On Hastings' counterexamples to the minimum output entropy additivity conjecture

Fernando G.S.L. Brandão[*]

Institute for Mathematical Sciences, Imperial College London, London SW7 2BW, UK and
QOLS, Blackett Laboratory, Imperial College London, London SW7 2BW, UK

Michał Horodecki[†]

Institute for Theoretical Physics and Astrophysics,
University of Gdańsk, 80-952 Gdańsk, Poland

Hastings recently reported a randomized construction of channels violating the minimum output entropy additivity conjecture. Here we revisit his argument, presenting a simplified proof. In particular, we do not resort to the exact probability distribution of the Schmidt coefficients of a random bipartite pure state, as in the original proof, but rather derive the necessary large deviation bounds by a concentration of measure argument. Furthermore, we prove non-additivity for the overwhelming majority of channels consisting of a Haar random isometry followed by partial trace over the environment, for an environment dimension much bigger than the output dimension. This makes Hastings' original reasoning clearer and extends the class of channels for which additivity can be shown to be violated.

I. INTRODUCTION

The oldest problem in quantum information theory is probably the determination of the capacity of a quantum-mechanical channel for classical information transmission. Given a quantum channel from a sender to a receiver, characterized by a trace preserving completely positive map $\mathcal{E}$, its classical capacity is defined as the maximum number of bits which can be reliably sent per use of the channel, in the limit of infinitely many realizations of the channel. Holevo [1] and Schumacher-Westmoreland [2] proved the following formula for the classical information transmission capacity:

$$C(\mathcal{E}) = \chi^{\infty}(\mathcal{E}) := \lim_{n \to \infty} \frac{\chi^{\infty}(\mathcal{E}^{\otimes n})}{n}, \tag{1}$$

where the Holevo χ-quantity [3] is defined by

$$\chi(\mathcal{E}) := \max_{\{p_i, \rho_i\}} S\left(\mathcal{E}\left(\sum_i p_i \rho_i\right)\right) - \sum_i p_i S\left(\mathcal{E}\left(\rho_i\right)\right), \tag{2}$$

with S being the von Neumann entropy and the maximization ranging over all ensembles $\{p_i, \rho_i\}$.

An important question concerning the capacity formula given by Eq. (1) is whether the *regularization* of the χ quantity to infinitely many uses of the channel is really needed in the right-hand-side of Eq. (1). Indeed, such necessity would render the evaluation of the formula given by Eq. (1) in general intractable; moreover, it would show that we do not fully understand the structure of the optimal coding strategy, since from Eq. (1) we cannot say anything about the -

---

[*]Electronic address: fernando.brandao@imperial.ac.uk
[†]Electronic address: fizmh@ug.edu.pl

in general entangled - states ρ_i appearing in the optimal ensemble. On a more positive note, the need of regularization would also show that we can boost the information transmission capacity by using entangled encoding states.

Based on numerical evidence in low dimensions and several results for particular classes of channels (e.g. [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]), it was conjectured the χ-quantity is additive, i.e. for every pair of channels $\mathcal{E}_1, \mathcal{E}_2$,

$$\chi(\mathcal{E}_1 \otimes \mathcal{E}_2) = \chi(\mathcal{E}_1) + \chi(\mathcal{E}_2). \tag{3}$$

The validity of this conjecture would imply that the classical capacity of a quantum channel is given simply by its Holevo χ-quantity, which would constitute a *single-letter* formula for the capacity. It turns out that Eq. (3) is in fact equivalent to the to the non-necessity of the limit in Eq. (1) [16]: $C(\mathcal{E}) = \chi(\mathcal{E})$ for every channel $\mathcal{E}$ if, and only if, Eq. (3) holds true for every pair of channels $\mathcal{E}_1, \mathcal{E}_2$ (see also [17]).

The additivity of the χ-quantity can be related to other additivity questions. The first concerns the entanglement cost of a bipartite quantum state ρ shared by Alice and Bob. It is defined as the optimal rate of EPR pairs needed for the formation of ρ, in the limit of asymptotically many copies of the state, by local operations and classical communication between Alice and Bob. It was shown in [18] that the entanglement cost is given by

$$E_C(\rho) := \lim_{n \to \infty} \frac{E_F(\rho^{\otimes n})}{n}, \tag{4}$$

where the entanglement of formation [19] is defined as

$$E_F(\rho) := \min_{\{p_i, \rho_i\}} \sum_i p_i S\left(\text{tr}_A\left(|\psi_i\rangle\langle\psi_i|\right)\right), \tag{5}$$

with the minimization taken over all pure state ensembles of ρ. As shown by Shor in Ref. [20] (building on [21, 22, 23]), the additivity of the entanglement of formation is equivalent to the additivity of χ as given by Eq. (3).

The second additivity question concerns the distillable common randomness of a bipartite state, given by the optimal rate of maximally correlated classical bits that can be extracted from a bipartite quantum state, when classical communication is allowed from Alice to Bob (the rate of bits communicated being subtracted from the rate of maximally correlated bits obtained in the end of the protocol). Devetak and Winter proved that [24]

$$C_D^{\rightarrow}(\rho) := \lim_{n \to \infty} \frac{I^{\rightarrow}(\rho^{\otimes n})}{n}, \tag{6}$$

with

$$I^{\rightarrow}(\rho) := \max_{\{M_i\}} \left(S(\rho_A) - \sum_i p_i S(\rho_i)\right), \tag{7}$$

where the maximization runs over POVMs $\{M_i\}$ applied to Alice's system, $p_i := \text{tr}(\rho(M_i \otimes \mathbb{I}))$ and $\rho_i := \text{tr}_A(\rho(M_i \otimes \mathbb{I}))/p_i$ [25]. In Ref. [26] Koashi and Winter derived a beautiful relation between the entanglement of formation and the quantity given in Eq. (7), showing in particular the equivalence of the need of the limit in Eq. (6) to the validity of Eq. (3) for every pair of channels.

An important simplification of the additivity problem, due to Shor [20], shows that the additivity of the χ-quantity is equivalent to a simpler question: the additivity of the minimum output entropy, defined as [20]

$$S_{\min}(\mathcal{E}) := \min_{\rho} S(\mathcal{E}(\rho)). \tag{8}$$

It turns out that Eq. (3) holds true if, and only if, for every pair of channels $\mathcal{E}_1, \mathcal{E}_2$

$$S_{\min}(\mathcal{E}_1 \otimes \mathcal{E}_2) = S_{\min}(\mathcal{E}_1) + S_{\min}(\mathcal{E}_2). \tag{9}$$

Recently, based on similar results on Rényi entropies by Winter [29] and Hayden [28] (see also [30, 31, 32]), Hastings proved the breakthrough result that the minimum output entropy is not additive [27]: in general, Eq. (9) does not hold true. This in turn implies that the limits in Eqs. (1), (4), and (6) are needed and thus that we are unfortunately further away from grasping these three capacities than what we might have expected.

Hastings argument combines the approach of Winter [29] and Hayden [28] to the problem with powerful new ideas and techniques to construct randomized examples of channels violating Eq. (9). In particular, his argument is heavily based on an exact expression for the eigenvalue probability distribution of the reduced density matrix of a Haar distributed bipartite state [33]. The main goal of the present paper is to revisit Hastings' proof by employing instead more general properties of the Haar distribution, such as large deviations bounds for the concentration of well-behaved functions around their mean-values in high dimensions. This allows us to present the proof in a relatively concise form. Moreover, we will be able to strengthen slightly Hastings' result and prove non-additivity of the overwhelming majority of Haar random channels (for appropriate input, output, and environment dimensions). As a by-product, we also obtain a new result concerning the concentration of measure phenomenon in high dimensional quantum states, which may be of independent interest.

We would like to refer the reader to an earlier paper by Fukuda, King, and Moser of a similar spirit [34], where Hastings' original argument is explained in great detail and rigor. In particular, the authors derived explicit lower bounds to the input, output and environment dimensions for which channels violating additivity can be constructed. Our approach is unlikely to provide better estimates than the ones found in Ref. [34], as it does not rely on the exact probability distribution of the Schmidt coefficients of a Haar bipartite state. However, as our proof differs from the original in a few places, the optimization of the dimensions in our version of the proof may still be an interesting task (which we do not pursue here however).

