# Entanglement Cost of Quantum Channels

Mario Berta,[1,*] Fernando G.S.L. Brandão,[1,2,3,†] Matthias Christandl,[1,‡] and Stephanie Wehner[3,§]

[1] *Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland.*
[2] *Departamento de Fisica, Universidade Federal de Minas Gerais, Belo Horizonte 30123-970, Brazil.*
[3] *Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, 117543 Singapore.*

(Dated: November 3, 2015)

The entanglement cost of a quantum channel is the minimal rate at which entanglement (between sender and receiver) is needed in order to simulate many copies of a quantum channel in the presence of free classical communication. In this paper we show how to express this quantity as a regularised optimization of the entanglement formation over states that can be generated between sender and receiver. Our formula is the channel analog of a well-known formula for the entanglement cost of quantum states in terms of the entanglement of formation; and shares a similar relation to the recently shattered hope for additivity.

The entanglement cost of a quantum channel can be seen as the analog of the quantum reverse Shannon theorem in the case where free classical communication is allowed. The techniques used in the proof of our result are then also inspired by a recent proof of the quantum reverse Shannon theorem and feature the one-shot formalism for quantum information theory, the post-selection technique for quantum channels as well as Sion's minimax theorem.

We discuss two applications of our result. First, we are able to link the security in the noisy-storage model to a problem of sending quantum rather than classical information through the adversary's storage device. This not only improves the range of parameters where security can be shown, but also allows us to prove security for storage devices for which no results were known before. Second, our result has consequences for the study of the strong converse quantum capacity. Here, we show that any coding scheme that sends quantum information through a quantum channel at a rate larger than the entanglement cost of the channel has an exponentially small fidelity.



* berta@caltech.edu
† fgslbrandao@gmail.com
‡ christandl@math.ku.dk
§ s.d.c.wehner@tudelft.nl

## I. INTRODUCTION

The quantification of the information theoretic power of quantum channels is one of the most fundamental problems in quantum information theory. Of particular interest is thereby the study of a channel's capacity for information transmission. This quantity corresponds to the number of bits $m$ that can be sent reliably when using the channel $n$ times using optimal encoding and decoding operations. Unlike classical channels, quantum channels have various distinct capacities, depending on the kind of information that is sent (e.g. classical or quantum) or on the kind of assistance that is allowed (e.g. free entanglement or free classical communication). Important examples of quantum channel capacities include the entanglement assisted classical capacity $C_E$ [1], and the classical communication assisted quantum capacities $Q_{\rightarrow}$, $Q_{\leftarrow}$ and $Q_{\leftrightarrow}$ depending on the direction of the assisting communication [2–4].

One way of tackling the problem of capacities is to think more broadly in terms of channel simulations. For example, the process of sending $m$ bits reliably using $n$ uses of a channel $\mathcal{E}$ can be understood as a simulation of $m$ perfect, noise-free, channels using $n$ copies of $\mathcal{E}$. The capacity of the channel $\mathcal{E}$ is then simply the rate $m/n$ at which such a simulation is possible in the limit of large $n$. One can also turn the problem upside down and ask: What is the optimal rate at which a perfect channel can simulate a noisy one? When the simulation can consume free entanglement between the sender and the receiver, this question is answered by the quantum reverse Shannon theorem. It states that the optimal rate is given by the entanglement assisted classical capacity $C_E$ [5, 6]. Apart from its deep conceptual appeal, the quantum reverse Shannon theorem led to the proof that the $C_E$ is in fact a strong converse capacity.

It is natural to ask how these capacities change in the presence of other free resources. In this work, we consider the simulation of a noisy quantum channel $\mathcal{E}$ by a noise-free channel in the presence of free classical communication. It turns out not to matter whether we allow free classical forward, backward, or even two-way communication, the capacity is the same in all scenarios. The problem we are considering can therefore be understood as the 'reverse problem' for all three classical communication assisted quantum capacities. Note that by quantum teleportation [7], the perfect quantum channel can equivalently be replaced with perfect entanglement. The central question of this paper can thus be summarized as

> *At what rate is entanglement, in the form of ebits, needed in order to asymptotically simulate a quantum channel $\mathcal{E}$, when classical communication is given for free?*

We call this rate the entanglement cost $E_C$ of a quantum channel. Our main contribution in this paper is to prove the following formula

$$E_C(\mathcal{E}) = \lim_{n\to\infty} \frac{1}{n} \max_{\psi^n} E_F\left(\left(\mathcal{E}^{\otimes n} \otimes \mathcal{I}\right)(\psi^n)\right) \ , \tag{1}$$

where the maximization is over all purifications $\psi^n$ of input states to the $n$-fold tensor product quantum channel $\mathcal{E}^{\otimes n}$ and $\mathcal{I}$ stands for the identity channel on the purifying system. The entanglement of formation $E_F$ is computed between purifying system and channel output; it is defined as

$$E_F(\rho_{AB}) = \inf_{\{p_i,\rho^i\}} \sum_i p_i H(A)_{\rho^i} \ , \tag{2}$$

where the infimum ranges over all pure state decompositions $\rho_{AB} = \sum_i p_i \rho^i_{AB}$, and $H(\cdot)$ denotes the von Neumann entropy. Note that expression (1) involves a regularization, and is therefore not a single-letter formula. Even if we would know that we can restrict the maximization to non-entangled input states, equation (1) would still not reduce to such a formula, due to Hasting's counterexample for the additivity of the entanglement of formation [8, 9].[1] Note also that $E_C$ is generally larger than $Q_{\leftrightarrow}$,[2] in fact more strikingly, there exist so-called bound entangled channels $\mathcal{E}$ (for instance entangling PPT channels) for which $E_C(\mathcal{E}) > Q_{\leftrightarrow}(\mathcal{E}) = 0$. This fact highlights an important difference compared to the case of free entanglement where the quantum reverse Shannon theorem implies that the corresponding rates are equal. In particular, when $E_C(\mathcal{E}) > Q_{\leftrightarrow}(\mathcal{E})$, the concatenated protocol which first simulates $\mathcal{E}$ from a noiseless channel and then the noiseless channel from $\mathcal{E}$ will result in a net loss.

---

[1] However we want to emphasize that we can compute explicit upper bounds for $E_C$, which are particularly useful for the applications given below.

[2] The same applies to $Q_{\rightarrow}$ and $Q_{\leftarrow}$ since both are smaller or equal to $Q_{\leftrightarrow}$.

As the name entanglement cost suggests, $E_C(\mathcal{E})$ is the quantum channel analog of the entanglement cost of quantum states $E_C(\rho_{AB})$, which corresponds to the rate of entanglement needed in order to generate a bipartite quantum state $\rho_{AB}$ [10]. Our formula (1) can be seen as the channel analog of the following well-known result for the quantum state problem [11]

$$E_C(\rho_{AB}) = \lim_{n\to\infty} \frac{1}{n} E_F(\rho_{AB}^{\otimes n}) \ . \tag{3}$$

and the gap between $E_C(\mathcal{E})$ and $Q_{\leftrightarrow}(\mathcal{E})$ has its analog in the gap between $E_C(\rho_{AB})$ and $E_D(\rho_{AB})$, the distillable entanglement.

We present two applications. The first one concerns the security in the noisy-storage model [12–14]. For the first time, we relate security in this model to a problem of sending quantum rather than classical information through the adversary's storage device. In particular, we show that any two-party cryptographic primitive can be implemented securely whenever

$$E_C(\mathcal{E}) \cdot \nu < \frac{1}{2} \ , \tag{4}$$

where the adversary's storage is of the form $\mathcal{E}^{\otimes \nu \cdot m}$, $m$ is the number of qubits transmitted during the protocol, and $\nu$ is the storage rate (see Section IV.1 for precise definitions). Our analysis improves the range of parameters when security can be obtained. We illustrate our results with explicit calculations for a number of specific channels. In particular, we obtain non-trivial bounds for dephasing noise and for any qubit channel - for instance the amplitude damping channel.

The second application of our result is an upper bound on the strong converse capacity for sending quantum information. The strong converse capacity is the minimal rate above which any attempt to send information necessarily has exponentially small fidelity.[3] The strong converse capacity for sending classical information is known to be equal to the classical capacity for a selected number of channels [15], or under additional assumptions [16, 17]. For many channels there are also upper bounds known [5, 18, 19], but a general formula for the strong converse classical capacity is not known. Understanding the strong converse capacity for sending quantum information turns out to be an even more elusive problem, and the only previous result relies on a statement involving the transmission of classical information [5]. Here, we make progress by showing that any coding scheme sending quantum information (using free forward, backward or two-way classical communication) at an asymptotic rate higher than the entanglement cost $E_C$, must have an exponentially small fidelity.

The proof of our main result (1) is based on one-shot information theory, which makes statements about structureless resources avoiding the usual requirement of independence and identical distribution (i.i.d.). The role of von Neumann entropies in the i.i.d. scenario is taken by min- and max-entropies from the smooth entropy formalism [20–26]. We work in this formalism and the proof of our main result is conceptually very similar to the proof of the quantum reverse Shannon theorem given in [6]. In order to prove the direct part of (1), we need to show the existence of a channel simulation for $\mathcal{E}^{\otimes n}$, whose asymptotic rate of entanglement consumption is upper bounded by $E_C(\mathcal{E})$. That is, we need to construct a completely positive and trace preserving (CPTP) map that is arbitrarily close to $\mathcal{E}^{\otimes n}$ in the diamond norm[4] and that uses local operations and classical communication as well as ebits at a rate of at most $E_C(\mathcal{E})$. Here it is worth noting that even though the channel we wish to simulate has i.i.d. structure, the channel simulation also has to work on non-i.i.d. inputs. The crucial idea in order to deal with this fact is to employ the post-selection technique for quantum channels [28], which is a tool to bound the distance in diamond norm between two completely positive and trace preserving (CPTP) maps. The technique upper bounds this distance by the distance arising from the purification of a special de Finetti input state.[5] With this, it is sufficient to find a CPTP map that does the channel simulation on the purification of this special de Finetti state, and to quantify how much entanglement this consumes. Since the state is a purification of a de Finetti state (and not a de Finetti state itself) it does not have i.i.d. structure. In order to deal with this fact we employ ideas from the one-shot entanglement cost for quantum states $E_C^{(1)}(\rho_{AB}, \varepsilon)$, which quantifies how much entanglement is needed in order to create one single copy of a bipartite quantum state $\rho_{AB}$ using local operations and classical communication [29, 30].[6] The resulting entanglement cost of the channel simulation is then upper bounded by an expression similar to (1), but with the

[3] Note that the strong converse capacity is greater or equal than the standard capacity (which is defined as the minimal rate above which the fidelity does not approach one).

[4] The diamond norm is the dual of the completely bounded norm [27].

[5] A de Finetti state consists of n identical and independent copies of an (unknown) state on a single subsystem.

[6] This is in contrast to the quantity $E_C(\rho_{AB})$ mentioned before, which answers the question of how much entanglement is needed in the asymptotic i.i.d. regime.

maximization over input states and the minimization in the definition of the entanglement of formation interchanged. Finally, in order to arrive at (1), we discretize the set of Kraus decompositions of $\mathcal{E}$ and apply von Sion's minimax theorem to swap the minimization and the maximization [31]. The proof of the converse follows a standard argument applied to the one-shot entanglement cost.

This paper is structured as follows. In Section II we introduce notation, definitions and state some basic lemmas. In particular, we review the results of [29] about the one-shot entanglement cost of quantum states. In Section III we derive our main result; we define and quantify the entanglement cost of quantum channels. This is followed by a discussion of applications in Section IV. Finally we end with a summary and give an outlook (Section V). The arguments are based on various technical statements, which are proven in Appendices A - C.

## II. PRELIMINARIES

We assume that all Hilbert spaces, in the following denoted $\mathcal{H}$, are finite-dimensional. The dimension of $\mathcal{H}_A$ is denoted by $|A|$. The set of linear operators on $\mathcal{H}$ is denoted by $\mathcal{L}(\mathcal{H})$ and the set of positive semi-definite operators on $\mathcal{H}$ is denoted by $\mathcal{P}(\mathcal{H})$. We define the sets of sub-normalized states $\mathcal{S}_{\leq}(\mathcal{H}) = \{\rho \in \mathcal{P}(\mathcal{H}) : \mathrm{tr}[\rho] \leq 1\}$, normalized states $\mathcal{S}(\mathcal{H}) = \{\rho \in \mathcal{P}(\mathcal{H}) : \mathrm{tr}[\rho] = 1\}$), and normalized pure states $\mathcal{V}(\mathcal{H}) = \{|\psi\rangle\langle\psi| \in \mathcal{S}(\mathcal{H}) : |\psi\rangle \in \mathcal{H}\}$. The tensor product of $\mathcal{H}_A$ and $\mathcal{H}_B$ is denoted by $\mathcal{H}_{AB} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$. Given a multipartite operator $\rho_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$, we write $\rho_A = \mathrm{tr}_B[\rho_{AB}]$ for the corresponding reduced operator. For $M_A \in \mathcal{L}(\mathcal{H}_A)$, we write $M_A \equiv M_A \otimes \mathbb{1}_B$ for the enlargement on any $\mathcal{H}_{AB}$, where $\mathbb{1}_B$ denotes the identity in $\mathcal{L}(\mathcal{H}_B)$. For $\mathcal{H}_A$, $\mathcal{H}_B$ with orthonormal bases $\{|i\rangle_A\}_{i=1}^{|A|}$, $\{|i\rangle_B\}_{i=1}^{|B|}$ and $|A| = |B|$, the canonical identity mapping from $\mathcal{L}(\mathcal{H}_A)$ to $\mathcal{L}(\mathcal{H}_B)$ with respect to these bases is denoted by $\mathcal{I}_{A\to B}$, i.e. $\mathcal{I}_{A\to B}(|i\rangle\langle j|_A) = |i\rangle\langle j|_B$. A linear map $\mathcal{E}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is positive if $\mathcal{E}_{A\to B}(\rho_A) \in \mathcal{P}(\mathcal{H}_B)$ for all $\rho_A \in \mathcal{P}(\mathcal{H}_A)$. It is completely positive if the map $(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{C\to C})$ is positive for all $\mathcal{H}_C$. Completely positive and trace preserving maps are called CPTP maps or quantum channels. The support of $\rho \in \mathcal{P}(\mathcal{H})$ is denoted by $\mathrm{supp}(\rho)$, the projector onto $\mathrm{supp}(\rho)$ is denoted by $\rho^0$ and $\mathrm{tr}\left[\rho^0\right] = \mathrm{rank}(\rho)$, the rank of $\rho$. For $\rho \in \mathcal{P}(\mathcal{H})$ we write $\|\rho\|_\infty$ for the operator norm of $\rho$, which is equal to the maximum eigenvalue of $\rho$. The trace norm of $M \in \mathcal{L}(\mathcal{H})$ is defined as $\|M\|_1 = \mathrm{tr}\left[\sqrt{M^\dagger M}\right]$, and the Hilbert-Schmidt norm of $M$ is given by $\|M\|_2 = \sqrt{\mathrm{tr}\left[M^\dagger M\right]}$.

