A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels

Nilanjana Datta
Statistical Laboratory, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, United Kingdom

Milán Mosonyi
School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK
and
Department of Analysis, Budapest University of Technology and Economics, Egry József u. 1., Budapest, 1111 Hungary

Min-Hsiu Hsieh
Centre for Quantum Computation & Intelligent Systems (QCIS), Faculty of Engineering and Information Technology (FEIT) University of Technology Sydney (UTS), NSW 2007, Australia

Fernando G.S.L. Brandão
Institute for Theoretical Physics, ETH Zürich, 8093 Zurich, Switzerland

We use the smooth entropy approach to treat the problems of binary quantum hypothesis testing and the transmission of classical information through a quantum channel. We provide lower and upper bounds on the optimal type II error of quantum hypothesis testing in terms of the smooth max-relative entropy of the two states representing the two hypotheses. Using then a relative entropy version of the Quantum Asymptotic Equipartition Property (QAEP), we can recover the strong converse rate of the i.i.d. hypothesis testing problem in the asymptotics. On the other hand, combining Stein's lemma with our bounds, we obtain a stronger (ε-independent) version of the relative entropy-QAEP. Similarly, we provide bounds on the one-shot ε-error classical capacity of a quantum channel in terms of a smooth max-relative entropy variant of its Holevo capacity. Using these bounds and the ε-independent version of the relative entropy-QAEP, we can recover both the Holevo-Schumacher-Westmoreland theorem about the optimal direct rate of a memoryless quantum channel with product state encoding, as well as its strong converse counterpart.

I. INTRODUCTION

Transmission of information through noisy channels is an essential requirement in various information-processing tasks. A channel can be characterized by its capacity, which quantifies the maximum amount of information which can be transmitted reliably per use of the channel. If the sender (Alice) encodes information at a rate less than the capacity, then the receiver (Bob) can recover the information with a probability of error which vanishes asymptotically in the number of uses of the channel. For rates above the capacity, the asymptotic probability of error is bounded away from zero. Another quantity of interest characterizing a channel is its strong converse capacity, which is the rate threshold above which information transmission fails with certainty, in the sense that the asymptotic probability of error is equal to one. Wolfowitz [49] proved that, for a memoryless classical channel, i.e., a classical channel for which there are no correlations in the noise acting on successive inputs, the strong converse capacity is equal to the capacity. This is referred to as the strong converse property (see e.g. [16]). The capacity of the channel hence provides a sharp threshold on its information-carrying power.

The classical capacity of memoryless quantum channels (with weak converse) was shown to be equal to the regularized Holevo capacity in [22, 39]. If codewords are restricted to product state inputs then regularization is not necessary, and the capacity is equal to the single-shot Holevo capacity of the channel. Moreover, in this case the strong converse property holds, as was proved independently by Ogawa and Nagaoka [34], and by Winter [48]. It is known that the classical capacity with general inputs can be strictly larger than the product-state capacity [17], and it is an open question whether the strong converse property still holds in this case. The only known result in this direction was obtained recently by König and Wehner [24], who proved the strong converse property for the unconstrained classical capacity of a class of quantum channels for which the Holevo capacity is additive. These include all unital qubit channels, the d-dimensional depolarizing channel and the Werner-Holevo channel.

It is well known that the problems of sending classical information through a quantum channel and binary state discrimination (hypothesis testing) are closely related to each other; in particular, the direct part of the channel coding theorem [22, 39] can be obtained from the direct part of Stein's lemma [18, 36]. Moreover, the direct part of the quantum analogue of fixed-length source compression is an immediate consequence of the direct part of Stein's lemma. The optimal asymptotic direct and strong converse rates coincide for binary state discrimination in the i.i.d. case, and are equal to the relative entropy of the two states [20, 35].

Coding theorems are typically obtained in two steps:

1. Establishing a trade-off relation between the rate and the error for finite n (where n denotes the number of channel uses, or the number of copies of the states in the above examples). These trade-off relations are given in terms of some entropic quantities.

2. Evaluating the asymptotics of these entropic quantities in the $n \to \infty$ limit to obtain the limiting optimal rate.

One standard way to do this is to use Rényi relative entropies or related quantities for the trade-off relations, and obtain the asymptotics by using additivity properties of these quantities (see, e.g., [1, 19, 24, 31, 33–35]), or by expressing the asymptotic rate as a regularized entropic quantity. Another approach, which has gained a lot of popularity recently, is to use smooth entropies and related quantities. Smooth entropies for quantum information theory were introduced in [38], and their theory further developed in [11, 12, 42–44]. Smooth entropies interpolate between the operational and the entropic sides of the coding problems, and hence provide a different insight into these poblems compared to the Rényi entropy approach. Due to this interpolation property, the finite-size trade-off relations are typically more straightforward to obtain, and the asymptotic results follow by the application of robust, problem-independent techniques, like the so-called Quantum Asymptotic Equipartition Property (QAEP) [42, 44]. This not only enables a unified treatment of many coding problems, but the techniques developed on the way provide a new set of tools to attack such problems; see, for instance the recent result about the strong converse of the quantum capacity of degradable channels [27].

In this paper, we show how the smooth entropy approach can be used to obtain the direct and the strong converse capacities for classical information transmission through memoryless quantum channels with product encoding. The structure of the paper is as follows. In Section II, we give the necessary technical background on smooth relative entropies. In Section III, we derive two-sided bounds on the optimal type II error of quantum hypothesis testing in terms of the smoothed max-relative entropy of the two states representing the two hypotheses (Theorem 11). Using a suitable version of the QAEP

(Corollary 9), we can derive the strong converse rate for the asymptotic hypothesis testing problem with i.i.d. hypotheses (Theorem 13). On the other hand, when combined with a recent result on finite-size corrections in Stein's lemma [2], the bounds of Theorem 11 yield a strengthening (an ε-independent version) of Corollary 9 (Theorem 14). (A similar result appeared recently in [45], after the submission of the first version of this paper.) In Section IV, we consider a slightly more general channel model than usual quantum channels, and give two-sided bounds on the ε-error capacity for one single use of such a channel. These bounds are given in terms of a generalization of the Holevo capacity, where the correlations are measured by the smooth max-relative entropy instead of the usual relative entropy. Using Theorem 14, we show in Section V that these bounds are asymptotically tight in the sense that one can recover from them the asymptotic direct and strong converse capacities of a memoryless channel. In particular, we get the direct and strong converse capacities of a quantum channel with product-state encoding. We conclude in Section VI by comparing our results to related results in the literature.

II. PRELIMINARIES

For a Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})$ denote the algebra of linear operators acting on $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})_+$ denote the set of positive semi-definite operators on $\mathcal{H}$, and let $\mathcal{D}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})_+$ denote the set of density matrices (or states), i.e., positive semi-definite operators of unit trace. Unless otherwise stated, we assume all Hilbert spaces to be finite-dimensional.

For self-adjoint operators $A, B \in \mathcal{B}(\mathcal{H})$, let $\{A \geq B\}$ denote the spectral projection of $A - B$ corresponding to the interval $[0, +\infty)$; the spectral projections $\{A > B\}$, $\{A \leq B\}$ and $\{A < B\}$ are defined in a similar way. For a self-adjoint operator $A \in \mathcal{B}(\mathcal{H})$, we use the notations $A_+ = A\{A > 0\}$ and $A_- = A\{A < 0\}$ for its positive and negative parts, respectively. Note that for any self-adjoint $A, X \in \mathcal{B}(\mathcal{H})$ such that $0 \leq X \leq I$, we have

$$\operatorname{Tr} XA = \operatorname{Tr} XA_+ - \operatorname{Tr} XA_- \leq \operatorname{Tr} XA_+ \leq \operatorname{Tr} A_+. \tag{1}$$

The *trace distance* between two operators A and B is given by

$$\|A - B\|_1 := \operatorname{Tr} |A - B| = \operatorname{Tr} \left((A - B)_+ + (A - B)_- \right).$$

For $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$, their *fidelity* is

$$F(\rho, \sigma) := \operatorname{Tr} \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = \max_{\psi_\rho, \psi_\sigma} |\langle \psi_\rho, \psi_\sigma \rangle|,$$

where the last expression is due to Uhlmann's theorem [46], and the maximum is taken over all purifications ψ_ρ, ψ_σ of ρ and σ, respectively. It is easy to see that

$$d_{\mathrm{op}}(\rho, \sigma) := \min_{\psi_\rho, \psi_\sigma} \frac{1}{2} \left\| |\psi_\rho\rangle\langle\psi_\rho| - |\psi_\sigma\rangle\langle\psi_\sigma| \right\|_1 = \sqrt{\left(\operatorname{Tr} \rho + \operatorname{Tr} \sigma\right)^2 / 4 - F(\rho, \sigma)^2},$$

and that d_{op} is a metric on $\mathcal{B}(\mathcal{H})_+$. The same arguments as in [14, 46] (see also [32]) yield that

$$\frac{d_{\mathrm{op}}(\rho, \sigma)^2}{\operatorname{Tr} \rho + \operatorname{Tr} \sigma} \leq \frac{1}{2} \left(\operatorname{Tr} \rho + \operatorname{Tr} \sigma \right) - \sqrt{\left(\operatorname{Tr} \rho + \operatorname{Tr} \sigma\right)^2 / 4 - d_{\mathrm{op}}(\rho, \sigma)^2} \leq \frac{1}{2} \left\| \rho - \sigma \right\|_1 \leq d_{\mathrm{op}}(\rho, \sigma)$$

(where the first expression should be replaced by 0 if $\rho = \sigma = 0$). For density operators $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, the above expressions simplify to

$$d_{\mathrm{op}}(\rho, \sigma) = \sqrt{1 - F^2(\rho, \sigma)}, \qquad \text{and} \qquad \frac{1}{2} d_{\mathrm{op}}(\rho, \sigma)^2 \leq 1 - F(\rho, \sigma) \leq \frac{1}{2} \left\| \rho - \sigma \right\|_1 \leq d_{\mathrm{op}}(\rho, \sigma). \tag{2}$$

The distance $d_s(\rho, \sigma) := \sqrt{1 - F^2(\rho, \sigma)}$ on density operators was introduced in [15] under the name *sine distance*, and our definition provides a natural extension of it to the set of positive semidefinite operators. The sine distance was extended to a metric on subnormalized states in a different way under the name *purified distance* in [43]. To distinguish it from the purified distance while reflecting the fact that it is the minimal distance of purifications, we will use the terminology *distance of optimal purifications*.

For $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$ and $\alpha \in (1, +\infty)$, the *Rényi α-relative entropy* of ρ with respect to σ is

$$D_\alpha(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1} \log \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}, & \operatorname{supp}\rho \subseteq \operatorname{supp}\sigma, \\ +\infty, & \text{otherwise.} \end{cases}$$

Here and henceforth logarithms are taken to base 2. It is easily seen that $\alpha \mapsto D_\alpha(\rho\|\sigma)$ is monotone increasing for fixed ρ and σ, and if ρ is a density operator then $\lim_{\alpha \searrow 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$, where $D(\rho\|\sigma)$ is the relative entropy, defined as

$$D(\rho\|\sigma) := \begin{cases} \operatorname{Tr}\rho(\log\rho - \log\sigma), & \operatorname{supp}\rho \subseteq \operatorname{supp}\sigma, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3}$$

The *von Neumann entropy* of a state ρ is given by $S(\rho) = -\operatorname{Tr}(\rho \log \rho)$.