Notation: We denote the set of density matrices acting on a Hilbert space $\mathcal{H}$ by $D(\mathcal{H})$. Moreover, we will often write A and B for finite dimensional Hilbert spaces, $A \otimes B$ or AB for their tensor product, and $|A|, |B|$ for their dimensions. For a pure state $|\psi^{AB}\rangle \in AB$, we define $\psi^{AB} := |\psi^{AB}\rangle\langle\psi^{AB}|$, while ψ^A will denote $\mathrm{tr}_B(\psi^{AB})$, where tr_B is the partial trace over subsystem B. We denote the d-dimensional unitary group by $\mathbb{U}(d)$. We define the entropy deviation from its maximal value of a state $\rho \in \mathcal{D}(\mathbb{C}^d)$ by $\delta S(\rho) := \log(d) - S(\rho)$. Let $\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} : ||x||_2 = 1\}$ denote the Euclidean sphere in $\mathbb{R}^{n+1}$ and μ denote the normalized rotationally invariant measure in $\mathbb{S}^n$ (the Haar measure). Finally, the Bachmann-Landau notation $g(n) = o(f(n))$ stands for $\forall k > 0, \exists n_0 : \forall n > n_0, g(n) \leq kf(n)$.

Structure of the paper: In section II we present the main results of the paper as well as the key definitions used in the proofs. The counterexamples to the additivity conjecture are given by the combination of three propositions II.6, II.7, and II.8, which are proven in sections III, IV, and V, respectively.

II.   DEFINITIONS AND MAIN RESULTS

We will consider channels from A to B of the form

$$\mathcal{E}(\rho) = \text{tr}_A\left(U\left(\rho^A \otimes |0\rangle\langle 0|^B\right)U^\dagger\right) \tag{10}$$

for a unitary $U \in \mathbb{U}(|A||B|)$. The channels thus have input and environment dimensions equal to $|A|$ and output dimension equals to $|B|$. Moreover, we will make use of the conjugate channel of $\mathcal{E}$, defined as

$$\overline{\mathcal{E}}(\rho) = \text{tr}_A\left(U^*\left(\rho^A \otimes |0\rangle\langle 0|^B\right)U^T\right). \tag{11}$$

The counterexamples to the minimum output entropy additivity conjecture will be constructed by selecting the unitary U at random from the Haar measure in $\mathbb{U}(|A||B|)$ and considering the regime of a very large environment dimension $|A| \gg |B|$.

Throughout the paper $c_0 > 0$ will denote a fixed constant which can be taken to be e.g. $c_0 = 1333$, while the Landau notation $o(1)$ will stand for a term which can be taken as small as desired by choosing $|A|$ large enough. Hastings theorem can be stated as follows.

Theorem I *For U drawn from the Haar measure in $\mathbb{U}(|A||B|)$, consider a channel as in Eq. (10). Then, for $c \geq c_0$, with probability $1 - o(1)$,*

$$S_{\min}(\mathcal{E} \otimes \overline{\mathcal{E}}) \leq S_{\min}(\mathcal{E}) + S_{\min}(\overline{\mathcal{E}}) - \frac{\log|B| - 2c}{|B|}. \tag{12}$$

We will prove Theorem I by the combination of two results. The first, analogous to a similar result of Winter and Hayden [28, 29, 30] on Rényi entropies, delivers an upper bound on the minimum output entropy of $\mathcal{E} \otimes \overline{\mathcal{E}}$ by considering the output entropy of the canonical maximally entangled state in $A \otimes B$ as an input.

Lemma II.1 *For a channel given by Eq. (10),*

$$S_{\min}(\mathcal{E} \otimes \overline{\mathcal{E}}) \leq 2\log|B| - \frac{\log|B|}{|B|}. \tag{13}$$

For completeness, we reproduce the proof of Lemma (II.1) in Appendix D.

The second result is a probabilistic argument for the existence of channels with high minimum output entropy. This is Hastings breakthrough contribution to the problem [27].

Lemma II.2 *For U drawn from the Haar measure in $\mathbb{U}(|A||B|)$, consider a channel as in Eq. (10). Then, for $c \geq c_0$, with probability $1 - o(1)$,*

$$S_{\min}(\mathcal{E}) \geq \log|B| - \frac{c}{|B|}. \tag{14}$$

The main idea in the proof of Lemma II.2 is to look at the probability that the output of a Haar random input state is close to a low entropy state (with entropy smaller than $\log|B| - c/|B|$). On one hand, we will show that for $|B|/|A| = o(1)$, this probability is upper bounded by $\exp(-cK|A|)$ (with $K > 0$ a constant), for a Haar random choice of the channel unitary. On the other hand, we will compute a lower bound on this probability, conditioned on the minimum output entropy of

the channel being small; in this way we will get a lower bound of order $\exp(-\ln(2)|A|)$. Putting these two estimates together we obtain Lemma II.2.

There are two key conceptual insights necessary to turn the idea of the previous paragraph into a proof. The first is to define an appropriate notion of closeness, when quantifying how close a state is to a low entropy one. For this, Hastings introduced the concept of a tube around a state [40], which will take a central role in the proof of Lemma II.2.

Definition II.3 *We define the tube around $\sigma \in \mathcal{D}(\mathbb{C}^D)$ with width parameter $N > 0$ as*

$$TUBE(\sigma, N) := \left\{ \pi \in \mathcal{D}(\mathbb{C}^D) : \exists \ \frac{1}{2} \leq p \leq 1 \quad s.t. \quad \left\| \pi - \left(p\sigma + (1-p)\frac{\mathbb{I}}{D} \right) \right\|_\infty \leq \sqrt{\frac{\log(N)}{N}} \right\}. \tag{15}$$

We will be interested in the probability that the output of a random input state, over a random choice of the channel, is in the tube of a low entropy state. The set of such states is formalized in the next definition.

Definition II.4 *For constants $N, c > 0$, we define the set of states in the tube of a low entropy state as*

$$X_{D,N,c} := \left\{ \rho \in \mathcal{D}(\mathbb{C}^D) : \exists \ \sigma \in \mathcal{D}(\mathbb{C}^D) \ \text{with} \ \delta S(\sigma) \geq c/D \ s.t. \ \rho \in TUBE(\sigma, N) \ \right\}. \tag{16}$$

The second insight is to consider the probability only of a particular subset of the set of states close to a low entropy state. We will look at the intersection of $X_{D,N,c}$ with the set of states of small operator norm. While this restriction will affect only very mildly the lower bound on the probability we are ultimately interesting in analyzing, it will allow us to get a much improved upper bound on it.

Definition II.5 *For a constant $a > 1$, we define the set of states with bounded operator norm as*

$$Y_{D,a} := \left\{ \rho \in \mathcal{D}(\mathbb{C}^D) : \|\rho\|_\infty \leq \frac{a}{D} \right\}. \tag{17}$$

We are now in position to state precisely the two propositions which will be the focus of the remainder of the paper.

Let $|\chi\rangle \in A$ be a Haar random state and $\mathcal{E}$ be a channel given by Eq. (10) with U drawn from the Haar measure in $\mathbb{U}(|A||B|)$. Then, for $|A| \geq |B|^2$, we have

Proposition II.6

$$\Pr_{\mathcal{E}, \chi} \left(\mathcal{E}(\chi) \in X_{|B|,|A|,c} \cap Y_{|B|,a} \right) \leq \exp\left(-\frac{c|A|}{128a} + o(1)|A| \right). \tag{18}$$

Moreover, for $\log |A| \geq 8|B|^8$ and $a \geq 15$,

Proposition II.7

$$\Pr_{\mathcal{E}, \chi} \left(\mathcal{E}(\chi) \in X_{|B|,|A|,c} \cap Y_{|B|,a} \right) \geq \frac{1}{8|A|} \exp(-\ln(2)|A|) \left(\Pr_{\mathcal{E}} \left(\delta S_{\min} \geq \frac{c}{|B|} \right) - o(1) \right). \tag{19}$$

Combining these two results we get Lemma II.2 by choosing $c > 128 \ln(2)a$ and $a = 15$.

