Approximate Degradable Quantum Channels

David Sutter,[1, *] Volkher B. Scholz,[1, *] Andreas Winter,[2, †] and Renato Renner[1, *]

[1]*Institute for Theoretical Physics, ETH Zurich, Switzerland*
[2]*ICREA & Física Teòrica: Informació i Fenòmens Quàntics,*
Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain

Degradable quantum channels are an important class of completely positive trace-preserving maps. Among other properties, they offer a single-letter formula for the quantum and the private classical capacity and are characterized by the fact that a complementary channel can be obtained from the channel by applying a degrading channel. In this work we introduce the concept of approximate degradable channels, which satisfy this condition up to some finite $\varepsilon \geq 0$. That is, there exists a degrading channel which upon composition with the channel is ε-close in the diamond norm to the complementary channel. We show that for any fixed channel the smallest such ε can be efficiently determined via a semidefinite program. Moreover, these approximate degradable channels also approximately inherit all other properties of degradable channels. As an application, we derive improved upper bounds to the quantum and private classical capacity for certain channels of interest in quantum communication.

1. INTRODUCTION

The highest rate at which quantum information can be transmitted asymptotically reliably per channel use is called *quantum capacity*. The *private classical capacity* of a quantum channel characterizes the highest possible rate at which classical information can be transmitted asymptotically reliably per channel use such that no information about the message leaks to the environment. Both of these quantities are mathematically characterized by a multi-letter expression, using regularization, that is complicated to evaluate — as a matter of fact, it is not even known to be computable [11, 17]. In general, it is even difficult to derive good upper and lower bounds that can be evaluated efficiently for the two capacities.

For *degradable channels*, which are characterized by the feature that the complementary channel can be written as a composition of the main channel with a degrading channel, the channel's coherent and private classical information are additive and coincide. As a result, the regularized expressions describing the quantum and private classical capacity reduce to the same single-letter formula for degradable channels [14, 35]. This simplifies the task of computing the capacity enormously and it happens that for some degradable channels the two capacities can be computed analytically.

Degradable channels form an important class of channels for which, thanks to the induced additivity properties, there is a good understanding of their quantum and private classical capacity. At the same time, the notion of a degradable channel seems to be fragile as a tiny perturbation of a degradable channel may not be degradable anymore. Furthermore, it is unknown whether a channel for which the degradability condition is approximately satisfied (up to some $\varepsilon \geq 0$ with respect to the diamond norm) is close to a degradable channel or not. Here, we introduce a robust generalization of the concept of a degradable channel. We call a channel ε*-degradable* if the degradability condition with respect to the diamond norm is satisfied up to some $\varepsilon \geq 0$. (The precise definition is given in Definition 3.1.) We show that these ε-degradable channels approximately inherit all the desirable properties that degradable channels have, such as additivity

* {suttedav, scholz, renner}@phys.ethz.ch
† andreas.winter@uab.cat

of the channel's coherent and channel private information. We further show that for an arbitrary channel, the smallest $\varepsilon \geq 0$ such that the channel is ε-degradable can be efficiently computed via a semidefinite program. This offers a universal method to compute efficiently upper bounds to the quantum and private classical capacity. This will be demonstrated by concrete examples, including the depolarizing channel.

Structure. The remainder of this article is structured as follows. Section 2 introduces a few preliminary results and gives an overview of what is known for degradable channels. Section 3 presents our main contribution which is a definition of approximate degradable channels that approximately inherits all the desirable properties degradable channels have. We show that for an arbitrary channel the smallest possible ε such that the channel is ε-degradable can be computed efficiently via a semidefinite program. Section 4 shows how the concept of approximate degradable channels can be used to derive powerful upper bounds to the quantum and private classical capacity. In Section 5 we discuss some examples and show that the upper bounds based on approximate degradable channels can be very tight. In Appendix A we discuss an alternative definition of approximate degradability, based on closeness to a degradable channel, and compare this to our definition, based on the degradability condition being approximately satisfied. We will argue that our choice leads to a more natural and ultimately more useful notion.

2. PRELIMINARIES

Notation. The logarithm with base 2 is denoted by $\log(\cdot)$ and the natural logarithm by $\ln(\cdot)$. For $k \in \mathbb{N}$, let $[k] := \{1, \ldots, k\}$. The space of Hermitian operators on a finite-dimensional Hilbert space $\mathcal{H}$ is denoted by H^n, where $|\mathcal{H}|$ is the dimension of $\mathcal{H}$. The cone of positive semidefinite Hermitian operators of dimension n is denoted by H^n_+. The space of trace class operators acting on some Hilbert space $\mathcal{H}$ is denoted by $\mathcal{S}(\mathcal{H})$. A quantum channel from a system A to a system B is represented by a completely positive trace-preserving (cptp) linear map $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$. For $\rho \in \mathcal{D}(\mathcal{H})$, where $\mathcal{D}(\mathcal{H}) := \{\rho \in \mathrm{H}^n_+ : \mathrm{tr}[\rho] = 1\}$ denotes the space of density operators on $\mathcal{H}$, the von Neumann entropy is defined as $H(\rho) := -\mathrm{tr}[\rho \log \rho]$. For a bipartite state $\rho_{AB} \in \mathcal{D}(A \otimes B)$ we define the conditional entropy as $H(A|B)_\rho := H(AB)_\rho - H(B)_\rho$ and the mutual information as $I(A : B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho$. For a state $\phi_{AA'} \in \mathcal{D}(A \otimes A')$ and a channel $\Phi : \mathcal{S}(A') \to \mathcal{S}(B)$ let $\rho_{AB} := \Phi(\phi_{AA'})$. Then an equivalent characterization of the coherent information defined by $I_{\mathsf{c}}(\rho, \Phi) := H(\Phi(\rho)) - H(\Phi^{\mathsf{c}}(\rho))$, where Φ^{c} is a complementary channel (that is defined just below) is $I(A\rangle B)_\rho := H(B)_\rho - H(AB)_\rho$. For a matrix $A, B \in \mathbb{C}^{m \times n}$, we denote the Frobenius inner product by $\langle A, B \rangle_F := \mathrm{tr}[A^\dagger B]$ and the induced Frobenius norm by $\|A\|_F := \sqrt{\langle A, A \rangle_F}$. The trace norm is defined as $\|A\|_{\mathrm{tr}} := \mathrm{tr}[\sqrt{A^\dagger A}]$. The operator norm is denoted by $\|A\|_{\mathrm{op}} := \sup_X \{\|AX\|_F : \|X\|_F = 1\}$. For a linear map $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ its diamond norm is defined by $\|\Phi\|_\diamond := \|\Phi \otimes \mathcal{I}_A\|_{\mathrm{tr}}$, where $\|\cdot\|_{\mathrm{tr}}$ denotes the trace norm for resources which is defined as $\|\Phi\|_{\mathrm{tr}} := \max_{\rho \in \mathcal{D}(A)} \|\Phi(\rho)\|_{\mathrm{tr}}$ and $\mathcal{I}_A$ denotes the identity map on A. We denote the standard n-simplex by $\Delta_n := \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i = 1\}$. The binary entropy function is defined as $h(\alpha) := -\alpha \log(\alpha) - (1 - \alpha) \log(1 - \alpha)$, for $\alpha \in [0, 1]$.

Quantum channels. A completely positive trace-preserving map $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ can be represented in different ways. In this article we will use three different representations that are known as *Stinespring*, *Kraus operator*, and *Choi-Jamiołkowski* representation. Stinespring's representation theorem [40] ensures that every quantum channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ can be written in terms of an isometry V from A to the joint system $B \otimes E$ followed by a partial trace such that $\Phi(\rho) = \mathrm{tr}_E(V \rho V^\dagger)$ for all $\rho \in \mathcal{S}(A)$. Tracing out system B instead of E defines a *complementary channel* $\Phi^{\mathsf{c}}(\rho) = \mathrm{tr}_B(V \rho V^\dagger)$ for all $\rho \in \mathcal{S}(\mathcal{H}_A)$. Let $|A| := \dim A$ and $|B| := \dim B$ denote the

input and output dimension of the quantum channel, respectively, and suppose the environment has dimension $|E| := \dim E$. The Kraus representation theorem ensures that for every cptp map $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ there exists a family $\{F_x\}_x$ of operators F_x that map from A to the B system, such that

$$\Phi : \mathcal{S}(\mathcal{H}_A) \ni \rho \mapsto \Phi(\rho) = \sum_x F_x \rho F_x^\dagger \in \mathcal{S}(\mathcal{H}_B)$$

and

$$\sum_x F_x^\dagger F_x = \mathrm{id}_A \ .$$

The Choi-Jamiołkowski representation of the channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is the operator $J(\Phi) \in \mathcal{S}(B \otimes A)$ that is defined as

$$J(\Phi) := |A| \, (\Phi \otimes \mathcal{I}_A) \, (|\Omega\rangle\langle\Omega|) = \sum_{1 \le i,j \le |A|} \Phi(E_{ij}) \otimes E_{ij} \, ,$$

where $|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^d |jj\rangle$ denotes a maximally entangled state and E_{ij} is a $(|A| \times |A|)$ matrix with a one entry at position (i,j) and zeros everywhere else. It is well known that the mapping Φ is completely positive if and only if $J(\Phi) \ge 0$ and that Φ is trace-preserving if and only if $\mathrm{tr}_B(J(\Phi)) = \mathbb{1}_A$. Using the Choi-Jamiołkowski representation, the action of the channel Φ can be written as

$$\Phi(\rho) = \mathrm{tr}_A(J(\Phi) \, (\mathrm{id}_B \otimes \rho^\top)) \, ,$$

for $\rho \in \mathcal{S}(A)$ where the transpose is with respect to the basis chosen for the maximally entangled state $|\Omega\rangle$.

An important class of cptp maps with beneficial properties are the so-called degradable and anti-degradable channels that were introduced in [14].

Definition 2.1 (Degradable and anti-degradable channels). A channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is called *degradable* if there exists a cptp map $\Xi : \mathcal{S}(B) \to \mathcal{S}(E)$ such that $\Phi^c = \Xi \circ \Phi$. The channel Φ is called *anti-degradable* if there exists a cptp map $\Theta : \mathcal{S}(E) \to \mathcal{S}(B)$ such that $\Phi = \Theta \circ \Phi^c$.

Note that Φ is anti-degradable if and only if Φ^c is degradable. Furthermore, the set of anti-degradable channels is convex and contains the set of entanglement-breaking channels (which is a set with positive volume under the set of all channels) [12].

Quantum and private classical capacity. The highest rate at which quantum information can be transmitted asymptotically reliably per channel use is called quantum capacity and is mathematically characterized by the celebrated LSD formula [13, 26, 34] (see also [4, 5, 32, 33])

$$Q(\Phi) = \lim_{k \to \infty} \frac{1}{k} Q^{(1)}(\Phi^{\otimes k}) \,, \tag{1}$$

with

$$Q^{(1)}(\Phi) := \max_{\rho \in \mathcal{D}(A)} I_c(\rho, \Phi) \tag{2}$$

$$:= \max_{\rho \in \mathcal{D}(A)} H\big(\Phi(\rho)\big) - H\big(\Phi^c(\rho)\big) \,, \tag{3}$$

where $I_{\mathsf{c}}(\rho, \Phi)$ denotes the coherent information, $Q^{(1)}(\Phi)$ is called channel coherent information, and Φ^{c} is a complementary channel to Φ.