Lemma 1 *Given a state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$, let $\rho_{A(B)} = \operatorname{Tr}_{B(A)} \rho_{AB}$. Then for any operator $\sigma_A \in \mathcal{B}(\mathcal{H}_A)_+$,*

$$\min_{\omega_B \in \mathcal{D}(\mathcal{H}_B)} D(\rho_{AB}\|\sigma_A \otimes \omega_B) = D(\rho_{AB}\|\sigma_A \otimes \rho_B). \tag{4}$$

Proof. Note that the left-hand side of (4) is equal to $+\infty$ if and only if the right-hand side is equal to $+\infty$. When both sides are finite, the assertion follows immediately from $D(\rho_{AB}\|\sigma_A \otimes \omega_B) - D(\rho_{AB}\|\sigma_A \otimes \rho_B) = D(\rho_B\|\omega_B) \geq 0$ (see Lemma 6 in [7] for more details). ■

The notion of the (smoothed) min-entropy was introduced in [38], which in our terminology would correspond to the conditional version of the (smoothed) max-relative entropy. The relative entropy version of the min-entropy has been introduced in [12] under the name of (smoothed) max-relative entropy, which is defined as follows. Note also that various symmetrised versions of the max-relative entropy have been known in operator theory as the Hilbert projective metric [21] and the Thompson distance [41].

Definition 2 *The max-relative entropy of two positive semi-definite operators ρ and σ is defined as*

$$D_{\max}(\rho\|\sigma) := \inf\{\gamma : \rho \leq 2^\gamma \sigma\}.$$

For any $0 \leq \varepsilon \leq 1$, the ε-smooth max-relative entropy of a state ρ and a positive semi-definite operator σ is defined as

$$D^\varepsilon_{\max}(\rho\|\sigma) := \min_{\bar\rho \in B_\varepsilon(\rho)} D_{\max}(\bar\rho\|\sigma),$$

where

$$B_\varepsilon(\rho) := \{\bar\rho \geq 0, \operatorname{Tr}\bar\rho = 1; d_{\mathrm{op}}(\bar\rho, \rho) \leq \varepsilon\} \tag{5}$$

is the ε-ball around ρ with respect to d_{op}.

Remark 3 *There are various, slightly different, ways to define the smoothing in the literature. One common choice is to use the purified distance d_p [43] instead of d_{op}, and allow subnormalized states in the sense of replacing $B_\varepsilon(\rho)$ in (5) with $\tilde{B}_\varepsilon(\rho) = \{\bar\rho \geq 0, \operatorname{Tr}\bar\rho \leq 1; d_p(\bar\rho, \rho) \leq \varepsilon\}$. We give some comments on our choice and its relation to that of [43] in Section VI.*

Note that for $\varepsilon = 0$ we have $D^0_{\max}(\rho\|\sigma) = D_{\max}(\rho\|\sigma)$, the function $\varepsilon \mapsto D^\varepsilon_{\max}(\rho\|\sigma)$ is monotone decreasing, and $D^\varepsilon_{\max}(\rho\|\sigma) = 0$ if and only if $\varepsilon \geq d_{\mathrm{op}}(\rho, \sigma)$. A quantitative bound on the effect of smoothing is given by the following:

Lemma 4 *Let $\rho \in \mathcal{D}(\mathcal{H})$ and $\sigma \in \mathcal{B}(\mathcal{H})_+$ be such that $\rho\sigma \neq 0$, let $\lambda > 0$, and let $\Delta_+(\lambda) := \big(\rho - \lambda\sigma\big)_+$ be the positive part of the operator $(\rho - \lambda\sigma)$. Then*

$$D^{\varepsilon(\lambda)}_{\max}(\rho\|\sigma) \leq \log \frac{\lambda}{\sqrt{1 - \varepsilon(\lambda)^2}}, \qquad \text{where} \qquad \varepsilon(\lambda) := \sqrt{\operatorname{Tr}\Delta_+(\lambda)\,(2 - \operatorname{Tr}\Delta_+(\lambda))}. \tag{6}$$

The function $\lambda \mapsto \operatorname{Tr}\Delta_+(\lambda)$ is convex on the whole real line. If, moreover, $\operatorname{supp}\rho \le \operatorname{supp}\sigma$ then the function $\lambda \mapsto \operatorname{Tr}\Delta_+(\lambda)$ is strictly decreasing and continuous on $\left[0, 2^{D_{\max}(\rho\|\sigma)}\right]$ with range $[0,1]$, and the function

$$\lambda \mapsto \varepsilon(\lambda) = \sqrt{\operatorname{Tr}\Delta_+(\lambda)\left(2 - \operatorname{Tr}\Delta_+(\lambda)\right)}$$

is strictly decreasing and continuous on $\left[0, 2^{D_{\max}(\rho\|\sigma)}\right]$ with range $[0,1]$.

Proof. By definition, $\rho \le \lambda\sigma + \Delta_+(\lambda)$. Using Lemma C.5 in [5] (see also Lemma 5 in [11]), we obtain the existence of a state $\widetilde{\rho}$ such that $\widetilde{\rho} \le (1-\operatorname{Tr}\Delta_+(\lambda))^{-1}\lambda\sigma$ and $F(\rho,\widetilde{\rho}) \ge 1 - \operatorname{Tr}\Delta_+(\lambda)$. By the former, we have $D_{\max}(\widetilde{\rho}\|\sigma) \le \log\lambda(1-\operatorname{Tr}\Delta_+(\lambda))^{-1}$, and by the latter, $\sqrt{1-F^2(\rho,\widetilde{\rho})} \le \varepsilon(\lambda)$. Thus, $\widetilde{\rho} \in B_{\varepsilon(\lambda)}$, and hence $D_{\max}^{\varepsilon(\lambda)}(\rho\|\sigma) \le D_{\max}(\widetilde{\rho}\|\sigma)$, from which (6) follows.

Let $\lambda_0, \lambda_1 > 0$ and let $\lambda_p := (1-p)\lambda_0 + p\lambda_1$ for every $p \in [0,1]$. Then

$$\begin{aligned}
\operatorname{Tr}\left[\Delta_+(p\lambda_0 + (1-p)\lambda_1)\right] &= \operatorname{Tr}\left[\{\rho - \lambda_p\sigma > 0\}(\rho - \lambda_p\sigma)\right] \\
&= \operatorname{Tr}\left[\{\rho - \lambda_p\sigma > 0\}\left[(1-p)(\rho - \lambda_0\sigma) + p(\rho - \lambda_1\sigma)\right]\right] \\
&= (1-p)\operatorname{Tr}\left[\{\rho - \lambda_p\sigma > 0\}(\rho - \lambda_0\sigma) + p\operatorname{Tr}\{\rho - \lambda_p\sigma > 0\}(\rho - \lambda_1\sigma)\right] \\
&\le (1-p)\operatorname{Tr}\left(\Delta_+(\lambda_0)\right) + p\operatorname{Tr}\left(\Delta_+(\lambda_1)\right),
\end{aligned}$$

where the last inequality follows from (1). This proves the assertion on the convexity.

Assume now that $\operatorname{supp}\rho \le \operatorname{supp}\sigma$. Note that $\operatorname{Tr}\Delta_+(\lambda) = 0$ if and only if $\lambda \ge 2^{D_{\max}(\rho\|\sigma)}$. If $\operatorname{supp}\rho \le \operatorname{supp}\sigma$ then $D_{\max}(\rho\|\sigma) < +\infty$, and, since $\operatorname{Tr}\Delta_+(\lambda) > 0$ when $\lambda < 2^{D_{\max}(\rho\|\sigma)}$, convexity yields that $\lambda \mapsto \operatorname{Tr}\Delta_+(\lambda)$ is strictly decreasing on $(-\infty, 2^{D_{\max}(\rho\|\sigma)}]$. In particular, it is strictly decreasing on $\left[0, 2^{D_{\max}(\rho\|\sigma)}\right]$, with range $[0,1]$, and convexity implies that it is also continuous. Since $x \mapsto \sqrt{x(2-x)}$ is strictly increasing and continuous on $[0,1]$ with range $[0,1]$, the statement follows. ■

Remark 5 *Note that $\operatorname{Tr}\Delta_+(\lambda) = 1$ for $\lambda = 0$, and the convexity of $\lambda \mapsto \operatorname{Tr}\Delta_+(\lambda)$ yields that $\lambda \mapsto \frac{\operatorname{Tr}\Delta_+(\lambda)-1}{\lambda} = \frac{-\sqrt{1-\varepsilon(\lambda)^2}}{\lambda}$ is monotone increasing, and hence so is the upper bound in* (6).

Remark 6 *A bound similar to our inequality* (6) *appeared in Lemma 6.1 of [44]. The difference between the two is the extra factor $\sqrt{1-\varepsilon(\lambda)^2}$ in our bound, which is due to our different choice of smoothing.*

Let $\rho_x, \sigma_x \in \mathcal{B}(\mathcal{H})_+$ for every $x \in \mathcal{X}$, where $\mathcal{X}$ is a finite set, let $\{p_x\}_{x\in\mathcal{X}}$ be a probability distribution on $\mathcal{X}$ and let $\{|x\rangle\langle x|\}_{x\in\mathcal{X}}$ be a set of orthonormal rank-1 projections on some Hilbert space $\mathcal{K}$. It follows immediately from Definition 2 that

$$\begin{aligned}
D_{\max}\left(\sum_{x\in\mathcal{X}} p_x\rho_x \middle\| \sum_{x\in\mathcal{X}} p_x\sigma_x\right) &\le \max_{x:\, p_x>0} D_{\max}\left(\rho_x\|\sigma_x\right) \\
&= D_{\max}\left(\sum_{x\in\mathcal{X}} p_x|x\rangle\langle x| \otimes \rho_x \middle\| \sum_{x\in\mathcal{X}} p_x|x\rangle\langle x| \otimes \sigma_x\right). \qquad (7)
\end{aligned}$$

The first inequality says that the max-relative entropy is jointly quasi-convex in its arguments (see also Lemma 9 in [12]). This in turn implies joint quasi-convexity of the ε-smooth max-relative entropy:

Lemma 7 *For any $0 \le \varepsilon \le 1$,*

$$D_{\max}^{\varepsilon}\left(\sum_i \gamma_i\rho_i \middle\| \sum_i \gamma_i\sigma_i\right) \le \max_i D_{\max}^{\varepsilon}(\rho_i\|\sigma_i), \qquad (8)$$

where for each i, $\gamma_i > 0$, ρ_i, σ_i are states, and $\sum_i \gamma_i = 1$.