We will derive Proposition II.6 from a new large deviation bound, which we believe might be of independent interest. It shows that with high probability the reduced state ψ^A of a random bipartite state $|\psi^{AB}\rangle$ is close, in two norm, to the maximally mixed state. Although similar results are well-known (see e.g. [35, 36]), the restriction to reduced states ψ^B with a small operator norm will allow us to sharpen the exponential bound essentially by a factor of $|B|$; this improvement turns out to be crucial in proving Proposition II.6. By a measure concentration argument we prove in section V the following

Proposition II.8 *For $|\psi^{AB}\rangle \in A \otimes B$ drawn from the Haar measure, $|A| \geq |B|^2$ and $a \geq 3$,*

$$\Pr\left(\left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2 \geq \varepsilon \;\; and \;\; \psi^B \in Y_{|B|,a}\right) \leq 4\exp\left(-\frac{|A||B|^2 \left(\varepsilon - 2|A|^{-\frac{1}{2}}\right)^2}{64a}\right). \tag{20}$$

III. PROOF OF PROPOSITION II.6

In this section we prove Proposition II.6. The idea is to combine Proposition II.8 and the following simple lemma relating the entropy deviation from its maximal value to the distance to the maximally mixed state.

Lemma III.1 *For every $\sigma \in \mathcal{D}(\mathbb{C}^D)$,*

$$\left\|\sigma - \frac{\mathbb{I}}{D}\right\|_2^2 \geq \frac{\log(D) - S(\sigma)}{D}. \tag{21}$$

Proof We have

$$\begin{aligned} S(\rho) &\geq -\log(\mathrm{tr}(\rho^2)) \\ &= -\log(D\mathrm{tr}(\rho^2)) + \log(D) \\ &\geq 1 - D\mathrm{tr}(\rho^2) + \log(D), \end{aligned} \tag{22}$$

where the first inequality follows from the concavity of the $\log$ and the second from the relation $\log(x) \leq x - 1$, valid for $x \geq 1$. Rearranging terms in Eq. (22), we find Eq. (21). $\square$

Proof (Proposition II.6)

Let $|\psi^{AB}\rangle$ be such that $\psi^B \in X_{|B|,|A|,c}$. Then there is a σ with $\delta S(\sigma) \geq c/|B|$ such that $\psi^B \in$ TUBE$(\sigma, |A|)$. From Lemma III.1 we get

$$\left\|\sigma - \frac{\mathbb{I}}{|B|}\right\|_2 \geq \frac{\sqrt{c}}{|B|}. \tag{23}$$

As $\|\psi^B - (p\sigma + (1-p)\mathbb{I}/|B|)\|_\infty \leq \sqrt{\frac{\log|A|}{|A|}}$, with $p \geq 1/2$, Eq. (23) gives

$$\left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2 \geq \frac{\sqrt{c}}{2|B|} - \sqrt{\frac{\log|A|}{|A|}}. \tag{24}$$

where we used $|| \cdot ||_2 \leq || \cdot ||_\infty$.

A moment of thought reveals that the distribution of $\mathcal{E}(\chi)$, for random $\mathcal{E}$ and χ, is the same as the distribution of the reduced density matrix ψ^B of a random bipartite state $|\psi\rangle^{AB} \in A \otimes B$. Therefore, from the argument of the previous paragraph

$$\Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \in X_{|B|,|A|,c} \cap Y_{|B|,a} \right) = \Pr_\psi \left(\psi^B \in X_{|B|,|A|,c} \cap Y_{|B|,a} \right) \tag{25}$$

$$\leq \Pr_\psi \left(\left\| \psi^B - \frac{\mathbb{I}}{|B|} \right\|_2 \geq \frac{\sqrt{c}}{2|B|} - \sqrt{\frac{\log|A|}{|A|}} \ \text{ and } \ \psi^B \in Y_{|B|,a} \right).$$

The result now follows from Proposition II.8. $\qquad\square$

IV. PROOF OF PROPOSITION II.7

On general lines, the idea of the lower bound given by Proposition II.7 is the following. Let P be the probability that a random channel has minimum output entropy bigger than $\log|B| - c/|B|$. For a given channel $\mathcal{E}$, let $\chi_\mathcal{E}$ be a pure input state to $\mathcal{E}$ with minimum output entropy, i.e. a state which satisfies $S(\mathcal{E}(\chi_\mathcal{E})) = S_{\min}(\mathcal{E})$. We will show that with probability larger than $\Omega(\exp(-\ln(2)|A|))$, $\mathcal{E}(\chi)$ is in the tube of $\mathcal{E}(\chi_\mathcal{E})$, for a random choice of the input state $|\chi\rangle$. From this we can conclude that $\mathcal{E}(\chi)$ is in the tube of a low entropy state with probability bigger than $(1 - P)\Omega(\exp(-\ln(2)|A|))$.

This is almost all there is to show, except that from the argument of the previous paragraph, we have no guarantee that the states $\mathcal{E}(\chi)$ which we have proven to be in the tube of a low entropy state also belong to $Y_{|B|,a}$. To overcome this difficulty, we employ a large deviation bound due to Harrow, Hayden, and Leung [35] (Lemma C.1 in Appendix C) which shows that with probability bigger than $1 - \exp(-|A|)$, $\mathcal{E}(\chi)$ belongs $Y_{|B|,a}$. This lemma thus allows us to disregard states not in $Y_{|B|,a}$ for sufficiently large $|A|$.

Proof (Proposition II.7)

The first step in the proof is to eliminate the event $\mathcal{E}(\chi) \in Y_{|B|,a}$. For this, we first use Lemma A.1 of Appendix A to get

$$\Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \in X_{|B|,|A|,c} \cap Y_{|B|,a} \right) \geq \Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \in X_{|B|,|A|,c} \right) - \Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \notin Y_{|B|,a} \right). \tag{26}$$

Then, from Lemma C.1 of Apendix C,

$$\Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \notin Y_{|B|,a} \right) \leq \left(\frac{10|B|}{a-1} \right)^{2|B|} \exp\left(-|A| \frac{(a-1) - \log(a)}{14\ln(2)} \right) \leq \exp\left(-|A| \right), \tag{27}$$

for $a \geq 15$ and $|A| \geq 2|B|\ln(2|B|)$.

In the remainder of the proof we show that

$$\Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \in X_{|B|,|A|,c} \right) \geq \frac{1}{8|A|} \exp(-\ln(2)|A|) \left(\Pr_\mathcal{E} \left(\delta S_{\min} \geq \frac{c}{|B|} \right) - o(1) \right). \tag{28}$$

The result then follows from Eqs. (26), (27), and (28).

Let us define $\sigma_\mathcal{E} := \mathcal{E}(\chi_\mathcal{E})$, with $\chi_\mathcal{E}$ an input to $\mathcal{E}$ with minimum output entropy, i.e. a state such that $S(\mathcal{E}(\chi_\mathcal{E})) = S_{\min}(\mathcal{E})$. From the definition of the set $X_{|B|,|A|,c}$ we find

$$\Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \in X_{|B|,|A|,c} \right) \geq \Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \in \mathrm{TUBE}(\sigma_\mathcal{E}, |A|) \ \text{ and } \ \delta S_{\min}(\mathcal{E}) \geq \frac{c}{|B|} \right). \tag{29}$$

We now proceed to bound the right-hand-side of Eq. (29). Following [34],

$$\Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \in \mathrm{TUBE}(\sigma_\mathcal{E}, |A|) \ \text{ and } \ \delta S_{\min}(\mathcal{E}) \geq \frac{c}{|B|} \right) =$$
$$\mathbb{E}_\mathcal{E} \left(\mathbf{1} \left(\delta S_{\min}(\mathcal{E}) \geq \frac{c}{|B|} \right) \Pr_{\chi} \left(\mathcal{E}(\chi) \in \mathrm{TUBE}(\mathcal{E}(\sigma_\mathcal{E}), |A|) \right) \right), \tag{30}$$

where $\mathbf{1}(\delta S_{\min}(\mathcal{E}) \geq c/|B|))$ is the indicator function of the event (only over channels): $\{\delta S_{\min}(\mathcal{E}) \geq c/|B|\}$.