Due to the regularization in (1), the quantum capacity is difficult to compute. As a consequence, it is of interest to derive good lower and upper bounds to $Q(\Phi)$. It is immediate to verify that $Q^{(1)}(\Phi) \leq Q(\Phi)$ is valid for every channel Φ, i.e., the channel coherent information is always a lower bound for the quantum capacity. However in general, this lower bound is not tight, i.e., there exist channels Φ such that $Q^{(1)}(\Phi) < Q(\Phi)$ [15, 37]. To derive generic upper bounds for the quantum capacity that can be computed efficiently turns out to be difficult. Beside a few channel specific techniques [38, 39, 45] that will be discussed in Section 4, generic upper bounds have been introduced based on a no-cloning argument [6, 8, 10] or semidefinite programming bounds [28, 29]. However, none of these generic upper bounds is expected to be particularly tight as explained in [39]. It is thus fair to say that the quantum capacity is still poorly understood in general — even for very low-dimensional channels. For degradable channels it has been shown that the channel coherent information is additive, i.e., that $Q(\Phi) = Q^{(1)}(\Phi)$ [14]. In general for a given channel Φ, the function $\rho \mapsto I_{\mathsf{c}}(\rho, \Phi)$ is not concave which complicates the task of computing $Q^{(1)}(\Phi)$ defined in (3). However, Φ being degradable implies that $\rho \mapsto I_{\mathsf{c}}(\rho, \Phi)$ is concave [46, Lem. 5] and as such $Q^{(1)}(\Phi)$, and hence $Q(\Phi)$, is characterized via a finite-dimensional convex optimization problem. We also note that due to a no-cloning argument, anti-degradable channels must have a zero quantum capacity, i.e., $Q(\Phi) = 0$ [6, 18].

The private classical capacity of a quantum channel characterizes the highest possible rate at which classical information can be transmitted asymptotically reliably per channel use such that no information about the message leaks to the environment. It is mathematically characterized by the regularized private channel information [9, 13], i.e.,

$$P(\Phi) = \lim_{k \to \infty} \frac{1}{k} P^{(1)}(\Phi^{\otimes k}),\tag{4}$$

with channel private information

$$P^{(1)}(\Phi) := \max_{\{\rho_i, p_i\}} \left\{ H\left(\sum_i p_i \Phi(\rho_i) \right) - \sum_i p_i H\big(\Phi(\rho_i)\big) - H\left(\sum_i p_i \Phi^{\mathsf{c}}(\rho_i) \right) + \sum_i p_i H\big(\Phi^{\mathsf{c}}(\rho_i)\big) \right\}.$$

Similar to the quantum capacity, the regularization arising in (4) complicates the task of evaluating the private classical capacity and the channel private information $P^{(1)}(\Phi)$ is always a lower bound to $P(\Phi)$ which however in general is not tight [36]. Finding generic upper bounds to the private classical capacity again turns out to be difficult. For degradable channels it has been shown that $P^{(1)}(\Phi) = P(\Phi)$, whereas for anti-degradable channels $P(\Phi) = 0$ holds [35].

There is a close connection between the quantum capacity and the private classical capacity of a quantum channel. Since fully quantum communication is necessarily private, $Q(\Phi) \leq P(\Phi)$ for every channel Φ. It can further be shown that $Q^{(1)}(\Phi) \leq P^{(1)}(\Phi)$ for all channels Φ [9, 13] (see also [42, Thm. 12.6.3]). For degradable channels we have $Q(\Phi) = Q^{(1)}(\Phi) = P^{(1)}(\Phi) = P(\Phi)$ [14, 35], while for anti-degradable channels we have seen that $Q(\Phi) = P(\Phi) = 0$ [18, 35].

In the following we will oftentimes use two continuity results for the von Neumann entropy. The first one is known as the *Fannes-Audenaert inequality*.

Lemma 2.2 ([3]). *For any states $\rho \in \mathcal{D}(A)$ and $\sigma \in \mathcal{D}(A)$ such that $\frac{1}{2} \|\rho - \sigma\|_{\mathrm{tr}} \leq \varepsilon \leq 1$,*

$$|H(\rho) - H(\sigma)| \leq \varepsilon \log(|A| - 1) + h(\varepsilon).$$

The second statement we will oftentimes use is a continuity result of the conditional von Neumann entropy which is a recent strengthening of the *Alicki-Fannes inequality* [1] by one of the authors (AW), partly motivated by a first version of this contribution.

Lemma 2.3 ([43, Lemma 2]). *For any states $\rho_{AB} \in \mathcal{D}(A \otimes B)$ and $\sigma_{AB} \in \mathcal{D}(A \otimes B)$ such that* $\frac{1}{2} \|\rho_{AB} - \sigma_{AB}\|_{\mathrm{tr}} \leq \varepsilon \leq 1$,

$$|H(A|B)_\rho - H(A|B)_\sigma| \leq 2\varepsilon \log(|A|) + \left(1 + \varepsilon\right) h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$

3. APPROXIMATE DEGRADABLE CHANNELS

In this section we precisely define the concept of an approximate degradable channel. Theorem 3.4 then shows that the desirable additivity properties of degradable channels are approximately inherited by ε-degradable channels. Proposition 3.6 shows that the smallest possible ε such that a given channel is ε-degradable can be efficiently computed via a semidefinite program. Finally we show that in the same spirit we can also define ε-anti-degradable channels (as done in Definition 3.7) which approximately inherit the properties of anti-degradable channels (cf. Theorem 3.8).

Definition 3.1 (ε-degradable). A channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is said to be ε-*degradable* if there exists another channel $\Xi : \mathcal{S}(B) \to \mathcal{S}(E)$ such that $\|\Phi^{\mathsf{c}} - \Xi \circ \Phi\|_\diamond \leq \varepsilon$.

According to Definition 3.1, we call a channel ε-*degradable* if the degradability condition is approximately satisfied. We note that a possible alternative characterization of approximate degradable channels is to call a channel ε-*close-degradable* if it is close to a degradable channel (cf. Definition A.1). Within this article, we will focus on the first definition of approximate degradable channels. The latter approach is discussed in Appendix A where we also mention major differences between these two definitions of approximate degradable channels which will justify why we favor ε-degradable channels over ε-close-degradable channels.

By Definition 3.1, every channel is ε-degradable for some $\varepsilon \in [0, 2]$ since $\|\Phi^{\mathsf{c}} - \Xi \circ \Phi\|_\diamond \leq \|\Phi^{\mathsf{c}}\|_\diamond + \|\Xi \circ \Phi\|_\diamond \leq 2$. The following theorem (Theorem 3.4) ensures that ε-degradable channels inherit the desirable additivity properties of the channel coherent and the channel private information that degradable channels offer, up to an error term that vanishes in the limit $\varepsilon \to 0$.

For a channel Φ from A to B and a degrading channel Ξ from B to $\widetilde{E} \simeq E$, choose Stinespring isometric dilations $V : A \hookrightarrow B \otimes E$ and $W : B \hookrightarrow \widetilde{E} \otimes F$, respectively. Then, define

$$U_\Xi(\Phi) := \max_{\rho \in \mathcal{D}(A)} \left\{ H(F|\widetilde{E})_\omega \ : \ \omega_{E\widetilde{E}F} = (W \otimes \mathbb{1})V\rho V^\dagger (W \otimes \mathbb{1})^\dagger \right\} . \tag{5}$$

Proposition 3.2. *If* $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ *is an ε-degradable channel with a degrading channel* $\Xi : \mathcal{S}(B) \to \mathcal{S}(E)$, *then*

$$\left| Q^{(1)}(\Phi) - U_\Xi(\Phi) \right| \leq \frac{\varepsilon}{2} \log(|E| - 1) + h\left(\frac{\varepsilon}{2}\right) .$$

Note that for degradable Φ, i.e., $\varepsilon = 0$, this reproduces the main observation of Devetak and Shor [14]. The significance of $U_\Xi(\Phi)$ is hence that it approximates $Q^{(1)}(\Phi)$ — with equality for $\varepsilon = 0$ [14] —, and at the same time it is given by a convex optimization problem, which in general considerably simplifies the task of computing it [7]. Furthermore, $U_\Xi(\Phi)$ itself is additive.

Lemma 3.3. *Let* $\Phi_1, \Phi_2 : \mathcal{S}(A) \to \mathcal{S}(B)$ *be channels and* $\Xi_1, \Xi_2 : \mathcal{S}(A) \to \mathcal{S}(E)$ *be degrading channels. Then*

$$U_{\Xi_1 \otimes \Xi_2}(\Phi_1 \otimes \Phi_2) = U_{\Xi_1}(\Phi_1) + U_{\Xi_2}(\Phi_2) . \tag{6}$$

Proof. This proof follows the original argument of Devetak and Shor for degradable channels [14, App. B]. Indeed, first, for channels Φ_i and degrading channels Ξ_i ($i = 1, 2$), we have

$$U_{\Xi_1 \otimes \Xi_2}(\Phi_1 \otimes \Phi_2) \geq U_{\Xi_1}(\Phi_1) + U_{\Xi_2}(\Phi_2),$$

by choosing a product input state. Second, for a state $\rho_{A_1 A_2}$ and the corresponding $\omega_{E_1 \widetilde{E}_1 F_1 E_2 \widetilde{E}_2 F_2}$,

$$H(F_1 F_2 | \widetilde{E}_1 \widetilde{E}_2) \leq H(F_1 | \widetilde{E}_1 \widetilde{E}_2) + H(F_2 | \widetilde{E}_1 \widetilde{E}_2)$$
$$\leq H(F_1 | \widetilde{E}_1) + H(F_2 | \widetilde{E}_2),$$

by applying strong subadditivity[1] three times. Hence maximizing over input states we get (6). $\square$

Proof of Proposition 3.2. The single-letter coherent information is

$$Q^{(1)}(\Phi) = \max_{\rho \in \mathcal{D}(A)} H\big(\Phi(\rho)\big) - H\big(\Phi^{\mathsf{c}}(\rho)\big)$$
$$= \max_{\rho \in \mathcal{D}(A)} H(\widetilde{E}F)_\omega - H(E)_\omega,$$

with respect to the state ω introduced in (5). Since by assumption Φ is ε-degradable, $\|\omega_E - \omega_{\widetilde{E}}\|_1 \leq \varepsilon$, and invoking the Fannes-Audenaert inequality (see Lemma 2.2) to replace $H(E)$ by $H(\widetilde{E})$, the claim follows. $\square$

Theorem 3.4 (Properties of ε-degradable channels). *If $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is an ε-degradable channel with a degrading channel $\Xi : \mathcal{S}(B) \to \mathcal{S}(E)$, then*

(i) $Q^{(1)}(\Phi) \leq Q(\Phi) \leq Q^{(1)}(\Phi) + \frac{\varepsilon}{2} \log(|E| - 1) + h(\frac{\varepsilon}{2}) + \varepsilon \log |E| + \big(1 + \frac{\varepsilon}{2}\big) h\big(\frac{\varepsilon}{2+\varepsilon}\big)$,

(ii) $Q(\Phi) \leq U_\Xi(\Phi) + \varepsilon \log |E| + \big(1 + \frac{\varepsilon}{2}\big) h\big(\frac{\varepsilon}{2+\varepsilon}\big)$,

(iii) $P^{(1)}(\Phi) \leq P(\Phi) \leq P^{(1)}(\Phi) + \frac{\varepsilon}{2} \log(|E| - 1) + h(\frac{\varepsilon}{2}) + 3\varepsilon \log |E| + 3(1 + \frac{\varepsilon}{2}) h\big(\frac{\varepsilon}{2+\varepsilon}\big)$,

(iv) $Q^{(1)}(\Phi) \leq P^{(1)}(\Phi) \leq Q^{(1)}(\Phi) + \frac{\varepsilon}{2} \log(|E| - 1) + h(\frac{\varepsilon}{2}) + \varepsilon \log |E| + \big(1 + \frac{\varepsilon}{2}\big) h\big(\frac{\varepsilon}{2+\varepsilon}\big)$.