Recall the following standard definitions. The *von Neumann entropy* of $\rho \in \mathcal{P}(\mathcal{H})$ is defined as $H(\rho) = -\mathrm{tr}[\rho \log \rho]$,[7] and the *conditional von Neumann entropy* of $A$ given $B$ for $\rho_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$ is given by

$$H(A|B)_\rho = H(AB)_\rho - H(B)_\rho \ . \tag{5}$$

**Definition 1.** Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$. The *entanglement of formation* of $\rho_{AB}$ is defined as

$$E_F(\rho_{AB}) = \inf_{\{p_i,\rho^i\}} \sum_i p_i H(A)_{\rho^i} = \inf_{\{p_i,\rho^i\}} H(A|R)_\rho \ , \tag{6}$$

where the infimum ranges over all pure states decompositions $\rho_{AB} = \sum_i p_i \rho^i_{AB}$ and $\rho_{AR} = \sum_i p_i \rho^i_A \otimes |i\rangle\langle i|_R$.

**Definition 2.** Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$. The *alternative max-entropy* of $A$ conditioned on $B$ is defined as

$$H_0(A|B)_\rho = \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \log \mathrm{tr}\left[\rho^0_{AB}(\mathbb{1}_A \otimes \sigma_B)\right] \ . \tag{7}$$

In the literature this quantity is also known as conditional max-entropy [20, 26] or conditional zero-Rényi entropy [29]. We will evaluate the alternative conditional max-entropy in particular on quantum-classical states.

**Lemma 3.** Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$ with $\rho_{AB} = \sum_{k\in K} \rho^k_A \otimes |k\rangle\langle k|_B$, $\rho^k_A \in \mathcal{P}(\mathcal{H}_A)$, and the $|k\rangle_B$ mutually orthogonal (i.e. the state is classical on $B$). Then,

$$H_0(A|B)_\rho = \max_{k\in K} H_0(A)_{\rho^k} \ . \tag{8}$$

[7] log denotes the logarithm to base 2.

*Proof.* We calculate

$$H_0(A|B)_\rho = \max_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \log \operatorname{tr}\left[\rho_{AB}^0 (\mathbb{1}_A \otimes \sigma_B)\right] \tag{9}$$

$$= \log \max_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \operatorname{tr}\left[\left(\sum_k \left(\rho_A^k\right)^0 \otimes |k\rangle\langle k|_B\right)(\mathbb{1}_A \otimes \sigma_B)\right] \tag{10}$$

$$= \log \max_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \operatorname{tr}\left[\sigma_B \cdot \left(\sum_k |k\rangle\langle k|_B \cdot \operatorname{tr}\left[\left(\rho_A^k\right)^0\right]\right)\right] \tag{11}$$

$$= \log \left\| \sum_k |k\rangle\langle k|_B \cdot \operatorname{tr}\left[\left(\rho_A^k\right)^0\right] \right\|_\infty = \log \max_k \operatorname{tr}\left[\left(\rho_A^k\right)^0\right] = \max_k H_0(A)_{\rho^k} \ . \tag{12}$$

□

Smooth entropy measures are defined by extremizing the non-smooth measures over a set of nearby states. Since we will later use some of the ideas from [29], we use the same definitions as in [29].

**Definition 4.** Let $\varepsilon \geq 0$, and $\rho_{AB} = \sum_k \rho_A^k \otimes |k\rangle\langle k|_B \in \mathcal{S}(\mathcal{H}_{AB})$. The *smooth alternative max-entropy* of $A$ conditioned on $B$ is defined as

$$H_0^\varepsilon(A|B)_\rho = \sup_{\overline{\rho}_{AB} \in \mathcal{B}_{qc}^\varepsilon(\rho_{AB})} H_0(A|B)_{\overline{\rho}} \ , \tag{13}$$

where

$$\mathcal{B}_{qc}^\varepsilon(\rho_{AB}) = \{\bar{\rho}_{AB} \in \mathcal{P}(\mathcal{H}) : \bar{\rho}_{AB} = \sum_k \bar{\rho}_A^k \otimes |k\rangle\langle k|_B, \|\rho_{AB} - \bar{\rho}_{AB}\|_1 \leq \varepsilon\} \ . \tag{14}$$

In the technical part of this paper we will need distance measures. For $\rho, \sigma \in \mathcal{S}_\leq(\mathcal{H})$ the purified distance is defined as [25, Definition 4]

$$P(\rho, \sigma) = \sqrt{1 - \bar{F}^2(\rho, \sigma)} \ , \tag{15}$$

where $\bar{F}(\cdot, \cdot)$ denotes the generalized fidelity (which equals the standard fidelity[8] if at least one of the states is normalized),

$$\bar{F}(\rho, \sigma) = F(\rho, \sigma) + \sqrt{(1 - \operatorname{tr}[\rho])\,(1 - \operatorname{tr}[\sigma])} \ . \tag{16}$$

The purified distance is a metric on $\mathcal{S}_\leq(\mathcal{H})$ [25, Lemma 5]. Henceforth we call $\rho$, $\sigma \in \mathcal{S}_\leq(\mathcal{H})$ $\varepsilon$-close if $P(\rho, \sigma) \leq \varepsilon$ and denote this by $\rho \approx_\varepsilon \sigma$. Furthermore, we will also need a distance measure for quantum channels. We use a norm on the set of CPTP maps which measures the probability by which two such mappings can be distinguished. The norm is known as the diamond norm in quantum information theory [27]. Here, we present it in a formulation which highlights that it is dual to the well-known completely bounded (cb) norm [32].

**Definition 5.** Let $\mathcal{E}_A : \mathcal{L}(\mathcal{H}_A) \mapsto \mathcal{L}(\mathcal{H}_B)$ be a linear map. The diamond norm of $\mathcal{E}_A$ is defined as

$$\|\mathcal{E}_A\|_\diamond = \sup_{k \in \mathbb{N}} \|\mathcal{E}_A \otimes \mathcal{I}_k\|_1 \ , \tag{17}$$

The supremum in Definition 5 is reached for $k = |A|$ [27, 32]. We call two CPTP maps $\mathcal{E}$ and $\mathcal{F}$ $\varepsilon$-close if they are $\varepsilon$-close in the metric induced by the diamond norm.

It is the main of result of [29] to quantify how much entanglement is needed in order to create a single copy of a bipartite state $\rho_{AB}$ [30], a scenario previously studied in the asymptotic i.i.d. setting [10, 11].

---

[8] The fidelity between $\rho, \sigma \in \mathcal{S}_\leq(\mathcal{H})$ is defined as $F(\rho, \sigma) = \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1$.

**Definition 6.** Consider a bipartite system with parties Alice and Bob, where Alice controls a system $\mathcal{H}_A$ and Bob $\mathcal{H}_B$. Let $\varepsilon \geq 0$, $\Phi_{\bar{A}\bar{B}}$ be a maximally entangled state between Alice and Bob, and $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$. An $\varepsilon$-faithful one-shot entanglement dilution protocol for $\rho_{AB}$ is a local operation and classical communication (LOCC) operation $\Lambda$ between Alice and Bob with $\bar{A} \to A$ at Alice's side and $\bar{B} \to B$ at Bob's side, such that

$$\Lambda(\Phi_{\bar{A}\bar{B}}) \approx_\varepsilon \rho_{AB} \ . \tag{18}$$

If $\Phi_{\bar{A}\bar{B}}$ has Schmidt rank $R$, $\log R$ is the dilution cost of the one-shot entanglement dilution protocol.

**Definition 7.** Let $\varepsilon \geq 0$ and $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$. The minimal dilution cost of all $\varepsilon$-faithful one-shot entanglement dilution protocols for $\rho_{AB}$ is called $\varepsilon$-faithful *one-shot entanglement cost* of $\rho_{AB}$ and is denoted by $E_C^{(1)}(\rho_{AB}, \varepsilon)$.

**Proposition 8.** [29, Theorem 1] Let $\varepsilon \geq 0$ and $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$. Then,

$$\min_{\{p_i, \rho^i\}} H_0^{2\sqrt{\varepsilon}}(A|R)_\rho \leq E_C^{(1)}(\rho_{AB}, \varepsilon) \leq \min_{\{p_i, \rho^i\}} H_0^{\varepsilon/2}(A|R)_\rho \ , \tag{19}$$

where the minimum ranges over all pure states decompositions $\rho_{AB} = \sum_i p_i \rho_{AB}^i$ and $\rho_{AR} = \sum_i p_i \rho_A^i \otimes |i\rangle\langle i|_R$.

The idea for the achievability is as follows. For any pure state decomposition $\rho_{AB} = \sum_i p_i \rho_{AB}^i$ Alice can locally create the classical-quantum state $\rho_{ABR} = \sum_i p_i \rho_{AB}^i \otimes |i\rangle\langle i|_R$, and then, conditioned on the index $i$, teleport the $B$-part of the pure states $\rho_{AB}^i$ to Bob. Minimizing over all pure state decompositions, a straightforward analysis shows that the resulting entanglement cost is bounded as in Proposition 8. We will make use of these ideas for the proof of our main theorem.

**Remark 9.** The bounds given in (19) also hold if we only allow one-way classical communication (forward or backward).

## III. ENTANGLEMENT COST OF QUANTUM CHANNELS

### III.1. Main Result

We are now in the position to define the entanglement cost of quantum channels and prove the main result of this paper, Theorem 12.

**Definition 10.** Consider a bipartite system with parties Alice and Bob. Let $\varepsilon \geq 0$, $\Phi_{\bar{A}\bar{B}}$ be a maximally entangled state between Alice and Bob, and $\mathcal{E} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map, where Alice controls $\mathcal{H}_A$ and Bob $\mathcal{H}_B$. A one-shot channel simulation for $\mathcal{E}$ with error $\varepsilon$ is a quantum protocol

$$\begin{aligned} \mathcal{F} : \quad & \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B) \\ & \rho_A \qquad \mapsto \Lambda(\rho_A \otimes \Phi_{\bar{A}\bar{B}}) \ , \end{aligned} \tag{20}$$

where $\Lambda$ is a LOCC operation between Alice and Bob with $A\bar{A} \to 0$ (no output) at Alice's side and $\bar{B} \to B$ at Bob's side, as well as

$$\|\mathcal{F} - \mathcal{E}\|_\diamond \leq \varepsilon \ . \tag{21}$$

If $\Phi_{\bar{A}\bar{B}}$ has Schmidt rank $R$, $\log R$ is the entanglement cost of the one-shot channel simulation.

By the definition of the diamond norm (Definition 5), this assures that for any possible input state, the output of the channel simulation $\mathcal{F}$ can only distinguished with small probability from the corresponding output of $\mathcal{E}$.

**Definition 11.** Let $\mathcal{E} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map. An asymptotic channel simulation for $\mathcal{E}$ is a sequence of one-shot channel simulations $\mathcal{F}^n$ for $\mathcal{E}^{\otimes n}$ with error $\varepsilon_n$, such that that $\lim_{n\to\infty} \varepsilon_n = 0$. The entanglement cost of the simulation is $\limsup_{n\to\infty} \frac{\log R_n}{n}$.

In the language of general channel simulations this corresponds to a so-called non-feedback simulation, since Alice does not obtain the output of the complementary channel [5].

**Theorem 12.** Let $\mathcal{E}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map. Then, the minimal entanglement cost $E_C(\mathcal{E}_{A\to B})$ of an asymptotic channel simulation for $\mathcal{E}_{A\to B}$ is given by

$$E_C(\mathcal{E}_{A\to B}) = \lim_{n\to\infty} \frac{1}{n} \max_{\psi^n_{AA'}} E_F\left(\left(\mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi^n_{AA'})\right) \ , \tag{22}$$

where $\psi^n_{AA'} = \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$.

*Proof.* We first show that the right-hand side of (22) can be achieved (Proposition 15), and thereafter that it is also a lower bound (Proposition 16). □

### III.2. Proof: Achievability

The proof proceeds in three steps leading to Proposition 15. The basic idea is as follows. Given a quantum channel $\mathcal{E}$, we need to show the existence of a sequence of one-shot channel simulations with asymptotically vanishing error, and an asymptotic entanglement cost upper bounded by the right-hand side of (22). The crucial step is that by the post-selection technique for quantum channels (Proposition 32), it is sufficient to come up with CPTP map (which consists of using maximally entangled states, local operations, and classical communication) that works for the purification of one special de Finetti input state. For this we use ideas from the one-shot entanglement cost of quantum states (Proposition 8).

**Lemma 13.** Let $\mathcal{E}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map. Then,

$$E_C(\mathcal{E}_{A\to B}) \leq \inf_{\{M^k_{A\to B}\}} \sup_{\psi_A \in \mathcal{S}(\mathcal{H}_A)} \sum_k p_k H(B)_{\psi^k} \ , \tag{23}$$

where the infimum is over all Kraus decompositions $\{M^k_{A\to B}\}$ of $\mathcal{E}_{A\to B}$, $\psi^k_B = \frac{1}{p_k} M^k_{A\to B} \psi_A \ {M^k_{A\to B}}^\dagger$ and $p_k = \mathrm{tr}\left[M^k_{A\to B}\psi_A \ {M^k_{A\to B}}^\dagger\right]$.

*Proof.* We construct a sequence of one-shot channel simulations $\mathcal{F}^n$ with asymptotically vanishing error $\varepsilon_n$, and an asymptotic entanglement cost $\frac{\log R_n}{n}$ as in (23). Without lost of generality we choose $\mathcal{F}^n$ to be permutation-covariant.[9] The post-selection technique (Proposition 32) applies to permutation-covariant quantum channels and upper bounds the error by

$$\varepsilon_n = \left\|\mathcal{E}^{\otimes n}_{A\to B} - \mathcal{F}^n_{A\to B}\right\|_\diamondsuit \leq (n+1)^{|A|^2-1} \cdot \left\|\left(\left(\mathcal{E}^{\otimes n}_{A\to B} - \mathcal{F}^n_{A\to B}\right) \otimes \mathcal{I}^{\otimes n}_{A'} \otimes \mathcal{I}_E\right)(\zeta^n_{AA'E})\right\|_1 \ , \tag{24}$$

where $\zeta^n_{AA'E}$ is a purification of the de Finetti state $\zeta^n_{AA'} = \int \psi^{\otimes n}_{AA'} d(\psi_{AA'})$ with $\psi_{AA'} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$, $\mathcal{H}_{A'} \cong \mathcal{H}_A$ and $d(\cdot)$ the measure on the normalized pure states on $\mathcal{H}_A \otimes \mathcal{H}_{A'}$ induced by the Haar measure on the unitary group acting on $\mathcal{H}_A \otimes \mathcal{H}_{A'}$, normalized to $\int d(\cdot) = 1$. Hence it is sufficient that the channel simulation $\mathcal{F}^n$ works on the state $\zeta^n_{AA'E}$ leading to

$$\omega^n_{BA'E} = \left(\mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}^{\otimes n}_{A'} \otimes \mathcal{I}_E\right)(\zeta^n_{AA'E}) \ , \tag{25}$$

up to an error $o\left((n+1)^{1-|A|^2}\right)$ in trace distance, for an asymptotic entanglement cost smaller than (23).