Proof. For every i, let $\nu_i \in B_\varepsilon(\rho_i)$ such that

$$D_{\max}^{\varepsilon}(\rho_i\|\sigma_i) = D_{\max}(\nu_i\|\sigma_i). \qquad (9)$$

Due to the joint concavity of the fidelity [32], we have

$$F\left(\sum_i \gamma_i \nu_i, \sum_i \gamma_i \sigma_i\right) \geq \sum_i \gamma_i F(\nu_i, \sigma_i) \geq \sum_i \gamma_i \sqrt{1-\varepsilon^2} = \sqrt{1-\varepsilon^2},$$

i.e., $\sum_i \gamma_i \nu_i \in \mathcal{B}_\varepsilon(\sum_i \gamma_i \sigma_i)$. Thus,

$$\begin{aligned} D_{\max}^{\varepsilon}\left(\sum_i \gamma_i \rho_i \middle\| \sum_i \gamma_i \sigma_i\right) &\leq D_{\max}\left(\sum_i \gamma_i \nu_i \middle\| \sum_i \gamma_i \sigma_i\right) \\ &\leq \max_i D_{\max}(\nu_i \| \sigma_i) = \max_i D_{\max}^{\varepsilon}(\rho_i \| \sigma_i), \end{aligned}$$

where in the second inequality we used the quasi-convexity of the max-relative entropy. ■

Lemma 8 and Lemma 10 given below relate the smoothed max-relative entropy to the quantum relative entropy and the α-relative entropies, respectively:

Lemma 8 *Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$ be a bipartite state and let $\sigma_A \in \mathcal{D}(\mathcal{H}_A)$ be such that* $\operatorname{supp} \rho_A \subseteq \operatorname{supp} \sigma_A$. *Then*

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \min_{\omega_n \in \mathcal{D}(\mathcal{H}_B^{\otimes n})} D_{\max}^{\varepsilon}(\rho_{AB}^{\otimes n} \| \sigma_A^{\otimes n} \otimes \omega_n) = D(\rho_{AB} \| \sigma_A \otimes \rho_B).$$

Proof. The assertion follows immediately from Proposition II.1 in [5] by choosing $\mathcal{M}_n = \sigma_A^{\otimes n} \otimes \mathcal{D}(\mathcal{H}_B^{\otimes n})$ for every $n \in \mathbb{N}$, and by taking into account Lemma 1. Note that in [5], it was assumed that $\mathcal{M}_1$ contains a state with full rank, but it is obviously enough to assume that $\mathcal{M}_1$ contains a state with support larger than or equal to that of ρ_{AB}. ■

By choosing system B to be one-dimensional in the above lemma, we obtain the following:

Corollary 9 *For states ρ and σ such that* $\operatorname{supp} \rho \subseteq \operatorname{supp} \sigma$,

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} D_{\max}^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) = D(\rho \| \sigma). \tag{10}$$

A related result was obtained in Theorem 2 in [12], where it was shown that the left-hand side of (10) is equal to the sup spectral divergence rate. A conditional entropy version of the above Theorem was proved in [42], under the name of *fully quantum asymptotic equipartition property*; see also [44].

Lemma 10 *Let $\rho \in \mathcal{D}(\mathcal{H})$, $\sigma \in \mathcal{B}(\mathcal{H})_+$, $\varepsilon \in (0,1)$ and $\alpha \in (1,2]$. Then*

$$D_{\max}^{\varepsilon}(\rho \| \sigma) \leq D_\alpha(\rho \| \sigma) + \frac{1}{\alpha - 1} \log \frac{2}{\varepsilon^2} - \log \sqrt{1-\varepsilon^2} \tag{11}$$

Proof. The proof is exactly analogous to that of Theorem 7 in [42]. Let $\lambda > 0$ be such that $\sqrt{1-\varepsilon^2} = 1 - \operatorname{Tr}(\rho - \lambda\sigma)_+$ (cf. Lemma 4). Following the proof of Theorem 7 in [42], we obtain

$$1 - \sqrt{1-\varepsilon^2} = \operatorname{Tr}(\rho - \lambda\sigma)_+ \leq \lambda^{1-\alpha} \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha},$$

and using Lemma 4 we get

$$D_{\max}^{\varepsilon}(\rho \| \sigma) \leq \log \frac{\lambda}{\sqrt{1-\varepsilon^2}} \leq D_\alpha(\rho \| \sigma) + \frac{1}{\alpha - 1} \log \left(1 - \sqrt{1-\varepsilon^2}\right)^{-1} - \log \sqrt{1-\varepsilon^2}.$$

Using that $\left(1 - \sqrt{1-\varepsilon^2}\right)^{-1} \leq 2/\varepsilon^2$, we finally get (11). ■

III. QUANTUM HYPOTHESIS TESTING

Consider the quantum hypothesis testing problem with the null hypothesis $H_0 : \rho$ versus the alternative hypothesis $H_1 : \sigma$, where ρ and σ are density operators on some finite-dimensional Hilbert space $\mathcal{H}$. We can decide which hypothesis is true based on the POVM $\{\Pi, I - \Pi\}$, where $0 \le \Pi \le I$. For a test Π, the error probability of the first kind (or type I error) and the second kind (or type II error) are defined as

$$\alpha(\Pi) := \mathrm{Tr}[(I - \Pi)\rho], \tag{12}$$
$$\beta(\Pi) := \mathrm{Tr}[\Pi\sigma], \tag{13}$$

respectively, where $\alpha(\Pi)$ is the probability of accepting σ when ρ is true while $\beta(\Pi)$ is the probability of accepting ρ when σ is true. Obviously, there is a trade-off between the two error probabilities, and there are various ways to jointly optimize them. In the asymmetric setting of Stein's lemma [8, 20, 35], the error probability of the second kind is optimized under the constraint that the error probability of the first kind stays below a threshold $\varepsilon \in (0, 1)$; the optimal error of the second kind is then given by

$$\beta_\varepsilon(\rho\|\sigma) := \min\{\beta(\Pi) \,:\, \alpha(\Pi) \le \varepsilon\},$$

where the minimization is over all POVMs $\{\Pi, I - \Pi\}$. In general, there is no closed formula known for $\beta_\varepsilon(\rho\|\sigma)$ or for the optimal POVM attaining it. However, we can give the following bounds in terms of the smoothed max-relative entropy of ρ and σ:

Theorem 11 *Assume that* $\mathrm{supp}\,\rho \subseteq \mathrm{supp}\,\sigma$. *For any* $0 < \varepsilon' < \varepsilon < 1$,

$$D_{\max}^{g(\varepsilon)}(\rho\|\sigma) \le -\log \beta_{1-\varepsilon}(\rho\|\sigma) \le D_{\max}^{\varepsilon'}(\rho\|\sigma) + \log\frac{1}{\varepsilon - \varepsilon'},$$

where $g(\varepsilon) := \sqrt{\varepsilon(2 - \varepsilon)}$.

Proof. *(Upper bound)* Let $\varepsilon \in (0, 1)$ be fixed. The assertion will follow if we can show that for any $0 \le \Pi \le I$ such that

$$\log \beta(\Pi) < -D_{\max}^{\varepsilon'}(\rho\|\sigma) - \log\frac{1}{\varepsilon - \varepsilon'} \tag{14}$$

we have

$$\alpha(\Pi) > 1 - \varepsilon.$$

Thus, let Π be such that (14) holds. By the definition of $D_{\max}^{\varepsilon'}(\rho\|\sigma)$, there exists a state $\bar\rho \in B_{\varepsilon'}(\rho)$ for which

$$\bar\rho \le 2^{D_{\max}^{\varepsilon'}(\rho\|\sigma)}\sigma, \tag{15}$$

and therefore

$$\begin{aligned}
\mathrm{Tr}\,\Pi\bar\rho &\le 2^{D_{\max}^{\varepsilon'}(\rho\|\sigma)}\,\mathrm{Tr}(\Pi\sigma) \\
&= 2^{D_{\max}^{\varepsilon'}(\rho\|\sigma)}\beta(\Pi) \\
&< 2^{D_{\max}^{\varepsilon'}(\rho\|\sigma)}2^{-D_{\max}^{\varepsilon'}(\rho\|\sigma)+\log(\varepsilon-\varepsilon')} \\
&= \varepsilon - \varepsilon'.
\end{aligned} \tag{16}$$

The first inequality follows from (15), and the second inequality follows from (14). Hence,

$$1 - \alpha(\Pi) = \mathrm{Tr}(\Pi\rho) = \mathrm{Tr}(\Pi\bar\rho) + \mathrm{Tr}(\Pi(\rho - \bar\rho)) < \varepsilon - \varepsilon' + \|\rho - \bar\rho\|_1/2 \le \varepsilon,$$

where the first inequality follows from (16) and the second inequality holds because $\bar\rho \in B_{\varepsilon'}(\rho)$.

(Lower bound) By Lemma 4, there exists a $\lambda > 0$ such that $\mathrm{Tr}(\rho - \lambda\sigma)_+ = \varepsilon$. For this λ, let $\Pi := \{\rho \ge \lambda\sigma\}$. Then

$$\mathrm{Tr}\,\Pi\rho \ge \mathrm{Tr}\,\Pi(\rho - \lambda\sigma) = \mathrm{Tr}(\rho - \lambda\sigma)_+ = \varepsilon,$$

or equivalently, $\alpha(\Pi) \leq 1-\varepsilon$, and hence

$$-\log \beta_{1-\varepsilon}(\rho\|\sigma) \geq -\log \beta(\Pi).$$

On the other hand, $\varepsilon = \operatorname{Tr}(\rho-\lambda\sigma)_+ = \operatorname{Tr}\Pi(\rho-\lambda\sigma) \leq 1-\lambda \operatorname{Tr}\Pi\sigma$ yields

$$\beta(\Pi) = \operatorname{Tr}\Pi\sigma \leq \frac{1-\varepsilon}{\lambda},$$

and hence,

$$-\log \beta(\Pi) \geq \log\lambda - \log(1-\varepsilon) \geq D_{\max}^{g(\varepsilon)}(\rho\|\sigma) + \log\sqrt{1-g(\varepsilon)^2} - \log(1-\varepsilon) = D_{\max}^{g(\varepsilon)}(\rho\|\sigma),$$

where we have used Lemma 4. ■

The above bounds, combined with Corollary 9, can be used to derive the strong converse theorem for hypothesis testing:

Definition 12 *The asymptotic strong converse rate R_{sc} of the quantum hypothesis testing problem for the null hypothesis $H_0 : \rho$ versus the alternative hypothesis $H_1 : \sigma$ is defined to be the smallest number R such that if*

$$\limsup_{n\to\infty} \frac{1}{n} \log \operatorname{Tr}\Pi_n \sigma^{\otimes n} \leq -R$$

for some sequence of tests $\{\Pi_n\}_{n\in\mathbb{N}}$ then

$$\lim_{n\to\infty} \operatorname{Tr}(I_n - \Pi_n)\rho^{\otimes n} = 1.$$

Theorem 13 ([35]) *The asymptotic strong converse rate R_{sc} of the quantum hypothesis testing problem for the null hypothesis $H_0 : \rho$ versus the alternative hypothesis $H_1 : \sigma$ is given by*

$$R_{sc} = D\left(\rho\|\sigma\right), \tag{17}$$

where $D\left(\rho\|\sigma\right)$ is the quantum relative entropy (3)*.*

Proof. It is easy to see that

$$R_{sc} = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \beta_{1-\varepsilon}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right).$$

The assertion then follows from Theorem 11 and Corollary 9. ■

In Theorem 11 we have derived bounds on the optimal type II error in terms of the smoothed max-relative entropy, and used Corollary 9 to obtain the strong converse rate for Stein's lemma. Proceeding the other way around, we can use the bounds of Theorem 11 together with a recent result from [2], to obtain a significantly stronger version of Corollary 9. Indeed, the bounds in Theorem 11 can be rewritten as

$$-\log \beta_{1-\varepsilon'}(\rho\|\sigma) - \log\frac{1}{\varepsilon'-\varepsilon} \leq D_{\max}^{\varepsilon}(\rho\|\sigma) \leq -\log \beta_{\sqrt{1-\varepsilon^2}}(\rho\|\sigma) \tag{18}$$

for every $0 < \varepsilon < \varepsilon' < 1$. Theorem 3.3 in [2] says that for every $\varepsilon \in (0,1)$ and $n \in \mathbb{N}$,

$$D\left(\rho\|\sigma\right) - \frac{f_1(\varepsilon)}{\sqrt{n}} \leq -\frac{1}{n}\log \beta_{1-\varepsilon}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) \leq D\left(\rho\|\sigma\right) + \frac{f_2(\varepsilon)}{\sqrt{n}},$$

where $f_1(\varepsilon), f_2(\varepsilon) > 0$ are defined as $f_1(\varepsilon) := 4\sqrt{2}\log(1-\varepsilon)^{-1}\log\eta$, $f_2(\varepsilon) := 4\sqrt{2}\log\varepsilon^{-1}\log\eta$ and $\eta := 1 + \operatorname{Tr}\rho^{3/2}\sigma^{-1/2} + \operatorname{Tr}\rho^{1/2}\sigma^{1/2}$. Comparing it with (18), we obtain the following:

Theorem 14 *For every $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ such that $\operatorname{supp}\rho \leq \operatorname{supp}\sigma$, every $0 < \varepsilon < \varepsilon' < 1$, and every $n \in \mathbb{N}$, we have*

$$\frac{1}{n} D^{\varepsilon}_{\max}\left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) \leq D\left(\rho \| \sigma\right) + \frac{1}{\sqrt{n}} 4\sqrt{2} (\log \eta) \log(1 - \sqrt{1-\varepsilon^2})^{-1}, \tag{19}$$

$$\frac{1}{n} D^{\varepsilon}_{\max}\left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) \geq D\left(\rho \| \sigma\right) - \frac{1}{\sqrt{n}} 4\sqrt{2} (\log \eta) \log(1 - \varepsilon')^{-1} - \frac{1}{n} \log \frac{1}{\varepsilon' - \varepsilon}. \tag{20}$$

In particular,

$$\lim_{n\to\infty} \frac{1}{n} D^{\varepsilon}_{\max}\left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) = D\left(\rho \| \sigma\right) \qquad \text{for every } \ \varepsilon \in (0,1). \tag{21}$$

Remark 15 *An analogy of the upper bound* (19) *has been obtained before in [42] for conditional entropies, and it was extended to relative entropies in [44], where the upper bound*

$$\frac{1}{n} \tilde{D}^{\varepsilon}_{\max}\left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) \leq D\left(\rho \| \sigma\right) + \frac{1}{\sqrt{n}} 4 (\log \eta) \sqrt{\log(1 - \sqrt{1-\varepsilon^2})^{-1}}, \tag{22}$$

was obtained for all $n \geq \frac{8}{5} \log(1 - \sqrt{1-\varepsilon^2})^{-1}$. Here, $\tilde{D}^{\varepsilon}_{\max}$ is the smoothed max-relative entropy according to the smoothing convention of [43]; see Section VI for its definition and its relation to our $D^{\varepsilon}_{\max}$. The difference between the two definitions yields a correction of order $1/n$, which is negligible compared to the $1/\sqrt{n}$ term. Note that the $\log(1 - \sqrt{1-\varepsilon^2})^{-1}$ is under the square root in (22)*, which is better than in* (19) *when $\varepsilon < \sqrt{3}/2$ and worse for $\varepsilon > \sqrt{3}/2$. On the other hand,* (19) *holds for every $n \in \mathbb{N}$, while* (22) *only holds for large enough n, depending on ε.*

The exact second order asymptotics of $-\log \beta_\varepsilon(\rho^{\otimes n} \| \sigma^{\otimes n})$, i.e, the limit

$$\lim_{n\to+\infty} \sqrt{n} \left(-\frac{1}{n} \log \beta_\varepsilon(\rho^{\otimes n} \| \sigma^{\otimes n}) - D(\rho \| \sigma) \right)$$

has been evaluated very recently in [26], and independently in [45]. For large n, this yields sharper bounds than the ones in Theorem 14. The advantage of the bounds in Theorem 14 is, however, that they hold for every $n \in \mathbb{N}$, and hence they provide easily computable bounds for any finite value of n.

The limit relation (21) *has also been obtained in the recent paper [45].*

IV. ONE-SHOT CAPACITY FOR TRANSMISSION OF CLASSICAL INFORMATION

A quantum channel is usually defined as a CPTP (completely positive and trace-preserving) linear map from $\mathcal{D}(\mathcal{H}_A)$ to $\mathcal{D}(\mathcal{H}_B)$, where $\mathcal{H}_A$ and $\mathcal{H}_B$ are (finite-dimensional) Hilbert spaces. Here we consider a more general channel model, where by a *channel* W we mean a map $W : \mathcal{X} \to \mathcal{D}(\mathcal{H}_B)$, where $\mathcal{H}_B$ is a finite-dimensional Hilbert space, and $\mathcal{X}$ is an arbitrary set, with no particular assumption on its cardinality or any mathematical structure. Obviously, usual quantum channels form a special subclass of this channel model, where the input set $\mathcal{X}$ is chosen to be the state space of some finite-dimensional Hilbert space, and W is assumed to be linear and CPTP. The channel $W : \mathcal{X} \to \mathcal{D}(\mathcal{H}_B)$ is *classical* if its image $\operatorname{ran} W := \{W(x)\}_{x\in\mathcal{X}}$ is a commutative subset of $\mathcal{B}(\mathcal{H}_B)$.

Suppose that Alice (the sender) wants to communicate with Bob (the receiver) using the channel W. To do this, they agree on a finite set of possible messages, labelled by natural numbers from 1 to M. To send the message labelled by $m \in \{1, \ldots, M\}$, Alice has to encode her message into an input signal of the channel, $\varphi(m) \in \mathcal{X}$, and send it through the channel W, resulting in the quantum state $W(\varphi(m))$ at Bob's side. Bob then performs a POVM (positive operator-valued measure) $\Pi := \{\Pi_i\}_{i=1}^M$, and if the outcome corresponding to Π_k happens, he concludes that the message with label k was sent. The probability of this event is $\operatorname{Tr}\left(W(\varphi(m))\Pi_k\right)$. A triple (M, φ, Π), as above is called a *code*. More precisely, a code $\mathcal{C}$ is a triple $\mathcal{C} = (M, \varphi, \Pi)$, where

- $M \in \mathbb{N}$ is the number of possible messages;
- $\varphi : \{1, 2, \cdots, M\} \to \mathcal{X}$ is Alice's encoding of possible messages into input signals of the channel;
- $\Pi := \{\Pi_m\}_{m=1}^M$ (with $\Pi_m \geq 0 \, \forall \, m = 1, 2, \ldots M$, and $\sum_{m=1}^M \Pi_m = I$) is a POVM on $\mathcal{H}_B$, performed by Bob to identify the message (decoding).

The *average error probability* $p_e(\mathcal{C}, W)$ of a code $\mathcal{C} = (M, \varphi, \Pi)$ is defined as

$$p_e(\mathcal{C}, W) := \frac{1}{M} \sum_{i=1}^{M} \left[1 - \operatorname{Tr}\left(W(\varphi(i))\Pi_i\right)\right]. \tag{23}$$

Definition 16 *For a given $\varepsilon > 0$, the **one-shot ε-error capacity**, $C_\varepsilon^{(1)}(W)$, of a channel W is defined as follows:*

$$C_\varepsilon^{(1)}(W) := \sup\{\log M : \exists \mathcal{C} := (M, \varphi, \Pi) \text{ s.t. } p_e(\mathcal{C}, W) \leq \varepsilon\}. \tag{24}$$

Note that it denotes the maximum number of bits that can be transmitted through a single use of the channel with average error probability of at most ε.

Our aim is to give bounds on the above defined operational capacities in terms of entropic quantities. To this end, we will need the notions of the α-capacities and ε-max capacities of a channel, which we define below.

For a set $\mathcal{X}$, let $\mathcal{P}_f(\mathcal{X})$ denote the set of finitely supported probability distributions on $\mathcal{X}$. Note that if $\mathcal{X} = \mathcal{D}(\mathcal{H}_A)$ for some Hilbert space $\mathcal{H}_A$ then specifying a $p \in \mathcal{P}_f(\mathcal{D}(\mathcal{H}_A))$ is equivalent to specifying an ensemble of states $\{\rho_k, p_k\}_{k=1}^r$, where $\rho_k \in \mathcal{D}(\mathcal{H}_A)$, $p_k \geq 0$, $k = 1, \ldots r$, and $p_1 + \ldots + p_r = 1$. For every set $\mathcal{X}$, let $\mathcal{H}_\mathcal{X}$ be a Hilbert space with $\dim \mathcal{H}_\mathcal{X} = |\mathcal{X}|$, and let $\{|x\rangle\}_{x\in\mathcal{X}}$, be an orthonormal basis in $\mathcal{H}_\mathcal{X}$. For any divergence measure $\mathcal{M}$, we define the corresponding capacity $\chi^*_\mathcal{M}(W)$ of a channel $W : \mathcal{X} \to \mathcal{D}(\mathcal{H}_B)$ as

$$\chi^*_\mathcal{M}(W) := \sup_{p\in\mathcal{P}_f(\mathcal{X})} \chi_\mathcal{M}(W, p), \tag{25}$$

$$\text{with} \qquad \chi_\mathcal{M}(W, p) := \inf_{\sigma_B\in\mathcal{D}(\mathcal{H}_B)} \mathcal{M}\left(\rho_{\mathcal{X}B}(p) \| \rho_\mathcal{X}(p) \otimes \sigma_B\right), \tag{26}$$

$$\text{where} \qquad \rho_{\mathcal{X}B}(p) := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x| \otimes W(x), \qquad p \in \mathcal{P}_f(\mathcal{X}), \tag{27}$$

and $\rho_\mathcal{X}(p) := \operatorname{Tr}_B \rho_{\mathcal{X}B}(p) = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|$. Note that $\chi_\mathcal{M}(W, p)$ measures the amount of correlation in the classical-quantum state $\rho_{\mathcal{X}B}(p)$, with respect to the divergence measure $\mathcal{M}$.

In particular, the α-capacities [9, 24, 29] and the ε-max capacities of a channel are defined by choosing $\mathcal{M} = D_\alpha$ and $\mathcal{M} = D^\varepsilon_{\max}$ in (25), respectively. We use the short-hand notations $\chi_\alpha(W, p)$, $\chi^*_\alpha(W)$, $\chi_{\max,\varepsilon}(W, p)$ and $\chi^*_{\max,\varepsilon}(W)$ for the corresponding quantities. A quantity related to our $\chi_{\max,\varepsilon}(W, p)$ appeared in [3], under the name *smooth max-information*. In the case of $\mathcal{M} = D_\alpha$, there is an explicit expression for the infimum in (26), and for the optimal σ_B achieving it; see, e.g., [9, 24, 40].

Lemma 10 yields the following inequality between the ε-max capacity and the α-capacities:

Lemma 17 *For any channel W, any $\varepsilon \in (0, 1)$ and any $\alpha \in (1, 2]$, we have*

$$\chi^*_{\max,\varepsilon}(W) \leq \chi^*_\alpha(W) + \frac{1}{\alpha - 1} \log \frac{2}{\varepsilon^2} - \log\sqrt{1 - \varepsilon^2}.$$

In the limit $\alpha \to 1$, the α-capacities yield the Holevo capacity $\chi^*(W)$ [29, 34]:

$$\lim_{\alpha\to 1} \chi^*_\alpha(W) = \chi^*(W) := \chi^*_D(W) = \sup_{p\in\mathcal{M}_f(\mathcal{X})} D\left(\rho_{\mathcal{X}B}(p) \| \rho_\mathcal{X}(p) \otimes \rho_B(p)\right), \tag{28}$$

where D stands for the relative entropy (3).

The ε-max capacity is quasi-convex as a function of the channel, as is stated in the following lemma.