Let us consider the probability over states inside the expectation value in Eq. (30). For a Haar random $|\chi\rangle \in A$, we can write

$$|\chi\rangle = \sqrt{x}|\chi_\mathcal{E}\rangle + \sqrt{1-x}|\phi\rangle, \tag{31}$$

where $x = |\langle \chi_\mathcal{E} | \chi \rangle|^2$ and $|\phi\rangle$ is a state orthogonal to $|\chi_\mathcal{E}\rangle$. In Lemma A.2 of Appendix A we prove that x and $|\phi\rangle$ are independent random variables and that $|\phi\rangle$ is distributed accordingly to the Haar measure in the subspace of A orthogonal to $|\psi\rangle$. Therefore,

$$\Pr_{\chi} \left(\mathcal{E}(\chi) \in \mathrm{TUBE}(\mathcal{E}(\sigma_\mathcal{E}), |A|) \right) \geq \Pr \left(x \geq 1/2 \right) \Pr_{\phi} \left(F \cap G \right), \tag{32}$$

where

$$F := \left\{ \|\mathcal{E}(|\chi_\mathcal{E}\rangle\langle\phi|)\|_\infty \leq \frac{1}{4}\sqrt{\frac{\log|A|}{|A|}} \right\}, \quad G := \left\{ \left\| \mathcal{E}(|\phi\rangle\langle\phi|) - \frac{\mathbb{I}}{|B|} \right\|_\infty \leq \frac{1}{2}\sqrt{\frac{\log|A|}{|A|}} \right\}. \tag{33}$$

Indeed, note that if $x \geq 1/2$ and F, G hold true

$$\left\| \mathcal{E}(\chi) - x\mathcal{E}(\chi_\mathcal{E}) - (1-x)\frac{\mathbb{I}}{|B|} \right\|_\infty \leq 2\|\mathcal{E}(|\chi_\mathcal{E}\rangle\langle\phi|)\|_\infty + \left\| \mathcal{E}(|\phi\rangle\langle\phi|) - \frac{\mathbb{I}}{|B|} \right\|_\infty \leq \sqrt{\frac{\log|A|}{|A|}}, \tag{34}$$

which implies $\mathcal{E}(\chi) \in \mathrm{TUBE}(\mathcal{E}(\sigma_\mathcal{E}), |A|)$.

In Lemma IV.1 we use a simple geometric argument to show

$$\Pr \left(|\langle \chi_\mathcal{E} | \chi \rangle|^2 \geq \frac{1}{2} \right) \geq \frac{1}{8|A|} \exp\left(-\ln(2)|A| \right). \tag{35}$$

Then, from Eqs. (30) and (32)

$$\Pr_{\mathcal{E},\chi} \left(\mathcal{E}(\chi) \in \mathrm{TUBE}(\mathcal{E}(\sigma_\mathcal{E}), |A|) \ \text{ and } \ \delta S_{\min}(\mathcal{E}) \geq \frac{c}{|B|} \right)$$
$$\geq \frac{1}{8|A|} \exp\left(-\ln(2)|A| \right) \mathbb{E}_\mathcal{E} \left(\mathbf{1} \left(\delta S_{\min}(\mathcal{E}) \geq \frac{c}{|B|} \right) \Pr_{\chi} \left(F \cap G \right) \right),$$
$$= \frac{1}{8|A|} \exp\left(-\ln(2)|A| \right) \Pr_{\mathcal{E},\chi} \left(F \cap G \cap \left(\delta S_{\min}(\mathcal{E}) \geq \frac{c}{|B|} \right) \right). \tag{36}$$

From Lemma A.1 we can bound the second term in the last line of the equation above as

$$\Pr_{\mathcal{E},\chi}\left(F\cap G\cap\left(\delta S_{\min}(\mathcal{E})\geq\frac{c}{|B|}\right)\right)\geq\Pr_{\mathcal{E},\chi}\left(\delta S_{\min}(\mathcal{E})\geq\frac{c}{|B|}\right)-\Pr_{\mathcal{E},\chi}\left(F^c\right)-\Pr_{\mathcal{E},\chi}\left(G^c\right). \tag{37}$$

Eq. (28) now follows from Lemma IV.2, where we prove that $\Pr_{\mathcal{E},\chi}(F^c), \Pr_{\mathcal{E},\chi}(G^c) = o(1)$, asymptotically in $|A|$. $\qquad\square$

Lemma IV.1 *Let $|\psi\rangle \in A$ be a fixed state and $|\chi\rangle \in A$ be drawn from the Haar measure. Then,*

$$\Pr\left(|\langle\psi|\chi\rangle|^2\geq\frac{1}{2}\right)\geq\frac{1}{8|A|}\exp\left(-|A|\ln 2\right) \tag{38}$$

Proof

The vectors $|\chi\rangle$ can be seen as points $(x_1,\ldots,x_n)$ on real unit sphere $\mathbb{S}^{n-1}$ with $n = 2|A|$. The Haar measure is thus the normalized area of the sphere and the condition $|\langle\psi|\chi\rangle|^2 \geq 1/2$ reads as $x_1^2 + x_2^2 \geq 1/2$.

Clearly $\Pr(x_1^2 + x_2^2 \geq 1/2)$ is lower bounded by $\Pr(x_1^2 \geq 1/2)$, which equals to the ratio of the area of a polar cap determined by the condition $x_1^2 \geq 1/2$ and the volume of the sphere. The area of the cap is in turn lower bounded by the volume of an $(n-1)$-dimensional ball given by the condition $x_2^2 + \ldots + x_n^2 \leq 1/2$ (the projection of the cap onto a subspace perpendicular to the x_1 axis). Invoking explicit formulas for the volume of a ball and the area of a sphere (see e.g. [37]), we obtain

$$\Pr(|\langle\psi|\chi\rangle|^2\geq 1/2)\geq\frac{1}{n\pi(\sqrt{2})^{n-1}}\geq\frac{1}{8|A|}e^{-\ln(2)|A|}. \tag{39}$$

$\qquad\square$

Lemma IV.2 *Let $|\psi\rangle$ be a fixed state in A, $|\phi\rangle$ be drawn from the Haar measure in the subspace of A orthogonal to $|\psi\rangle$ and $\mathcal{E}$ be a channel as in Eq. (10), with U drawn from the Haar measure in $\mathbb{U}(|A||B|)$. Define*

$$F:=\left\{\|\mathcal{E}(|\psi\rangle\langle\phi|)\|_\infty\leq\frac{1}{4}\sqrt{\frac{\log|A|}{|A|}}\right\},\quad G:=\left\{\left\|\mathcal{E}(|\phi\rangle\langle\phi|)-\frac{\mathbb{I}}{|B|}\right\|_\infty\leq\frac{1}{2}\sqrt{\frac{\log|A|}{|A|}}\right\}, \tag{40}$$

Then, for $\log|A|\geq 8|B|^8$ there are constants $C_1, C_2 > 0$ such that

$$\Pr_{\mathcal{E},\phi}\left(F\right)\geq 1-\exp\left(-\frac{C_1\log|A|}{|B|^8}\right),\quad\Pr_{\mathcal{E},\phi}\left(G\right)\geq 1-\exp\left(-C_2\log|A|\right). \tag{41}$$

Proof Let us start with the bound on the probability of F^c. Consider the complementary channel of $\mathcal{E}$, defined by $\mathcal{E}^c(\rho) := \text{tr}_B\left(U\left(\rho^A\otimes|0\rangle\langle 0|^B\right)U^\dagger\right)$. Noting that $\mathcal{E}^c$ is a channel with input and output dimension $|A|$ and environment dimension $|B|$, we can write

$$\mathcal{E}^c(\rho)=\sum_{k=1}^{|B|^2}A_k\rho A_k^\dagger, \tag{42}$$

for Kraus operators A_k such that $\sum_k A_k^\dagger A_k = \mathbb{I}$. Thus

$$\mathcal{E}(\rho)=\sum_{k=1}^{|B|^2}\sum_{k'=1}^{|B|^2}\text{tr}(A_{k'}^\dagger A_k\rho)|k\rangle\langle k'|, \tag{43}$$

from which we find

$$\|\mathcal{E}(|\psi\rangle\langle\phi|)\|_\infty \leq |B|^4 \max_{k,k'} |\langle\phi|A_{k'}^\dagger A_k|\psi\rangle|. \tag{44}$$