For $\{M^{n,k}_{A\to B}\}$ a Kraus decomposition of $\mathcal{E}^{\otimes n}_{A\to B}$, Alice locally applies the CPTP map with Kraus operators $M^{n,k}_{A\to B} \otimes |k\rangle_R$ to the state $\zeta^n_{AA'E}$ and sends a copy of the classical register $k$ to Bob creating the state

$$\omega^n_{A'BER} = \sum_k M^{n,k}_{A\to B} \zeta^n_{AA'E} \ {M^{n,k}_{A\to B}}^\dagger \otimes |k\rangle\langle k|_R \ . \tag{26}$$

---

[9] This can be seen as follows. First, Alice and Bob create shared randomness using classical communication. Then, Alice applies a random permutation $\pi$ on the input system chosen according to the shared randomness. This is followed by the original map (which might not yet be permutation-covariant), and Bob who undoes the permutation by applying $\pi^{-1}$ on the output system. If needed, the classical communication cost of this procedure can be kept sub-linear in $n$ by using randomness recycling, as discussed in [5, Section IV. D]. Alternatively, one could also use a sub-linear amount of entanglement to assure the permutation covariance.

Conditioned on $k$ Alice and Bob can now use $\log \operatorname{rank}\left(M^{n,k}_{A\to B}\right)$ many ebits to teleport $B$ from Alice to Bob (since $\zeta^n_{AA'E}$ is pure). By a property of the alternative conditional max-entropy on quantum-classical states (Lemma 3) this then leads to a total entanglement cost of

$$H_0(B|R)_{\omega^n} = \log \max_k \operatorname{rank}\left(M^{n,k}_{A\to B}\right) \ . \tag{27}$$

Moreover, at the cost of an approximation error $\delta_n \geq 0$ in purified distance, this can be reduced to $H_0^{\delta_n/4}(B|R)_{\omega^n}$. This is achieved by pretending that we have another quantum-classical state $\bar{\omega}^n_{A'BER}$ which is $\delta_n/4$ close to $\omega^n_{A'BER}$ in trace distance, and then applying the teleportation protocol defined by $\bar{\omega}^n_{A'BER}$.[10] Now by taking an infimum over all Kraus decomposition $\{M^{n,k}_{A\to B}\}$ of $\mathcal{E}^{\otimes n}_{A\to B}$ we get a $\delta_n$-faithful (measured in purified distance) channel simulation of $\mathcal{F}^n$ on the input state $\zeta^n_{AA'E}$ for an entanglement cost upper bounded by

$$\inf_{\{M^{n,k}_{A\to B}\}} H_0^{\delta_n/4}(B|R)_{\omega^n} \ . \tag{28}$$

Using a corollary of Carathéodory's theorem (Lemma 33), we know that

$$\zeta^n_{AA'} \equiv \int \psi^{\otimes n}_{AA'} d(\psi_{AA'}) = \sum_{j=1}^N q_j \left(\psi^j_{AA'}\right)^{\otimes n} \ , \tag{29}$$

with $\psi^j_{AA'} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$, $N = (n+1)^{2\left(|A|^2-1\right)}$ and $\{q_j\}_{j=1}^N$ a probability distribution. This allows us to write

$$\omega^n_{BR} = \sum_{j=1}^N q_j \sum_k M^{n,k}_{A\to B}\left(\psi^j_A\right)^{\otimes n} {M^{n,k}_{A\to B}}^\dagger \otimes |k\rangle\langle k|_R \ . \tag{30}$$

One particular choice for a Kraus decomposition $\{M^{n,k}_{A\to B}\}$ of $\mathcal{E}^{\otimes n}_{A\to B}$ in (28) is then to choose a Kraus decomposition $\{M^k_{A\to B}\}$ for $\mathcal{E}_{A\to B}$ and take this decomposition for every tensor product factor. Thus we find a $\delta_n$-faithful (measured in purified distance) channel simulation of $\mathcal{F}^n$ on the input state $\zeta^n_{AA'E}$ for an entanglement cost upper bounded by

$$\inf_{\{M^k_{A\to B}\}} H_0^{\delta_n/4}(B|R)_{\omega^n} \ , \tag{31}$$

where the infimum ranges over all Kraus decompositions $\{M^k_{A\to B}\}$ of $\mathcal{E}_{A\to B}$, and $\omega^n_{BR} = \sum_{j=1}^N q_j \omega^j_{BR}$ with

$$\omega^j_{BR} = \sum_k M^k_{A\to B}\psi^j_A \ {M^k_{A\to B}}^\dagger \otimes |k\rangle\langle k|_R \ . \tag{32}$$

But by a property of the smooth alternative conditional max-entropy (Lemma 27) we have

$$\inf_{\{M^k_{A\to B}\}} H_0^{\delta_n/4}(B|R)_{\omega^n} \leq \inf_{\{M^k_{A\to B}\}} \max_j H_0^{\delta_n/4}(B|R)_{(\omega^j)^{\otimes n}} + 2\left(|A|^2-1\right)\cdot\log(n+1) \ . \tag{33}$$

Using the asymptotic equipartition property for the smooth alternative conditional max-entropy (Lemma 31) we arrive at an entanglement cost of

$$n\cdot\left\{\inf_{\{M^k_{A\to B}\}} \max_j H(B|R)_{\omega^j}\right\} + \sqrt{n}\cdot\log\left(|B|+3\right)\cdot\sqrt{\log\left(\frac{16}{\delta_n^2}\right)} + 2\left(|A|^2-1\right)\cdot\log(n+1) \ . \tag{34}$$

Now we choose $\delta_n = \frac{1}{2}(n+1)^{2\left(1-|A|^2\right)}$ and the entanglement cost becomes

$$n\cdot\left\{\inf_{\{M^k_{A\to B}\}} \max_j H(B|R)_{\omega^j}\right\} + \sqrt{n}\cdot\log\left(|B|+3\right)\cdot\sqrt{2+4\cdot\log(n+1)\cdot\left(|A|^2-1\right)} + 2\left(|A|^2-1\right)\cdot\log(n+1) \ . \tag{35}$$

---

[10] See Lemma 34 for the equivalence of distance measures.

By the equivalence of the purified distance and the trace distance (Lemma 34), the error measured in the trace distance is then upper bounded by $(n+1)^{2\left(1-|A|^2\right)}$. This together with (24) implies that there exists a sequence of one-shot channel simulations $\mathcal{F}^n$ for $\mathcal{E}^{\otimes n}$ with error

$$\lim_{n\to\infty}\varepsilon_n = \lim_{n\to\infty}\|\mathcal{E}_{A\to B}^{\otimes n} - \mathcal{F}_{A\to B}^n\|_\diamond \leq \lim_{n\to\infty}(n+1)^{1-|A|^2} = 0 \ , \tag{36}$$

where the entanglement cost of this asymptotic channel simulation is bounded by

$$\inf_{\{M_{A\to B}^k\}}\max_j H(B|R)_{\omega^j} \leq \inf_{\{M_{A\to B}^k\}}\sup_{\psi_A\in\mathcal{S}(\mathcal{H}_A)}\sum_k p_k H(B)_{\psi^k} \ , \tag{37}$$

where the infimum ranges over all Kraus decompositions $\{M_{A\to B}^k\}$ of $\mathcal{E}_{A\to B}$, $\psi_B^k = \frac{1}{p_k}M_{A\to B}^k\psi_A\ {M_{A\to B}^k}^\dagger$ and $p_k = \mathrm{tr}\left[M_{A\to B}^k\psi_A\ {M_{A\to B}^k}^\dagger\right]$.

□

**Lemma 14.** Let $\mathcal{E}_{A\to B}:\mathcal{L}(\mathcal{H}_A)\to\mathcal{L}(\mathcal{H}_B)$ be a CPTP map. Then

$$E_C(\mathcal{E}_{A\to B}) \leq E_C^1(\mathcal{E}_{A\to B}) \equiv \max_{\psi_{AA'}} E_F\left(\left(\mathcal{E}_{A\to B}\otimes\mathcal{I}_{A'}\right)(\psi_{AA'})\right) \ , \tag{38}$$

where $\psi_{AA'}\in\mathcal{V}(\mathcal{H}_A\otimes\mathcal{H}_{A'})$ and $\mathcal{H}_{A'}\cong\mathcal{H}_A$.

*Proof.* The basic idea is to use a minimax theorem (Lemma 38) to interchange the infimum with the supremum in the preceding lemma (Lemma 13). To start with, we want to discretize the set of Kraus decompositions $\{M_k\}$ of $\mathcal{E}$ with at most $\chi$ Kraus operators. For this we note that every such Kraus decomposition $\{M_k\}$ can be seen as a vector $v_\chi\in\mathbb{C}^{\chi\cdot|A||B|}$, by just writing all Kraus operators one after another in a vector.[11] Furthermore, we have $\sum_k M_k^\dagger M_k = \mathbb{1}_B$ and therefore $v_\chi\in\mathcal{N}_\chi = \{w\in\mathbb{C}^{\chi\cdot|A||B|}\mid\|w\|_2=\sqrt{|B|}\}$.[12] We now discretize the set $\mathcal{T}_\chi\subseteq\mathcal{N}_\chi$ of all $v_\chi$ that correspond to a Kraus decomposition $\{M_k\}$ of $\mathcal{E}$ with at most $\chi$ Kraus operators, using a lemma about $\varepsilon$-nets (Lemma 37). The lemma states that there exists a set $\mathcal{T}_{\chi,\varepsilon}\subseteq\mathcal{T}_\chi$ with $|\mathcal{T}_{\chi,\varepsilon}|\leq\left(\frac{2\sqrt{|B|}}{\varepsilon}+1\right)^{2\chi\cdot|A||B|}\equiv M(\chi,\varepsilon)$, such that for every $v_\chi\in\mathcal{T}_\chi$, there exists a $v_{\chi,\varepsilon}\in\mathcal{T}_{\chi,\varepsilon}$ with $\|v_\chi - v_{\chi,\varepsilon}\|_2\leq\varepsilon$.

As the next step we consider the set $\Gamma_{\chi,\varepsilon}$ of probability distributions $\{q_j\}_{j=1}^N$ over $\mathcal{T}_{\chi,\varepsilon}$, and note for every such probability distribution, there exists a corresponding Kraus decomposition $\{\sqrt{q_j}\cdot M_{j,k}\}_{j,k=1}^{N,\chi}$ of $\mathcal{E}$. Restricting the infimum in (23) to $\Gamma_{\chi,\varepsilon}$, we find

$$E_C(\mathcal{E}_{A\to B}) \leq \inf_{\Gamma_{\chi,\varepsilon}}\sup_\psi\sum_j q_j\sum_k p_{j,k}H(B)_{\psi^{j,k}} \ , \tag{39}$$

where $\psi^{j,k} = \frac{1}{p_{j,k}}M_{j,k}\psi M_{j,k}^\dagger$, and $p_{j,k} = \mathrm{tr}\left[M_{j,k}\psi M_{j,k}^\dagger\right]$.

To apply the minimax theorem (Lemma 38) to interchange the infimum and the supremum in (39), we need to check all the conditions of Lemma 38.

$\mathcal{S}(\mathcal{H}_A)$ is compact, convex set. To see that $\sum_j q_j\sum_k p_{j,k}H(B)_{\psi^{j,k}}$ is concave in $\psi_A$, we consider $\psi_A = r^{(1)}\psi_A^1 + r^{(2)}\psi_A^2$ with $\psi_A^1,\psi_A^2\in\mathcal{S}_=(\mathcal{H}_A)$ and $r^{(1)}+r^{(2)}=1$. We define $\tilde{r}_{j,k}^{(1)} = \frac{r^{(1)}\cdot p_{j,k}^{(1)}}{p_{j,k}}$, $\tilde{r}_{j,k}^{(2)} = \frac{r^{(2)}\cdot p_{j,k}^{(2)}}{p_{j,k}}$ with $p_{j,k}^{(1)} = \mathrm{tr}\left[M_{j,k}\psi_A^1 M_{j,k}^\dagger\right]$, $p_{j,k}^{(2)} = \mathrm{tr}\left[M_{j,k}\psi_A^2 M_{j,k}^\dagger\right]$. Since $\tilde{r}_{j,k}^{(1)}+\tilde{r}_{j,k}^{(2)}=1$, we have by the concavity of the von Neumann entropy for $\psi_B^{1,j,k} = \frac{M_{j,k}\psi_A^1 M_{j,k}^\dagger}{p_{j,k}^{(1)}}$, $\psi_B^{2,j,k} = \frac{M_{j,k}\psi_A^2 M_{j,k}^\dagger}{p_{j,k}^{(2)}}$ that

$$H(B)_{\psi^{j,k}} \geq \tilde{r}_{j,k}^{(1)}H(B)_{\psi^{1,j,k}} + \tilde{r}_{j,k}^{(2)}H(B)_{\psi^{2,j,k}} \ . \tag{40}$$

---

[11] Kraus decompositions with less than $\chi$ Kraus operators can just be filled up with zeros.

[12] For this note that $\|v_\chi\|_2 = \|\sum_k M_k^\dagger M_k\|_2$, where the norm on the let hand side denotes the euclidean vector norm and the norm on the right hand side denotes the Hilbert-Schmidt matrix norm.

By multiplying this with $q_j \cdot p_{j,k}$ and taking the sum over all $j,k$ we conclude

$$\sum_j q_j \sum_k p_{j,k} H(B)_{\psi^{j,k}} \geq r^{(1)} \cdot \sum_j q_j \sum_k p^{(1)}_{j,k} H(B)_{\psi^{1,j,k}} + r^{(2)} \cdot \sum_j q_j \sum_k p^{(2)}_{j,k} H(B)_{\psi^{2,j,k}} \ . \tag{41}$$

The function $\sum_j q_j \sum_k p_{j,k} H(B)_{\psi^{j,k}}$ is also continuous in $\psi_A$, since for any $\psi^1_A, \psi^2_A \in \mathcal{S}(\mathcal{H}_A)$ with $\|\psi^1_A - \psi^2_A\|_1 \leq \delta$ for some $\delta > 0$, it follows from the monotonicity of the trace norm under CPTP maps and the continuity of the conditional von Neumann entropy (Lemma 41) that

$$|\sum_j q_j \sum_k p^{(1)}_{j,k} H(B)_{\psi^{1,j,k}} - \sum_j q_j \sum_k p^{(2)}_{j,k} H(B)_{\psi^{2,j,k}}| \leq 4\delta \log |B| + 2h(\delta) \ , \tag{42}$$

where $h(\cdot)$ denotes the binary Shannon entropy.