By combining the four statements given in the theorem above we can generate other interesting upper bounds such as

$$P(\Phi) \leq P^{(1)}(\Phi) + \frac{\varepsilon}{2} \log(|E| - 1) + h\left(\frac{\varepsilon}{2}\right) + 3\varepsilon \log |E| + 3\left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right)$$
$$\leq Q^{(1)}(\Phi) + \varepsilon \log(|E| - 1) + 2h\left(\frac{\varepsilon}{2}\right) + 4\varepsilon \log |E| + 4\left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right)$$
$$\leq Q(\Phi) + \frac{3\varepsilon}{2} \log(|E| - 1) + 3h\left(\frac{\varepsilon}{2}\right) + 5\varepsilon \log |E| + 5\left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right)$$
$$\leq U_\Xi(\Phi) + \frac{3\varepsilon}{2} \log(|E| - 1) + 3h\left(\frac{\varepsilon}{2}\right) + 6\varepsilon \log |E| + 6\left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right),$$

where we used statements (iii), (iv), (i), and (ii).

We note that there exist different ways to prove the statements (i)-(iv) of Theorem 3.4. On the one hand, they follow easily by exploiting the properties of the quantity $U_\Xi(\Phi)$. On the other hand, they could also be shown by a telescoping sum technique, which we here use to prove the

---

[1] Recall that the celebrated strong subadditivity of quantum entropy [24, 25] ensures that for any state ρ_{ABC} we have $H(A|B)_\rho \geq H(A|BC)_\rho$.

statement (iii). A similar result would also follow by combining (ii) and (iv), but would give slightly worse bounds. To prove statement (iii), we need two preliminary lemmas. To simplify notation, let us define a multivariate mutual information for n quantum systems $A_1, \ldots, A_n$ and a state $\rho_{A_1 A_2 \ldots A_n} \in \mathcal{D}(\otimes_{i=1}^n A_i)$ as $I(A_1 : A_2 : \ldots : A_n)_\rho := \sum_{i=1}^n H(A_i)_\rho - H(A_1 A_2 \ldots A_n)_\rho$.

Lemma 3.5. *Let $\varepsilon \geq 0$, $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ and $\Xi : \mathcal{S}(B) \to \mathcal{S}(E)$ be two channels such that $\|\Phi^{\mathsf{c}} - \Xi \circ \Phi\|_\diamond \leq \varepsilon$. Let $j \in [n]$ and consider states $\rho_{A_1 \ldots A_n}$, $\sigma_{A_1 \ldots A_{j-1} B_j A_{j+1} \ldots A_n} := \Phi_{A_j \to B_j}(\rho_{A_1 \ldots A_n})$, $\eta_{A_1 \ldots A_{j-1} E_j A_{j+1} \ldots A_n} := \Phi^{\mathsf{c}}_{A_j \to E_j}(\rho_{A_1 \ldots A_n})$ and $\nu_{A_1 \ldots A_{j-1} E_j A_{j+1} \ldots A_n} := \Xi_{B_j \to E_j} \circ \Phi_{A_j \to B_j}(\rho_{A_1 \ldots A_n}) = \Xi_{B_j \to E_j}(\sigma_{A_1 \ldots A_{j-1} B_j A_{j+1} \ldots A_n})$. Then,*

$$I(A_1 : \ldots : A_{j-1} : B_j : A_{j+1} : \ldots : A_n)_\sigma - I(A_1 : \ldots : A_{j-1} : E_j : A_{j+1} : \ldots : A_n)_\eta$$
$$\geq -\frac{\varepsilon}{2} \log(|E| - 1) - h\left(\frac{\varepsilon}{2}\right) - \varepsilon \log(|E|) - \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right).$$

Proof. By definition of the multivariate mutual information and since $H(A_i)_\sigma = H(A_i)_\eta$ for $i \neq j$ we obtain

$$I(A_1 : \ldots : A_{j-1} : B_j : A_{j+1} : \ldots : A_n)_\sigma - I(A_1 : \ldots : A_{j-1} : E_j : A_{j+1} : \ldots : A_n)_\eta$$
$$= H(B_j)_\sigma - H(E_j)_\eta - H(A_1 \ldots A_{j-1} B_j A_{j+1} \ldots A_n)_\sigma + H(A_1 \ldots A_{j-1} E_j A_{j+1} \ldots A_n)_\eta$$
$$= H(B_j)_\sigma - H(E_j)_\eta - H(B_j | A_1 \ldots A_{j-1} A_{j+1} \ldots A_n)_\sigma + H(E_j | A_1 \ldots A_{j-1} A_{j+1} \ldots A_n)_\eta$$
$$\geq H(B_j)_\sigma - H(E_j)_\nu - \frac{\varepsilon}{2} \log(|E| - 1) - H(B_j | A_1 \ldots A_{j-1} A_{j+1} \ldots A_n)_\sigma - h(\varepsilon/2)$$
$$\quad + H(E_j | A_1 \ldots A_{j-1} A_{j+1} \ldots A_n)_\nu - \varepsilon \log |E| - \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right)$$
$$= I(B_j : A_1 \ldots A_{j-1} A_{j+1} \ldots A_n)_\sigma - I(E_j : A_1 \ldots A_{j-1} A_{j+1} \ldots A_n)_\nu - \frac{\varepsilon}{2} \log(|E| - 1)$$
$$\quad - h(\varepsilon/2) - \varepsilon \log |E| - \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right)$$
$$\geq -\frac{\varepsilon}{2} \log(|E| - 1) - h(\varepsilon/2) - \varepsilon \log |E| - \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right),$$

where the second equality uses that

$$H(A_1 \ldots A_{j-1} A_{j+1} \ldots A_n)_\sigma = H(A_1 \ldots A_{j-1} A_{j+1} \ldots A_n)_\eta.$$

The first inequality follows from Fannes-Audenaert inequality (cf. Lemma 2.2) and the Alicki-Fannes inequality (cf. Lemma 2.3). The final inequality follows from monotonicity of mutual information under local channels, which is also known as the data processing inequality for mutual information. □

Proof of Theorem 3.4. We start by proving claim (ii). For n uses of the channel Φ, consider an input state ρ on $A^n = A_1 \ldots A_n$ and the corresponding ω on $E^n \widetilde{E}^n F^n$. Then, the coherent information of the channel equals

$$H(\widetilde{E}^n F^n) - H(E^n) = H(\widetilde{E}^n F^n) - H(\widetilde{E}^n) + H(\widetilde{E}^n) - H(E^n)$$
$$= H(F^n | \widetilde{E}^n) + \sum_{t=1}^n \left[H(\widetilde{E}_t | \widetilde{E}_{<t} E_{>t}) - H(E_t | \widetilde{E}_{<t} E_{>t}) \right],$$

"telescoping" the difference into a sum of "local" differences. Each term in the latter sum has its modulus bounded by $\delta := \varepsilon \log |E| + \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right)$, via Lemma 2.3. Maximizing over input states and using the additivity of $U_\Xi(\Phi)$, Equation (6), yields

$$Q^{(1)}(\Phi^{\otimes n}) \leq n \, U_\Xi(\Phi) + n\delta.$$

Dividing by n and taking the limit $n \to \infty$ gives the statement of claim (ii).

We next prove claim (i). The lower bound $Q^{(1)}(\Phi) \le Q(\Phi)$ is immediate. Combining claim (ii) with Proposition 3.2 immediately proves claim (i). Appendix B presents an alternative proof for claim (i).

We next prove claim (iii). Note that $P^{(1)}(\Phi) \le P(\Phi)$ for an arbitrary channel Φ is proven in [42, p. 323], therefore only the upper bound on $P(\Phi)$ needs to be shown. Let $\xi_1 := 2\varepsilon \log |E| + 2(1 + \frac{\varepsilon}{2})h(\frac{\varepsilon}{2+\varepsilon})$, $\xi_2 := \frac{\varepsilon}{2}\log(|E| - 1) + h(\varepsilon/2) + \varepsilon \log |E| + \left(1 + \frac{\varepsilon}{2}\right)h\left(\frac{\varepsilon}{2+\varepsilon}\right)$ and suppose $\rho_{XA_1'\dots A_n'}$ is the state that maximizes $P^{(1)}(\Phi^{\otimes n})$, where

$$\rho_{XA_1'\dots A_n'} := \sum_{x \in \mathcal{X}} P_X(x) |x\rangle\langle x|_X \otimes \rho_{A_1'\dots A_n'}^x ,$$

and let

$$\rho_{XB_1E_1\dots B_nE_n} := \left(\bigotimes_{i=1}^n V_{A_i \to B_iE_i}^i\right) \rho_{XA_1'\dots A_n'} \left(\bigotimes_{i=1}^n V_{A_i \to B_iE_i}^i\right)^\dagger$$

be the state that arises when sending $\rho_{XA_1'\dots A_n'}$ through $\Phi^{\otimes n}$, with $V_{A_i \to B_iE_i}^i$ denoting the isometric extension of the i-th channel Φ. Consider a spectral decomposition of each state $\rho_{A_1'\dots A_n'}^x$ as $\rho_{A_1'\dots A_n'}^x = \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x)\varphi_{A_1'\dots A_n'}^{x,y}$, where each state $\varphi_{A_1'\dots A_n'}^{x,y}$ is pure. Let $\sigma_{XYA_1'\dots A_n'}$ be an extension of $\rho_{XA_1'\dots A_n'}$ with

$$\sigma_{XYA_1'\dots A_n'} := \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Y|X}(y|x)P_X(x) |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \varphi_{A_1'\dots A_n'}^{x,y}$$

and let

$$\sigma_{XYB_1E_1\dots B_nE_n} := \left(\bigotimes_{i=1}^n V_{A_i \to B_iE_i}^i\right) \sigma_{XYA_1'\dots A_n'} \left(\bigotimes_{i=1}^n V_{A_i \to B_iE_i}^i\right)^\dagger$$

be the state that arises when sending $\sigma_{XYA_1'\dots A_n'}$ through $\Phi^{\otimes n}$. By assumption