Lemma 18 *Let $W_i : \mathcal{X} \to \mathcal{D}(\mathcal{H}_B)$ be channels for $i = 1, \ldots, r$, and let $\{\gamma_i\}_{i=1}^r$ be a probability distribution. For every $\varepsilon \in [0, 1]$,*

$$\chi^*_{\max,\varepsilon}\left(\sum_i \gamma_i W_i\right) \leq \max_i \chi^*_{\max,\varepsilon}(W_i). \tag{29}$$

Proof. Let $p \in \mathcal{P}_f(\mathcal{X})$, $\rho^i_{\mathcal{X}B} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x| \otimes W_i(x)$ and $\rho_{\mathcal{X}B} := \sum_i \gamma_i \rho^i_{\mathcal{X}B}$. Note that $\rho_{\mathcal{X}} = \rho^i_{\mathcal{X}} = \sum_x p(x)|x\rangle\langle x|$ for every i. For every i, let $\sigma_i \in \mathcal{D}(\mathcal{H}_B)$ be such that $\chi_{\max,\varepsilon}(W_i, p) = D^\varepsilon_{\max}(\rho^i_{\mathcal{X}B} \| \rho^i_{\mathcal{X}} \otimes \sigma_i)$. Then

$$\begin{aligned}
\chi_{\max,\varepsilon}\left(\sum_i \gamma_i W_i, p\right) &\leq D^\varepsilon_{\max}\left(\rho_{\mathcal{X}B}\middle\|\rho_{\mathcal{X}} \otimes \sum_i \gamma_i \sigma_i\right) \\
&= D^\varepsilon_{\max}\left(\sum_i \gamma_i \rho^i_{\mathcal{X}B}\middle\|\rho_{\mathcal{X}} \otimes \sum_i \gamma_i \sigma_i\right) \\
&\leq \max_i D^\varepsilon_{\max}\left(\rho^i_{\mathcal{X}B} \| \rho_{\mathcal{X}} \otimes \sigma_i\right) \\
&= \max_i D^\varepsilon_{\max}\left(\rho^i_{\mathcal{X}B} \| \rho^i_{\mathcal{X}} \otimes \sigma_i\right) \\
&= \max_i \chi_{\max,\varepsilon}\left(W_i, p\right), \qquad (30)
\end{aligned}$$

where the first inequality is due to the definition (25) and the second is due to Lemma 7. The inequality (29) follows immediately from (30). ∎

After this preparation, we are ready to give the main result of the paper:

Theorem 19 *For any $0 < \varepsilon' < \varepsilon < \varepsilon'' < 1$, the one-shot ε-error capacity of a channel W satisfies the following bounds:*

$$\chi^*_{\max,\sqrt{1-(\varepsilon')^2}}(W) + \log\frac{\varepsilon'(\varepsilon-\varepsilon')^2}{8\varepsilon} \leq C^{(1)}_\varepsilon(W) \leq \chi^*_{\max,1-\varepsilon''}(W) - \log(\varepsilon''-\varepsilon). \qquad (31)$$

Before proving Theorem 19, we give the following corollaries:

Corollary 20 *In the setting of Theorem 19, we have*

$$C^{(1)}_\varepsilon(W) \leq \chi^*_\alpha(W) + \frac{1}{\alpha-1}\log\frac{2}{(1-\varepsilon'')^2} - \log(\varepsilon''-\varepsilon). \qquad (32)$$

Proof. Immediate from the second inequality in (31) and Lemma 17. ∎

Corollary 21 *Let $W_i : \mathcal{X} \to \mathcal{D}(\mathcal{H}_B)$ be channels for $i = 1, \ldots, r$, and let $\{\gamma_i\}_{i=1}^r$ be a probability distribution. For every $0 < \varepsilon < \varepsilon'' < 1$ and every $\alpha \in (1, 2]$,*

$$C^{(1)}_\varepsilon\left(\sum_i \gamma_i W_i\right) \leq \max_i \chi^*_\alpha(W_i) + \frac{1}{\alpha-1}\log\frac{2}{(1-\varepsilon'')^2} - \log(\varepsilon''-\varepsilon).$$

Proof. Immediate from the second inequality in (31) and Lemmas 18 and 17. ∎

To prove the lower bound in (31) we will need the following lemma from [18]:

Lemma 22 *Consider any channel $W : \mathcal{X} \mapsto \mathcal{D}(\mathcal{H}_B)$. For any $\lambda > 0$, $M \in \mathbb{N}$, $p \in \mathcal{P}_f(\mathcal{X})$ and $c > 0$, there exists a code $\mathcal{C} = (M, \varphi, \Pi)$ such that*

$$p_e(\mathcal{C}, W) \leq (1+c)\left(1 - \sum_x p_x \operatorname{Tr}[\{W(x) > \lambda W(p)\}W(x)]\right) + (2 + c + c^{-1})\frac{M}{\lambda},$$

where $W(p) := \sum_x p(x)W(x)$.

The following Proposition yields the lower bound in (31):

Proposition 23 *In the setting of Theorem 19, we have, for any $p \in \mathcal{P}_f(\mathcal{X})$,*

$$C^{(1)}_\varepsilon(W) \geq D^{\sqrt{1-(\varepsilon')^2}}_{\max}(\rho_{\mathcal{X}B}(p) \| \rho_{\mathcal{X}}(p) \otimes \rho_B(p)) + \log\frac{\varepsilon'(\varepsilon-\varepsilon')^2}{8\varepsilon}. \qquad (33)$$

Proof. Let $0 < \varepsilon' < \varepsilon < 1$, let $p \in \mathcal{P}_f(\mathcal{X})$, and $\rho_{\mathcal{X}B} := \rho_{\mathcal{X}B}(p)$ as in (27). To prove the inequality in (33), it is sufficient to prove that there exists a code $\mathcal{C} = (M, \varphi, \Pi)$ such that

$$\log M \geq D + \log \frac{\varepsilon'(\varepsilon - \varepsilon')^2}{8\varepsilon}, \qquad D := D_{\max}^{\sqrt{1-(\varepsilon')^2}}(\rho_{\mathcal{X}B} \| \rho_{\mathcal{X}} \otimes \rho_B), \tag{34}$$

and $p_e(\mathcal{C}, W) \leq \varepsilon$. Note that if the lower bound in (34) is negative then there is nothing to prove, and hence for the rest we assume the contrary.

Let λ be such that $1 - \varepsilon' = \operatorname{Tr} \Delta_+(\lambda)$, where $\Delta_+(\lambda) := \big(\rho_{\mathcal{X}B} - \lambda \rho_{\mathcal{X}} \otimes \rho_B\big)_+$. Then

$$1 - \varepsilon' = \operatorname{Tr} \Delta_+(\lambda) \leq \operatorname{Tr}[\{\rho_{\mathcal{X}B} > \lambda \rho_X \otimes \rho_B\}\rho_{\mathcal{X}B}] = \sum_x p_x \operatorname{Tr}[\{W(x) > \lambda W(p)\}W(x)]. \tag{35}$$

Moreover, Lemma 4 yields that

$$D = D_{\max}^{g(1-\varepsilon')}(\rho_{\mathcal{X}B} \| \rho_{\mathcal{X}} \otimes \rho_B) \leq \log \lambda - \log \sqrt{1 - g(1-\varepsilon')^2} = \log \lambda - \log \varepsilon'. \tag{36}$$

By Lemma 22, for any $c > 0$ and $M \in \mathbb{N}$, there exists a code $\mathcal{C}$ of size M such that

$$\begin{aligned} p_e(\mathcal{C}, W) &\leq (1+c)\left(1 - \sum_x p_x \operatorname{Tr}[\{W(x) > \lambda W(p)\}W(x)]\right) + \frac{(1+c)^2}{c}\frac{M}{\lambda} \\ &\leq (1+c)\varepsilon' + \frac{(1+c)^2}{c} M \frac{2^{-D}}{\varepsilon'}, \end{aligned}$$

where the second inequality follows from the choice of λ. Such a code surely satisfies $p_e(\mathcal{C}, W) \leq \varepsilon$ if the RHS above is upper bounded by ε, or equivalently,

$$M \leq \frac{c}{(1+c)^2} 2^D \varepsilon \varepsilon' - \frac{c}{1+c} 2^D (\varepsilon')^2.$$

The RHS of the above inequality is maximal if $c = \frac{\varepsilon - \varepsilon'}{\varepsilon + \varepsilon'}$, which yields the bound

$$M \leq 2^D \frac{\varepsilon'(\varepsilon - \varepsilon')^2}{4\varepsilon} =: M_\varepsilon.$$

Hence,

$$C_\varepsilon^{(1)}(W) \geq \log \lfloor M_\varepsilon \rfloor \geq \log M_\varepsilon - 1 = D + \log \frac{\varepsilon'(\varepsilon - \varepsilon')^2}{8\varepsilon}.$$

■

Proof of Theorem 19: The lower bound in (31) follows immediately (33), by taking the supremum over $p \in \mathcal{P}(\mathcal{X})$.

To prove the upper bound in (31), fix $0 < \varepsilon < \varepsilon'' < 1$, and define

$$\gamma := \chi^*_{\max, 1-\varepsilon''}(W) = \sup_{p \in \mathcal{P}_f(\mathcal{X})} \chi_{\max, 1-\varepsilon''}(W, p).$$

We need to prove that if $\mathcal{C} = (M, \varphi, \Pi)$ is a code such that $\log M > \gamma - \log(\varepsilon'' - \varepsilon)$ then $p_e(\mathcal{C}, W) > \varepsilon$.

Thus, let $\mathcal{C} = (M, \varphi, \Pi)$ be a code with $\log M > \gamma - \log(\varepsilon'' - \varepsilon)$; then, there exists a $c > 1$ such that

$$\frac{2^\gamma}{M} = \frac{\varepsilon'' - \varepsilon}{c}.$$

Let $x_k = \varphi(k)$, $k = 1, \ldots, M$ be the codewords, and let $\rho_k = W(x_k)$ be the output states of the channel. Let $p \in \mathcal{P}_f(\mathcal{X})$ be the uniform distribution on the codewords, i.e., $p(x) = 1/M$ if $x = x_k$ for some $k = 1, \ldots, M$, and $p(x) = 0$ otherwise. For this p, we have

$$\rho_{\mathcal{X}B} := \rho_{\mathcal{X}B}(p) = \frac{1}{M} \sum_{k=1}^M |x_k\rangle\langle x_k| \otimes \rho_k.$$

Let $0 < \delta < \log c$. By the definition of $\chi_{\max,1-\varepsilon''}(W,p)$, there exist $\bar{\sigma}_B \in \mathcal{D}(\mathcal{H}_B)$ and $\bar{\rho}_{\mathcal{X}B} \in B_{1-\varepsilon''}(\rho_{\mathcal{X}B})$ such that

$$D_{\max}(\bar{\rho}_{\mathcal{X}B}\|\rho_{\mathcal{X}} \otimes \overline{\sigma}_B) \leq \chi_{\max,1-\varepsilon''}(W,p) + \delta \leq \gamma + \delta.$$

Using the definition of $\bar{\rho}_{\mathcal{X}B}$ and (2), we have

$$\frac{1}{2}\left\|\bar{\rho}_{\mathcal{X}B} - \rho_{\mathcal{X}B}\right\|_1 \leq d_{\mathrm{op}}(\bar{\rho}_{\mathcal{X}B}, \rho_{\mathcal{X}B}) \leq 1 - \varepsilon'', \tag{37}$$

and

$$\bar{\rho}_{\mathcal{X}B} \leq 2^{D_{\max}(\bar{\rho}_{\mathcal{X}B}\|\rho_{\mathcal{X}}\otimes\overline{\sigma}_B)}\left(\rho_{\mathcal{X}} \otimes \bar{\sigma}_B\right) \leq 2^{\gamma+\delta}\rho_{\mathcal{X}} \otimes \bar{\sigma}_B = \frac{2^{\gamma+\delta}}{M}\sum_{k=1}^{M}|x_k\rangle\langle x_k| \otimes \bar{\sigma}_B. \tag{38}$$