Let $k_{\max}, k'_{\max}$ be the optimal indices in the equation above and define $|\theta\rangle := A_{k'_{\max}}^\dagger A_{k_{\max}}|\psi\rangle/\|A_{k'_{\max}}^\dagger A_{k_{\max}}|\psi\rangle\|^{1/2}$. As $\|A_k\|_\infty \leq 1$ for all k, we get $\|A_{k'_{\max}}^\dagger A_{k_{\max}}|\psi\rangle\| \leq 1$ and hence

$$\|\mathcal{E}(|\psi\rangle\langle\phi|)\|_\infty \leq |B|^4 |\langle\theta|\phi\rangle|. \tag{45}$$

We thus have

$$\Pr_\phi (F^c) \leq \Pr_\phi \left(|\langle\theta|\phi\rangle| \geq \frac{1}{4|B|^4} \sqrt{\frac{\log|A|}{|A|}} \right). \tag{46}$$

Applying Lemma IV.3 to the equation above we find

$$\Pr_\phi \left(|\langle\theta|\phi\rangle| \geq \frac{1}{4|B|^4} \sqrt{\frac{\log|A|}{|A|}} \right) \leq 2 \exp\left(-\frac{K \log|A|}{|B|^8} \right), \tag{47}$$

for $\log|A| \geq 8|B|^8$ and a constant $K > 0$. This gives the bound on $\Pr(F)$ given in Eq. (41).

Let us now turn to the bound on the probability of G. From Lemma A.2 of Appendix A, we can select $|\phi\rangle$ by drawing $|\chi\rangle \in A$ from the Haar measure and setting $|\chi\rangle = \sqrt{x}|\psi\rangle + \sqrt{1-x}|\phi\rangle$. Then we have

$$\left\| \mathcal{E}(\phi) - \frac{\mathbb{I}}{|B|} \right\|_\infty \leq \|\mathcal{E}(\phi) - \mathcal{E}(\chi)\|_\infty + \left\| \mathcal{E}(\chi) - \frac{\mathbb{I}}{|B|} \right\|_\infty$$

$$\leq \|\phi - \chi\|_1 + \left\| \mathcal{E}(\chi) - \frac{\mathbb{I}}{|B|} \right\|_\infty, \tag{48}$$

where the first inequality follows from the triangle inequality and the second from the fact that $\|X\|_\infty \leq \|X\|_1$ and the monotonicity of the trace norm under trace preserving CP maps. Therefore,

$$\Pr_{\mathcal{E},\phi} (G) \geq \Pr_{\mathcal{E},\chi} \left(\|\phi - \chi\|_1 \leq \frac{1}{4}\sqrt{\frac{\log|A|}{|A|}} \text{ and } \left\| \mathcal{E}(\chi) - \frac{\mathbb{I}}{|B|} \right\|_\infty \leq \frac{1}{4}\sqrt{\frac{\log|A|}{|A|}} \right). \tag{49}$$

From Lemma A.1 of Appendix A, in turn,

$$\Pr_{\mathcal{E},\phi} (G) \geq 1 - \Pr_{\mathcal{E},\chi} \left(\|\phi - \chi\|_1 \geq \frac{1}{4}\sqrt{\frac{\log|A|}{|A|}} \right) - \Pr_{\mathcal{E},\chi} \left(\left\| \mathcal{E}(\chi) - \frac{\mathbb{I}}{|B|} \right\|_\infty \geq \frac{1}{4}\sqrt{\frac{\log|A|}{|A|}} \right). \tag{50}$$

One one hand, we have $\|\phi - \chi\|_1 \leq \sqrt{2 - 2|\langle\phi|\chi\rangle|^2} = \sqrt{2x(2-x)} \leq 2\sqrt{x} = 2|\langle\psi|\chi\rangle|$. Following [34], we find that if we replace $|\phi\rangle$ by $|\chi\rangle$, then with high probability it will only incur in a small error. Indeed, from Lemma IV.3

$$\Pr_{\mathcal{E},\chi} \left(\|\phi - \chi\|_1 \geq \frac{1}{4}\sqrt{\frac{\log|A|}{|A|}} \right) \leq 2 \exp\left(-K \log|A| \right), \tag{51}$$

for a constant $K > 0$.

On the other hand, from Lemma C.1 of section C,

$$\Pr_{\mathcal{E},\chi}\left(\left\|\mathcal{E}(\chi) - \frac{\mathbb{I}}{|B|}\right\|_\infty \geq \frac{1}{4}\sqrt{\frac{\log|A|}{|A|}}\right) \leq \exp\left(-\frac{\log|A|}{560\ln(2)}\right). \tag{52}$$

Combining these two last equations with Eq. (50), we find the lower bound on $\Pr(G)$ given in Eq. (41). $\qquad\square$

Lemma IV.3 *Let $S \subseteq A$ be a $|S|$-dimensional subspace of A and let P_S be the projector onto S. For $|\phi\rangle \in S$ drawn from the Haar measure in S and a fixed $|\theta\rangle \in A$,*

$$\Pr_\phi\left(|\langle\theta|\phi\rangle| \geq \frac{1}{\sqrt{|S|}} + \varepsilon\right) \leq 4\exp\left(-\frac{|S|\varepsilon^2}{16}\right), \tag{53}$$

Proof We prove the lemma by applying Levy's lemma, given in Lemma V.1 of section V, with $f(|\phi\rangle) := |\langle\theta|\phi\rangle|$. On one hand, we have

$$\mathbb{E}\left(f(|\phi\rangle)^2\right) = \langle\theta|\left(\frac{P_S}{|S|}\right)|\theta\rangle \leq \frac{1}{|S|}. \tag{54}$$

Then, from the convexity of x^2, $\mathbb{E}\left(f(|\phi\rangle)\right)^2 \leq \mathbb{E}\left(f(|\phi\rangle)^2\right) \leq |S|^{-1}$. On the other hand, the Lipschitz constant of f is easily seen to be unity. The result then follows easily from Lemma V.1. $\qquad\square$

Remark: We note that in the proof of Proposition II.7 we set the input dimension $|A|$ to be exponentially larger than the output dimension $|B|$; this is due to the factor of $\log|A|/|A|$ in the definition of the tube. We could have instead defined the width of the tube as $f(|A|)/|A|$ for any function f sublinear in $|A|$. In this way we can get a much better dependence of the input dimension $|A|$ with the output dimension $|B|$. Besides that, as in Hastings' original proof, we have used equal input and environment dimensions. However, our approach allow us to consider the general case in essentially the same fashion. In pricinple, this could lead to a better scaling of the minimal dimensions for which counterexamples can be shown to exist.

V. PROOF OF PROPOSITION II.8

Let $\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} : ||x||_2 = 1\}$ denote the Euclidean sphere in $\mathbb{R}^{n+1}$ and μ denote the normalized rotationally invariant measure in $\mathbb{S}^n$ (the Haar measure). Our strategy to prove Proposition II.8 is to explore the measure concentration phenomenon in high dimensional spheres [37, 38]. For a subset $A \subset \mathbb{S}^n$, define the ε-neighborhood of A as

$$A_\varepsilon := \{y \in \mathbb{S}^n : \exists\ x \in A\ \ s.t.\ \ ||x - y||_2 \leq \varepsilon\}. \tag{55}$$

Theorem II *(Concentration of Measure in $\mathbb{S}^n$ [37, 38]) Let $A \subset \mathbb{S}^n$ and $0 \leq \epsilon \leq 1$. If $\mu(A) \geq 1/2$, then $\mu(A_\varepsilon) \geq 1 - 4\exp\left(-\frac{(n+1)\epsilon^2}{16}\right)$.*

This theorem says that the area of $\mathbb{S}^n$ is sharply concentrated around any set with measure bigger than $1/2$. A simple but very powerful corollary of Theorem II says that slowly varying functions on $\mathbb{S}^n$ attain a value very close to its average almost everywhere (see e.g. [36] for applications to quantum information theory). This is the content of Levy's Lemma.