$$
\begin{aligned}
P^{(1)}(\Phi^{\otimes n}) &= I(X : B_1 \dots B_n)_\rho - I(X : E_1 \dots E_n)_\rho \\
&= I(X : B_1 \dots B_n)_\sigma - I(X : E_1 \dots E_n)_\sigma \quad (7) \\
&= I(XY : B_1 \dots B_n)_\sigma - I(XY : E_1 \dots E_n)_\sigma \\
&\quad - \big(I(Y : B_1 \dots B_n | X)_\sigma - I(Y : E_1 \dots E_n | X)_\sigma\big) \\
&\le I(XY : B_1 \dots B_n)_\sigma - I(XY : E_1 \dots E_n)_\sigma + n\xi_1 \quad (8) \\
&= H(B_1 \dots B_n)_\sigma - H(B_1 \dots B_n | XY)_\sigma \\
&\quad - H(E_1 \dots E_n)_\sigma + H(E_1 \dots E_n | XY)_\sigma + n\xi_1 \\
&= H(B_1 \dots B_n)_\sigma - H(B_1 \dots B_n | XY)_\sigma \\
&\quad - H(E_1 \dots E_n)_\sigma + H(B_1 \dots B_n | XY)_\sigma + n\xi_1 \quad (9) \\
&= H(B_1 \dots B_n)_\sigma - H(E_1 \dots E_n)_\sigma + n\xi_1 \\
&= \sum_{i=1}^n \big(H(B_i)_\sigma - H(E_i)_\sigma\big) - \big(I(B_1 : \dots : B_n)_\sigma - I(E_1 : \dots : E_n)_\sigma\big) + n\xi_1 \\
&\le \sum_{i=1}^n \big(H(B_i)_\sigma - H(E_i)_\sigma\big) + n(\xi_1 + \xi_2) \quad (10)
\end{aligned}
$$

$$\leq n Q^{(1)}(\Phi) + n(\xi_1 + \xi_2) \tag{11}$$

$$\leq n P^{(1)}(\Phi) + n(\xi_1 + \xi_2), \tag{12}$$

where (7) is valid since $\rho_{XB_1E_1\ldots B_nE_n} = \mathrm{tr}_Y(\sigma_{XYB_1E_1\ldots B_nE_n})$. Inequality (8) follows by applying n times in sequence the strengthened Alicki-Fannes inequality (cf. Lemma 2.3) for the quantum mutual information, followed by the data processing inequality together with the fact that X is classical and the assumption that $\|\Phi^{\mathsf{c}} - \Xi \circ \Phi\|_\diamond \leq \varepsilon$. More precisely, for $i \in [n]$ let

$$\eta^i_{XYE_1\ldots E_n} := \left(\left(\bigotimes_{j=1}^{n-1} \Phi^{\mathsf{c}} \right) \otimes \left(\bigotimes_{k=1}^{i} \xi \circ \Phi \right) \right) (\sigma_{XYA'_1\ldots A'_n}).$$

By assumption, for all $i = 2, \ldots, n$ we have $\|\eta_i - \eta_{i-1}\|_{\mathrm{tr}} \leq \varepsilon$. Thus

$$|I(Y : E_1 \ldots E_n | X)_{\eta_i} - I(Y : E_1 \ldots E_n | X)_{\eta_{i-1}}|$$
$$= |I(Y : E_i | X, E_1 \ldots E_{i-1} E_{i+1} \ldots E_n)_{\eta^i} - I(Y : E_i | X, E_1 \ldots E_{i-1} E_{i+1} \ldots E_n)_{\eta^{i-1}}| \tag{13}$$
$$\leq \xi_1, \tag{14}$$

where (13) follows since by construction the states η_i and η_{i-1} differ only on subsystem E_i and (14) is a consequence of applying the strengthened Alicki-Fannes inequality (cf. Lemma 2.3) twice. Applying the argument described by (13) and (14) n times in sequence shows that

$$-I(Y : B_1 \ldots B_n | X)_\sigma + I(Y : E_1 \ldots E_n | X)_\sigma$$
$$\leq -I(Y : B_1 \ldots B_n | X)_\sigma + I(Y : E_1 \ldots E_n | X)_{\eta^n} + n\xi_1 \tag{15}$$
$$\leq n\xi_1, \tag{16}$$

where the final step uses the data processing inequality. Equation (9) holds as σ is pure on $B_1E_1 \ldots B_nE_n$ when conditioning on XY. The second inequality (10) follows by n times applying Lemma 3.5 and (11) holds as ρ is not necessarily that state maximizing the coherent information. Finally, (12) follows from the fact that $Q^{(1)}(\Phi) \leq P^{(1)}(\Phi)$ is true for all channels Φ [42, Theorem 12.6.3]. This then proves claim (iii).

We finally prove claim (iv) of Theorem 3.4. Note that $Q^{(1)}(\Phi) \leq P^{(1)}(\Phi)$ is true for all channels Φ [42, Theorem 12.6.3], hence only the upper bound for $P^{(1)}(\Phi)$ needs to be shown. Consider a classical-quantum state $\rho_{XA'} = \sum_{x \in \mathcal{X}} P_X(x) |x\rangle\langle x|_X \otimes \rho^x_{A'}$ and let $\sigma_{XBE} = V_{A' \to BE} \rho_{XA'} (V_{A' \to BE})^\dagger$, where $V_{A' \to BE}$ is the isometric extension of Φ. Each state $\rho^x_{A'}$ can be decomposed as $\rho^x_{A'} = \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \varphi^{x,y}_{A'}$, where each state $\varphi^{x,y}_{A'}$ is pure. Consider the following extension of the state σ_{XBE}

$$\eta_{XYBE} := \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Y|X}(y|x) P_X(x) |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes V_{A' \to BE} \varphi^{x,y}_{A'} (V_{A' \to BE})^\dagger.$$

Suppose that σ_{XBE} maximizes the private information and let $\xi := \frac{\varepsilon}{2} \log(|E|-1) + h(\frac{\varepsilon}{2}) + \varepsilon \log |E| + (1 + \frac{\varepsilon}{2}) h(\frac{\varepsilon}{2+\varepsilon})$, then

$$P^{(1)}(\Phi) = I(X : B)_\sigma - I(X : E)_\sigma$$
$$= I(X : B)_\eta - I(X : E)_\eta \tag{17}$$
$$= I(XY : B)_\eta - I(Y : B|X)_\eta - I(XY : E)_\eta + I(Y : E|X)_\eta \tag{18}$$
$$= I(XY : B)_\eta - I(XY : E)_\eta - \left(I(Y : B|X)_\eta - I(Y : E|X)_\eta \right)$$

$$\leq I(XY:B)_\eta - I(XY:E)_\eta + \xi$$
$$= H(B)_\eta - H(B|XY)_\eta - H(E)_\eta + H(E|XY)_\eta + \xi$$
$$= H(B)_\eta - H(B|XY)_\eta - H(E)_\eta + H(B|XY)_\eta + \xi \tag{19}$$
$$= H(B)_\eta - H(E)_\eta + \xi$$
$$\leq Q^{(1)}(\Phi) + \xi\,,$$

where (17) follows since $\sigma_{XBE} = \mathrm{tr}_Y(\eta_{XYBE})$ and (18) is a simple application of the chain rule. The inequality step follows from Lemma 3.5 for $n = 2$ and since X is a classical system. Equation (19) is true since η is pure on EB when conditioning on XY. The final inequality follows since η is not necessarily the state that maximize the coherent information. This proves statement (iv) of Theorem 3.4.

$\square$

By Definition 3.1 it can be verified immediately that if a channel Φ is ε-degradable it is also ε'-degradable for all $\varepsilon' \geq \varepsilon$. The smallest possible parameter ε such that Φ is ε-degradable is given by

$$\varepsilon_\Phi := \inf_\Xi \|\Phi^{\mathsf{c}} - \Xi \circ \Phi\|_\diamond$$
$$\text{s.t. } \Xi : \mathcal{S}(B) \to \mathcal{S}(E) \text{ is cptp}\,. \tag{20}$$

Proposition 3.6. *The optimization problem* (20) *can be expressed as a semidefinite program.*

Proof. Watrous proved [41, Sec. 4] that for two channels $\Theta_1, \Theta_2 : \mathcal{S}(A) \to S(B)$ the diamond norm of their difference, i.e., $\|\Theta_1 - \Theta_2\|_\diamond$ can be expressed as a semidefinite program (SDP) of the form

$$\|\Theta_1 - \Theta_2\|_\diamond = 2 \inf_Z \|\mathrm{tr}_B(Z)\|_\infty$$
$$\text{s.t. } Z \geq J(\Theta_1 - \Theta_2) \tag{21}$$
$$Z \geq 0\,,$$

where $J(\Theta_1 - \Theta_2)$ denotes the Choi-Jamiołkowski representation of $\Theta_1 - \Theta_2$.[2] Since the Choi-Jamiołkowski representation is linear we obtain

$$\varepsilon_\Phi = \inf_\Xi \|\Phi^{\mathsf{c}} - \Xi \circ \Phi\|_\diamond$$
$$\text{s.t. } \Xi : \mathcal{S}(\mathcal{H}_B) \to \mathcal{S}(\mathcal{H}_E) \text{ is cptp} \tag{22}$$

By using (21) this can be rewritten as

$$\varepsilon_\Phi = 2 \inf_\Xi \inf_Z \|\mathrm{tr}_E(Z)\|_\infty$$
$$\text{s.t. } Z \geq J(\Phi^{\mathsf{c}}) - J(\Xi \circ \Phi)$$
$$Z \geq 0 \tag{23}$$
$$J(\Xi) \geq 0$$
$$\mathrm{tr}_E(J(\Xi)) = \mathbb{1}_B$$

---

[2] Note that $\|X\|_\infty$ can be expressed as $\inf\{\mu \in \mathbb{R} : X \leq \mu\mathbb{1}\}$.

where the two final constraints in (23), i.e., $J(\Xi) \geq 0$ and $\mathrm{tr}_E(J(\Xi)) = \mathbb{1}_B$ ensure that Ξ is completely positive and trace-preserving. Since two minimizations can be always interchanged we can reformulate the optimization problem such that we obtain

$$
\begin{aligned}
\varepsilon_\Phi = 2 \inf_{Z, J(\Xi)} &\; \|\mathrm{tr}_E(Z)\|_\infty \\
\mathrm{s.\,t.} &\; Z \geq J(\Phi^{\mathbf{c}}) - J(\Xi \circ \Phi) \\
&\; Z \geq 0 \\
&\; J(\Xi) \geq 0 \\
&\; \mathrm{tr}_E(J(\Xi)) = \mathbb{1}_B \,.
\end{aligned}
\tag{24}
$$

It is now easy to see that (24) is a semidefinite program. Note that for any cptp map $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ we can reshuffle the Choi-Jamiołkowski operator $J(\Phi)$ to a *transfer matrix* $T(\Phi)$ defined by the involution $\langle ij| \, T(\Phi) \, |k\ell\rangle = \langle ik| \, J(\Phi) \, |j\ell\rangle$. Concatenating channels can be reduced to multiplying transfer matrices, i.e., the transfer matrix corresponding to the channel $\Xi \circ \Phi$ can be written as $T(\Xi \circ \Phi) = T(\Xi) \, T(\Phi)$ [44]. Expressing (24) in terms of transfer matrices thus shows that it is a semidefinite program.

