_{XR'} \| \rho_X \otimes \sigma_{R'})
\]

\[
= \sup_{\{p_X(x), \rho_x\}} \inf_{\sigma_{R'} \in \mathcal{S}(\mathcal{H}_{R'})} \sup_{\{p_X(x), \rho_x\}} \sup_{\alpha \in (0,1)} D(\rho_{XR'} \| \rho_X \otimes \sigma_{R'}) = \sup_{\{p_X(x), \rho_x\}} \sup_{\{p_X(x), \rho_x\}} \sup_{\alpha \in (0,1)} I(X : R')_\rho = \chi(\mathcal{N}),
\]

where

\[
\rho_{XR'} \equiv \sum_x p_X(x)|x\rangle\langle x| \otimes [\mathcal{N}(\rho_x)]_{R'_\rho}.
\]

The fourth equality follows from applying Lemma 5. The fifth follows from (6) and the fact that the sandwiched Rényi relative entropy is monotone nondecreasing in \( \alpha \).

To find the other limit, we note that for fixed \( \rho_R \) and \( \sigma_{R'} \), \( \alpha \mapsto \tilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'}) \) is monotone nondecreasing. We then use an idea from [43, Appendix A]. Note that for fixed \( \alpha \) and positive definite \( \sigma_{R'} \), \( \rho_R \mapsto \tilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'}) \) is continuous (and thus upper semicontinuous). Let \( \varepsilon \in (0,1) \)
and define $\sigma(\varepsilon) \equiv (1 - \varepsilon)\sigma + \varepsilon \pi$, where $\pi$ is the maximally mixed state. Thus $\sigma(\varepsilon)$ is positive definite for $\sigma$ positive semidefinite. We have

$$\sigma(\varepsilon) \geq (1 - \varepsilon)\sigma,$$

which implies that

$$D(\rho \| \sigma(\varepsilon)) \leq D(\rho \| (1 - \varepsilon)\sigma) = D(\rho \| \sigma) - \log(1 - \varepsilon).$$

(This is because of the well-known fact that $\sigma \leq \sigma'$ implies that $D(\rho \| \sigma) \geq D(\rho \| \sigma')$. Hence, using Lemma 5, Lemma 3, and the identification of the Holevo information as an information radius [25, 26], we can conclude the following chain of (in)equalities:

$$\lim_{\alpha \downarrow 1} \tilde{\chi}_\alpha(\mathcal{N}) = \inf_{\alpha > 1} \tilde{\chi}_\alpha(\mathcal{N}) = \inf_{\alpha > 1} \tilde{K}_\alpha(\mathcal{N}) = \inf_{\alpha > 1} \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} \sup_{\rho_R \in \mathcal{S}(H_R)} \tilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'})$$

$$\leq \inf_{\alpha > 1} \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} \sup_{\rho_R \in \mathcal{S}(H_R)} \tilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma(\varepsilon)_{R'})$$

$$= \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} \sup_{\rho_R \in \mathcal{S}(H_R)} \tilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma(\varepsilon)_{R'})$$

$$= \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} \sup_{\rho_R \in \mathcal{S}(H_R)} D(\mathcal{N}(\rho_R) \| \sigma_{R'})$$

$$\leq \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} \sup_{\rho_R \in \mathcal{S}(H_R)} D(\mathcal{N}(\rho_R) \| \sigma_{R'}) - \log(1 - \varepsilon) = \chi(\mathcal{N}) - \log(1 - \varepsilon).$$

The first inequality follows because there is an infimum with respect to $\sigma_{R'}$, such that replacing $\sigma_{R'}$ with $\sigma(\varepsilon)_{R'}$ can never decrease the quantity. The fifth equality follows from applying Lemma 5. The sixth equality follows from (6) and the fact that the sandwiched Rényi relative entropy is monotone nondecreasing in $\alpha$. The last inequality follows from (14). The last equality uses $\chi(\mathcal{N}) = K(\mathcal{N})$.

Given that $\varepsilon > 0$ was arbitrary, we can conclude that $\lim_{\alpha \downarrow 1} \tilde{\chi}_\alpha(\mathcal{N}) \leq \chi(\mathcal{N})$. Combined with the fact that $\tilde{\chi}_\alpha(\mathcal{N}) \geq \chi(\mathcal{N})$ for all $\alpha \geq 1$, we conclude that

$$\lim_{\alpha \downarrow 1} \tilde{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N}).$$