$\Gamma_{\chi,\varepsilon}$ is a compact, convex set. Moreover $\sum_j q_j \sum_k p_{j,k} H(B)_{\psi^{j,k}}$ is linear in $\{q_j\}$ and therefore in particular convex and continuous. By finally applying the minimax theorem (Lemma 38) in (39), we find

$$E_C(\mathcal{E}_{A\to B}) \leq \sup_{\psi} \inf_{\Gamma_{\chi,\varepsilon}} \sum_j q_j \sum_k p_{j,k} H(B)_{\psi^{j,k}} \ . \tag{43}$$

Since the function is concave, the infimum is taken on an extreme point and hence

$$\inf_{\Gamma_{\chi,\varepsilon}} \sum_j q_j \sum_k p_{j,k} H(B)_{\psi^{j,k}} = \inf_{\{M_k\}} \sum_k p_k H(B)_{\psi^k} \ , \tag{44}$$

where the second infimum ranges over all Kraus decompositions $\{M_k\} \cong v_{\chi,\varepsilon} \in \mathcal{T}_{\chi,\varepsilon}$ of $\mathcal{E}$.

Now let $0 < \varepsilon \leq \frac{1}{2\chi|B|}$. As the next step we show that for every Kraus decomposition $\{M_k\} \cong v_\chi \in \mathcal{T}$ of $\mathcal{E}$, there exists a Kraus decomposition $\{M_{k,\varepsilon}\} \cong v_{\chi,\varepsilon} \in \mathcal{T}_{\chi,\varepsilon}$ of $\mathcal{E}$, such that

$$|\sum_k p_{k,\varepsilon} H(B)_{\psi^{k,\varepsilon}} - \sum_k p_k H(B)_{\psi^k}| \leq 8\varepsilon\chi|B| \log |B| + 2h(2\varepsilon\chi|B|) \ , \tag{45}$$

where $\psi^{k,\varepsilon} = \frac{1}{p_{k,\varepsilon}} M_{k,\varepsilon} \psi M^\dagger_{k,\varepsilon}$, $p_{k,\varepsilon} = \mathrm{tr}\left[M_{k,\varepsilon} \psi M^\dagger_{k,\varepsilon}\right]$, and $h(\cdot)$ denotes the binary Shannon entropy. To see this, we rewrite (45), using Definition 1, to

$$|\sum_k p_{k,\varepsilon} H(B)_{\psi^{k,\varepsilon}} - \sum_k p_k H(B)_{\psi^k}| = |H(B|R)_{\psi^{k,\varepsilon}} - H(B|R)_{\psi^k}| \ , \tag{46}$$

where $\psi^{k,\varepsilon}_{BR} = \sum_k p_{k,\varepsilon} \psi^{k,\varepsilon}_B \otimes |k\rangle\langle k|_R$ and $\psi^k_{BR} = \sum_k p_k \psi^k_B \otimes |k\rangle\langle k|_R$. To estimate (46) we want to use the continuity of the conditional von Neumann entropy (Lemma 41), and for this we analyze

$$\left\| \sum_k p_{k,\varepsilon} \psi^{k,\varepsilon}_B \otimes |k\rangle\langle k|_R - \sum_k p_k \psi^k_B \otimes |k\rangle\langle k|_R \right\|_1 = \sum_k \left\| M_{k,\varepsilon} \psi M^\dagger_{k,\varepsilon} - M_k \psi M^\dagger_k \right\|_1 \ . \tag{47}$$

By the triangle inequality for the trace norm, the equivalence of the trace norm and the Hilbert-Schmidt norm (Lemma 35), and the sub-multiplicativity of the Hilbert-Schmidt norm (Lemma 36), we get

$$\begin{aligned}
\sum_k \left\| M_{k,\varepsilon} \psi M^\dagger_{k,\varepsilon} - M_k \psi M^\dagger_k \right\|_1 &\leq \sum_k \left\| M_{k,\varepsilon} \psi \left( M^\dagger_{k,\varepsilon} - M^\dagger_k \right) \right\|_1 + \left\| \left( M_{k,\varepsilon} - M_k \right) \psi M^\dagger_k \right\|_1 && (48)\\
&\leq \sqrt{|B|} \left( \sum_k \left\| M_{k,\varepsilon} \psi \left( M^\dagger_{k,\varepsilon} - M^\dagger_k \right) \right\|_2 + \left\| \left( M_{k,\varepsilon} - M_k \right) \psi M^\dagger_k \right\|_2 \right) && (49)\\
&\leq \sqrt{|B|} \left( \sum_k \|M_{k,\varepsilon}\|_2 \cdot \|\psi\|_2 \cdot \left\| M^\dagger_{k,\varepsilon} - M^\dagger_{k,\varepsilon} \right\|_2 + \|M_{k,\varepsilon} - M_k\|_2 \cdot \|\psi\|_2 \cdot \|M^\dagger_{k,\varepsilon}\|_2 \right) && (50)\\
&\leq \sqrt{|B|} \left( \varepsilon\chi\sqrt{|B|} + \chi\sqrt{|B|}\varepsilon \right) = 2\varepsilon\chi|B| \ . && (51)
\end{aligned}$$

Finally (45) follows by the continuity of the conditional von Neumann entropy (Lemma 41). Thus we find together with (43) and (44) that

$$E_C(\mathcal{E}_{A\to B}) \leq \sup_{\psi} \inf_{\{M_k\}} \sum_k p_k H(B)_{\psi^k} + 8\varepsilon\chi|B|\log|B| + 2h(2\varepsilon\chi|B|) \ , \tag{52}$$

where the infimum goes over all Kraus decompositions $\{M_k\} \cong v_\chi \in \mathcal{T}$ of $\mathcal{E}$ and $h(\cdot)$ denotes the binary Shannon entropy. Finally note that

$$\inf_{\{M_k\}} \sum_k p_k H(B)_{\psi^k} = E_F\left(\sum_k (M^k_{A\to B})\psi_{AA'}(M^k_{A\to B})^\dagger\right) \ , \tag{53}$$

where the infimum ranges over all Kraus decompositions $\{M_k\}$ of $\mathcal{E}$, $\psi_{AA'} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$, and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. But this infimum is actually taken for a decomposition of size at most $|A|^2|B|^2$ (Lemma 43). Thus, if we set $\chi = |A|^2|B|^2$ and let $\varepsilon \to 0$, we find

$$E_C(\mathcal{E}_{A\to B}) \leq \sup_{\psi_{AA'}} E_F\left(\left(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi_{AA'})\right) \ , \tag{54}$$

where $\psi_{AA'} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. Since the entanglement of formation is continuous (Lemma 42) and $\mathcal{S}(\mathcal{H}_A)$ is compact, the supremum can be turned into a maximum. □

**Proposition 15.** Let $\mathcal{E}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map. Then,

$$E_C(\mathcal{E}_{A\to B}) \leq \lim_{n\to\infty} \frac{1}{n} \max_{\psi^n_{AA'}} E_F\left(\left(\mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi^n_{AA'})\right) \ , \tag{55}$$

where $\psi^n_{AA'} \in \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$.

*Proof.* This follows from standard blocking arguments as in [33]. Namely, by applying the non-regularized achievability (Lemma 14) to the quantum channel $\mathcal{E}^{\otimes n}_{A\to B}$ for some $n > 1$, we get

$$E_C(\mathcal{E}^{\otimes n}_{A\to B}) \leq \frac{1}{n} \max_{\psi^n_{AA'}} E_F\left(\left(\mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi^n_{AA'})\right) \ , \tag{56}$$

where $\psi^n_{AA'} \in \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. Since $n \cdot E_C(\mathcal{E}_{A\to B}) \leq E_C(\mathcal{E}^{\otimes n}_{A\to B})$,[13] we get the claim by letting $n \to \infty$. □

### III.3. Proof: Converse

The idea of the proof of the converse is that any asymptotic channel simulation for $\mathcal{E}_{A\to B}$ must be able to produce any states of the form $\left(\mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}^{\otimes n}_{A'}\right)(\psi^n_{AA'})$ for $n \to \infty$. But by the converse for the one-shot entanglement cost for quantum states (Proposition 8) we have a lower bound on the entanglement that is needed to do this.

**Proposition 16.** Let $\mathcal{E}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map. Then,

$$E_C(\mathcal{E}_{A\to B}) \geq \lim_{n\to\infty} \frac{1}{n} \max_{\psi^n_{AA'}} E_F\left(\left(\mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi^n_{AA'})\right) \ . \tag{57}$$

where $\psi^n_{AA'} \in \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$.

*Proof.* By the definition of an $\varepsilon$-faithful one-shot channel simulation $\mathcal{F}^n$ for $\mathcal{E}^{\otimes n}$ (Definition 10), we have that

$$\left\|\mathcal{F}^n - \mathcal{E}^{\otimes n}\right\|_\diamond \leq \varepsilon \ . \tag{58}$$

---

[13] This is immediate since a channel simulation for $\mathcal{E}^{\otimes n}_{A\to B}$ is a channel simulation for $n$ copies of $\mathcal{E}_{A\to B}$.

This implies in particular that

$$\max_{\psi^n_{AA'}} \left\| \left( \left( \mathcal{F}^n_{A\to B} - \mathcal{E}^{\otimes n}_{A\to B} \right) \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right) \right\|_1 \leq \varepsilon \ , \tag{59}$$

where $\psi^n_{AA'} \in \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. Hence every $\varepsilon$-faithful one-shot channel simulation $\mathcal{F}^n$ for $\mathcal{E}^{\otimes n}$ needs to be able to produce any state of the form $\left( \mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right)$ up to an error $\varepsilon$ (measured in trace distance). But by the definition of the one-shot entanglement cost for quantum states (Definition 7), the entanglement that is needed for this, is given by

$$\max_{\psi^n_{AA'}} E_C^{(1)} \left( \left( \mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right), \varepsilon/2 \right) \ , \tag{60}$$

where $\psi^n_{AA'} \in \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$.[14] Thus we find for the entanglement cost of asymptotic channel simulations for $\mathcal{E}_{A\to B}$ that

$$E_C(\mathcal{E}_{A\to B}) \geq \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \max_{\psi^n_{AA'}} E_C^{(1)} \left( \left( \mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right), \varepsilon/2 \right) \ , \tag{61}$$

where $\psi^n_{AA'} \in \mathcal{V}_{\leq}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. But for $\omega^n_{BA'} = \left( \mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right)$, the converse for the one-shot entanglement cost for quantum states (Proposition 8) says that

$$E_C^{(1)}(\omega^n_{BA'}, \varepsilon/2) \geq \min_{\{p_i, \omega^i\}} H_0^{\sqrt{2\varepsilon}}(B|R)_{\omega^n} \ , \tag{62}$$

where the minimum ranges over all pure states decompositions $\omega^n_{BA'} = \sum_i p_i^n \omega^{n,i}_{BA'}$ and $\omega^n_{BR} = \sum_i p_i^n \omega^{n,i}_B \otimes |i\rangle\langle i|_R$. Now let $\bar{\omega}^n_{BR} \in \mathcal{B}^{\sqrt{2\varepsilon}}_{qc}(\omega^n_{BR})$ such that $H_0^{\sqrt{2\varepsilon}}(B|R)_{\omega^n} = H_0(B|R)_{\bar{\omega}^n}$. Because the alternative conditional max-entropy is lower bounded by the conditional von Neumann entropy (Lemma 25), and since the conditional von Neumann entropy is continuous (Lemma 41), we find

$$H_0^{\sqrt{2\varepsilon}}(B|R)_{\omega^n} = H_0(B|R)_{\bar{\omega}^n} \geq H(B|R)_{\bar{\omega}^n} \geq H(B|R)_{\omega^n} - 4n\sqrt{2\varepsilon} \log |B| - 2h(\sqrt{2\varepsilon}) \ , \tag{63}$$

where $h(\cdot)$ denotes the binary Shannon entropy. Thus, we conclude by the definition of the entanglement of formation (Definition 1) that

$$\begin{aligned} \min_{\{p_i^n, \omega^{n,i}\}} H_0^{\sqrt{2\varepsilon}}(B|R)_{\omega^n} &\geq \min_{\{p_i^n, \omega^{n,i}\}} H(B|R)_{\omega^n} - 4n\sqrt{2\varepsilon} \log |B| - 2h(\sqrt{2\varepsilon}) & (64) \\ &= E_F(\omega^n_{BA'}) - 4n\sqrt{\varepsilon} \log |B| - 2h(\sqrt{2\varepsilon}) & (65) \\ &= E_F \left( \left( \mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right) \right) - 4n\sqrt{2\varepsilon} \log |B| - 2h(\sqrt{2\varepsilon}) \ , & (66) \end{aligned}$$

where the minimum ranges over all pure states decompositions $\omega^n_{BA'} = \sum_i p_i^n \omega^{n,i}_{BA'}$ and $\omega^n_{BR} = \sum_i p_i^n \omega^{n,i}_B \otimes |i\rangle\langle i|_R$, as well as $\psi^n_{AA'} \in \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ with $\mathcal{H}_{A'} \cong \mathcal{H}_A$. Together with (61) and (62) this then implies

$$\begin{aligned} E_C(\mathcal{E}_{A\to B}) &\geq \lim_{\varepsilon\to 0} \lim_{n\to\infty} \left\{ \frac{1}{n} \max_{\psi^n_{AA'}} E_F \left( \left( \mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right) \right) - 4\sqrt{2\varepsilon} \log |B| - \frac{2}{n} h(\sqrt{2\varepsilon}) \right\} & (67) \\ &= \lim_{n\to\infty} \frac{1}{n} \max_{\psi^n_{AA'}} E_F \left( \left( \mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right) \right) \ , & (68) \end{aligned}$$

where $\psi^n_{AA'} \in \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. □

### III.4. Properties

Our main result (Theorem 12) remains true if we restrict the classical communication to be one-way (forward or backward). This follows from the corresponding result about the entanglement cost of quantum states (Remark 9).[15] We also note that the non-regularized achievability (Lemma 14) together with the converse (Proposition 16) imply the following bounds.

---

[14] The factor 1/2 appears because the one shot entanglement cost for quantum states is defined in terms of the purified distance (Definition 7), cf. Lemma 34 about the equivalence of distance measures.

[15] This is also true we think of the problem as simulating a noisy quantum channel from a perfect quantum channel (instead of simulating a noisy quantum channel from perfect entanglement), since in this case a maximally entangled state can always be distributed by the ideal channel.

**Corollary 17.** Let $\mathcal{E}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map. Then, we have that

$$\max_{\psi_{AA'}} E_C\left(\left(\mathcal{E}_{A\to B}\otimes\mathcal{I}_{A'}\right)(\psi_{AA'})\right) \leq E_C(\mathcal{E}_{A\to B}) \leq \max_{\psi_{AA'}} E_F\left(\left(\mathcal{E}_{A\to B}\otimes\mathcal{I}_{A'}\right)(\psi_{AA'})\right) \ , \tag{69}$$

where $\psi_{AA'} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$, and $\mathcal{H}_{A'} \cong \mathcal{H}_A$.

Since the right-hand side of (69) vanishes for every entanglement breaking channel,[16] and since the left-hand side of (69) is greater than zero if the channel is not entanglement breaking [34], this results in the following corollary.

**Corollary 18.** Let $\mathcal{E}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map. Then $E_C(\mathcal{E}_{A\to B}) = 0$ if and only if $\mathcal{E}_{A\to B}$ is entanglement breaking.