□

Semidefinite programs (SDPs) can be solved efficiently, i.e., in time that is polynomial in the program description size [19]. We note that nowadays there exist several different algorithms that in practice solve SDPs very efficiently. A good overview can be found, e.g., in [7, 31]. Therefore, for an arbitrary channel Φ its parameter ε_Φ (given in (20)) that defines how close it is to being degradable can be evaluated efficiently.

The conceptual idea we used above to derive upper bounds on the quantum and the private classical capacity is that a channel that is close to being degradable should have a channel coherent and channel private information that is nearly additive. The same idea can be applied to approximate anti-degradable channels.

Definition 3.7 (ε-anti-degradable). A channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is said to be *ε-anti-degradable* if there exists a channel $\Xi : \mathcal{S}(E) \to \mathcal{S}(B)$ such that $\|\Phi - \Xi \circ \Phi^{\mathbf{c}}\|_\diamond \leq \varepsilon$.

Theorem 3.8 (Properties of ε-anti-degradable channels). *If $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is an ε-anti-degradable channel, then*

$$
\begin{aligned}
Q(\Phi) &\leq P(\Phi) \\
&\leq \frac{\varepsilon}{2} \log(|B| - 1) + \varepsilon \log |B| + h\left(\frac{\varepsilon}{2}\right) + \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right).
\end{aligned}
$$

Proof. The inequality $Q(\Phi) \leq P(\Phi)$ is straightforward since full quantum communication is necessarily private [42, Thm. 12.6.3]. Consider a cq state $\phi_{XA'_1 \ldots A'_n}$ and for $i \in [n]$ let $V^i_{A'_i \to B_i E_i}$ denote the isometric extension of the i-th channel Φ and $U^i_{A'_i \to B_i E_i}$ denote the isometric extension of the i-th channel $\Xi \circ \Phi^{\mathbf{c}}$. Let $\rho_{XB_1 E_1 \ldots B_n E_n} := (\bigotimes_{i=1}^n V^i) \phi (\bigotimes_{i=1}^n V^{i\dagger})$ and $\sigma^{(i)} := (U^1 \otimes \ldots \otimes U^i \otimes V^{i+1} \otimes \ldots \otimes V^n) \phi (U^{1\dagger} \otimes \ldots \otimes U^{i\dagger} \otimes V^{i+1\dagger} \otimes \ldots \otimes V^{n\dagger})$. Suppose $\rho_{XB_1 E_1 \ldots B_n E_n}$ is the state that maximizes $P^{(1)}(\Phi^{\otimes n})$ and let $\xi := \frac{\varepsilon}{2} \log(|B| - 1) + \varepsilon \log(|B|) + h(\varepsilon/2) + (1 + \frac{\varepsilon}{2}) h(\frac{\varepsilon}{2+\varepsilon})$, then

$$
\begin{aligned}
P^{(1)}(\Phi^{\otimes n}) = I(X : B_1 \ldots B_n)_\rho &- I(X : E_1 \ldots E_n)_\rho \\
&\leq I(X : B_1 \ldots B_n)_{\sigma^{(1)}} - I(X : E_1 \ldots E_n)_{\sigma^{(1)}} + \xi \\
&\leq I(X : B_1 \ldots B_n)_{\sigma^{(n)}} - I(X : E_1 \ldots E_n)_{\sigma^{(n)}} + n\xi
\end{aligned}
\tag{25}
$$

$$\leq n\xi\,,$$

where the first inequality follows by the strengthened Alicki-Fannes inequality (see Lemma 2.3) together with the Fannes-Audenaert inequality (see Lemma 2.2) and the fact that $\left\|\rho - \sigma^{(1)}\right\|_{\mathrm{tr}} \leq \varepsilon$ which follows by the assumption $\|\Phi - \Xi \circ \Phi^{\mathsf{c}}\|_{\diamond}$. The inequality (25) follows by applying the same argument in sequence for all $1 \leq i \leq n$. Note that by assumption $\left\|\sigma^{(i)} - \sigma^{(i+1)}\right\|_{\mathrm{tr}} \leq \varepsilon$ for all $i \in [n-1]$. The final inequality uses that by construction the state $\sigma^{(n)}$ is generated by sending ϕ through n copies of an anti-degradable channel. Anti-degradable channels are known to have a private capacity that is zero [35]. $\qquad\square$

Similar as for ε-degradable channels, given a channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ we can consider

$$\bar{\varepsilon}_{\Phi} := \left\{ \begin{array}{ll} \inf_{\Xi} & \|\Phi - \Xi \circ \Phi^{\mathsf{c}}\|_{\diamond} \\ \mathrm{s.t.} & \Xi : \mathcal{S}(B) \to \mathcal{S}(E) \text{ is cptp} \,, \end{array} \right. \tag{26}$$

which defines the smallest possible parameter ε such that the channel Φ is ε-anti-degradable. Since $(\Phi^{\mathsf{c}})^{\mathsf{c}} = \Phi$ we have $\bar{\varepsilon}_{\Phi} = \varepsilon_{\Phi^{\mathsf{c}}}$. Proposition 3.6 thus implies that (26) can also be phrased as an SDP. Theorem 3.8 of course is valid for $\bar{\varepsilon}_{\Phi}$.

4. UPPER BOUNDS VIA CONVEX DECOMPOSITIONS OF CHANNELS

In this section we show how to combine the concept of ε-degradable channels with a standard technique to derive upper bounds to the quantum capacity that is based on the idea of decomposing an arbitrary channel Φ into a convex sum of approximate degradable channels.

It has been shown that for an arbitrary quantum channel Φ the mapping $\Phi \mapsto Q(\Phi)$ is convex if Φ is (anti-)degradable [45]. Therefore, if a channel Φ can be written as a convex combination of (anti-)degradable channels, i.e., $\Phi = \sum_{i=1}^{n} p_i \Xi_i$, where $p \in \Delta_n$ and $\{\Xi_i\}_{i=1}^{n}$ are (anti-)degradable, $Q(\Phi) \leq \sum_{i=1}^{n} p_i Q(\Xi_i) = \sum_{i=1}^{n} p_i Q^{(1)}(\Xi_i)$ which describes a single-letter upper bound to the quantum capacity of Φ that can be powerful as demonstrated in [38, 39]. A drawback of this technique is that it is channel specific, i.e., the convex decomposition into degradable channels has to be reconstructed from scratch for every different channel. In addition, for an arbitrary channel, it is unclear how to efficiently find a convex decomposition of degradable channels — even worse it is highly questionable if this is even possible in general. The extreme points of the set of all qubit channels have been shown to be degradable or anti-degradable channels [12, 45] and therefore for qubit channels a convex decomposition into (anti-)degradable channels does exist, even if it might be difficult to find. However, a characterization of the extreme points of quantum channels with an input dimension larger than two is unknown [30] and as such there is no reason to believe that an arbitrary quantum channel can be written as a convex combination of (anti-)degradable channels.

Recall the definitions of the symmetric side-channel assisted quantum and private classical capacities [35, 39]:

$$Q_{ss}(\Phi) := \sup_{\Theta} Q(\Phi \otimes \Theta) = \sup_{\Theta} Q^{(1)}(\Phi \otimes \Theta) \tag{27}$$

$$P_{ss}(\Phi) := \sup_{\Theta} P(\Phi \otimes \Theta) = \sup_{\Theta} P^{(1)}(\Phi \otimes \Theta) \,, \tag{28}$$

where Θ ranges over so-called symmetric channels, i.e. those with $\Theta = \Theta^{\mathsf{c}}$. By definition, $Q(\Phi) \leq Q_{ss}(\Phi)$ and $P(\Phi) \leq P_{ss}(\Phi)$. The important insights of [35, 39] were that both Q_{ss} and P_{ss} have single-letter formulas, respectively; both are convex in the channel; and both coincide with $Q^{(1)}$

for degradable channels. The following theorem extends this insight to approximate degradable channels.

Theorem 4.1. *If Φ is an ε-degradable channel, with a degrading channel Ξ, then*

(i) $Q_{ss}(\Phi) \leq U_\Xi(\Phi) + \varepsilon \log |E| + (1 + \frac{\varepsilon}{2})h(\frac{\varepsilon}{2+\varepsilon})$,

(ii) $P_{ss}(\Phi) \leq U_\Xi(\Phi) + \varepsilon(2\log |E| + \frac{1}{2}\log |F|) + \frac{5}{2}(1 + \frac{\varepsilon}{2})h(\frac{\varepsilon}{2+\varepsilon})$.

We note that the statement (i) of Theorem 4.1 implies the statement (ii) of Theorem 3.4.

Proof. Optimizing over symmetric side channels with generic isometry $U : A' \hookrightarrow B' \otimes E'$, and input states $\rho_{AA'}$ (or rather its purification), we have (using the same notation as in the proof of Theorem 3.4 otherwise for the approximate degrading channel):

$$Q_{ss}(\Phi) = \sup I(AA'\rangle BB') , \tag{29}$$

where the supremum is over all input states $\rho_{AA'}$ and over all symmetric side channels (see (27)). With the same notation we then find

$$Q_{ss}(\Phi) = \sup H(BB') - H(EE') \tag{30}$$

$$= \sup H(B|E') - H(E|E') \tag{31}$$

$$\leq \sup H\big(F\widetilde{E}|E'\big) - H\big(\widetilde{E}|E'\big) + \varepsilon \log |E| + \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2+\varepsilon}\right) \tag{32}$$

$$= \sup H\big(F|\widetilde{E}E'\big) + \varepsilon \log |E| + \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2+\varepsilon}\right) \tag{33}$$

$$= \sup H\big(F|\widetilde{E}\big) + \varepsilon \log |E| + \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2+\varepsilon}\right) . \tag{34}$$

Here, the first two lines are by definition, the third by the symmetry between B' and E'; in the fourth we use the Stinespring isometry W of the approximate degrading channel and Lemma 2.3; in the fifth line we rewrite the difference of conditional entropies using the chain rule, and in the last line we have "$\leq$" by strong subadditivity, but equality is achieved with a trivial symmetric side channel. But now, $\sup H\big(F|\widetilde{E}\big)$ contains only the maximization over input states ρ_A, giving $U_\Xi(\Phi)$.

The proof for P_{ss} is similar, cf. Theorem 3.4. $\qquad\square$

The significance of Theorem 4.1 is that Q_{ss} is convex, unlike Q (and likewise P_{ss}, in contrast to P). As a result, we even get strengthened upper bounds by taking the convex hull of the bound in Theorem 3.4 and other upper bounds on Q_{ss} [27, 39]. This is done in Section 5 to derive upper bounds for the capacity of a depolarizing and a BB84 channel.

5. APPLICATIONS

We now illustrate the power of the bounds derived in the previous sections on three examples. Recall that the upper bounds for the quantum and the private classical capacity derived in Sections 3 and 4 are valid for arbitrary channels Φ. As the parameter ε_Φ given in (20) is described via an SDP, it can be evaluated efficiently for every possible channel. Thus our upper bounds can be immediately applied and efficiently evaluated for arbitrary channels, whereas most previous upper bounds rely on channel specific constructions which can be different for each channel and are usually difficult to find [38, 39]. As we will see in this section, we can also combine different upper bounds, i.e., taking the best known upper bounds for every scenario.