This also implies that

$$\lim_{\alpha \downarrow 1} \tilde{K}_\alpha(\mathcal{N}) = K(\mathcal{N}).$$

We now show that $\lim_{\alpha \downarrow 1} \tilde{K}_\alpha(\mathcal{N}) = K(\mathcal{N})$, which follows from a similar line of reasoning:

$$\lim_{\alpha \downarrow 1} \tilde{K}_\alpha(\mathcal{N}) = \sup_{\alpha \in (0, 1)} \tilde{K}_\alpha(\mathcal{N}) = \sup_{\alpha \in (0, 1)} \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} \sup_{\rho_R \in \mathcal{S}(H_R)} \tilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'})$$

$$= \inf_{\rho_R \in \mathcal{S}(H_R)} \sup_{\alpha \in (0, 1)} \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} \sup_{\rho_R \in \mathcal{S}(H_R)} \tilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'})$$

$$= \inf_{\rho_R \in \mathcal{S}(H_R)} \sup_{\alpha \in (0, 1)} \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} \sup_{\rho_R \in \mathcal{S}(H_R)} D(\mathcal{N}(\rho_R) \| \sigma_{R'})$$

$$= \inf_{\rho_R \in \mathcal{S}(H_R)} \sup_{\alpha \in (0, 1)} \inf_{\sigma_{R'} \in \mathcal{S}(H_{R'})} D(\mathcal{N}(\rho_R) \| \sigma_{R'}) = K(\mathcal{N}),$$

where the third equality follows from Lemma 5 and the fact that for fixed $\alpha$ and $\rho_R$, $\sigma_{R'} \mapsto \tilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'})$ is lower semicontinuous and that the pointwise supremum of lower semicontinuous functions is lower semicontinuous. \(\triangle\)
We are grateful to Patrick Hayden, Milan Mosonyi, and Graeme Smith for insightful discussions about the topic of this paper. We also thank an anonymous referee for many helpful suggestions for improving the paper.

REFERENCES

1. Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, New York: Wiley, 1991.

2. El Gamal, A. and Kim, Y.-H., *Network Information Theory*, Cambridge: Cambridge Univ. Press, 2011.

3. Holevo, A.S., The Capacity of the Quantum Channel with General Signal States, *IEEE Trans. Inform. Theory*, 1998, vol. 44, no. 1, pp. 269–273.

4. Schumacher, B. and Westmoreland, M.D., Sending Classical Information via Noisy Quantum Channels, *Phys. Rev. A*, 1997, vol. 56, no. 1, pp. 131–138.

5. Hastings, M.B., Superadditivity of Communication Capacity Using Entangled Inputs, *Nat. Phys.*, 2009, vol. 5, no. 4, pp. 255–257.

6. Horodecki, M., Shor, P.W., and Ruskai, M.B., Entanglement Breaking Channels, *Rev. Math. Phys.*, 2003, vol. 15, no. 6, pp. 629–641.

7. Shor, P.W., Additivity of the Classical Capacity of Entanglement-Breaking Quantum Channels, *J. Math. Phys.*, 2002, vol. 43, no. 9, pp. 4334–4340.

8. Bowen, G. and Nagarajan, R., On Feedback and the Classical Capacity of a Noisy Quantum Channel, *IEEE Trans. Inform. Theory*, 2005, vol. 51, no. 1, pp. 320–324.

9. Hayashi, M., Error Exponent in Asymmetric Quantum Hypothesis Testing and Its Application to Classical-Quantum Channel Coding, *Phys. Rev. A*, 2007, vol. 76, no. 6, p. 062301.

10. Wolfowitz, J., *Coding Theorems of Information Theory*, Berlin: Springer, 1978, 3rd ed.

11. Arimoto, S., On the Converse to the Coding Theorem for Discrete Memoryless Channels, *IEEE Trans. Inform. Theory*, 1973, vol. 19, no. 3, pp. 357–359.

12. Ogawa, T. and Nagaoka, H., Strong Converse to the Quantum Channel Coding Theorem, *IEEE Trans. Inform. Theory*, 1999, vol. 45, no. 7, pp. 2486–2489.

13. Winter, A., Coding Theorem and Strong Converse for Quantum Channels, *IEEE Trans. Inform. Theory*, 1999, vol. 45, no. 7, pp. 2481–2485.

14. Koenig, R. and Wehner, S., A Strong Converse for Classical Channel Coding Using Entangled Inputs, *Phys. Rev. Lett.*, 2009, vol. 103, no. 7, p. 070504.

15. Wilde, M.M., Winter, A., and Yang, D., Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy, *Comm. Math. Phys.*, 2014, vol. 331, no. 2, pp. 593–622.

16. Bardhan, B.R., García-Patrón, R., Wilde, M.M., and Winter, A., Strong Converse for the Classical Capacity of All Phase-Insensitive Bosonic Gaussian Channels, *IEEE Trans. Inform. Theory*, 2015, vol. 61, no. 4, pp. 1842–1850.

17. Nagaoka, H., Strong Converse Theorems in Quantum Information Theory, *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, Hayashi, M., Ed., Singapore: World Sci., 2005, pp. 64–65.

18. Müller-Lennert, M., Dupuis, F., Szehr, O., Fehr, S., and Tomamichel, M., On Quantum Rényi Entropies: A New Generalization and Some Properties, *J. Math. Phys.*, 2003, vol. 54, no. 12, p. 122203.

19. King, C., Maximal $p$-Norms of Entanglement Breaking Channels, *Quantum Inf. Comput.*, 2003, vol. 3, no. 2, pp. 186–190.