## IV. APPLICATIONS AND EXAMPLES

In this section we present two applications of our formula for the entanglement cost of channels and calculate some examples. We start with problem of proving security in the noisy storage model and then turn to the problem of deriving bounds for the strong converse of quantum capacities.

### IV.1. Security in the Noisy Storage Model

We will see below that $E_C$ forms a natural quantity when considering security in the noisy-storage model [12, 13, 35]. It will enable us to extend the parameter regime where security of all existing protocols [12–14, 36–39] can be proven. The appeal of this model is that it allows to solve any cryptographic problem involving two mutually distrustful parties, such as bit commitment, oblivious transfer [13] or secure identification [40, 41]. This is impossible without imposing any assumptions, such as a noisy quantum memory, on the adversary [42–46]. Proposed protocols can thereby be implemented with any hardware suitable for quantum key distribution.

Let us first provide a brief overview of the noisy-storage model as illustrated in Figure 1 - details can be found in e.g. [13]. The central assumption of the noisy-storage model is that the adversary can only store quantum information in a memory described by a particular channel $\mathcal{F} : \mathcal{L}(\mathcal{H}_{in}) \to \mathcal{L}(\mathcal{H}_{out})$. In practice, the use of the memory device is enforced by introducing waiting times $\Delta t$ into the protocol. This is the only restriction imposed on the adversary who is otherwise all-powerful. In particular, he can store an unlimited amount of classical information, and all his actions are instantaneous. This includes any computations, communications, measurements and state preparation that may be necessary to perform an error-correcting encoding and decoding before and after using his noisy memory device.

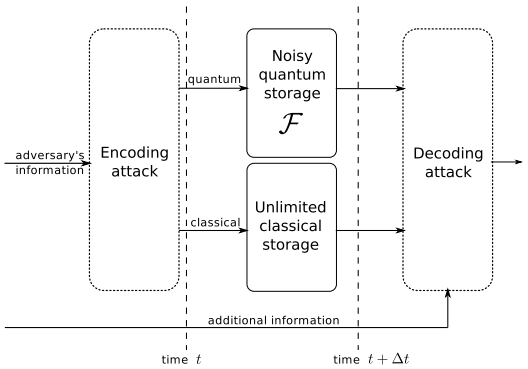


FIG. 1. Noisy-storage assumption: During waiting times $\Delta t$, the adversary can only use his noisy memory device to store quantum information. However, he is otherwise all powerful, and storage of classical information is free.

[16] A quantum channel $\mathcal{E}_{A\to B}$ is called entanglement breaking if $\left(\mathcal{E}_{A\to B}\otimes\mathcal{I}_{A'}\right)(\psi_{AA'})$ is separable for all $\psi_{AA'} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$.

In [13], a natural link was formed between security in the noisy-storage model, and the information carrying capacity of the storage channel $\mathcal{F}$. Of particular interest were thereby memory assumptions that scale with the number $m$ of qubits transmitted during the protocol.[17] That is, the channel is of the form $\mathcal{F} = \mathcal{E}^{\otimes \nu \cdot m}$, where $\nu$ is referred to as the storage rate. It was shown that any two-party cryptographic problem can in principle[18] be implemented securely if [13]

$$C(\mathcal{E}) \cdot \nu < \frac{1}{2} \ , \tag{70}$$

where $C(\mathcal{E})$ denotes the strong converse classical capacity of the channel $\mathcal{E}$ (which is known to equal the classical capacity for certain classes of channels [15]). For the special case of $\mathcal{E} = \mathcal{I}_2$, i.e. the one qubit identity channel, the condition simplifies to

$$\nu < \frac{1}{2} \ . \tag{71}$$

This case is also known as bounded-storage [14, 38, 47]. For protocols involving qubits in a simple BB84 like scheme this is the best bound known today, although using a protocol with very high dimensional encodings can lead to an improvement up to $\nu < 1$ [48].

When considering storing quantum information exchanged during the protocol, it may come as a surprise that the classical capacity should be relevant. Indeed, looking at Figure 1 it becomes clear that a much more natural quantity would be the quantum capacity of $\mathcal{E}$. Whereas we do not accomplish this goal, we make significant progress by linking the security to $E_C(\mathcal{E})$.

**Lemma 19.** Let $m$ be the number of qubits transmitted in the protocol, and let the adversary's storage be of the form $\mathcal{F} = \mathcal{E}^{\otimes \nu \cdot m}$. Then for sufficiently large $m$ any two-party cryptographic primitive can be implemented securely in the noisy-storage model if

$$E_C(\mathcal{E}) \cdot \nu < \frac{1}{2} \ . \tag{72}$$

*Proof.* Consider the case of bounded, noise-free, memory. Note that (71) from [13] tells us that security can be achieved for large enough $m$ if the dimension $d$ of the adversary's storage device is strictly smaller than $d < 2^{m/2}$. Now, suppose by contradiction that security could not be achieved with a storage of the form $\mathcal{F} = \mathcal{E}^{\otimes n}$, where $n = \nu \cdot m$ and $E_C(\mathcal{E}) \cdot n \leq \log d$. However, then there exists a successful cheating strategy also in the case of bounded storage of dimension $d$: the adversary could simply simulate $\mathcal{E}^{\otimes n}$ using an entangled state of dimension $d$ with $\log d = E_C(\mathcal{E}) \cdot n$, possibly using additional classical forward communication provided by his unlimited classical storage device. Hence for large enough $m$, security can be achieved if $E_C(\mathcal{E}) \cdot \nu < \frac{1}{2}$ as claimed. □

Note that for small $m$, a corresponding one-shot quantity $E_C^{(1)}$ is relevant (but is not discussed in this work).[19] It should also be noted that our bound provides a further improvement apart from replacing $C$ by $E_C$, as we no longer explicitly require any strong converse behavior. This is implicitly provided by our simulation argument.

At first glance, our improved bound may appear rather unsatisfying. How could we hope to use this bound to make explicit statements when the formula for $E_C$ involves regularization? First of all, note that for any entanglement breaking channel $\mathcal{E}$, $E_C(\mathcal{E}) = 0$, which leads to immediate security bounds: security can then be attained for any storage rate $\nu$. However, we can show security even for a much larger class of entanglement preserving channels. We now show that even though it is unclear how to calculate $E_C$ explicitly, we can nevertheless obtain improved bounds. The key to such bounds is Lemma 14, which gives us

$$E_C(\mathcal{E}_{A\to B}) \leq E_C^1(\mathcal{E}_{A\to B}) = \max_{\psi_{AA'}} E_F\left(\left(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi_{AA'})\right) \ , \tag{73}$$

where $\psi_{AA'} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. Most channels considered in the noisy-storage model are qubit channels, and for these an exact formula for the entanglement of formation was shown in [49]

$$E_F\left(\left(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi_{AA'})\right) = h\left(\frac{1}{2} + \frac{1}{2} \cdot \sqrt{1 - C^2\left(\left(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi_{AA'})\right)}\right) \ , \tag{74}$$

---

[17] In turn, this tells us how many qubits need to be send in order to achieve security against an attacker with a certain amount of storage.
[18] That is, by transmitting a sufficiently large number $m$ of qubits.
[19] However, statements for any finite $m$ can be made using our results (although the resulting bounds might not be optimal).

with $h(\cdot)$ the binary Shannon entropy, and the concurrence

$$C(\rho) = \max\left\{0, \sqrt{\lambda_1} - \sqrt{\lambda_2} - \sqrt{\lambda_3} - \sqrt{\lambda_4}\right\} \ , \tag{75}$$

with $\lambda_i$'s the eigenvalues of $\rho\tilde{\rho}$ in decreasing order, $\tilde{\rho} = (\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$ with $\rho^*$ the complex conjugate of $\rho$ in the canonical basis, and $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$. Furthermore we know from [50, 51] that for $\psi_{AA'}$ pure

$$C\left(\left(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\psi_{AA'})\right) = C\left(\left(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\phi_{AA'})\right) \cdot C(\psi_{AA'}) \ , \tag{76}$$

where $\phi_{AA'}$ denotes the maximally entangled state. Since $C(\psi_{AA'}) \leq 1$, it follows

$$E_C^1(\mathcal{E}_{A\to B}) = h\left(\frac{1}{2} + \frac{1}{2} \cdot \sqrt{1 - C^2\left(\left(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\phi_{AA'})\right)}\right) \ , \tag{77}$$

that is, it only remains to compute $C(\cdot)$ for the Choi-Jamiolkowski state of the channel. This can be done explicitly using (75) for any qubit channel of interest. To obtain a bound for when security can be achieved we thus can calculate when the condition

$$\nu \cdot h\left(\frac{1}{2} + \frac{1}{2} \cdot \sqrt{1 - C^2\left(\left(\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'}\right)(\phi_{AA'})\right)}\right) < \frac{1}{2} \tag{78}$$

is fulfilled. Figures 2 and 3 illustrate the improvements obtained for depolarizing and dephasing noise respectively. Note that since previous bounds involved the classical capacity, dephasing noise was no better than mere bounded storage. Using our new bound, however, we obtain non-trivial bounds even for this case. Figure 4 provides security bounds for the one qubit amplitude damping channel $\mathcal{E}_{\text{damp}}(\rho) = E_0\rho E_0 + E_1\rho E_1$ where $E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{r} \end{pmatrix}$ and $E_1 = \begin{pmatrix} 0 & \sqrt{1-r} \\ 0 & 0 \end{pmatrix}$. No previous security bound was known for this channel.

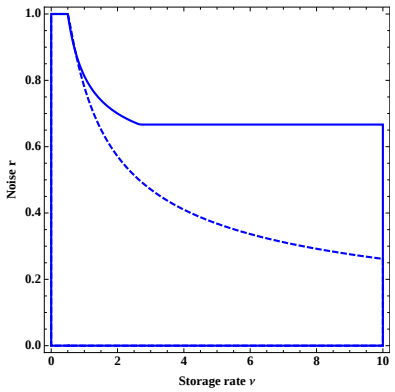


FIG. 2. Depolarizing channel. Security was previously known below the dashed line. Now for $(r, \nu)$ inside the solid line.

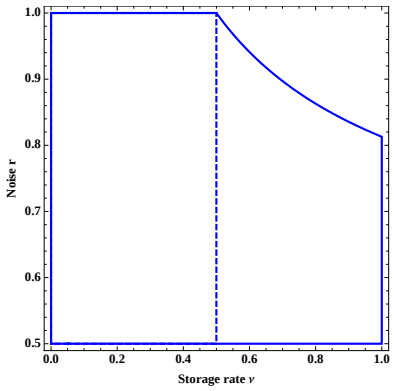


FIG. 3. Dephasing channel. Before security was no better than for bounded storage, left of dashed line. Now for $(r, \nu)$ inside the solid line.

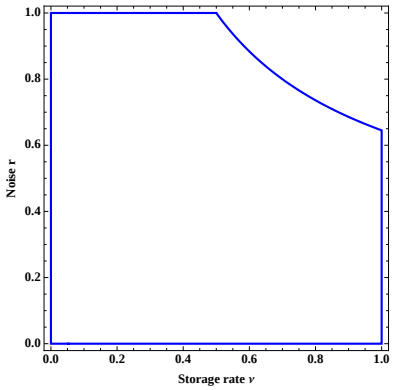


FIG. 4. Amplitude damping channel. No security statement was known previously. Now for $(r, \nu)$ inside the solid line.

### IV.2. An Upper Bound on the Strong Converse Quantum Capacity

To determine a quantum channel's capacity for sending information, two aspects need to be addressed. First of all, one needs to show that the capacity can be achieved. That is, there exists some coding scheme that allows to transmit information reliably at any rate up to the capacity. Second, however, the capacity should really form a threshold for information transmission. That is, if one tries to send information at a rate above the capacity, then there exists no coding scheme that allows to send information without any error. Such a statement is also known as a weak converse.

This however, does not yet exclude the possibility of sending information with a small error at a rate that exceeds the capacity. The minimal rate for which the success in transmitting information drops exponentially with the number of channel uses, is known as the strong converse capacity. The strong converse capacity is appealing since it really gives a sharp threshold for information transmission. But to determine the strong converse capacity forms a challenge even when it comes to sending classical information. Only when restricted to non-entangled input states [16, 17] or certain classes of quantum channels [15], it is known that the strong converse classical capacity is actually the same as the classical capacity. However, various upper bounds on the strong converse classical capacity are known [5, 18, 19]. For example, the quantum reverse Shannon theorem shows that the entanglement assisted classical capacity $C_E$ and its strong converse version are identical [5]. Of course $C_E$ is then also an upper bound on the unassisted strong converse classical capacity. In addition, the result immediately implies that the entanglement assisted quantum capacity $Q_E = C_E/2$ and its strong converse version are identical. Thus, $Q_E$ is an upper bound on the unassisted strong converse quantum capacity.

As the second application of our result, we prove a new upper bound on the strong converse quantum capacity. Similar to the quantum reverse Shannon theorem [5], we employ the idea of a channel simulation to prove that when we send quantum information at a rate exceeding $E_C$, then the fidelity gets exponentially small. Our bound holds for all channels. To start with, let us first define the notion of quantum capacity more formally.

**Definition 20.** Consider a bipartite system with parties Alice and Bob. Let $\varepsilon \geq 0$ and $\mathcal{E} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map, where Alice controls $\mathcal{H}_A$ and Bob $\mathcal{H}_B$. An $\varepsilon$-error code for $\mathcal{E}$ consists of an encoding CPTP map $\Lambda_{\text{enc}} : \left(\mathbb{C}^2\right)^{\otimes R} \to \mathcal{H}_A$ on Alice's side, and a decoding CPTP map $\Lambda_{\text{dec}} : \mathcal{H}_B \to \left(\mathbb{C}^2\right)^{\otimes R}$ on Bob's side such that

$$\|\Lambda_{\text{dec}} \circ \mathcal{E} \circ \Lambda_{\text{enc}} - \mathcal{I}\|_\diamond \leq \varepsilon \ , \tag{79}$$

where $\mathcal{I} : \left(\mathbb{C}^2\right)^{\otimes R} \to \left(\mathbb{C}^2\right)^{\otimes R}$ is the identity channel, and the rate of the code is given by $R$. Furthermore, an asymptotic code for $\mathcal{E}$ is a sequence of $\varepsilon_n$-error codes for $\mathcal{E}^{\otimes n}$ with rate $R_n$ such that $\lim_{n\to\infty} \varepsilon_n = 0$, and the corresponding asymptotic rate is given by $R = \limsup_{n\to\infty} \frac{R_n}{n}$. The quantum capacity $Q(\mathcal{E})$ is then defined as the minimal asymptotic rate of asymptotic codes for $\mathcal{E}$.

Note that there are slightly different ways to define the quantum capacity, and we could use other distance measures (like the entanglement fidelity or the channel fidelity) in (79). Yet, it was as shown that all definitions lead to the same capacity (see Lemma 44, taken from [52]). Similarly, we can define the quantum capacity in the presence of free classical forward communication from the sender to the receiver, denoted by $Q_\rightarrow$, the quantum capacity in the presence of free classical backward communication from the receiver to the sender, denoted by $Q_\leftarrow$, and the two-way classical communication assisted quantum capacity $Q_\leftrightarrow$.