A. Depolarizing channel

Consider a depolarizing channel $\mathcal{D}_p : \mathcal{S}(A) \ni \rho \mapsto (1-p)\rho + \frac{p}{3}(X\rho X + Z\rho Z + Y\rho Y) \in \mathcal{S}(B)$ with $\dim A = \dim B = 2$ and $p \in [0,1]$. Its channel coherent information is maximized on a Bell state as input, and hence evaluates to [6, 15] (see also [42, p. 575])

$$Q^{(1)}(\mathcal{D}_p) = 1 + (1-p)\log(1-p) + p\log\left(\frac{p}{3}\right). \tag{35}$$

A well-known upper bound to $Q(\Phi)$ has been derived in [38, Cor. 7] (see also [27, Thm. 5.5]) and is given for $0 \le p \le \frac{1}{4}$ and $\gamma(p) := 4(\sqrt{1-p} - 1 + p)$ by

$$Q(\mathcal{D}_p) \le \text{conv}\left\{1 - h(p), h\left(\frac{1+\gamma(p)}{2}\right) - h\left(\frac{\gamma(p)}{2}\right), 1 - 4p\right\}. \tag{36}$$

Combining Theorem 3.4 and Theorem 4.1 we obtain the upper bound

$$Q(\mathcal{D}_p) \le \text{conv}\left\{U_\Xi(\mathcal{D}_p), 1 - h(p), h\left(\frac{1+\gamma(p)}{2}\right) - h\left(\frac{\gamma(p)}{2}\right), 1 - 4p\right\}. \tag{37}$$

We recall that $U_\Xi(\mathcal{D}_p)$ is given via a convex optimization problem. In order to further simplify our upper bound, using the structure of the depolarizing channel we further bound $U_\Xi(\mathcal{D}_p)$ with the help of Proposition 3.2.

$$Q(\mathcal{D}_p) \le \text{conv}\left\{Q^{(1)}(\Phi) + \frac{\varepsilon_p}{2}\log(|E| - 1) + h\left(\frac{\varepsilon_p}{2}\right) + \varepsilon_p \log|E| + \left(1 + \frac{\varepsilon_p}{2}\right) h\left(\frac{\varepsilon_p}{2 + \varepsilon_p}\right),\right.$$
$$\left. 1 - h(p), h\left(\frac{1+\gamma(p)}{2}\right) - h\left(\frac{\gamma(p)}{2}\right), 1 - 4p\right\}. \tag{38}$$

Figure 1 compares the new upper bound given by (38), for ε_p as in (20), with the upper bound given by (36). We note that the upper bound (37) can be potentially considerably better than (38), however one has to solve a convex optimization problem that defines $U_\Xi(\mathcal{D}_p)$.[3]

It can be shown that for two copies of an arbitrary quantum channel Φ we have $\varepsilon_{\Phi\otimes\Phi} \le 2\varepsilon_\Phi$. To see this, we assume without loss of generality that Ξ denotes the optimizer in (20). Since $\Xi \otimes \Xi$ is a feasible degrading channel, we obtain

$$\begin{aligned}
\varepsilon_{\Phi\otimes\Phi} &\le \|\Phi^{\mathbf{c}} \otimes \Phi^{\mathbf{c}} - (\Xi \circ \Phi) \otimes (\Xi \circ \Phi)\|_\diamond \\
&= \|(\Phi^{\mathbf{c}} - \Xi \circ \Phi) \otimes \Phi^{\mathbf{c}} - (\Xi \circ \Phi) \otimes (\Xi \circ \Phi - \Phi^{\mathbf{c}})\|_\diamond \\
&\le \|\Phi^{\mathbf{c}} - \Xi \circ \Phi\|_\diamond \|\Phi^{\mathbf{c}}\|_\diamond + \|\Xi \circ \Phi\|_\diamond \|\Phi^{\mathbf{c}} - \Xi \circ \Phi\|_\diamond \\
&\le 2\varepsilon_\Phi,
\end{aligned}$$

where we used that the diamond norm is multiplicative under the tensor product. It could happen that $\varepsilon_{\Phi\otimes\Phi}$ is considerably smaller than $2\varepsilon_\Phi$, i.e., the tensor channel $\Phi \otimes \Phi$ would be noticeably closer to being degradable than Φ. Numerics for the depolarizing channel however show that this is not the case (we observe that $\varepsilon_{\Phi\otimes\Phi}$ is always close to $2\varepsilon_\Phi$).

---

[3] Very recently after this paper, Leditzky *et al.* presented a new upper bound for the depolarizing channel [21] that outperforms (38) in the high noise regime. See [21] for a comparison of the new bound with (38).
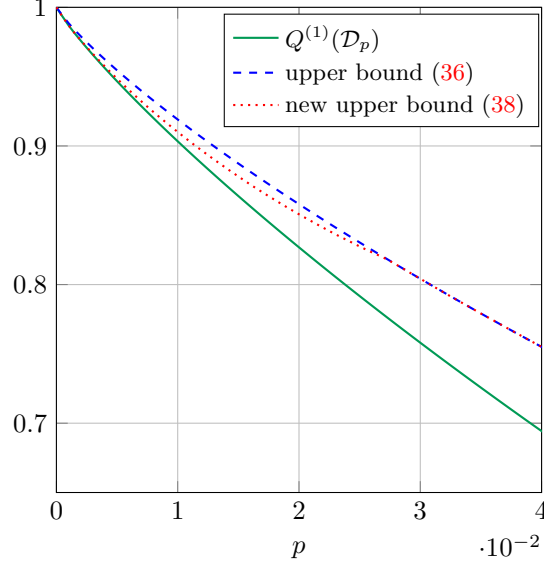
FIG. 1. This plot depicts upper and lower bounds for the quantum capacity $Q(\mathcal{D}_p)$ of a qubit depolarizing channel $\mathcal{D}_p$. The channel coherent information given in (35) denotes a lower bound on $Q(\mathcal{D}_p)$ (cf. solid green curve). The dashed blue curve denotes an upper bound on $Q(\mathcal{D}_p)$ given in (36). The dotted red line depicts the upper bound given by (38) with ε_p as given in (20).

B. BB84 channel

Consider a qubit Pauli channel with independent bit flip and phase flip error probability where $p_X \in [0, \frac{1}{2}]$ denotes the bit flip and $p_Z \in [0, \frac{1}{2}]$ the phase flip probability. Due to its relevance for the BB84 protocol this channel is often called BB84 channel. More formally this is a channel $\mathcal{B}_{p_X, p_Z} : \mathcal{S}(A) \ni \rho \mapsto (1 - p_X - p_Z + p_X p_Z)\rho + (p_X - p_X p_Z)X\rho X + (p_Z - p_Z p_X)Z\rho Z + p_X p_Z Y\rho Y \in \mathcal{S}(B)$ with $\dim A = \dim B = 2$. It is immediate to verify that a Bell state maximizes the coherent information and therefore the channel coherent information of the BB84 channel is given by $Q^{(1)}(\mathcal{B}_{p_X, p_Z}) = 1 - h(p_X) - h(p_Z)$. For the case where $p_X = p_Z =: p$ it has been shown that [38]

$$Q(\mathcal{B}_{p,p}) \leq h\left(\frac{1}{2} - 2p(1-p)\right) - h\left(2p(1-p)\right). \tag{39}$$

Combining Theorem 3.4 and Theorem 4.1 we obtain the upper bound

$$Q(\mathcal{B}_{p,p}) \leq \operatorname{conv}\left\{Q^{(1)}(\mathcal{B}_{p,p}) + \frac{\varepsilon_p}{2}\log(|E| - 1) + h\left(\frac{\varepsilon_p}{2}\right) + \varepsilon_p \log|E| + \left(1 + \frac{\varepsilon_p}{2}\right)h\left(\frac{\varepsilon_p}{2 + \varepsilon_p}\right),\right.$$
$$\left. h\left(\frac{1}{2} - 2p(1-p)\right) - h\left(2p(1-p)\right)\right\}, \tag{40}$$

which is strictly better than (39). Figure 2 compares the upper bound of the quantum capacity derived in Theorem 3.4, for ε_p as in (20), with previously known upper bounds. In the high-noise regime (39) outperforms (40). In most quantum key distribution protocols the secret-key rate is given by the difference of a min-entropy term and a term that comes from the error correction step (which corresponds to the quantum capacity). Oftentimes these two terms are of the same magnitude and therefore the improvement of (40) compared to (39) which looks small on Figure 2 can be very relevant for bounding the secret-key rate. We note that the upper bound (40) can be slightly improved by using the $U_\equiv(\mathcal{B}_{p,p})$ quantity.
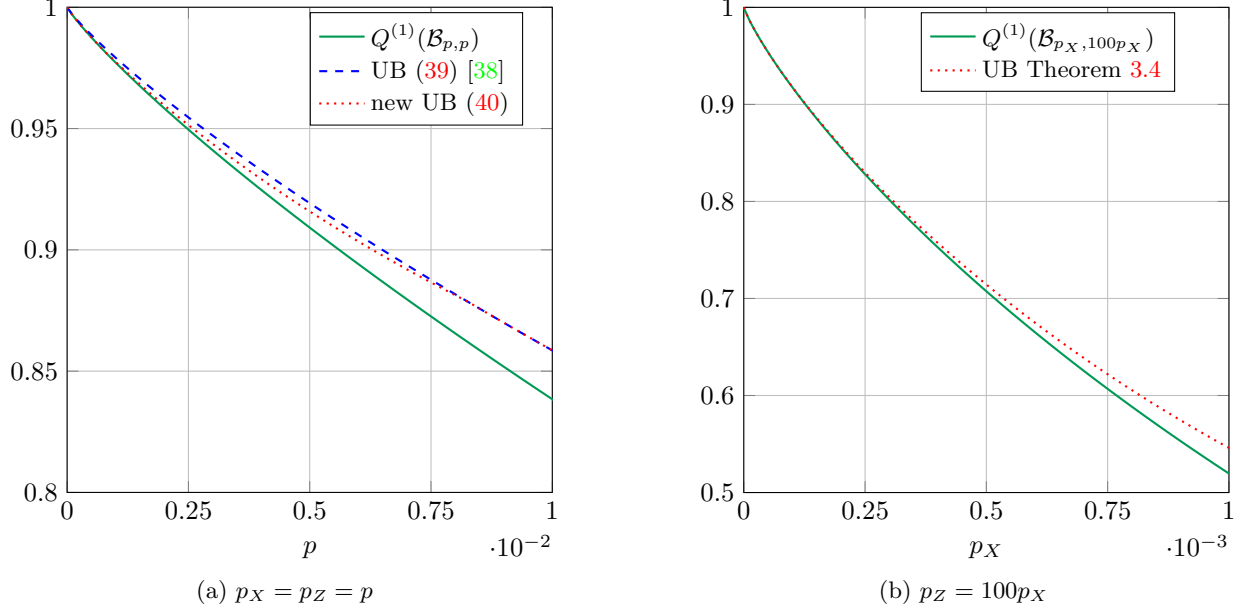
(a) $p_X = p_Z = p$ (b) $p_Z = 100p_X$

FIG. 2. This plot compares upper and lower bounds on $Q(\mathcal{B}_{p_X,p_Z})$ derived in (40) and in Theorem 3.4, for ε_{p_X,p_Z} as in (20), with the best previously known upper bounds. We consider two different setups of p_X/p_Z.