Let $\hat{\Pi} := \sum_{k=1}^{M}|x_k\rangle\langle x_k| \otimes \Pi_k$, which is a projection on $\mathcal{H}_{\mathcal{X}B}$. Then

$$\begin{aligned}
1 - p_e(\mathcal{C}, W) &= \frac{1}{M}\sum_{k=1}^{M}\mathrm{Tr}(\rho_k\Pi_k) = \mathrm{Tr}\,\rho_{\mathcal{X}B}\hat{\Pi} = \mathrm{Tr}(\rho_{\mathcal{X}B} - \bar{\rho}_{\mathcal{X}B})\hat{\Pi} + \mathrm{Tr}\,\bar{\rho}_{\mathcal{X}B}\hat{\Pi} \\
&\leq \frac{1}{2}\left\|\bar{\rho}_{\mathcal{X}B} - \rho_{\mathcal{X}B}\right\|_1 + \frac{2^{\gamma+\delta}}{M}\sum_{k=1}^{M}\mathrm{Tr}\,\Pi_k\bar{\sigma}_B \\
&\leq 1 - \varepsilon'' + \frac{2^{\gamma+\delta}}{M} < 1 - \varepsilon'' + \varepsilon'' - \varepsilon = 1 - \varepsilon,
\end{aligned}$$

where the first inequality follows from (38), the second from (37), and the last one from the initial assumption on M and the choice of δ. ■

V. FROM ONE-SHOT TO ASYMPTOTICS

In the asymptotic scenario, one considers a sequence of channels $\mathbf{W} := \{W^{(n)}\}_{n\in\mathbb{N}}$, where $W^{(n)} : \mathcal{X}^{(n)} \to \mathcal{D}(\mathcal{H}_B^{(n)})$. A code $\mathcal{C}^{(n)} = (M^{(n)}, \varphi^{(n)}, \Pi^{(n)})$ for $W^{(n)}$ and its average error probability $p_e(\mathcal{C}^{(n)}, W^{(n)})$ are defined the same way as before, i.e., $M^{(n)}$ is a natural number, $\varphi^{(n)} : \{1,\ldots,M^{(n)}\} \to \mathcal{X}^n$ is the encoding map, $\Pi^{(n)} := \{\Pi_i^{(n)}\}_{i=1}^{M_n}$ is the decoding POVM, with each $\Pi_i^{(n)} \in \mathcal{B}(\mathcal{H}_B^{\otimes n})$, and

$$p_e(\mathcal{C}^{(n)}, W^{(n)}) = \frac{1}{M^{(n)}}\sum_{i=1}^{M^{(n)}}\left[1 - \mathrm{Tr}\,W^{(n)}(\varphi^{(n)}(i))\Pi_i^{(n)}\right].$$

If there exists a sequence of codes $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ for which the average probability of error $p_e(\mathcal{C}^{(n)}, W^{(n)}) \to 0$ as $n \to \infty$, then $R := \liminf_n \frac{1}{n}\log|\mathcal{C}^{(n)}|$ is said to be an *achievable rate*. The *(direct) capacity* $C(\mathbf{W})$ of the sequence of channels $\mathbf{W}$ is defined as the supremum of all achievable rates. The corresponding *strong converse capacity* $C^*(\mathbf{W})$ is defined as the infimum of R such that for any sequence of codes $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ with rate $\liminf_n \frac{1}{n}\log|\mathcal{C}^{(n)}| \geq R$, we have $p_e(\mathcal{C}^{(n)}) \to 1$ as $n \to \infty$. It is obvious that

$$C(\mathbf{W}) \leq C^*(\mathbf{W}). \tag{39}$$

The channel is said to satisfy the *strong converse property* if $C(\mathbf{W}) = C^*(\mathbf{W})$.

A. Memoryless channels

We say that $\mathbf{W}$ is *memoryless* if for every $n \in \mathbb{N}$, $\mathcal{X}^{(n)} = \mathcal{X}^n := \times_{k=1}^{n}\mathcal{X}$, $\mathcal{H}_B^{(n)} = \mathcal{H}_B^{\otimes n}$, and

$$W^{(n)}(x_1,\ldots,x_n) = W^{\otimes n}(x_1,\ldots,x_n) := W(x_1) \otimes \ldots \otimes W(x_n) \tag{40}$$

for any sequence $(x_1,\ldots,x_n) \in \mathcal{X}^n$, where for simplicity we denote $W^{(1)}$ by W.

Remark 24 *Note that if W is a usual quantum channel, ie., a CPTP map from $\mathcal{D}(\mathcal{H}_A)$ to $\mathcal{D}(\mathcal{H}_B)$ then the memoryless extensions $W^{\otimes n}$, defined in* (40)*, are different from the usual tensor product extensions of W. Indeed, one can easily see that our definition of $W^{\otimes n}$ coincides with the nth tensor product extension of W* with the restriction *that only product-state codewords are allowed at the input of the channel. Hence, in this case the above defined direct capacity (strong converse capacity) is the so-called* product-state *classical capacity (strong converse classical capacity) of the channel.*

Note also that the usual tensor product extension of W is nothing else but the unique factorization of the n-linear map given in (40) *through $\mathcal{D}(\mathcal{H}_A)^{\otimes n}$ (note that $\mathcal{X} = \mathcal{D}(\mathcal{H}_A)$ in this case).*

For a memoryless channel $\mathbf{W}$, we denote the capacity and the strong converse capacity simply as $C(W)$ and $C^*(W)$ respectively, since in this case the sequence of channels, $\mathbf{W}$, is given solely in terms of W. The capacity of such a channel is given by its Holevo capacity $\chi^*(W)$ [22, 39], and it satisfies the strong converse property [34, 48], i.e.,

$$C(W) = C^*(W) = \chi^*(W). \tag{41}$$

Here we show how the above identity can be obtained from our one-shot bounds in Theorem 19.

Let $\mathbf{W}$ be a memoryless channel. By (39), it is sufficient to show that

$$C^*(W) \leq \chi^*(W) \qquad \text{and} \qquad C(W) \geq \chi^*(W). \tag{42}$$

Note that the α-capacities are weakly additive, in the sense that [34]

$$\chi^*_\alpha(W^{\otimes n}) = n\chi^*_\alpha(W), \qquad n \in \mathbb{N}. \tag{43}$$

Hence, by Corollary 20, we have

$$C^{(1)}_\varepsilon(W^{\otimes n}) \leq n\chi^*_\alpha(W) + \frac{1}{\alpha - 1}\log\frac{2}{(1-\varepsilon'')^2} - \log(\varepsilon'' - \varepsilon).$$

for every $0 < \varepsilon < \varepsilon'' < 1$ and $\alpha \in (1, 2]$, and $n \in \mathbb{N}$. It is easy to verify that

$$C^*(W) = \lim_{\varepsilon \to 1} \limsup_{n\to\infty} \frac{1}{n} C^{(1)}_\varepsilon(W^{\otimes n}), \tag{44}$$

and hence we obtain

$$C^*(W) \leq \lim_{\varepsilon\to 1} \chi^*_\alpha(W) = \chi^*_\alpha(W).$$

Finally, taking the limit $\lim_{\alpha \searrow 1}$ and using (28), we obtain

$$C^*(W) \leq \chi^*(W).$$

To show the second inequality in (42), we first note that

$$C(W) = \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} C^{(1)}_\varepsilon(W^{\otimes n}).$$

Let $p \in \mathcal{P}_f(\mathcal{X})$, and for every $n \in \mathbb{N}$, let $p^{\otimes n} \in \mathcal{P}_f(\mathcal{X}^n)$ be the nth i.i.d. extension of p, given by $p^{\otimes n}(x_1, \ldots, x_n) = p(x_1) \cdot \ldots \cdot p(x_n)$, $x_1, \ldots, x_n \in \mathcal{X}$. One can easily see that

$$\rho_{\mathcal{X}^n B^n}(p^{\otimes n}) = \rho_{\mathcal{X}B}(p)^{\otimes n},$$

and the lower bound in Theorem 19 yields that

$$C^{(1)}_\varepsilon(W^{\otimes n}) \geq D^{\sqrt{1-(\varepsilon')^2}}_{\max}(\rho_{\mathcal{X}B}(p)^{\otimes n} \| \rho_{\mathcal{X}}(p)^{\otimes n} \otimes \rho_B(p)^{\otimes n}) + \log\frac{\varepsilon'(\varepsilon - \varepsilon')^2}{8\varepsilon}$$

for every $0 < \varepsilon' < \varepsilon < 1$. Hence, we have

$$\liminf_{n\to\infty} \frac{1}{n} C^{(1)}_\varepsilon(W^{\otimes n}) \geq \liminf_{n\to\infty} \frac{1}{n} D^{\sqrt{1-(\varepsilon')^2}}_{\max}(\rho_{\mathcal{X}B}(p)^{\otimes n} \| \rho_{\mathcal{X}}(p)^{\otimes n} \otimes \rho_B(p)^{\otimes n}) = D(\rho_{\mathcal{X}B}(p) \| \rho_{\mathcal{X}}(p) \otimes \rho_B(p)),$$

where we used (21) for the last identity. Taking the supremum over $p \in \mathcal{P}_f(\mathcal{X})$ then yields

$$C(W) \geq \chi^*(W). \tag{45}$$

Remark 25 *Using the standard block coding argument,* (45) *yields immediately that the classical capacity of a memoryless quantum channel (without the product-state restriction) is lower bounded by the regularized Holevo capacity, as in the Holevo-Schumacher-Westmoreland theorem [22, 39].*

B. Averaged channels

We consider a class of channels which are convex combinations of a finite number of memoryless channels. For a channel in this class, n successive uses is given by the map $W^{(n)} : \mathcal{X}^n \to \mathcal{D}(\mathcal{H}_B^{\otimes n})$, defined as

$$W^{(n)} = \sum_{i=1}^{K} \gamma_i W_i^{\otimes n}, \tag{46}$$

where $\{\gamma_i\}_{i=1}^K$ is a probability distribution (we assume that all the γ_i are strictly positive), and for each $W_i : \mathcal{X} \to \mathcal{D}(\mathcal{H}_B)$, $W_i^{\otimes n}$ is the memoryless extension defined in (40), i.e., $W_i^{\otimes n}(x_1, \ldots, x_n) = W_i(x_1) \otimes \ldots \otimes W_i(x_n)$, $x_j \in \mathcal{X}$, $j = 1, \ldots, n$ and $n \in \mathbb{N}$. This model describes a scenario in which Alice and Bob know that they are communicating through a memoryless channel, but instead of knowing the exact identity of this channel (as in the memoryless case), they only know that they are using the channel W_i with probability γ_i. Note that if the first input is sent through the channel W_i then all successive inputs are also sent through the same channel. Hence the channel has long-term memory. It is an analogue of the classical averaged channel first introduced by Jacobs [23]. Let $C(\mathbf{W})$ and $C^*(\mathbf{W})$ denote the capacity and strong converse capacity of the sequence of channels $\mathbf{W} := \{W^{(n)}\}_{n\in\mathbb{N}}$, respectively.

This long-term memory channel was introduced in [10], where the authors evaluated $C(\mathbf{W})$ as

$$C(\mathbf{W}) = \sup_{p\in\mathcal{P}_f(\mathcal{X})} \min_{1\le i\le K} \chi(W_i, p). \tag{47}$$

This result was later generalized to more general forms of averaged channels in [4].

Using the fact that the error probability is an affine function of the channel, it can be seen that the strong converse capacity of an averaged channel $\mathbf{W}$ is given by

$$C^*(\mathbf{W}) = \max_{1\le i\le K} C^*(W_i) = \sup_{p\in\mathcal{P}_f(\mathcal{X})} \max_{1\le i\le K} \chi(W_i, p),$$

where the second identity follows from the memoryless case. Below we show how the one-shot upper bound of Theorem 19 yields an upper bound on the one-shot capacity of an averaged channel, which in turn yields the inequality $C^*(\mathbf{W}) \le \max_{1\le i\le K} C^*(W_i)$. For completeness, we give a proof for the converse inequality, too.