Lemma V.1 *(Levy's Lemma [37, 38]) Let $f : \mathbb{S}^n \to \mathbb{R}$ be a function with Lipschitz constant η and a point $x \in \mathbb{S}^n$ be chosen uniformly at random. Then*

$$\Pr\left(|f(x) - \mathbb{E}f| \geq \alpha\right) \leq 4\exp\left(-\frac{(n+1)\alpha^2}{16\eta^2}\right). \tag{56}$$

Given a Haar distributed state $|\psi\rangle \in A$, we can see it as an Haar distributed point in $\mathbb{S}^{2|A|-1}$. Therefore the lemma above applies to Haar pure states as well.

The proof of Proposition II.8 will follow closely the standard argument for deriving Levy's Lemma (see e.g. [37, 38]). An important difference is that we are only interested in establishing a large deviation bound for a particular subset of the state space, namely for states $|\psi^{AB}\rangle$ whose the reduced state ψ^B has operator norm bounded by $a/|B|$. Such a restriction will allow us to use an improved bound on the Lipschitz constant of the function $g(|\psi^{AB}\rangle) := \left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2$ and sharpen the exponential bound appearing in Levy's Lemma by a factor of $|B|/(4a)$.

Proof (Proposition II.8) Define

$$g(|\psi^{AB}\rangle) := \left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2. \tag{57}$$

Note that g is a function from $\mathbb{S}^{2|A||B|-1}$ to $\mathbb{R}$. Let $m(g)$ be the median of g and set $M := \{|\psi^{AB}\rangle : g(|\psi^{AB}\rangle) \leq m(g)\}$. In Lemma V.3 we show $m(g) \leq 2|A|^{-\frac{1}{2}}$. Thus for every $|\psi^{AB}\rangle \in M$, we have

$$\|\psi^B\|_2^2 \leq \frac{1}{|B|} + m(g)^2 \leq \frac{1}{|B|} + \frac{4}{|A|}. \tag{58}$$

An application of Lemma B.1 of Appendix B with $\lambda = a/|B|$ then gives the following bound on the operator norm of states in M,

$$\|\psi^B\|_\infty \leq \frac{3}{|B|} \leq \frac{a}{|B|}, \tag{59}$$

for every $\psi^{AB} \in M$ and $|A| \geq |B|^2$ and $a \geq 3$.

Consider a state $|\psi^{AB}\rangle$ such that

$$g(|\psi^{AB}\rangle) \geq m(g) + \beta \quad \text{and} \quad \|\psi^B\|_\infty \leq a/|B|. \tag{60}$$

Because of the bound on the operator norm of ψ^B, we can use Lemma V.2 to find from the first inequality of Eq. (60) that ψ^{AB} must be at least $\beta\sqrt{\frac{|B|}{4a}}$ away from M. Furthermore, by definition of the median, $\mu(M) \geq 1/2$. Therefore from Theorem II

$$\Pr\left(\left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2 \geq \varepsilon \quad \text{and} \quad \psi^B \in Y_{|B|,a}\right) \leq 1 - \mu\left(A_{(\epsilon-m(g))\sqrt{|B|/4a}}\right) \leq \exp\left(-\frac{|A||B|^2(\varepsilon - m(g))^2}{64a}\right), \tag{61}$$

and we are done. □

The next lemma shows that for states with operator norm bounded by a/B, the Lipschitz constant of the function g is improved by a factor of $\sqrt{|B|/(4a)}$.

Lemma V.2 *Let $|\psi^{AB}\rangle, |\phi^{AB}\rangle \in A \otimes B$ be such that $\|\psi^B\|_\infty, \|\phi^B\|_\infty \leq a/|B|$. Then*

$$\left|\left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2 - \left\|\phi^B - \frac{\mathbb{I}}{|B|}\right\|_2\right| \leq \sqrt{\frac{4a}{|B|}}\||\psi^{AB}\rangle - |\phi^{AB}\rangle\|_2. \tag{62}$$

Proof We assume without loss of generality that $\left\|\psi^B - \mathbb{I}/|B|\right\|_2 \geq \left\|\phi^B - \mathbb{I}/|B|\right\|_2$. Let $\{|i\rangle\}_{i=1}^{\mathrm{rank}(\psi^B)}$ be an eigenbasis for ψ^B and define $M(\rho) := \sum_i \langle i|\rho|i\rangle |i\rangle\langle i|$. Then,

$$
\begin{aligned}
\left| \left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2 - \left\|\phi^B - \frac{\mathbb{I}}{|B|}\right\|_2 \right| &= \left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2 - \left\|\phi^B - \frac{\mathbb{I}}{|B|}\right\|_2 \\
&\leq \left\|M(\psi^B) - \frac{\mathbb{I}}{|B|}\right\|_2 - \left\|M(\phi^B) - \frac{\mathbb{I}}{|B|}\right\|_2 \\
&\leq \|M(\psi^B) - M(\phi^B)\|_2,
\end{aligned}
\tag{63}
$$

where the first inequality follows from Lemma V.4, and the second inequality from the triangle inequality.

Let $\{p_k\}_k$ and $\{q_k\}_k$ be the eigenvalues of $M(\psi^B) = \psi^B$ and $M(\phi^B)$, respectively. Since $\|\psi^B\|_\infty, \|\phi^B\|_\infty \leq a/|B|$, we find from Lemma V.4 that $(\max_k p_k), (\max_k q_k) \leq a/|B|$. Hence

$$
\begin{aligned}
\|M(\psi^B) - M(\phi^B)\|_2^2 &= \sum_k (p_k - q_k)^2 \tag{64}\\
&= \sum_k (\sqrt{p_k} - \sqrt{q_k})^2 (\sqrt{p_k} + \sqrt{q_k})^2 \\
&\leq \frac{4a}{|B|} \sum_k (\sqrt{p_k} - \sqrt{q_k})^2 \\
&= \frac{4a}{|B|} \left(2 - 2F(M(\psi^B), M(\phi^B))\right) \\
&\leq \frac{4a}{|B|} \left(2 - 2F(\psi^B, \phi^B)\right) \\
&\leq \frac{4a}{|B|} \left(2 - 2F(\psi^{AB}, \phi^{AB})\right) = \frac{4a}{|B|} \||\psi^{AB}\rangle - |\phi^{AB}\rangle\|_2^2,
\end{aligned}
$$

where the last two inequalities follows from the monotonicity of the fidelity under trace preserving CP maps. Putting Eqs. (63) and (64) together gives the result. $\qquad\square$

The next lemma gives an upper bound on the median of the function g.

Lemma V.3 *Let $g : \mathbb{S}^{2|A||B|} \to \mathbb{R}$ be such that*

$$
g(|\psi^{AB}\rangle) := \left\|\psi^B - \frac{\mathbb{I}}{|B|}\right\|_2
\tag{65}
$$

and $m(g)$ be the median of g. Then $m(g) \leq 2|A|^{-\frac{1}{2}}$.

Proof We start by bounding the median by the expectation value of g as follows

$$
\mathbb{E}g = \int_{g \geq m(g)} g(\psi)\mu(d\psi) + \int_{g \leq m(g)} g(\psi)\mu(d\psi) \geq m(g) \int_{g \geq m(g)} \mu(d\psi) = \frac{m(g)}{2}.
\tag{66}
$$

We proceed by lower bounding the expectation value of $g(|\psi\rangle)$,

$$
(\mathbb{E}g)^2 \leq \mathbb{E}(g^2) = \mathbb{E}\left(\mathrm{tr}((\psi^B)^2)\right) - \frac{1}{|B|} = tr\left(\mathbb{E}\left(\psi^{AB} \otimes \psi^{A'B'}\right) \mathbb{I}_{AA'} \otimes \mathbb{F}^{BB'}\right) - \frac{1}{|B|},
\tag{67}
$$

where $\mathbb{F}^{BB'}$ is the swap operator the two systems BB'. The first inequality of the equation above follows from the convexity of x^2. From Schur's Lemma,

$$
\mathbb{E}\left(\psi^{AB} \otimes \psi^{A'B}\right) = \frac{\mathbb{I}^{AA'BB'} + \mathbb{F}^{AA'} \otimes \mathbb{F}^{BB'}}{|A||B|(|A||B| + 1)}.
\tag{68}
$$

Putting Eqs. (67) and (68) together gives $m(g) \leq 2|A|^{-\frac{1}{2}}$. $\qquad \square$

The final lemma of this section shows the monotonicity of the operator and two norms under pinching.