C. Randomizing channels

A quantum channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is called ε-*randomizing* if for any state $\rho \in \mathcal{D}(A)$, $\|\Phi(\rho) - \frac{1}{|B|}\mathbb{1}_B\|_{\mathrm{op}} \leq \frac{\varepsilon}{|B|}$, where $\frac{1}{|B|}\mathbb{1}_B$ is the maximally mixed state on the system B. Consider a channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ whose complementary channel is $\Phi^{\mathsf{c}} : \mathcal{S}(A) \ni \rho \mapsto \frac{1}{|B|}\sum_{i=1}^{|B|} U_i\,\rho\,U_i^{\dagger} \in \mathcal{S}(E)$, where $\{U_i\}_{i=1}^{|B|}$ are independent random matrices Haar-distributed on the unitary group $\mathcal{U}(|A|)$ with $|A| = |E|$. Consider the fully mixing channel $\Xi : \mathcal{S}(B) \ni \rho \mapsto \mathrm{tr}[\rho]\frac{1}{|E|}\mathbb{1}_E \in \mathcal{S}(E)$.

Proposition 5.1 ([2]). *If $|B| \geq C\frac{|E|^3}{\varepsilon^2}$ for some constant $C > 0$ and $0 < \varepsilon < 1$, then with high probability the channel Φ is ε-degradable.*

Proof. Consider the two channels Φ and Ξ as defined above. We show that if $|B| \geq C\frac{|E|^3}{\varepsilon^2}$, with high probability $\|\Phi^{\mathsf{c}} - \Xi \circ \Phi\|_{\diamond} \leq \varepsilon$. This follows directly from [2, Thm. 1] together with the fact that for two arbitrary cptp maps $\Theta_1, \Theta_2 : \mathcal{S}(A) \to \mathcal{S}(B)$,

$$\|\Theta_1 - \Theta_2\|_{\diamond} \leq |B| \max_{\rho \in \mathcal{D}(A)} \|\Theta_1(\rho) - \Theta_2(\rho)\|_{\mathrm{tr}}$$

$$\leq |B|^2 \max_{\rho \in \mathcal{D}(A)} \|\Theta_1(\rho) - \Theta_2(\rho)\|_{\mathrm{op}} \ .$$

$\square$

We thus can use Theorem 3.4 to estimate $Q(\Phi)$ and $P(\Phi)$ from above for most of the random unitary channels Φ as defined above that have an environment that is considerably smaller than the output system. As shown in [2], it can be verified that the constant C in Proposition 5.1 can be chosen, e.g., as $C = 150$.

6. DISCUSSION

We have seen that the concept of degradable channels can be generalized to the more robust notion of approximate degradable channels such that the beneficial additivity properties degradable channels offer are approximately preserved. As it can be efficiently determined how close in the diamond norm an arbitrary channel satisfies the degradability condition (by solving an SDP), the framework of approximate degradable channels can be used to derive upper bounds to the quantum and private classical capacity that can be evaluated efficiently. Unlike previous attempts to derive upper bounds, our method does not rely on channel specific arguments and therefore can be applied to all channels.

For future work it would be of interest to better understand the differences between ε-degradable channels and ε-close-degradable channels (as introduced in Appendix A). A problem that is left open is the question if a converse statement to Proposition A.5 is possible, i.e., if an ε-degradable channel must be also $\theta(\varepsilon)$-close-degradable for some function $\theta : \mathbb{R}_{\geq 0} \ni \varepsilon \mapsto \theta(\varepsilon) \in \mathbb{R}_{\geq 0}$. Another question that deserves further investigation is if the optimization problem (A1) can be solved (or at least approximated) efficiently.

The concept of approximate degradable channels could also be useful in classical information theory. We note that to some extent, the current understanding about the capacity region of a classical broadcast channel is comparable to the knowledge about the quantum and private classical capacity — no single-letter formula is known, except in the case of a degradable broadcast channel [16]. As such it could be promising to apply the framework of approximate degradable quantum channels, introduced in this article, to classical broadcast channels.

Very recently, the bounds for approximate degradable channels derived in this article have been applied to determine the quantum and the private capacity of low-noise quantum channels to leading orders in the channels distance to the perfect channel [22].

ACKNOWLEDGMENTS

Appendix A: Approximate degradabiliy versus closeness to degradable channels

The operational meaning behind Definition 3.1 is that a channel is called ε-degradable if the degradability condition is approximately (up to an ε) satisfied. An alternative approach is to consider the distance to degradable channels.

Definition A.1 (ε-close-degradable). A channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is said to be *ε-close-degradable* if there exists a degradable channel $\Psi : \mathcal{S}(A) \to \mathcal{S}(B)$ such that $\|\Phi - \Psi\|_{\diamond} \leq \varepsilon$.

By Definition A.1 and since the identity channel is degradable, it follows that every channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is ε-close-degradable with respect to some $\varepsilon \in [0, 2]$. The following proposition proves that ε-close-degradable channels, similar as ε-degradable channels, inherit the additivity properties of degradable channels with an error term that vanishes in the limit $\varepsilon \to 0$.

Proposition A.2 (Properties of ε-close-degradable channels). *Let $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ be a quantum channel that is ε-close-degradable (with respect to a degradable channel $\Psi : \mathcal{S}(A) \to \mathcal{S}(B)$), then*

(i) $|Q(\Phi) - Q^{(1)}(\Psi)| \leq \varepsilon \log |B| + (2 + \varepsilon) h\left(\frac{\varepsilon}{2 + \varepsilon}\right)$,

(ii) $|P(\Phi) - Q^{(1)}(\Psi)| \leq 2\varepsilon \log |B| + 2(2 + \varepsilon) h\left(\frac{\varepsilon}{2 + \varepsilon}\right)$.

Proof. We first prove statement (i) of the proposition. Since the coherent information is additive for degradable channels,

$$
|Q(\Phi) - Q^{(1)}(\Psi)| = |Q(\Phi) - Q(\Psi)|
$$
$$
\leq \varepsilon \log |B| + (2 + \varepsilon) h\left(\frac{\varepsilon}{2 + \varepsilon}\right),
$$

where the inequality is due to [23, Cor. 2] together with Lemma 2.3.

Statement (ii) of the proposition can be proven as follows. As Ψ is degradable we have $P(\Psi) = Q^{(1)}(\Psi)$ which gives

$$
|P(\Phi) - Q^{(1)}(\Psi)| = |P(\Phi) - P(\Psi)|
$$
$$
\leq 2\varepsilon \log |B| + 2(2 + \varepsilon) h\left(\frac{\varepsilon}{2 + \varepsilon}\right),
$$

where the inequality follows from the proof of [23, Cor. 3] with Lemma 2.3. $\qquad\square$

As mentioned in Section 2, the function $\rho \mapsto I_{\mathsf{c}}(\rho, \Psi)$ is concave if Ψ is degradable which can be helpful when computing the channel coherent information given in (3) (see Proposition A.2). We emphasize that the proofs of Theorem 3.4 and Proposition A.2 are different although they both prove a similar statement, however under different assumptions. The proof of Theorem 3.4 generalizes Devetak and Shor's proof for additivity of degradable channels [14], whereas the proof of Proposition A.2 is based on continuity properties of channel capacities, following [23] and using the improved Alicki-Fannes inequality (Lemma 2.3).

Definition A.1 directly implies that if a channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is ε-close-degradable it is also ε'-close-degradable for all $\varepsilon' \geq \varepsilon$. The smallest possible value ε such that Φ is ε-close-degradable is given by

$$
\hat{\varepsilon}_\Phi := \begin{cases} \inf_{\Psi, \Theta} & \|\Phi - \Psi\|_\diamond \\ \text{s.t.} & \Psi^{\mathsf{c}} = \Theta \circ \Psi \\ & \Psi : \mathcal{S}(A) \to \mathcal{S}(B) \text{ is cptp} \\ & \Theta : \mathcal{S}(B) \to \mathcal{S}(E) \text{ is cptp} . \end{cases} \tag{A1}
$$

Note that unlike the optimization problem (20) which can be phrased as an SDP and as a consequence can be solved efficiently, it is unclear if $\hat{\varepsilon}_\Phi$ can be computed efficiently. The optimization problem (A1) is clearly not an SDP as the constraint $\Psi^{\mathsf{c}} = \Theta \circ \Psi$ is not linear in (Ψ, Θ).

Similarly to an ε-close-degradable channel being defined via being close to a degradable channel we can define an ε-close-anti degradable channel as being close to an anti-degradable channel.

Definition A.3 (ε-close-anti-degradable). *A channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ is said to be ε-close-anti-degradable if there exists an anti-degradable channel $\Xi : \mathcal{S}(A) \to \mathcal{S}(B)$ such that $\|\Phi - \Xi\|_\diamond \leq \varepsilon$.*

Corollary A.4 (Properties of ε-close-anti-degradable channels)**.** *Let* $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ *be a quantum channel that is* ε*-close-anti-degradable, then* $Q(\Phi) \leq P(\Phi) \leq 2\varepsilon \log |B| + 2(2 + \varepsilon)h\left(\frac{\varepsilon}{2+\varepsilon}\right).$

Proof. This corollary follows immediately from [23, Cor. 3] together with Lemma 2.3 and the fact that anti-degradable are known to have a private capacity that is zero [35]. □

Similar as above, given a channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ the smallest parameter ε such that Φ is ε-close-anti-degradable is given by

$$\tilde{\varepsilon}_\Phi := \left\{ \begin{array}{ll} \inf_{\Xi} & \|\Phi - \Xi\|_\diamond \\ \text{s.\,t.} & \Xi = \Theta \circ \Xi^{\mathsf{c}} \\ & \Xi : \mathcal{S}(A) \to \mathcal{S}(B) \text{ is cptp} \\ & \Theta : \mathcal{S}(E) \to \mathcal{S}(B) \text{ is cptp} . \end{array} \right.$$

The close connection between anti-degradability and 2-extendibility may be helpful to (efficiently) compute the quanitity $\tilde{\varepsilon}_\Phi$ (see [21, Lemma B.1]).[4] Corollary A.4 implies that for an arbitrary channel $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ with $|B| := \dim B$, we have $Q(\Phi) \leq 2\tilde{\varepsilon}_\Phi \log |B| + 2(2 + \tilde{\varepsilon}_\Phi)h\left(\frac{\tilde{\varepsilon}_\Phi}{2+\tilde{\varepsilon}_\Phi}\right).$ There is a close connection between the two concepts of an ε-degradable and an ε-close-degradable channel as stated in the following proposition.