Applying Corollary 21 to $W^{(n)} = \sum_{i=1}^K \gamma_i W_i^{\otimes n}$, we obtain

$$C_\varepsilon^{(1)}\left(\sum_i \gamma_i W_i\right) \le n \max_{1\le i\le K} \chi_\alpha^*(W_i) + \frac{1}{\alpha-1}\log\frac{2}{(1-\varepsilon'')^2} - \log(\varepsilon''-\varepsilon)$$

for any $0 < \varepsilon < \varepsilon'' < 1$, where we have used the additivity of the α-capacities (43). By the same argument as in Section V A, we obtain that

$$C^*(\mathbf{W}) = \lim_{\varepsilon\to 1} \limsup_{n\to\infty} \frac{1}{n} C_\varepsilon^{(1)}\left(W^{(n)}\right) \le \lim_{\alpha\to 1} \max_{1\le i\le K} \chi_\alpha^*(W_i) = \max_{1\le i\le K} \chi^*(W_i).$$

To show that $C^*(\mathbf{W}) \ge \max_{1\le i\le K} \chi^*(W_i)$, it suffices to prove that for any $0 \le R < \max_{1\le i\le K} \chi^*(W_i)$, there exists a sequence of codes $\{\mathcal{C}_n\}_{n=1}^\infty$ with rate at least R such that

$$p_e(\mathcal{C}_n, W^{(n)}) \not\to 1 \quad \text{as} \quad n \to \infty. \tag{48}$$

Thus, let R be as above, and let j be such that

$$\chi^*(W_j) = \max_{1\le i\le K} \chi^*(W_i).$$

Then it follows from the HSW theorem ([22, 39]; see also [18]) that there exists a sequence of codes $\mathcal{C}^{(n)}$ such that $\liminf_{n\to\infty} \frac{1}{n}\log|\mathcal{C}_n| \ge R$ and

$$\lim_{n\to\infty} p_e(\mathcal{C}^{(n)}, W_j^{\otimes n}) = 0.$$

Hence, if and Alice and Bob use this code to communicate over the long-term memory channel $\mathbf{W}$ then

$$\limsup_{n\to\infty} p_e(\mathcal{C}^{(n)}, W^{(n)}) = \limsup_{n\to\infty} \sum_{i=1}^{n} \gamma_i p_e(\mathcal{C}^{(n)}, W_i^{\otimes n}) \leq 1 - \gamma_j,$$

and the statement follows.

VI. DISCUSSION

We have given bounds on the optimal type II error of Stein's lemma in terms of the smoothed max-relative entropy of the two states, and on the one-shot capacity of a channel with error threshold in terms of a quantity analogous to the Holevo capacity, defined again using the smoothed max-relative entropy. The smoothed max-relative entropy is a central notion in the so-called one-shot information theory, which has been a very active and quickly evolving research field in the past few years. The aim of this section is to relate and compare our results to existing results in the field.

First, a few comments about the choice of the distance measure for smoothing. In the original definition of the smoothed min-entropy [38], smoothing was defined with respect to the variational distance d_v (half the trace distance), which was replaced in much of the recent works with the so-called purified distance d_p [43], defined as

$$d_p(\rho,\sigma) := \sqrt{1 - [F(\rho,\sigma) + \sqrt{(1-\operatorname{Tr}\rho)(1-\operatorname{Tr}\sigma)}]^2}$$

for subnormalized states ρ, σ. In fact, for the type of bounds we considered here, it is quite irrelevant what distance d is used for the smoothing, as long as it is equivalent to the variational distance (in the sense that there exist strictly monotone functions $f, g : [0,+\infty) \to [0,+\infty)$ such that $g(0) = 0$ and $f(d_v(\rho,\sigma)) \leq d(\rho,\sigma) \leq g(d_v(\rho,\sigma))$ for every subnormalized states ρ and σ). Indeed, while the concrete form of the smoothing parameter as a function of the error threshold, as well as the form of the additive constants (e.g., in Theorem 11), may be different for different distance measures, these differences disappear in the asymptotic limit as long the distances are equivalent. In particular, the variational distance, the purified distance d_p, the extension d_{op} of the sine distance used in this paper, and the Bures distance $d_B(\rho,\sigma) := \min_{\{\psi_\rho,\psi_\sigma\}} \|\psi_\rho - \psi_\sigma\| = \sqrt{\operatorname{Tr}\rho + \operatorname{Tr}\sigma - F(\rho,\sigma)}$ [6] are all equivalent on the set of (subnormalized) states, and hence they result in qualitatively equivalent smoothed entropies. The distances d_{op}, d_p and d_B, all derived from the fidelity, also seem equally useful for smoothing dual conditional entropies in the sense of [43].

There are also differences in the choice of the neighbourhood over which smoothing is performed; the main difference here is optimizing over subnormalized states in an ε-neighbourhood of the given state ρ, or restricting the optimization to normalized states. Again, the difference between the resulting quantities is irrelevant for the asymptotic properties of these quantities. We briefly show this here for our definition $D^{\varepsilon}_{\max}(\rho\|\sigma)$ of the smoothed max-relative entropy (where optimization is restricted to normalized states and the distance is d_{op}) and another common choice [43], defined as

$$\tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma) := \inf\{D_{\max}(\bar{\rho}\|\sigma) \,:\, \bar{\rho} \geq 0,\, \operatorname{Tr}\bar{\rho} \leq 1;\, d_p(\rho,\bar{\rho}) \leq \varepsilon\}$$

(where optimization is over subnormalized states and the distance is d_p). Indeed, let ρ be a state, and $\hat{\rho} \in B_\varepsilon(\rho)$, where $B_\varepsilon(\rho)$ is the ε-ball around the state ρ with respect to d_{op}. Then $d_{\mathrm{op}}(\rho,\hat{\rho}) = d_p(\rho,\hat{\rho})$ and hence $\hat{\rho} \in \tilde{B}_\varepsilon(\rho)$, which implies $\tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma) \leq D_{\max}(\hat{\rho}\|\sigma)$, and optimizing over $\hat{\rho} \in B_\varepsilon(\rho)$ yields $\tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma) \leq D^{\varepsilon}_{\max}(\rho\|\sigma)$. On the other hand, if $\bar{\rho} \in \tilde{B}_\varepsilon(\rho)$ then $\varepsilon \geq d_p(\rho,\bar{\rho}) = \sqrt{1 - F(\rho,\bar{\rho})^2}$, and hence $\sqrt{1-\varepsilon^2} \leq F(\rho,\bar{\rho}) \leq \sqrt{\operatorname{Tr}\bar{\rho}}$, where the last inequality is due to the monotonicity of the fidelity under the trace. Let $\hat{\rho} := \bar{\rho}/\operatorname{Tr}\bar{\rho}$. Then $F(\rho,\hat{\rho}) = F(\rho,\bar{\rho})/\sqrt{\operatorname{Tr}\bar{\rho}} \geq F(\rho,\bar{\rho}) \geq \sqrt{1-\varepsilon^2}$, and hence $d_{\mathrm{op}}(\rho,\hat{\rho}) = \sqrt{1 - F(\rho,\hat{\rho})^2} \leq \varepsilon$, i.e., $\hat{\rho} \in B_\varepsilon(\rho)$. Thus, $D^{\varepsilon}_{\max}(\rho\|\sigma) \leq D_{\max}(\hat{\rho}\|\sigma) = D_{\max}(\bar{\rho}\|\sigma) - \log\operatorname{Tr}\bar{\rho} \leq D_{\max}(\bar{\rho}\|\sigma) - \log(1-\varepsilon^2)$. Optimizing over $\bar{\rho} \in \tilde{B}_\varepsilon(\rho)$ yields $D^{\varepsilon}_{\max}(\rho\|\sigma) \leq \tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma) - \log(1-\varepsilon^2)$. Hence, we finally have

$$\tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma) \leq D^{\varepsilon}_{\max}(\rho\|\sigma) \leq \tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma) - \log(1-\varepsilon^2), \qquad \varepsilon \in (0,1). \tag{49}$$

In particular,

$$\lim_{\varepsilon \searrow 0} \left| \tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma) - D^{\varepsilon}_{\max}(\rho\|\sigma) \right| = 0.$$

Our main reason to restrict the optimization to normalized states is that otherwise the smoothed max-relative entropy can be negative; in fact, it is easy to see that $\lim_{\varepsilon\nearrow 1} \tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma) = -\infty$, while $D^{\varepsilon}_{\max}(\rho\|\sigma) \geq 0$ for any two states ρ and σ. Since the smoothed max-relative entropy is a kind of a statistical divergence, or generalized relative entropy, we prefer to keep it non-negative on pairs of normalized states.

In the first version of this paper [13], we used a different type of smoothing, defined as $\widehat{D}^{\varepsilon}_{\max}(\rho\|\sigma) := \inf\{D_{\max}(\bar{\rho}\|\sigma) : \bar{\rho} \geq 0, \operatorname{Tr}\bar{\rho} \leq 1; \|\rho - \bar{\rho}\|_1 \leq \varepsilon\}$, and gave the bounds

$$\widehat{D}^{4\sqrt{\varepsilon}}_{\max}(\rho\|\sigma) \leq -\log\beta_{1-\varepsilon}(\rho\|\sigma) \leq \widehat{D}^{\varepsilon/2}_{\max}(\rho\|\sigma) + \log\frac{2}{\varepsilon}$$

on the optimal type II error. Using similar arguments as above, this yields the bounds

$$D^{\sqrt{4\sqrt{\varepsilon}}}_{\max}(\rho\|\sigma) + \log(1-4\sqrt{\varepsilon}) \leq -\log\beta_{1-\varepsilon}(\rho\|\sigma) \leq D^{\varepsilon/4}_{\max}(\rho\|\sigma) + \log\frac{2}{\varepsilon}$$

and

$$\tilde{D}^{\sqrt{8\sqrt{\varepsilon}}}_{\max}(\rho\|\sigma) \leq -\log\beta_{1-\varepsilon}(\rho\|\sigma) \leq \tilde{D}^{\varepsilon/4}_{\max}(\rho\|\sigma) + \log\frac{2}{\varepsilon} \tag{50}$$

in terms of the alternative smoothed max-relative entropies discussed above. Using the quantum Stein's lemma, these yield the ε-independent version of Corollary 9 for $\widehat{D}^{\varepsilon}_{\max}(\rho\|\sigma)$ in the range $\varepsilon \in (0, 1/16)$, for $D^{\varepsilon}_{\max}(\rho\|\sigma)$ in the range $(0, 1/16)$ and for $\tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma)$ in the range $(0, 1/64)$. Similar bounds were obtained very recently in [45], of the form

$$\tilde{D}^{\sqrt{\varepsilon}}_{\max}(\rho\|\sigma) - \log\nu(\sigma) + \log\varepsilon \leq -\log\beta_{1-\varepsilon}(\rho\|\sigma) \leq \tilde{D}^{\sqrt{\varepsilon-\delta}}_{\max}(\rho\|\sigma) - 3\log\delta + 3\log 3 + \log(1-\varepsilon+\delta),$$

where $\nu(\sigma)$ is the number of different eigenvalues of σ. These bounds are valid for all $\varepsilon \in (0,1)$ and $\delta \in (0,\varepsilon)$, and hence the quantum Stein's lemma applied to these bounds yields the ε-independent version of Corollary 9 for $\tilde{D}^{\varepsilon}_{\max}(\rho\|\sigma)$ in the whole range $\varepsilon \in (0,1)$. Using (49), these bounds yield

$$D^{\sqrt{\varepsilon}}_{\max}(\rho\|\sigma) - \log\nu(\sigma) + \log\varepsilon(1-\varepsilon) \leq -\log\beta_{1-\varepsilon}(\rho\|\sigma) \leq D^{\sqrt{\varepsilon-\delta}}_{\max}(\rho\|\sigma) - 3\log\delta + 3\log 3 + \log(1-\varepsilon+\delta)$$

in terms of the smooth entropies used in this paper. Likewise, our bounds in Theorem 11 yield, with the help of (49), the bounds

$$\tilde{D}^{g(\varepsilon)}_{\max}(\rho\|\sigma) \leq -\log\beta_{1-\varepsilon}(\rho\|\sigma) \leq \tilde{D}^{\varepsilon'}_{\max}(\rho\|\sigma) + \log\frac{1}{\varepsilon-\varepsilon'} - \log(1-(\varepsilon')^2),$$

where $g(\varepsilon) := \sqrt{\varepsilon(2-\varepsilon)}$, and $0 < \varepsilon' < \varepsilon$. Apart from the different smoothing conventions, the difference between the bounds of [45] and our Theorem 11 stems from the different proof methods; while the bounds of [45] were derived using an intermediate quantity, the single-shot quantum information spectrum, we used a more direct approach in proving Theorem 11, which results in somewhat simpler expressions.