As our argument makes crucial use of the idea of simulating a noisy channel with perfect, noise-free, channels, we now first establish a strong converse for the identity channel. For the unassisted quantum capacity this is straightforward, and can be understood in terms of the impossibility of compressing $n$ qubits into a smaller storage device.

**Lemma 21.** Let $\mathcal{I}_2$ be the qubit identity channel. Then we have for every sequence of $\varepsilon_n$-error codes for $\mathcal{I}_2^{\otimes n}$ with asymptotic rate $R$ that

$$\varepsilon_n \geq 1 - 2^{-n(R-1)} \ . \tag{80}$$

*Proof.* For Kraus decompositions $\{E_j\}$, $\{D_k\}$ of the CPTP maps $\Lambda_{\text{enc}}$, $\Lambda_{\text{dec}}$ respectively, we get for the channel fidelity

$$F_c(\Lambda_{\text{dec}} \circ \mathcal{I} \circ \Lambda_{\text{enc}}) = \sum_{j,k} \left| \text{tr}\left[ D_k E_j \left( \frac{\mathbb{1}}{2^{nR}} \right) \right] \right|^2 \leq \sum_{j,k} \text{tr}\left[ D_k E_j \left( \frac{\mathbb{1}}{2^{nR}} \right) E_j^\dagger D_k^\dagger \right] \text{tr}\left[ \Pi_k \left( \frac{\mathbb{1}}{2^{nR}} \right) \right] \tag{81}$$

$$\leq \frac{1}{2^{nR}} \sum_{j,k} \text{tr}\left[ D_k E_j \left( \frac{\mathbb{1}}{2^{nR}} \right) E_j^\dagger D_k^\dagger \right] \text{tr}\left[\Pi_k\right] \leq 2^{-n(R-1)} \ , \tag{82}$$

where $\Pi_k$ denotes the projector onto the subspace to which $D_k$ maps, and the first inequality follows from the Cauchy-Schwarz inequality. By $F_c(\mathcal{E}) \geq 1 - \|\mathcal{E} - \mathcal{I}\|_\diamond$ (Lemma 44) this implies the claim. □

This can be generalized to the case of free classical communication assistance.

**Corollary 22.** Let $\mathcal{I}_2$ be the qubit identity channel. Then we have for every sequence of classical communication assisted $\varepsilon_n$-error codes for $\mathcal{I}_2^{\otimes n}$ with asymptotic rate $R$ that

$$\varepsilon_n \geq 1 - 2^{-n(R-1)} \ . \tag{83}$$

*Proof.* Since back communication is allowed, the general form of a protocol consists of potentially many rounds of forward quantum and classical communication as well as backward classical communication. We first analyze one such round, which has without lost of generality the following form:

1. CPTP map $\mathcal{D}^1$ at the receiver with Kraus operators $\{D_i^1\}$
2. Classical communication from the receiver to the sender, denoted by the register $B$
3. CPTP map $\mathcal{E}$ at the sender with Kraus operators $\{\hat{E}_{j,b}\} = \{E_{j,b} \otimes |b\rangle\langle b|_B\}$
4. Classical communication from the sender to the receiver, denoted by the register $F$
5. CPTP map $\mathcal{D}^2$ at the receiver with Kraus operators $\{\hat{D}_{k,f}^2\} = \{D_{k,f}^2 \otimes |f\rangle\langle f|_F\}$

The channel fidelity after this round can be estimated as before (Lemma 21)

$$F_c(\mathcal{D}^2 \circ (\mathcal{I}_2^{\otimes n} \otimes \mathcal{I}_F) \circ \mathcal{E} \circ \mathcal{I}_B \circ \mathcal{D}^1) = \sum_{ijkbf} \left| \mathrm{tr}\left[ \hat{D}_{k,f}^2 \hat{E}_{j,b} D_i^1 \left( \frac{\mathbb{1}}{2^{nR}} \right) \right] \right|^2 \tag{84}$$

$$\leq \sum_{ijkbf} \mathrm{tr}\left[ \hat{D}_{k,f}^2 \hat{E}_{j,b} D_i^1 \left( \frac{\mathbb{1}}{2^{nR}} \right) \left(D_i^1\right)^\dagger \hat{E}_{j,b}^\dagger \left( \hat{D}_{k,f}^2 \right)^\dagger \right] \mathrm{tr}\left[ \Pi_{k,f} \left( \frac{\mathbb{1}}{2^{nR}} \right) \right] \tag{85}$$

$$\leq 2^{-n(R-1)} \ , \tag{86}$$

where $\Pi_{k,f}$ denote the projector onto the subspace that $\hat{D}_{k,f}^2$ maps. It is now easily seen that adding more rounds does not affect the argument; the projectors $\Pi$ are just chosen such that they project on the subspaces to which the Kraus operators of the last CPTP map at the receiver map to. □

To generalize this to arbitrary quantum channels we need one more ingredient. We need to show that the asymptotic channel simulation for some quantum channel (as discussed in Theorem 12) can be done for an error rate which is exponentially small in $n$.

**Lemma 23.** Let $\mathcal{E} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map and $\delta_1 > 0$. Then, there exists an asymptotic channel simulation for $\mathcal{E}$ with an entanglement cost of $E_C + \delta_1$ and an error

$$\alpha_n = (n+1)^{|A|^2-1} \cdot 2^{-n \cdot \frac{\delta_1^2}{8(\log(|B|+3))^2}} \ . \tag{87}$$

*Proof.* In the proof of Lemma 13, we can choose the parameter $\delta_n$ as $\delta_n = \frac{1}{2} \cdot 2^{-n \cdot \frac{\delta_1^2}{8(\log(|B|+3))^2}}$. By (24) this leads to a total error rate of

$$\alpha_n = (n+1)^{|A|^2-1} \cdot 2^{-n \cdot \frac{\delta_1^2}{8(\log(|B|+3))^2}} \tag{88}$$

for the asymptotic channel simulation, and by (34) the entanglement cost for this is upper bounded by

$$n \cdot \min_{\{M_{A\to B}^k\}} \max_j H(B|R)_{\omega^j} + \sqrt{n} \cdot \log\left(|B|+3\right) \cdot \sqrt{\log\left(\frac{16}{\delta_n^2}\right)} + 2 \cdot \log(n+1) \cdot \left(|A|^2 - 1\right) \ . \tag{89}$$

Since

$$\lim_{n\to\infty} \frac{1}{n} \cdot \left( \sqrt{n} \cdot \log\left(|B|+3\right) \cdot \sqrt{\log\left(\frac{16}{\delta_n^2}\right)} \right) = \delta_1 \ , \tag{90}$$

we get an entanglement cost of $E_C + \delta_1$ (by considering the rest of the proof of the direct part of Theorem 12, that is, Lemma 14 and Proposition 15). □

Using this lemma, we can now finally prove the following upper bound on the strong converse quantum capacity. The main idea of our proof is argue by contraction: we show that if we were able to send quantum information at a rate exceeding $E_C$, then we could effectively send information through a perfect channel at a higher rate than is allowed by Corollary 22. Since our upper bound holds for any classical communication assistance, we henceforth only talk about $Q_\leftrightarrow$.

**Theorem 24.** Let $\mathcal{E} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a CPTP map and $\delta_2 > \delta_1 > 0$. Then for every sequence of two-way classical communication assisted $\varepsilon_n$-error codes for $\mathcal{E}^{\otimes n}$ with asymptotic rate $R = E_C(\mathcal{E}) + \delta_2$, we have

$$\varepsilon_n \geq 1 - (n+1)^{|A|^2-1} \cdot 2^{-n \cdot \frac{\delta_1^2}{8(\log(|B|+3))^2}} - 2^{-n \cdot \frac{\delta_2 - \delta_1}{E_C(\mathcal{E})+\delta_1} - 1} = 1 - 2^{-O(n)} \ . \tag{91}$$

*Proof.* We start with the perfect qubit identity channel $\mathcal{I}_2$ and do a channel simulation for $\mathcal{E}$ as defined in Definition 11. As we have just seen this can be done for an entanglement cost $E_C(\mathcal{E}) + \delta_1$ and an exponentially small error $\alpha_n = (n+1)^{|A|^2-1} \cdot 2^{-n \cdot \frac{\delta_1^2}{8(\log(|B|+3))^2}}$ (Lemma 23). Now suppose that there existed a hypothetical asymptotic code for $\mathcal{E}$ allowing us to send information at a rate $R = E_C + \delta_2$ for an error rate $\varepsilon_n \geq 0$. Hence, in total, we would have an asymptotic code for $\mathcal{I}_2$ at a rate $\frac{E_C(\mathcal{E})+\delta_2}{E_C(\mathcal{E})+\delta_1} > 1$ for some error rate $\gamma_n > 0$. But by the triangle inequality of the metric induced by the diamond norm and Corollary 22, we know that

$$(n+1)^{|A|^2-1} \cdot 2^{-n \cdot \frac{\delta_1^2}{8(\log(|B|+3))^2}} + \varepsilon_n \geq \gamma_n \geq 1 - \frac{1}{2} \cdot 2^{-n \cdot \left( \frac{E_C(\mathcal{E})+\delta_2}{E_C(\mathcal{E})+\delta_1} - 1 \right)} \ , \tag{92}$$

and thus we are done. □

As an easy example, we consider the qubit erasure channel $\mathcal{E}_{\text{eras}}(\rho) = (1-p)\rho + p \cdot |e\rangle\langle e|$ with $p \in [0,1]$. We immediately have $E_C(\mathcal{E}_{\text{eras}}) \geq 1 - p$, and calculate [53]

$$E_C(\mathcal{E}_{\text{eras}}) \leq \max_{\psi} E_F((\mathcal{E}_{\text{eras}} \otimes \mathcal{I})(\psi)) \leq E_F((\mathcal{E}_{\text{eras}} \otimes \mathcal{I})(\phi)) \leq E_F((1-p)\phi + p \cdot |e\rangle\langle e| \otimes \frac{\mathbb{1}}{2}) \tag{93}$$

$$\leq (1-p) \cdot E_F(\phi) + p \cdot E_F(|e\rangle\langle e| \otimes \frac{\mathbb{1}}{2}) \tag{94}$$

$$= 1 - p \ , \tag{95}$$

where $\Phi$ denotes the maximally entangled state, and we used the non-regularized converse for the entanglement cost (Corollary 17), as well as the convexity of the entanglement of formation [10]. Hence $E_C(\mathcal{E}_{\text{eras}}) = 1 - p$, and since it is also known that $Q_{\leftrightarrow}(\mathcal{E}_{\text{eras}}) = 1 - p$ [54], we get by Theorem 24 that $Q_{\leftrightarrow}(\mathcal{E}_{\text{eras}})$ is a strong converse capacity. Note that, this argument for the qubit erasure channel was basically already present in [54]. For generic quantum channels, we expect that the upper bound given by the entanglement cost is far from being tight. We can compare the quantum capacities of qubit channels with our upper bound from (77)

$$E_C(\mathcal{E}_{A\to B}) \leq E_C^1(\mathcal{E}_{A\to B}) = h\left( \frac{1}{2} + \frac{1}{2} \cdot \sqrt{1 - C^2\left( (\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'})(\phi_{AA'}) \right)} \right) \ , \tag{96}$$

where $h(\cdot)$ denotes the binary Shannon entropy, and $C(\cdot)$ is defined as in (75). For $Q_{\rightarrow}(\mathcal{E})$ this can e.g. be evaluated for all degradable qubit channels [55, 56]. As an example we mention the qubit dephasing channel $\mathcal{E}_{\text{deph}}(\rho) = (1-p)\rho + p \cdot \sigma_z \rho \sigma_z$ with $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, for which we get

$$Q_{\rightarrow}(\mathcal{E}_{\text{deph}}) = 1 - h(p) \leq h\left( \frac{1}{2} + \sqrt{p(1-p)} \right) = E_C^1(\mathcal{E}_{\text{deph}}) \leq 1 - \frac{1}{2} \cdot h(\frac{p}{2}) = Q_E(\mathcal{E}_{\text{deph}}) \ . \tag{97}$$

where $h(\cdot)$ denotes the binary Shannon entropy [57]. As shown in Fig. 5, this is far from being tight. However, since $Q_{\leftrightarrow}$ (and also $Q_{\leftarrow}$) can be much larger than $Q_{\rightarrow}$, and since not too much is known about these capacities, the following upper bound might be useful. We have for every qubit channel $\mathcal{E}_{A\to B}$ that

$$Q_{\leftrightarrow}(\mathcal{E}_{A\to B}) \leq h\left( \frac{1}{2} + \frac{1}{2} \cdot \sqrt{1 - C^2\left( (\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'})(\phi_{AA'}) \right)} \right) \ . \tag{98}$$

## V. DISCUSSION AND OUTLOOK

We calculated the rate of entanglement needed in order to asymptotically simulate a quantum channel when classical communication is for free. Because of the free classical communication, the problem is equivalent to the question about the rate of quantum communication needed in order to simulate a quantum channel. A natural subsequent

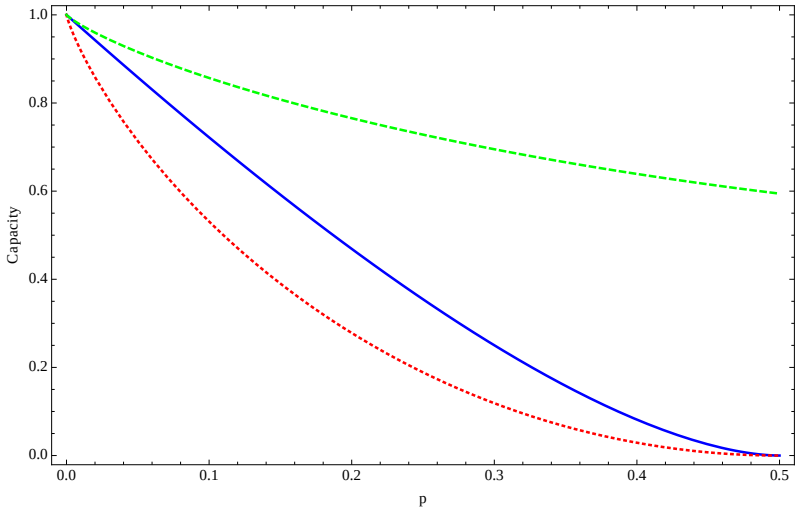


FIG. 5. The qubit dephasing channel with dephasing parameter $p$ - quantum capacity $Q$ (dotted line) vs. upper bound $E_C^1$ on the entanglement cost (solid line) vs. entanglement assisted quantum capacity $Q_E$ (dashed line).

question is to ask what rate of classical communication is actually needed. However, in the spirit of general quantum channel simulations, we might even want to ask more generally about rate triples $(q, e, c)$ needed in order to achieve the channel simulation. Here $q$ denotes quantum communication, $e$ entanglement, and $c$ classical communication. The quantum reverse Shannon theorem can then be understood as e.g. $(Q_E, \infty, 0)$ or $(0, \infty, C_E)$, whereas our entanglement cost corresponds to e.g. $(0, E_C, \infty)$ or $(E_C, 0, \infty)$. Some more examples are discussed in [5, Figure 2] and a particularly interesting case is the following. For $e = 0$, $c = 0$, and product state inputs, the channel simulation can be done for [5, Theorem 3]

$$q = \lim_{n\to\infty} \frac{1}{n} E_P \left( \left( \mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \phi_{AA'} \right)^{\otimes n} \right) \tag{99}$$

with $\phi_{AA'}$ the maximally entangled state, and $E_P$ the entanglement of purification [58]

$$E_P(\rho_{AB}) = \min_{\rho_{AA'BB'}:\mathrm{tr}_{A'B'}[|\rho\rangle\langle\rho|_{AA'BB'}]=\rho_{AB}} E_F(\rho_{AA'BB'}) \ . \tag{100}$$

Now one could hope to generalize this to a channel simulation for general input states using the techniques presented above, leading to

$$q = \lim_{n\to\infty} \frac{1}{n} \max_{\psi^n_{AA'}} E_P \left( \left( \mathcal{E}^{\otimes n}_{A\to B} \otimes \mathcal{I}_{A'} \right) \left( \psi^n_{AA'} \right) \right) \ , \tag{101}$$

where $\psi^n_{AA'} \in \mathcal{V}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{A'}^{\otimes n})$, and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. However, this does not work for same reason as the quantum reverse Shannon theorem can not be proven for general input states using only maximally entangled states; an issue known as entanglement spread [5, 59–61].