Proposition A.5 (Relation between ε-close-degradable and ε-degradable)**.** *Let* $\Phi : \mathcal{S}(A) \to \mathcal{S}(B)$ *be a quantum channel that is* ε*-close-degradable, then* Φ *is* $(\varepsilon + 2\sqrt{\varepsilon})$*-degradable.*

Proof. Let $\Phi_1, \Phi_2 : \mathcal{S}(A) \to \mathcal{S}(B)$ be two channels such that $\|\Phi_1 - \Phi_2\|_\diamond \leq \varepsilon$. Then, by the continuity of Stinespring's representation [20, Equation (2)] it follows that there exist two complementary channels Φ_1^{c} and Φ_2^{c} such that $\|\Phi_1^{\mathsf{c}} - \Phi_2^{\mathsf{c}}\|_\diamond \leq 2\sqrt{\varepsilon}$. By assumption there exist two channels $\Xi : \mathcal{S}(A) \to \mathcal{S}(B)$ and $\Theta : \mathcal{S}(B) \to \mathcal{S}(E)$ such that $\|\Phi - \Xi\|_\diamond \leq \varepsilon$ with $\Xi^{\mathsf{c}} = \Theta \circ \Xi$. As explained above, the continuity of Stinespring's representation implies that

$$\left\|\Phi^{\mathsf{c}} - \Theta \circ \Xi\right\|_\diamond = \left\|\Phi^{\mathsf{c}} - \Xi^{\mathsf{c}}\right\|_\diamond \leq 2\sqrt{\varepsilon} .$$

Using the triangle inequality gives

$$\begin{aligned} \left\|\Phi^{\mathsf{c}} - \Theta \circ \Phi\right\|_\diamond &\leq \left\|\Phi^{\mathsf{c}} - \Theta \circ \Xi\right\|_\diamond + \left\|\Theta \circ \Xi - \Theta \circ \Phi\right\|_\diamond \\ &\leq 2\sqrt{\varepsilon} + \left\|\Xi - \Phi\right\|_\diamond \\ &\leq 2\sqrt{\varepsilon} + \varepsilon , \end{aligned}$$

which proves the assertion. □

It is unclear whether a converse statement to the one given in Proposition A.5, i.e., that a channel being ε-degradable implies that it is $\theta(\varepsilon)$-close-degradable for some function $\theta : [0, 1] \to \mathbb{R}_{\geq 0}$, is valid. This seems to be difficult to prove (if possible at all) as the structure of the set of degradable channels is poorly understood. (It is known that the set of degradable channels is not convex [12], however it is unknown how large this set is.)

---

[4] We would like to thank one referee for pointing this out.

Appendix B: Alternative proof of claim (i) of Theorem 3.4

For the sake of completeness we present here an alternative proof of claim (i). The lower bound $Q^{(1)}(\Phi) \leq Q(\Phi)$ is immediate. Consider a pure state $\phi_{AA'_1 \ldots A'_n}$ and for $i \in [n]$ let $V^i_{A'_i \to B_i E_i}$ denote the isometric extension of the i-th channel Φ. For $i \in [n]$ let $\sigma^i := V^i \phi (V^i)^\dagger$, $\rho_{AB_1 E_1 \ldots B_n E_n} := (\bigotimes_{i=1}^n V^i) \phi (\bigotimes_{i=1}^n V^{i\dagger})$ and define

$$\xi := \frac{\varepsilon}{2} \log(|E| - 1) + h(\varepsilon/2) + \varepsilon \log|E| + \left(1 + \frac{\varepsilon}{2}\right) h\left(\frac{\varepsilon}{2 + \varepsilon}\right).$$

Assuming that $\rho_{AB_1 E_1 \ldots B_n E_n}$ is the state that maximizes $Q^{(1)}(\Phi^{\otimes n})$ gives

$$
\begin{aligned}
Q^{(1)}(\Phi^{\otimes n}) &= I(A\rangle B_1 \ldots B_n)_\rho \\
&= H(B_1 \ldots B_n)_\rho - H(AB_1 \ldots B_n)_\rho \\
&= H(B_1 \ldots B_n)_\rho - H(E_1 \ldots E_n)_\rho \qquad\qquad\qquad\qquad\qquad (\text{B1}) \\
&= \sum_{i=1}^n H(B_i)_\rho - H(E_i)_\rho - \left(I(B_1 : B_2 : \ldots : B_n)_\rho - I(E_1 : E_2 : \ldots : E_n)_\rho\right) \\
&\leq \sum_{i=1}^n H(B_i)_\rho - H(E_i)_\rho + n\xi \qquad\qquad\qquad\qquad\qquad\qquad (\text{B2}) \\
&= \sum_{i=1}^n H(B_i)_{\sigma^i} - H\left(AB_i A'_1 \ldots A'_{i-1} A'_{i+1} \ldots A'_n\right)_{\sigma^i} + n\xi \qquad (\text{B3}) \\
&= \sum_{i=1}^n I(AA'_1 \ldots A'_{i-1} A'_{i+1} \ldots A'_n \rangle B_i)_{\sigma^i} + n\xi \\
&\leq n Q^{(1)}(\Phi) + n\xi,
\end{aligned}
$$

where (B1) follows since the state ρ is pure on the system $AB_1 E_1 \ldots B_n E_n$. Inequality (B2) follows by n times applying Lemma 3.5. Equation (B3) is true since the entropies of ρ and $\{\sigma^i\}_{i=1}^n$ on the given reduced systems are equal and σ^i is pure on $AA'_1 \ldots A'_{i-1} B_i E_i A'_{i+1} \ldots A'_n$. The final inequality follows as the states $\{\sigma^i\}_{i=1}^n$ are not necessarily the optimizers for the corresponding coherent informations. This proves statement (i) of Theorem 3.4.

---

[1] R. Alicki and M. Fannes. Continuity of quantum conditional information. *Journal of Physics A: Mathematical and General*, 37(5):L55–L57, 2004-02-06T00:00:00.

[2] G. Aubrun. On almost randomizing channels with a short Kraus decomposition. *Communications in Mathematical Physics*, 288(3):1103–1116, 2009.

[3] K. M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127, 2007.

[4] H. Barnum, E. Knill, and M. A. Nielsen. On quantum fidelities and channel capacities. *IEEE Transactions on Information Theory*, 46(4):1317–1329, Jul 2000.

[5] H. Barnum, M. A. Nielsen, and B. Schumacher. Information transmission through a noisy quantum channel. *Physical Review A*, 57:4153–4175, Jun 1998.

[6] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824–3851, 1996.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[8] D. Bruss, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello, and J. A. Smolin. Optimal universal and state-dependent quantum cloning. *Physical Review A*, 57:2368–2378, Apr 1998.

[9] N. Cai, A. Winter, and R. W. Yeung. Quantum privacy and quantum wiretap channels. *Problems of Information Transmission*, 40(4):318–336, 2004.

[10] N. J. Cerf. Pauli cloning of a quantum bit. *Physical Review Letters*, 84:4497–4500, May 2000.

[11] T. Cubitt, D. Elkouss, W. Matthews, M. Ozols, D. Pérez-García, and S. Strelchuk. Unbounded number of channel uses may be required to detect quantum capacity. *Nature Communications*, 6:4, March 2015.

[12] T. S. Cubitt, M. B. Ruskai, and G. Smith. The structure of degradable quantum channels. *Journal of Mathematical Physics*, 49(10):102104, 2008.

[13] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51(1):44–55, 2005.

[14] I. Devetak and P. W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Communications in Mathematical Physics*, 256:287–303, 2005.

[15] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin. Quantum-channel capacity of very noisy channels. *Physical Review A*, 57:830–839, Feb 1998.

[16] A. El Gamal and Y.-H. Kim. *Network Information Theory.* Cambridge University Press, Jan 2012.

[17] D. Elkouss and S. Strelchuk. Superadditivity of private information for any number of uses of the channel. *Phys. Rev. Lett.*, 115:040501, Jul 2015.

[18] A. S. Holevo. *Quantum Systems, Channels, Information.* De Gruyter Studies in Mathematical Physics 16, 2012.

[19] L. G. Khachiyan. Polynomial algorithms in linear programming. {*USSR*} *Computational Mathematics and Mathematical Physics*, 20(1):53 – 72, 1980.

[20] D. Kretschmann, D. Schlingemann, and R. F. Werner. The information-disturbance tradeoff and the continuity of Stinespring's representation. *IEEE Transactions on Information Theory*, 54(4):1708–1717, April 2008.

[21] F. Leditzky, N. Datta, and G. Smith. Useful states and entanglement distillation, 2017. available at `arXiv:1701.03081`.

[22] F. Leditzky, D. Leung, and G. Smith. Quantum and private capacities of low-noise channels, 2017. available at `arXiv:1705.04335`.

[23] D. Leung and G. Smith. Continuity of quantum channel capacities. *Communications in Mathematical Physics*, 292(1):201–215, 2009.

[24] E. H. Lieb and M. B. Ruskai. A fundamental property of quantum-mechanical entropy. *Physical Review Letters*, 30:434–436, Mar 1973.

[25] E. H. Lieb and M. B. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14(12):1938–1941, 1973.

[26] S. Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55(3):1613–1622, 1997.

[27] Y. Ouyang. Channel covariance, twirling, contraction, and some upper bounds on the quantum capacity. *Quantum Info. Comput.*, 14(11-12):917–936, September 2014.

[28] E. M. Rains. Bound on distillable entanglement. *Physical Review A*, 60:179–184, Jul 1999.

[29] E. M. Rains. A semidefinite program for distillable entanglement. *IEEE Transactions on Information Theory*, 47(7):2921–2933, Nov 2001.

[30] M. B. Ruskai. Some open problems in quantum information theory, 2007. available at `arXiv:0708.1902v1`.

[31] A. Schrijver. *Theory of Linear and Integer Programming.* John Wiley & Sons, 1998.

[32] B. Schumacher. Sending entanglement through noisy quantum channels. *Physical Review A*, 54:2614–2628, Oct 1996.

[33] B. Schumacher and M. A. Nielsen. Quantum data processing and error correction. *Physical Review A*, 54:2629–2635, Oct 1996.

[34] P. W. Shor. The quantum channel capacity and coherent information. Presented at the MSRI Workshop on Quantum Computation, 2002.

[35] G. Smith. Private classical capacity with a symmetric side channel and its application to quantum cryptography. *Physical Review A*, 78:022306, Aug 2008.

[36] G. Smith, J. M. Renes, and J. A. Smolin. Structured codes improve the Bennett-Brassard-84 quantum key rate. *Phys. Rev. Lett.*, 100:170502, Apr 2008.

[37] G. Smith and J. A. Smolin. Degenerate quantum codes for Pauli channels. *Physical Review Letters*, 98:030501, Jan 2007.

[38] G. Smith and J. A. Smolin. Additive extensions of a quantum channel. pages 368–372. Proceedings Information Theory Workshop (ITW), Porto, Portugal, May 5–9, 2008.

[39] G. Smith, J. A. Smolin, and A. Winter. The quantum capacity with symmetric side channels. *IEEE Transactions on Information Theory*, 54(9):4208–4217, Sept 2008.

[40] W. F. Stinespring. Positive functions on c^*-algebras. *Proceedings of the American Mathematical Society*, 6(211):pp. 211–216, April 1955.

[41] J. Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11):217–238, 2009.

[42] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, June 2013.

[43] A. Winter. Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints. *Communications in Mathematical Physics*, 347(1):291–313, 2016.

[44] M. M. Wolf. Quantum channels & operations, 2012. Lecture notes; available at http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf.

[45] M. M. Wolf and D. Perez-Garcia. Quantum capacities of channels with small environment. *Physical Review A*, 75:012303, Jan 2007.

[46] J. Yard, P. Hayden, and I. Devetak. Capacity theorems for quantum multiple-access channels: classical-quantum and quantum-quantum capacity regions. *IEEE Transactions on Information Theory*, 54(7):3091–3113, 2008.