In Section IV we derived bounds on the one-shot ε-error capacities of a channel W in terms of its ε-max capacities, which in the asymptotics gave that the strong converse capacity of W is equal to its Holevo capacity. We emphasize here again that in the case where W is a quantum channel, our definition of the (strong converse) capacity gives the (strong converse) capacity for product state encoding [34, 48]. The error bound of [34] actually gives that the unconstrained strong converse rate for arbitrary (i.e., not necessarily product) encoding cannot exceed the infimum (over α) of the regularized α-capacities; in particular, when the α-capacities are additive in the sense that $\chi^*_\alpha(W^{\otimes n}) = n\chi^*_\alpha(W)$ for every n and α close enough to 1, then the unconstrained strong converse rate is equal to the Holevo capacity. Such additivity results were shown in [24] for a class of quantum channels, including the qudit depolarizing channels and unital qubit channels, thereby providing the first and so far the only examples for quantum channels with the strong converse property with unconstrained encoding. The error bound of [34] automatically yields an upper bound on the one-shot ε-error capacities in terms of the α-capacities with $\alpha > 1$ (cf. Corollary 20), as was already pointed out in Theorem V.1 of [29]. A counterpart of these bounds, i.e., lower bounds on the one-shot ε-error capacities in terms of the α-capacities with $\alpha \in (0,1)$, have been obtained in [28, 29].

It is well-known that channel coding (for classical information) and hypothesis testing are closely related to each other, and that the direct part of the channel coding theorem (the Holevo-Schumacher-Westmoreland (HSW) theorem [22, 39]) can be recovered using this relation and the quantum Stein's

lemma [18, 36]. Explicit bounds on the one-shot ε-error capacity of a channel W in terms of the optimal type II error for discriminating states of the form $\rho_{\mathcal{X}B} = \sum_x p(x)|x\rangle\langle x| \otimes W(x)$ (cf. (27)) from the product of its marginals, have been given in [47], which again yields in the asymptotic limit the HSW theorem, i.e., that the (direct) capacity of W is lower bounded by the Holevo capacity of W. The upper bound of [47] has been further improved in [25], using state discrimination with restricted measurements, and it has been shown in [47] that these bounds yield

$$C_\varepsilon(\mathbf{W}) := \sup_{\{\mathcal{C}_n\}} \left\{ \liminf_{n\to\infty} \frac{1}{n} \log |\mathcal{C}_n| \,:\, \limsup_{n\to\infty} p_e(\mathcal{C}_n, W^{(n)}) \leq \varepsilon \right\} \leq \frac{\chi^*(W)}{1-\varepsilon}$$

for a sequence of i.i.d. channels with product encoding. While this is sufficient to determine the direct capacity with weak converse ($\varepsilon \to 0$), it is not informative for the strong converse capacity ($\varepsilon \to 1$). In comparison, our approach yields $C_\varepsilon(\mathbf{W}) \leq \chi^*(W)$ for every $\varepsilon \in (0,1)$, which in particular gives that the strong converse capacity is upper bounded by the Holevo capacity, as we showed in Section V A.

Bounds on the one-shot ε-error classical capacity of a quantum channel have been given before in [37], in terms of a mixture of smoothed min- and max-relative entropies. While these bounds are suitable to obtain the direct capacity of a memoryless channel (with product encoding), they only provide upper bounds on the asymptotic ε-error capacity for ε up to $1/2$, and hence they cannot be used to obtain the strong converse capacity.

VII. ACKNOWLEDGMENTS

ND would like to thank Igor Bjelakovic for a helpful exchange and for pointing out related results for classical and quantum compound channels. MM was supported by the Marie Curie International Incoming Fellowship "QUANTSTAT". MH was supported by the UTS Chancellor's Postdoctoral Research Fellowship. FB acknowledges support from the Swiss National Science Foundation, via the National Centre of Competence in Research QSIT. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 213681.

[1] K.M.R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, F. Verstraete, "Discriminating states: the quantum Chernoff bound," *Phys. Rev. Lett.* **98** 160501, (2007).

[2] K.M.R. Audenaert, M. Mosonyi, F. Verstraete, "Quantum state discrimination bounds for finite sample size," *J. Math. Phys.*, **53**, issue 12, 122205, (2012).

[3] M. Berta, M. Christandl, and R. Renner, "The Quantum Reverse Shannon Theorem based on One-Shot Information Theory," *Commun. Math. Phys.* vol. 306, 579, 2011

[4] I. Bjelakovic and H. Boche, "Classical Capacities of Compound and Averaged Channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3360-3374, 2009.

[5] F.G.S.L. Brandao, M. Plenio, "A generalization of quantum Stein's lemma," *Comm. Math. Phys.*, vol. 295, 791–828, 2010.

[6] D. Bures, "An Extension of Kakutani's Theorem on Infinite Product Measures to the Tensor Product of Semifinite W*-Algebras," *Transactions of the American Mathematical Society* Vol. 135, pp. 199-212, (1969).

[7] F. Buscemi and N. Datta, "The quantum capacity of channels with arbitrarily correlated noise," *IEEE Transactions on Information Theory*, vol. 56, Issue 3, pp. 1447-1460, 2010

[8] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[9] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inf. Theory* vol. 41, 26–34, (1995)

[10] N. Datta and T.C. Dorlas, "The Coding Theorem for a Class of Quantum Channels with Long-Term Memory," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, p. 8147, Jul. 2007.

[11] N. Datta and R. Renner, "Smooth Entropies and the Quantum Information Spectrum," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2807-2815, June 2009.

[12] N. Datta, "Min- and Max-Relative Entropies and a New Entanglement Monotone," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2816-2826, June 2009.

[13] N. Datta, M-H. Hsieh, F.G.S.L. Brandao, "Strong converse rates and an example of violation of the strong converse property," *http://arxiv.org/abs/1106.3089v1*, June 2011.

[14] C.A. Fuchs, J. van de Graaf, "Cryptographic distinguishability measures for quantum-mechanical states," *IEEE Transactions on Information Theory*, vol. 45, pp. 1216–1227, May 1999.

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

[16] T.S. Han, *Information-Spectrum Methods in Information Theory*, Springer-Verlag, 2002.

[17] M.B. Hastings, "A Counterexample to Additivity of Minimum Output Entropy," *Nature Physics* **5**, 255 (2009).

[18] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1753-1768, Jul. 2003.

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

[20] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability, *Communications in Mathematical Physics*, vol. 143, no. 1, pp. 99-114, Dec. 1991.

[21] D. Hilbert, "Neue Begaündung der Bolya-Lobatschefskyschen Geometrie," *Math. Ann.* **57**, pp. 137–150, (1903).

[22] A.S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 269-273, Jan. 1998.

[23] K. Jacobs, "Almost periodic channels," *Colloquium on Combinatotial Methods in Probability Theory*, Aarhus, 1962.

[24] R. König and S. Wehner, "A strong converse for classical channel coding using entangled inputs.," *Physical Review Letters*, vol. 103, no. 7, p. 070504, Aug. 2009.

[25] W. Matthews, S. Wehner, "Finite blocklength converse bounds for quantum channels," *arXiv:1210.4722*, (2012).

[26] Ke Li, "Second Order Asymptotics for Quantum Hypothesis Testing," *arXiv:1208.1400*, 2012.

[27] C. Morgan, A. Winter, "Towards a strong converse for the quantum capacity (of degradable channels)," *arXiv:1301.4927*, (2013).

[28] M. Mosonyi, N. Datta, "Generalized relative entropies and the capacity of classical-quantum channels," *J. Math. Phys.* vol. 50, 072104, 2009.

[29] M. Mosonyi and F. Hiai, "On the quantum Rényi relative entropies and related capacity formulas," *IEEE Trans. Inform. Theory*, vol. 57, 2474-2487, (2011).

[30] H. Nagaoka, M. Hayashi, "An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses," *IEEE Trans. Inform. Theory*, vol. 53, issue 2, 534–549, (2007).

[31] H. Nagaoka, "The converse part of the theorem for quantum Hoeffding bound," *quant-ph/0611289*

[32] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, New York: Cambridge University Press, 2000.

[33] M. Nussbaum, A. Szkoła, "A lower bound of Chernoff type for symmetric quantum hypothesis testing," *Ann. Statist.* **37**, 1040–1057, (2009).

[34] T. Ogawa and H. Nagaoka, "Strong converse to the quantum channel coding theorem," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2486-2489, 1999.

[35] T. Ogawa and H. Nagaoka, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2428-2433, 2000.

[36] T. Ogawa, H. Nagaoka, "Making good codes for classical-quantum channel coding via quantum hypothesis testing," *IEEE Trans. Inform. Theory* **53** no. 6, pp. 2261–2266, (2007) (preprint: arXiv:quant-ph/0208139, 2002)

[37] J.M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation", *IEEE Trans. Inform. Theory*, vol. 57, pp. 7377–7385, Nov. 2011.

[38] R. Renner, *Security of Quantum Key Distribution*, PhD dissertation, Swiss Federal Institute of Technology Zurich, Diss. ETH No. 16242, (2005).

[39] B. Schumacher and M. Westmoreland, "Sending classical information via noisy quantum channels," *Physical Review A*, vol. 56, no. 1, pp. 131-138, Jul. 1997.

[40] R. Sibson, "Information radius," *Z. Wahrscheinlichkeitsth. Verw. Gebiete* **14**, 149–161, 1969.

[41] A.C. Thompson, "On certain contraction mappings in a partially ordered vector space," *Proc. Amer. Math. Soc.* **14**,pp. 438–443, (1963).

[42] M. Tomamichel, R. Colbeck, R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, 5840–5847, 2009.

[43] M. Tomamichel, R. Colbeck, R. Renner, "Duality Between Smooth Min- and Max-Entropies," *IEEE Trans. Inf. Theory* **56**, pp. 4674-4681, (2010).

[44] M. Tomamichel, "A Framework for Non-Asymptotic Quantum Information Theory," *PhD Thesis, Department of Physics, ETH Zurich, arXiv:1203.2142.*

[45] M. Tomamichel, M. Hayashi, "A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks," *preprint, arXiv:1208.1478*, (2012).

[46] A. Uhlmann, "The "transition probability" in the state space of a *-algebra," *Reports on Mathematical Physcis* vol. 9, pp. 273–278, (1976).

[47] L. Wang and R. Renner, "One-Shot Classical-Quantum Capacity and Hypothesis Testing," *Phys. Rev. Lett.* vol. 108, 200501, 2012.

[48] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2481-2485, 1999.
[49] J. Wolfowitz, "Coding Theorems of Information Theory," Springer, New York, 1964.