Lemma V.4 *For every X,*

$$\|X\|_2 \geq \left\|\sum_k P_k X P_k\right\|_2, \quad \|X\|_\infty \geq \left\|\sum_k P_k X P_k\right\|_\infty, \qquad (69)$$

for orthogonal projectors P_k with $\sum_k P_k = 1$.

Proof Direct calculation. $\qquad \square$

VI. ACKNOWLEDGEMENT

APPENDIX A: A FEW PROBABILITY FACTS

Lemma A.1 *For two events M, N, $\Pr(M \cap N) \geq \Pr(M) - \Pr(N^c)$, where N^c is the complement of N.*

Proof We have

$$\Pr(M) = \Pr(M \cap N) + \Pr(M \cap N^c) \leq Pr(M \cap N) + \Pr(N^c). \qquad (A1)$$

Rearranging terms in the equation above gives the result of the lemma. $\qquad \square$

Lemma A.2 *Let $|\chi\rangle \in A$ be drawn from the Haar measure. Write*

$$|\chi\rangle = \sqrt{x}|\psi\rangle + \sqrt{1-x}|\phi\rangle, \qquad (A2)$$

where $|\psi\rangle \in A$ is a fixed state, $x = |\langle\psi|\chi\rangle|^2$, and $|\phi\rangle$ is a state orthogonal to $|\psi\rangle$. Then x and $|\phi\rangle$ are independent random variables and $|\phi\rangle$ is distributed accordingly to the Haar measure in the subspace of A orthogonal to $|\psi\rangle$.

Proof Let $p_A(|\psi\rangle)$ be the probability density function associated with the Haar measure in A. We can write $p_A(|\psi\rangle) = p_A(x, |\phi\rangle)$. From the invariance of the Haar measure under unitary transformations, $p_A(U|\psi\rangle) = p_A(x, U|\phi\rangle)$, for every x and every unitary U which acts non-trivially only in the subspace of A orthogonal to $|\psi\rangle$, $A_{\psi^\perp}$. Therefore, the conditional probability density function

$$p_A(|\phi\rangle\,|\,x) = \frac{p_A(|\phi\rangle, x)}{p_A(x)} \qquad (A3)$$

is such that $p_A(U|\phi\rangle\,|\,x) = p_A(|\phi\rangle\,|\,x)$ for every x and unitary U acting on $A_{\psi^\perp}$. From the uniqueness of the Haar measure, we find that for every x, $p_A(|\phi\rangle\,|\,x) = p_{A_{\psi^\perp}}(|\phi\rangle)$. This shows both that $|\phi\rangle$ is independent of x and that it is Haar distributed. $\qquad \square$

APPENDIX B: RELATING OPERATOR NORM, TWO NORM, AND ENTROPY

Lemma B.1 *Let $\rho \in \mathcal{D}(\mathbb{C}^D)$ be such that $\|\rho\|_\infty \geq \lambda > 1/D$. Then*

$$S(\rho) \leq s(\lambda, D) := (1 - \lambda) \log(D - 1) + h(\lambda), \tag{B1}$$

and

$$\|\rho\|_2^2 \geq \lambda^2 + \frac{(1 - \lambda)^2}{D - 1}, \tag{B2}$$

where $h(x) := -x \log x - (1 - x) \log(1 - x)$ is the Shannon binary entropy.

Proof Let λ_i be the eigenvalues of ρ in decreasing order. Then, for every $N \in \{1, ..., D\}$,

$$\sum_{i=1}^N \lambda_i \geq \lambda_1 + (N - 1)\frac{(1 - \lambda_1)}{D - 1}, \tag{B3}$$

which shows that $\{\lambda_i\}$ is majorized by the probability distribution $q := \left\{\lambda_1, \frac{(1-\lambda_1)}{D-1}, ..., \frac{(1-\lambda_1)}{D-1}\right\}$. From the Schur convexity of $x \log x$,

$$S(\rho) \leq S(q) = s(\lambda_1, D) := (1 - \lambda_1) \log(D - 1) + h(\lambda_1). \tag{B4}$$

A simple calculation shows that $\frac{\partial s(\mu)}{\partial \lambda} \leq 0$ for all $\mu \geq 1/D$. Therefore, the function $s(\lambda, D)$ is monotonic decreasing in λ for $\lambda \geq 1/D$. As $\lambda_1 = \|\rho\|_\infty \geq \lambda$, we find that $S(\rho) \leq s(\lambda, D)$.

The bound on the two norm can be obtained in an analogous way. As x^2 is Schur convex, we get that

$$\|\rho\|_2^2 \geq \|q\|_2^2 = r(\lambda_1, D) := \lambda_1^2 + \frac{(1 - \lambda_1)^2}{D - 1}. \tag{B5}$$

A simple calculation shows that $r(\lambda_1, D)$ is monotonic increasing in λ_1, so that $r(\lambda_1, D) \geq r(\lambda, D)$.
□

APPENDIX C: LARGE DEVIATION BOUND FOR THE OPERATOR NORM

The following lemma, due to Harrow, Hayden, and Leung [35] is used twice in the proof of Proposition II.7.

Lemma C.1 *(Lemma III.4 of [36]) Let $|\psi^{AB}\rangle \in A \otimes B$ be drawn from the Haar measure. For every $0 < \varepsilon < 1$,*

$$\Pr_\psi \left(\|\psi^B\|_\infty \geq \frac{1}{|B|} + \frac{\varepsilon}{|B|} \right) \leq \left(\frac{10|B|}{\varepsilon} \right)^{2|B|} \exp\left(-|A| \frac{\varepsilon^2}{14 \ln(2)} \right), \tag{C1}$$

while for every $\varepsilon > 0$ [39]

$$\Pr_\psi \left(\|\psi^B\|_\infty \geq \frac{1}{|B|} + \frac{\varepsilon}{|B|} \right) \leq \left(\frac{10|B|}{\varepsilon} \right)^{2|B|} \exp\left(-|A| \frac{(\varepsilon - \log(1 + \varepsilon))}{14 \ln(2)} \right), \tag{C2}$$

APPENDIX D: PROOF OF LEMMA II.1

Following Refs. [28, 30], we use the canonical maximally entangled state $|\Phi^{AA'}\rangle :=$ $|A|^{-1}\sum_{i=1}^{|A|}|i\rangle^A|i\rangle^{A'}$ as an input state to

$$\mathcal{E}\otimes\overline{\mathcal{E}}(\rho) = \mathrm{tr}_{AA'}\left((U\otimes U^*)\left(\rho^{AA'}\otimes|0\rangle\langle0|^B\otimes|0\rangle\langle0|^{B'}\right)(U\otimes U^*)^\dagger\right), \tag{D1}$$

where U acts on AB and U^* on $A'B'$.