Another interesting question concerns the relation of $E_C(\mathcal{E})$ and $Q_\rightarrow(\mathcal{E})$. We know that $E_C(\mathcal{E}) \geq Q_\leftrightarrow(\mathcal{E})$, with the inequality typically being strict. Can we obtain a characterization of channels for which $E_C(\mathcal{E}) = Q_\leftrightarrow(\mathcal{E})$? This is an analog of the problem of characterizing bipartite states for which the distillable entanglement is equal the entanglement cost, which is still wide open.

*Note added.* After completion of this work, security in the noisy storage model was linked to the strong converse quantum capacity of the adversary's storage device [62]. This means that our bound on the strong converse from Section IV.2 can also be applied directly to calculate rates for security. However, our arguments from Section IV.1 apply to virtually any form of the noisy storage model, whereas the results from [62] are only applicable for the so-called six-state encoding.

## ACKNOWLEDGMENTS

We acknowledge discussions with Jonathan Oppenheim, Michael Walter, Reinhard R. Werner, Mark M. Wilde, and Andreas Winter. We thank Cupjin (Jiachen) Huang for pointing out an error in the proof of Lemma 13 in an earlier version of this manuscript and Mark M. Wilde for pointing out an error in Figure 5 in an earlier version of this manuscript. MB and MC are supported by the Swiss National Science Foundation (grant PP00P2-128455), the German Science Foundation (grants CH 843/1-1 and CH 843/2-1), and the National Centre of Competence in Research 'Quantum Science and Technology'. FB and SW are supported by the National Research Foundation and the Ministry of Education, Singapore. FB is supported by a "Conhecimento Novo" fellowship from the Brazilian agency Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG). MB, FB and SW would like to thank the Institute Mittag-Leffler (Djursholm, Sweden), where part of this work was done.


## Appendix A: Properties of Smooth Entropy Measures

**Lemma 25.** [26, Lemma 10] Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$. Then,

$$H_0(A|B)_\rho \geq H(A|B)_\rho \ . \tag{A1}$$

**Lemma 26.** Let $\rho_A = \sum_{j=1}^N p_j \rho_A^j \in \mathcal{P}(\mathcal{H}_A)$ with $\rho_A^j \in \mathcal{P}(\mathcal{H}_A)$ for $j = 1, \ldots, N$. Then,

$$H_0(A)_\rho \leq \max_j H_0(A)_{\rho^j} + \log N \ . \tag{A2}$$

*Proof.* We have $\operatorname{rank}(M+N) \leq \operatorname{rank}(M) + \operatorname{rank}(N)$ for $M, N \in \mathcal{L}(\mathcal{H}_A)$ [63, Proposition 0.4.5] and hence

$$H_0(A)_\rho = \log\left(\operatorname{rank}\left(\sum_{j=1}^N p_j \rho_A^j\right)\right) \leq \log\left(\sum_{j=1}^N \operatorname{rank}\left(p_j \rho_A^j\right)\right) = \log\left(\sum_{j=1}^N \operatorname{rank}\left(\rho_A^j\right)\right) \leq \log\left(N \cdot \max_j \operatorname{rank}\left(\rho_A^j\right)\right) \tag{A3}$$

$$= \max_j H_0(A)_{\rho^j} + \log N \ . \tag{A4}$$

□

**Lemma 27.** Let $\varepsilon \geq 0$, $\rho_{AB} = \sum_{j=1}^N p_j \rho_{AB}^j \in \mathcal{S}(\mathcal{H}_{AB})$, $p_j > 0$ for $j = 1, \ldots, N$, and $\rho_{AB}^j = \sum_{k \in K} \rho_A^{j,k} \otimes |k\rangle\langle k|_B$ with $\rho_A^{j,k} \in \mathcal{P}(\mathcal{H}_{AB})$ and the $|k\rangle_B$ mutually orthogonal. Then,

$$H_0^\varepsilon(A|B)_\rho \leq \max_j H_0^\varepsilon(A|B)_{\rho^j} + \log N \ . \tag{A5}$$

*Proof.* Let $\bar{\rho}_{AB}^j \in \mathcal{B}_{qc}^\varepsilon(\rho_{AB}^j)$ such that $H_0^\varepsilon(A|B)_{\rho^j} = H_0(A|B)_{\bar{\rho}^j}$ for each $j = 1, \ldots, N$. Now define $\bar{\rho}_{AB} = \sum_{j=1}^N p_j \bar{\rho}_{AB}^j$ and note that $\bar{\rho}_{AB} \in \mathcal{B}_{qc}^\varepsilon(\rho_{AB})$. Using the definition of the alternative smooth conditional max-entropy and its form on quantum-classical states (Lemma 3), it follows that

$$H_0^\varepsilon(A|B)_\rho \leq H_0(A|B)_{\bar{\rho}} = \max_{k \in K} H_0(A)_{\sum_j p_j \bar{\rho}^{j,k}} \ . \tag{A6}$$

Using the preceding lemma (Lemma 26) and again the structure of quantum-classical states (Lemma 3), we conclude

$$\max_{k \in K} H_0(A)_{\sum_j p_j \bar{\rho}_A^{j,k}} \leq \max_{k \in K} \max_j H_0(A)_{\bar{\rho}^{j,k}} + \log N = \max_j H_0(A|B)_{\bar{\rho}^j} + \log N = \max_j H_0^\varepsilon(A|B)_{\rho^j} + \log N \ . \tag{A7}$$

□

**Lemma 28.** Let $\varepsilon > 0$, $\rho_{AB} = \sum_k \rho_A^k \otimes |k\rangle\langle k|_B \in \mathcal{S}(\mathcal{H}_{AB})$, and the $|k\rangle_B$ mutually orthogonal. Then, the smoothing in $H_0^\varepsilon(A|B)_\rho$ can without lost of generality be restricted to states that commute with $\rho_{AB}$.

*Proof.* The crucial step is to see that for every $\sigma_{AB} = \sum_k \sigma_A^k \otimes |k\rangle\langle k|_B \in \mathcal{B}^\varepsilon_{qc}(\rho_{AB})$, there exists a unitary $U_{AB} = \sum_k U_A^k \otimes |k\rangle\langle k|_B$ such that $U_{AB}\sigma_{AB}U_{AB}^\dagger \in \mathcal{B}^\varepsilon_{qc}(\rho_{AB})$ and $[U_{AB}\sigma_{AB}U_{AB}^\dagger, \rho_{AB}] = 0$. For this, just choose $U_A^k$ to be the unitary that maps the eigenbasis of $\sigma_A^k$ to the eigenbasis of $\rho_A^k$. Therefore $[U_{AB}\sigma_{AB}U_{AB}^\dagger, \rho_{AB}] = 0$, and furthermore by Lemma 39

$$\varepsilon \geq \|\rho_{AB} - \sigma_{AB}\|_1 = \sum_k \|\rho_A^k - \sigma_A^k\|_1 \geq \sum_k \|P_A^k - Q_A^k\|_1 = \sum_k \|\rho_A^k - U_A^k\sigma_A^k(U_A^k)^\dagger\|_1 = \|\rho_{AB} - U_{AB}\sigma_{AB}U_{AB}^\dagger\|_1 \ , \tag{A8}$$

where $P_A^k, Q_A^k$ denote the eigenvalue distributions of $\rho_A^k, \sigma_A^k$ respectively. □

The definition of the smooth alternative conditional max-entropy can be specialized canonically to classical probability distributions.

**Definition 29.** Let $\varepsilon \geq 0$, $X$ and $Y$ be random variables with range $\mathcal{X}$ and $\mathcal{Y}$ respectively, and joint probability distribution $P_{XY}$. The *max-entropy* of $X$ conditioned on $Y$ is defined as

$$H_0(X|Y)_P = \max_{y\in\mathcal{Y}} \log |\mathrm{supp}\,(P_X^y)\,| \ , \tag{A9}$$

where $P_X^y$ denotes the function $P_X^y : x \mapsto P_{XY}(x, y)$. The *smooth max-entropy* of $X$ conditioned on $Y$ is defined as

$$H_0^\varepsilon(X|Y)_P = \inf_{\bar{P}_{XY}\in\mathcal{B}_c^\varepsilon(P_{XY})} H_0(X|Y)_{\bar{P}} \ , \tag{A10}$$

where $\mathcal{B}_c^\varepsilon(P_{XY})$ denotes the set of non-negative linear functions $\bar{P}_{XY} : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}^+$ such that $\|P_{XY} - \bar{P}_{XY}\|_1 \leq \varepsilon$.

The following is an entropic formulation of the classical asymptotic equipartition property.

**Lemma 30.** [64, Theorem 1] Let $X$ and $Y$ be random variables with range $\mathcal{X}$ and $\mathcal{Y}$ respectively, and joint probability distribution $P_{XY}$. Furthermore let $\varepsilon > 0$, $n \geq 1$, and let $P^n_{X^nY^n} = P_{X_1Y_1}\times\ldots\times P_{X_nY_n}$ be the $n$-fold product probability distribution over $\mathcal{X}^n \times \mathcal{Y}^n$. Then

$$\frac{1}{n}H_0^\varepsilon(X^n|Y^n)_{P^n} \leq H(X|Y)_P + \frac{\log\left(|X|+3\right)\cdot\sqrt{\log\left(\frac{1}{\varepsilon^2}\right)}}{\sqrt{n}} \ . \tag{A11}$$

This can be generalized to the following quantum-classical asymptotic equipartition property.

**Lemma 31.** Let $\varepsilon > 0$, $n \geq 1$, $\rho_{AB} = \sum_k \rho_A^k \otimes |k\rangle\langle k|_B \in \mathcal{S}(\mathcal{H}_{AB})$ and the $|k\rangle_B$ mutually orthogonal. Then,

$$\frac{1}{n}H_0^\varepsilon(A|B)_{\rho^{\otimes n}} \leq H(A|B)_\rho + \frac{\log\left(|A|+3\right)\cdot\sqrt{\log\left(\frac{1}{\varepsilon^2}\right)}}{\sqrt{n}} \ . \tag{A12}$$

*Proof.* The basic idea is that by Lemma 28, the smoothing of the alternative conditional max-entropy can be restricted to states that commute with the initial state, and hence all states that appear are diagonal in the same basis. Working in this basis, this then allows us to use the classical asymptotic equipartition property (Lemma 30). In more detail, we calculate

$$\frac{1}{n}H_0^\varepsilon(A|B)_{\rho^{\otimes n}} = \frac{1}{n}\min_{\bar{\rho}^n_{AB}\in\mathcal{B}^\varepsilon_{qc}(\rho_{AB}^{\otimes n})} H_0(A|B)_{\bar{\rho}^n} = \frac{1}{n}\min_{\bar{P}^n_{AB}\in\mathcal{B}^\varepsilon_c(P^n_{AB})} H_0(A|B)_{\bar{P}^n} \ , \tag{A13}$$

where the second equality is due to Lemma 28, $P^n_{AB}$ is the eigenvalue distribution of $\rho_{AB}^{\otimes n}$, and $\mathcal{B}_c^\varepsilon(\cdot)$ is defined as in Definition 29. Moreover, we conclude by the definition of the classical smooth conditional max-entropy (Definition 29), and the classical asymptotic equipartition property (Lemma 30)

$$\frac{1}{n}\min_{\bar{P}^n_{AB}\in\mathcal{B}^\varepsilon_c(P^n_{AB})} H_0(A|B)_{\bar{P}^n} = \frac{1}{n}H_0^\varepsilon(A|B)_{P^n} \leq H(A|B)_P + \frac{\log\left(|A|+3\right)\cdot\sqrt{\log\left(\frac{1}{\varepsilon^2}\right)}}{\sqrt{n}} \tag{A14}$$

$$= H(A|B)_\rho + \frac{\log\left(|A|+3\right)\cdot\sqrt{\log\left(\frac{1}{\varepsilon^2}\right)}}{\sqrt{n}} \ , \tag{A15}$$

where $P_{AB}$ denotes the eigenvalue distribution of $\rho_{AB}$. □

## Appendix B: The Post-Selection Technique

The following proposition lies at the heart of the *post-selection technique*.

**Proposition 32.** [28] Let $\varepsilon > 0$ and $\mathcal{E}_A^n$ and $\mathcal{F}_A^n$ be CPTP maps from $\mathcal{L}(\mathcal{H}_A^{\otimes n})$ to $\mathcal{L}(\mathcal{H}_B)$. If there exists a CPTP map $K_\pi$ for any permutation $\pi$ such that $(\mathcal{E}_A^n - \mathcal{F}_A^n) \circ \pi = K_\pi \circ (\mathcal{E}_A^n - \mathcal{F}_A^n)$, then $\mathcal{E}_A^n$ and $\mathcal{F}_A^n$ are $\varepsilon$-close whenever

$$\|((\mathcal{E}_A^n - \mathcal{F}_A^n) \otimes \mathcal{I}_{RR'})(\zeta_{ARR'}^n)\|_1 \leq \varepsilon (n+1)^{-(|A|^2-1)} \ , \tag{B1}$$

where $\zeta_{ARR'}^n$ is a purification of the de Finetti state $\zeta_{AR}^n = \int \sigma_{AR}^{\otimes n} d(\sigma_{AR})$ with $\sigma_{AR} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_R)$, $\mathcal{H}_A \cong \mathcal{H}_R$ and $d(\cdot)$ the measure on the normalized pure states on $\mathcal{H}_A \otimes \mathcal{H}_R$ induced by the Haar measure on the unitary group acting on $\mathcal{H}_A \otimes \mathcal{H}_R$, normalized to $\int d(\cdot) = 1$. Furthermore we can assume without loss of generality that $|R'| \leq (n+1)^{|A|^2-1}$.