20. Mosonyi, M. and Hiai, F., On the Quantum Rényi Relative Entropies and Related Capacity Formulas, *IEEE Trans. Inform. Theory*, 2011, vol. 57, no. 4, pp. 2474–2487.
21. Cooney, T., Mosonyi, M., and Wilde, M.M., Strong Converse Exponents for a Quantum Channel Discrimination Problem and Quantum-Feedback-Assisted Communication, *Comm. Math. Phys.*, 2016, vol. 344, no. 3, pp. 797–829.

22. Umegaki, H., Conditional Expectation in an Operator Algebra, IV (Entropy and Information), *Kôdai Math. Sem. Rep.*, 1962, vol. 14, no. 2, pp. 59–85.

23. Frank, R.L. and Lieb, E.H., Monotonicity of a Relative Rényi Entropy, *J. Math. Phys.*, 2013, vol. 54, no. 12, p. 122201.

24. Beigi, S., Sandwiched Rényi Divergence Satisfies Data Processing Inequality, *J. Math. Phys.*, 2013, vol. 54, no. 12, p. 122202.

25. Ohya, M., Petz, D., and Watanabe, N., On Capacities of Quantum Channels, *Probab. Math. Statist.*, 1997, vol. 17, no. 1, pp. 179–196.

26. Schumacher, B. and Westmoreland, M.D., Optimal Signal Ensembles, *Phys. Rev. A*, 2001, vol. 63, no. 2, p. 022308.

27. Schumacher, B. and Westmoreland, M.D., Relative Entropy in Quantum Information Theory, *Quantum Computation and Information (Proc. AMS Special Session, Jan. 19–21, 2000, Washington, DC)*, Lomonaco, S.J., Jr. and Brandt, H.E., Eds., Providence, RI: Amer. Math. Soc., 2002, pp. 265–290.

28. Holevo, A.S., Multiplicativity of $p$-Norms of Completely Positive Maps and the Additivity Problem in Quantum Information Theory, *Uspekhi Mat. Nauk*, 2006, vol. 61, no. 2(368), pp. 113–152 [Russian Math. Surveys (Engl. Transl.), 2006, vol. 61, no. 2, pp. 301–339].

29. Sion, M., On General Minimax Theorems, *Pacific J. Math.*, 1958, vol. 8, no. 1, pp. 171–176.

30. Audenaert, K.M.R., A Note on the $p \to q$ Norms of 2-Positive Maps, *Linear Algebra Appl.*, 2009, vol. 430, no. 4, pp. 1436–1440.

31. Bennett, C.H., Devetak, I., Shor, P.W., and Smolin, J.A., Inequalities and Separations among Assisted Capacities of Quantum Channels, *Phys. Rev. Lett.*, 2006, vol. 96, no. 15, p. 150502.

32. Smith, G. and Smolin, J.A., Extensive Nonadditivity of Privacy, *Phys. Rev. Lett.*, 2009, vol. 103, no. 12, p. 120503.

33. Korbicz, J.K., Horodecki, P., and Horodecki, R., Quantum-Correlation Breaking Channels, Broadcasting Scenarios, and Finite Markov Chains, *Phys. Rev. A*, 2012, vol. 86, no. 4, p. 042319.

34. Polyanskiy, Y. and Verdú, S., Arimoto Channel Coding Converse and Rényi Divergence, in *Proc. 48th Annual Allerton Conf. on Communication, Control, and Computation, Sept. 29 – Oct. 1, 2010, Allerton, IL, USA*, pp. 1327–1333.

35. Mosonyi, M. and Ogawa, T., Strong Converse Exponent for Classical-Quantum Channel Coding, *Comm. Math. Phys.*, 2017, vol. 355, no. 1, pp. 373–426.

36. Augustin, U., Noisy Channels, *Habilitation Thesis*, Univ. Erlangen-Nürnberg, Germany, 1978.

37. Dueck, G. and Körner, J., Reliability Function of a Discrete Memoryless Channel at Rates above Capacity, *IEEE Trans. Inform. Theory*, 1979, vol. 25, no. 1, pp. 82–85.

38. Csiszár, I. and Körner, J., Feedback Does Not Affect the Reliability Function of a DMC at Rates above Capacity, *IEEE Trans. Inform. Theory*, 1982, vol. 28, no. 1, pp. 92–93.

39. Brádler, K., Hayden, P., Touchette, D., and Wilde, M.M., Trade-off Capacities of the Quantum Hadamard Channels, *Phys. Rev. A*, 2010, vol. 81, no. 6, p. 062312.

40. Wilde, M.M., *Quantum Information Theory*, Cambridge: Cambridge Univ. Press, 2017, 2nd ed.

41. Fannes, M., A Continuity Property of the Entropy Density for Spin Lattice Systems, *Comm. Math. Phys.*, 1973, vol. 31, no. 4, pp. 291–294.

42. Audenaert, K.M.R., A Sharp Continuity Estimate for the von Neumann Entropy, *J. Phys. A*, 2007, vol. 40, no. 28, pp. 8127–8136.

43. Tomamichel, M., Wilde, M.M., and Winter, A., Strong Converse Rates for Quantum Communication, *IEEE Trans. Inform. Theory*, 2017, vol. 63, no. 1, pp. 715–727.