We can get a lower bound on the operator norm of $\mathcal{E}\otimes\overline{\mathcal{E}}(\Phi^{AA'})$ as follows

$$
\begin{aligned}
\left\|\mathcal{E}\otimes\overline{\mathcal{E}}(\Phi^{AA'})\right\|_\infty &\geq \mathrm{tr}\left(\Phi^{BB'}\mathcal{E}\otimes\overline{\mathcal{E}}(\Phi^{AA'})\right) \\
&= \mathrm{tr}\left(\Phi^{BB'}\mathrm{tr}_{AA'}\left((U\otimes U^*)\left(\Phi^{AA'}\otimes|0\rangle\langle0|^B\otimes|0\rangle\langle0|^{B'}\right)(U\otimes U^*)^\dagger\right)\right) \\
&= \mathrm{tr}\left(\mathbb{I}^{AA'}\otimes\Phi^{BB'}\left((U\otimes U^*)\left(\Phi^{AA'}\otimes|0\rangle\langle0|^B\otimes|0\rangle\langle0|^{B'}\right)(U\otimes U^*)^\dagger\right)\right) \\
&\overset{(i)}{\geq} \mathrm{tr}\left(\Phi^{AA'}\otimes\Phi^{BB'}\left((U\otimes U^*)\left(\Phi^{AA'}\otimes|0\rangle\langle0|^B\otimes|0\rangle\langle0|^{B'}\right)(U\otimes U^*)^\dagger\right)\right) \\
&= \mathrm{tr}\left((U\otimes U^*)^\dagger\Phi^{AA'}\otimes\Phi^{BB'}(U\otimes U^*)\left(\Phi^{AA'}\otimes|0\rangle\langle0|^B\otimes|0\rangle\langle0|^{B'}\right)\right) \\
&\overset{(ii)}{=} \mathrm{tr}\left(\Phi^{AA'}\otimes\Phi^{BB'}\left(\Phi^{AA'}\otimes|0\rangle\langle0|^B\otimes|0\rangle\langle0|^{B'}\right)\right)\geq\frac{1}{|B|}. \tag{D2}
\end{aligned}
$$

In (i) we used $\Phi^{BB'} \leq \mathbb{I}$, while (ii) follows from the identity $\left(\mathbb{I}^C\otimes X^{C'}\right)|\Phi^{CC'}\rangle = \left((X^C)^T\otimes\mathbb{I}^{C'}\right)|\Phi^{CC'}\rangle$.

Applying Lemma B.1 to $\mathcal{E}\otimes\overline{\mathcal{E}}(\Phi^{AA'})$, with $D = |B|^2$ and $\lambda = |B|^{-1}$ then gives

$$S\left(\mathcal{E}\otimes\overline{\mathcal{E}}(\Phi^{AA'})\right) \leq s(|B|^{-1},|B|^2) = 2\log|B| - \frac{\log|B|}{|B|}. \tag{D3}$$

[1] A.S. Holevo. The capacity of the quantum channel with general signal states. IEEE Trans. Inf. Theo. **44**, 269 (1998).
[2] B. Schumacher and M.D. Westmoreland. Sending classical information bia noisy quantum channels. Phys. Rev. A **56**, 131 (1997).
[3] A.S. Holevo. Information theoretical aspects of quantum measurements. Probl. Info. Transm. **9**, 177 (1973).
[4] G.G. Amosov, A.S. Holevo, and R.F. Werner. On some addivitity problems in quantum information theory. Probl. Inform. Transm. **36**, 25 (2000).
[5] C. King. Additivity for unital qubit channels. J. Math. Phys. **43** 4641-4653 (2002).
[6] P.W. Shor. Additivity of the Classical Capacity of Entanglement-Breaking Quantum Channels. J. Math. Phys. Vol. **43**, 4334 (2002).
[7] C. King. The capacity of the quantum depolarizing channel. IEEE Trans. Info. Theory 49, 221-229 (2003).
[8] R. Alicki. Isotropic quantum spin channels and additivity questions. arXiv:quant-ph/0402080.
[9] N. Datta. Multiplicativity of p-norms in Werner-Holevo channels for $1 < p < 2$. arXiv:quant-ph/0410063, 2004.
[10] Nilanjana Datta, Alexander S. Holevo, Yuri Suhov. Additivity for transpose depolarizing channels. arXiv:quant-ph/0412034.

[11] K. Matsumoto, F. Yura. Entanglement Cost of Antisymmetric States and Additivity of Capacity of Some Quantum Channel. J. Phys. A: Math. Gen. **37**, 167 (2004).

[12] R. Alicki, M. Fannes. Note on multiple additivity of minimal Renyi entropy output of the Werner-Holevo channels. arXiv:quant-ph/0407033

[13] C. King, K. Matsumoto, M. Nathason, and M.B. Ruskai. Properties of conjugate channels with applications to addivity and multiplicativity. Markov Process and Related Fields **13**, 391 (2007).

[14] N. Datta and M.B. Ruskai. Maximal output purity and capacity for asymmetric unital qudit channels. J. Phys. A: Math. Gen. **3**, 9785 (2005).

[15] M.M. Wolf and J. Eisert. Classical information capacity of a class of quantum channels. New J. Phys. **7**, 93 (2005).

[16] M. Fukuda and M.W. Wolf. Simplifying additivity problems using direct sum constructions. J. Math. Phys. **48**, 072101 (2007).

[17] F.G.S.L. Brandao, M. Horodecki, M.B. Plenio, S. Virmani. Remarks on the equivalence of full additivity and monotonicity for the entanglement cost. Open Sys. Inf. Dyn. **14**, 333 (2007).

[18] P.M. Hayden, M. Horodecki and B.M. Terhal. The asymptotic entanglement cost of preparing a quantum state. J. Phys. A: Math. Gen. **34**, 6891 (2001).

[19] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin and W.K. Wootters. Mixed State Entanglement and Quantum Error Correction. Phys. Rev. A **54**, 3824 (1996).

[20] P.W. Shor. Equivalence of Additivity Questions in Quantum Information Theory. Comm. Math. Phys. **246**, 453 (2004).

[21] K. Matsumoto, T. Shimono and A. Winter. Remarks on additivity of the Holevo channel capacity and of the entanglement of formation. Comm. Math. Phys. **246**, 427 (2004).

[22] K.M.R. Audenaert and S.L. Braunstein. On Strong Superadditivity of the Entanglement of Formation. Comm. Math. Phys. **243**, 443 (2004).

[23] A. Pomeransky. Strong superadditivity of the entanglement of formation follows from its additivity. Phys. Rev. A **68**, 032317 (2003).

[24] I. Devetak and A. Winter. Distilling common randomness from bipartite quantum states. quant-ph/0304196.

[25] L. Henderson and V. Vedral. J. Phys. A **34**, 6899 (2001).

[26] M. Koashi and A. Winter. Monogamy of entanglement and other correlations. Phys. Rev. A **69**, 022309 (2004).

[27] M.B. Hastings. Superadditivity of communication capacity using entangled inputs. Nature Physics **5**, 255 (2009).

[28] P. Hayden. The maximal p-norm multiplicativity conjecture is false. arXiv:0707.3291.

[29] A. Winter. The maximum output p-norm of quantum channels is not multiplicative for any $p > 2$. arXiv:0707.0402.

[30] P. Hayden and A. Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$. Comm. Math. Phys. **284**, 263 (2008).

[31] B. Collins and I. Nechita. Random quantum channels I: graphical calculus and the Bell state phenomenon. arXiv:0906.1877.

[32] B. Collins and I. Nechita. Random quantum channels II: Entanglement of random subspaces, Renyi entropy estimates and additivity problems. arXiv:0905.2313.

[33] S. Lloyd and H. Pagels. Complexity as thermodynamic depth. Ann. Phys. **188**, 186 (1988).

[34] M. Fukuda, C. King, and D. Moser. Comments on Hastings' Additivity Counterexamples. arXiv:0905.3697.

[35] A. Harrow, P. Hayden, D. Leung. Superdense coding of quantum states. Phys. Rev. Lett. **92**, 187901 (2004).

[36] P. Hayden, D.W. Leung, and A. Winter. Aspects of generic entanglement. Comm. Math. Phys. **265**, 95 (2006).

[37] V.D. Milman and G. Schechtman. Asymptotic theory of finite dimensional normed spaces, volume 1200 of Lectures Notes in Mathematics. Springer-Verlag, 1986.

[38] M. Ledoux. The concentration of measure phenomenon, vol. 89 of Mathematical Surveys and Monographs. American Mathematical Society, 2001.

[39] P. Hayden, D. Leung, and G. Smith. Multiparty data hiding of quantum information. Phys. Rev. A **71**, 062339 (2005).

[40] The name *tube* is taken from Ref. [34].