**Lemma 33.** [6, Corollary D.6] Let $\zeta_{AR}^n = \int \sigma_{AR}^{\otimes n} d(\sigma_{AR})$ as in Proposition 32. Then $\zeta_{AR}^n = \sum_i p_i \left(\omega_{AR}^i\right)^{\otimes n}$ with $\omega_{AR}^i \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_R)$, $i \in \{1, 2, \ldots, (n+1)^{2|A||R|-2}\}$, and $\{p_i\}$ a probability distribution.

## Appendix C: Technical Lemmas

**Lemma 34.** [25, Lemma 6] Let $\rho, \sigma \in \mathcal{S}_{\leq}(\mathcal{H})$. Then

$$\frac{1}{2} \cdot \|\rho - \sigma\|_1 \leq P(\rho, \sigma) \leq \sqrt{\|\rho - \sigma\|_1 + |\mathrm{tr}[\rho] - \mathrm{tr}[\sigma]|} \ . \tag{C1}$$

**Lemma 35.** [63] Let $M \in \mathbb{C}^{a \times b}$ for $a, b \in \mathbb{N}$. Then $\|M\|_2 \leq \|M\|_1 \leq \sqrt{\mathrm{rank}(M)} \cdot \|M\|_2$.

**Lemma 36.** [65, Section 5.2] Let $M \in \mathbb{C}^{a \times b}$ and $N \in \mathbb{C}^{b \times c}$ for $a, b, c \in \mathbb{N}$. Then $\|M \cdot N\|_2 \leq \|M\|_2 \|N\|_2$.

**Lemma 37.** Let $0 < \varepsilon < 1$ and $D, d > 0$. Furthermore let $\mathcal{N}_D^d = \left\{w \in \mathbb{C}^d \mid \|w\|_2 \leq D\right\}$ and let $\mathcal{T}$ be some subset of $\mathcal{N}_D^d$. Then, there exists a subset $\mathcal{T}_\varepsilon \subseteq \mathcal{T}$ with $|\mathcal{T}_\varepsilon| \leq \left(\frac{2D}{\varepsilon} + 1\right)^{2d}$, such that for every vector $v \in \mathcal{T}$, there exists a vector $v_\varepsilon \in \mathcal{T}_\varepsilon$ with $\|v - v_\varepsilon\|_2 \leq \varepsilon$.

*Proof.* The proof is inspired by [66, Lemma II.4]. Let $\mathcal{T}_\varepsilon = \{v_i\}_{i=1,\ldots,m}$ be a maximal subset of $v \in \mathcal{T}$ satisfying $\|v_i - v_j\|_2 \geq \varepsilon$ for all $i, j$.[20] It remains to estimate $m$. As subsets of $\mathbb{R}^{2d}$, the open balls of radius $\varepsilon/2$ about each $v_i \in \mathcal{T}_\varepsilon$ are pairwise disjoint, and all contained in the ball of radius $D + \varepsilon/2$ centered at the origin. Hence

$$m \cdot (\varepsilon/2)^{2d} \leq (D + \varepsilon/2)^{2d} \ . \tag{C2}$$

□

**Lemma 38.** [31, Corollary 3.3] Let $X$ and $Y$ be convex, compact sets and $f$ a real valued function on $X \times Y$, that is convex in the first argument, concave in the second argument and continuous in both. Then,

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) = \sup_{y \in Y} \inf_{x \in X} f(x, y) \ . \tag{C3}$$

**Lemma 39.** [67] Let $\rho, \sigma \in \mathcal{P}(\mathcal{H})$, and denote the corresponding eigenvalue distribution by $P_X, Q_X$ respectively. Then,

$$\|\rho - \sigma\|_1 \geq \|P_X - Q_X\|_1 \ . \tag{C4}$$

**Lemma 40.** [68, Theorem 1] Let $\rho_A, \sigma_A \in \mathcal{S}(\mathcal{H}_A)$ with $\rho_A \approx_\varepsilon \sigma_A$ for some $\varepsilon \geq 0$. Then,

$$|H(A)_\rho - H(A)_\sigma)| \leq \varepsilon \cdot \log(|A| - 1) + h(\varepsilon) \ , \tag{C5}$$

where $h(\cdot)$ denotes the binary Shannon entropy.

---

[20] Such a subset can be constructed by starting with an arbitrary vector $v_1 \in \mathcal{T}$, as a next step taking another vector $v_2 \in \mathcal{T}$ with $\|v_1 - v_2\|_2 \geq \varepsilon$, and then $v_3 \in \mathcal{T}$ with $\|v_1 - v_3\|_2 \geq \varepsilon$, $\|v_2 - v_3\|_2 \geq \varepsilon$ etc. A subset constructed like this becomes maximal as soon as it is not possible to add another vector $v_k \in \mathcal{T}$, such that $\|v_k - v_i\|_2 \geq \varepsilon$ for all vectors $v_i$ that are already in the subset.

**Lemma 41.** [69] Let $\rho_{AB}, \sigma_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$ with $\|\rho_{AB} - \sigma_{AB}\|_1 \leq \varepsilon$ for some $\varepsilon \geq 0$. Then,

$$|H(A|B)_\rho - H(A|B)_\sigma| \leq 4\varepsilon \cdot \log |A| + 2h(\varepsilon) \ , \tag{C6}$$

where $h(\cdot)$ denotes the binary Shannon entropy.

**Lemma 42.** Let $\rho_{AB}, \sigma_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$ with $\rho_{AB} \approx_\varepsilon \sigma_{AB}$ for some $\varepsilon \geq 0$. Then,

$$|E_F(\rho_{AB}) - E_F(\sigma_{AB})| \leq 8\varepsilon \cdot \log |A| + 2h(2\varepsilon) \ , \tag{C7}$$

where $h(\cdot)$ denotes the binary Shannon entropy.

*Proof.* The proof is the same as the original one [70], but uses the (improved) continuity of the conditional von Neumann entropy (Lemma 41) instead of the continuity of the unconditional von Neumann entropy (Lemma 40). □

**Lemma 43.** [71, Lemma 1] Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$. Then the minimization over all pure states decompositions $\rho_{AB} = \sum_i p_i \rho_{AB}^i$ in the entanglement of formation $E_F(\rho_{AB}) = \min_{\{p_i, \rho^i\}} \sum_i p_i H(A)_{\rho^i}$ (Definition 1), is taken for a decomposition with at least $\mathrm{rank}(\rho_{AB})$ and at most $\mathrm{rank}(\rho_{AB})^2$ elements.

**Lemma 44.** [52, Proposition 4.3] Let $\mathcal{E} : \mathcal{L}(\mathcal{H}_A) \mapsto \mathcal{L}(\mathcal{H}_B)$ be a quantum channel. Then,

$$1 - \min_{\rho \in \mathcal{S}(\mathcal{H}_A)} F_e(\rho, \mathcal{E}) \leq 4\sqrt{1 - F_c(\mathcal{E})} \leq 4\sqrt{\|\mathcal{E} - \mathcal{I}\|_\diamond} \leq 8 \left(1 - \min_{\rho \in \mathcal{S}(\mathcal{H}_A)} F_e(\rho, \mathcal{E})\right)^{1/4} \ , \tag{C8}$$

where $F_c(\mathcal{E}) = \langle \phi | (\mathcal{E} \otimes \mathcal{I})(\phi) | \phi \rangle$ with $\phi_{AA'}$ the maximally entangled state on $\mathcal{H}_A \otimes \mathcal{H}_{A'}$, and $F_e(\rho, \mathcal{E}) = \langle \rho | (\mathcal{E} \otimes \mathcal{I})(\rho) | \rho \rangle$ with $\rho_{AA'} \in \mathcal{V}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$ a purification of $\rho_A$.


[1] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, IEEE Transactions on Information Theory **48**, 2637 (2002).
[2] I. Devetak, IEEE Transactions on Information Theory **51**, 44 (2005).
[3] P. W. Shor, Lecture notes, MSRI Workshop on Quantum Computation (2002).
[4] S. Lloyd, Physics Review A **55**, 1613 (1997).
[5] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, IEEE Transactions on Information Theory **60**, 2926 (2014).
[6] M. Berta, M. Christandl, and R. Renner, Communications in Mathematical Physics **306**, 579 (2011).
[7] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Physical Review Letters **70**, 1895 (1993).
[8] M. B. Hastings, Nature Physics **5**, 255 (2009).
[9] P. W. Shor, Communications in Mathematical Physics **3**, 453 (2004).
[10] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Physical Review A **54**, 3824 (1996).
[11] P. M. Hayden, M. Horodecki, and B. T. Terhal, Journal of Physics A **34**, 6891 (2001).
[12] S. Wehner, C. Schaffner, and B. Terhal, Physical Review Letters **100**, 220502 (2008).
[13] R. König, S. Wehner, and J. Wullschleger, IEEE Transactions on Information Theory **58**, 1962 (2012).
[14] I. B. Damgård, S. Fehr, L. Salvail, and C. Schaffner, in *Proceedings of 46th IEEE FOCS* (2005) pp. 449–458.
[15] R. König and S. Wehner, Physical Review Letters **103**, 070504 (2009).
[16] A. Winter, IEEE Transactions on Information Theory **45**, 2481 (1999).
[17] T. Ogawa and H. Nagaoka, IEEE Transactions on Information Theory **45**, 2486 (1999).
[18] N. Datta, M. Hsieh, and F. Brandão, IEEE Transactions on Information Theory **59**, 8014 (2013).
[19] T. Dorlas and C. Morgan, Physical Review A **84** (2011).
[20] R. Renner, International Journal of Quantum Information **6**, 1 (2008).
[21] R. Renner and S. Wolf, in *Proceedings of IEEE International Symposium Information Theory* (2004) p. 233.
[22] R. Renner and R. König, Lecture Notes in Computer Science **3378**, 407 (2005).
[23] R. König, R. Renner, and C. Schaffner, IEEE Transactions on Information Theory **55**, 4674 (2009).
[24] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Transactions on Information Theory **55**, 5840 (2009).
[25] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Transactions on Information Theory **56**, 4674 (2010).
[26] N. Datta, IEEE Transactions on Information Theory **55**, 2816 (2009).
[27] A. Kitaev, Russian Mathematical Surveys **52**, 1191 (1997).
[28] M. Christandl, R. König, and R. Renner, Physics Review Letters **102**, 020504 (2009).
[29] F. Buscemi and N. Datta, Physical Review Letters **106**, 130503 (2011).
[30] M. Hayashi, *Quantum Information: An Introduction* (Springer, 2006).
[31] M. Sion, Pacific Journal of Mathematics **8**, 171 (1958).

[32] V. I. Paulsen, *Completely bounded maps and operator algebras* (Cambridge University Press, 2002).
[33] H. Barnum, M. A. Nielsen, and B. Schumacher, Physics Review A **57**, 4153 (1998).
[34] D. Yang, M. Horodecki, R. Horodecki, and B. Synak-Radtke, Physical Review Letters **95**, 190501 (2005).
[35] S. Wehner, *Cryptography in a Quantum World*, Ph.D. thesis, University of Amsterdam (2008), arXiv:0806.3483v1.
[36] S. Wehner, M. Curty, C. Schaffner, and H.-K. Lo, Physical Review A **81**, 052336 (2010).
[37] C. Schaffner, B. Terhal, and S. Wehner, Quantum Information & Computation **9**, 11 (2008).
[38] I. B. Damgård, S. Fehr, R. Renner, L. Salvail, and C. Schaffner, in *Advances in Cryptology - CRYPTO '07*, Lecture Notes in Computer Science, Vol. 4622 (Springer, 2007) pp. 360–378.
[39] C. Schaffner, Physical Review A **82**, 032308 (2010).
[40] I. B. Damgård, S. Fehr, L. Salvail, and C. Schaffner, in *Advances in Cryptology - CRYPTO '07*, Lecture Notes in Computer Science, Vol. 4622 (Springer, 2007) pp. 342–359.
[41] N. J. Bouman, S. Fehr, C. Gonzales-Guillen, and C. Schaffner, in *Theory of Quantum Computation, Communication, and Cryptography 2012*, Lecture Notes in Computer Science, Vol. 7582 (Springer, 2013) pp. 29–44.
[42] H.-K. Lo, Physical Review A **56**, 1154 (1997).
[43] D. Mayers, arXiv:quant-ph/9603015v3 (1996).
[44] H.-K. Lo and H. Chau, Physica D: Nonlinear Phenomena **120**, 177 (1996).
[45] H.-K. Lo and H. F. Chau, Physical Review Letters **78**, 3410 (1997).
[46] D. Mayers, Physical Review Letters **78**, 3414 (1997).
[47] C. Schaffner, *Cryptography in the Bounded-Quantum-Storage Model*, Ph.D. thesis, University of Aarhus (2007), arXiv:0709.0289v1.
[48] P. Mandayam and S. Wehner, Physical Review A **83**, 022329 (2011).
[49] W. K. Wootters, Physical Review Letters **80**, 2245 (1998).
[50] F. Verstraete, J. Dehaene, and B. DeMoor, Physical Review A **64**, 010101(R) (2001).
[51] T. Konrad, F. de Melo, M. Tiersch, C. Kasztelan, A. Aragao, and A. Buchleitner, Nature Physics **4**, 99 (2008).
[52] D. Kretschmann and R. F. Werner, New Journal of Physics **6**, 26 (2004).
[53] M. Wilde, Private Communication (2012).
[54] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, Physical Review Letters **78**, 3217 (1997).
[55] I. Devetak and P. W. Shor, Communications in Mathematical Physics **256**, 287 (2005).
[56] J. Yard, P. Hayden, and I. Devetak, IEEE Transactions on Information Theory **54**, 3091 (2008).
[57] M. W. Wilde, *From Classical to Quantum Shannon Theory* (2011) arXiv:1106.1445v2.
[58] B. M. Terhal, M. Horodecki, D. W. Leung, and D. P. DiVincenzo, Journal of Mathematical Physics **43**, 4286 (2002).
[59] A. W. Harrow, Proceedings of 16th International Congress on Mathematical Physics (2009).
[60] P. Hayden and A. Winter, Physical Review A **67**, 012326 (2003).
[61] A. W. Harrow and H.-K. Lo, IEEE Transactions on Information Theory **50**, 319 (2004).
[62] M. Berta, O. Fawzi, and S. Wehner, IEEE Transactions on Information Theory **60**, 1168 (2014).
[63] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, 1985).
[64] T. Holenstein and R. Renner, IEEE Transactions on Information Theory **77**, 1865 (2011).
[65] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra* (Cambridge University Press, 2000).
[66] P. Hayden, D. Leung, P. W. Shor, and A. Winter, Communications in Mathematical Physics **250**, 371 (2004).
[67] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, 2000).
[68] K. M. R. Audenaert, Journal of Physics A **40**, 8127 (2007).
[69] R. Alicki and M. Fannes, Journal of Physics A **37**, L55 (2004).
[70] M. A. Nielsen, Physical Review A **61**, 064301 (2000).
[71] A. Uhlmann, Open Systems & Information Dynamics **5**, 209 (1998).