

# On Covert Quantum Sensing and the Benefits of Entanglement

Mehrdad Tahmasbi<sup>1</sup> and Matthieu R. Bloch<sup>2</sup>, *Senior Member, IEEE*

**Abstract**—Motivated by applications to covert quantum radar, we analyze a covert quantum sensing problem, in which a legitimate user aims at estimating an unknown parameter taking finitely many values by probing a quantum channel while remaining undetectable from an adversary receiving the probing signals through another quantum channel. When channels are classical-quantum, we characterize the optimal error exponent under a covertness constraint for sensing strategies in which probing signals do not depend on past observations. When the legitimate user's channel is a unitary depending on the unknown parameter, we provide achievability and converse results that show how one can significantly improve covertness using an entangled input state.

**Index Terms**—Quantum sensing, covert communication.

## I. INTRODUCTION

WHILE much of the information-theoretic security literature focuses on ensuring secrecy and privacy, in the sense of preventing or minimizing the information content leaked by signals, there have been recent efforts geared at understanding the information-theoretic limits of covertness, defined as the ability to avoid detection by hiding the mere presence of signals themselves. In particular, such information-theoretic limits have been successfully characterized in the context of covert communication and covert sensing. Covert communications describe situations in which two legitimate parties attempt to communicate reliably over a noisy channel while avoiding detection by a third party. A problem closely related to covert communication is stealth communication [1], [2], in which it is required that the adversary's observations be distributed according to a pre-specified distribution, not necessarily associated to the absence of communication. This different constraint could result in different asymptotics of the optimal number of transmitted bits. Covert communications are governed by a square-root law [3], which limits the number of bits that can be reliably and covertly transmitted to the square root of the block length, and the channel-dependent pre-constant that governs the scaling is known for classical discrete-memoryless channels [4], [5], [6],

Gaussian channels [5], classical-quantum channels [7], [8], and lossy bosonic channels [9]. Covert sensing, in contrast, refers to scenarios in which the estimation of parameters of interest requires the use of probing signals that emit energy and are therefore detectable; if estimation could be achieved through purely passive sensing, covertness would automatically be guaranteed. Covert sensing is also governed by a form of square-root law. Specifically, [10], [11] have considered the problem of estimating an unknown phase over a bosonic channel while keeping the sensing undetectable by a passive quantum adversary. This operation is made possible by the presence of thermal noise, which allows one to hide the useful sensing signal in the background thermal noise and results in a mean-square phase estimation error scaling as  $\mathcal{O}(\frac{1}{\sqrt{n}})$  if  $n$  is the number of modes. Reference [12] has investigated a slightly different model in which the objective is to covertly estimate the impulse response of a linear system. One of the main results obtained is that the bandwidth of sensing signals must scale linearly with the time duration of these signals. Potential applications of covert sensing include covert radar and covert pilot estimation in wireless communications. We note that for the latter application, other models such as compound or arbitrarily varying channels have been studied [13] to handle channel state uncertainty.

The present work studies covert sensing by drawing on connections with *active hypothesis testing* [14], [15], also known as *controlled sensing* [16]. Active hypothesis testing differs from traditional hypothesis testing [17], by considering the possibility of changing the kernel through which unknown parameters are observed, which leads to estimation strategies that are potentially faster or more accurate. Recent studies of active hypothesis testing have built upon the pioneering work of Chernoff [18] on sequential design of experiments to provide new insights into the problem, including the benefits of sequentiality and adaptivity [14], [16], [19], the role of extrinsic Jensen-Shannon Divergence as an information utility metric [20], the unavoidable trade-off between reliability and resolution of estimation [15], and the identification of situations when pure (non-randomized) policies are optimal [16]. Examples of recent applications of active hypothesis testing include radar [21] and millimeter-wave beam alignment [22]. Active hypothesis testing offers a natural framework for studying covert sensing since covertness effectively requires one to use different observation kernels to hide the presence of probing signals.

The problem of quantum state or channel discrimination without covertness constraint has been intensively studied, see, e.g., [23, Ch. 3], [24]. The optimal error exponent

Manuscript received August 2, 2020; revised December 3, 2020; accepted January 14, 2021. Date of publication February 2, 2021; date of current version March 16, 2021. This work was supported by NSF under award 1910859. (Corresponding author: Mehrdad Tahmasbi.)

Mehrdad Tahmasbi was with the School of ECE, Georgia Institute of Technology, Atlanta, GA 30332 USA. He is now with the University of Amsterdam, 1012 WX Amsterdam, The Netherlands, and also with QuSoft, 1098 XG Amsterdam, The Netherlands (e-mail: m.tahmasbi@uva.nl).

Matthieu R. Bloch is with the School of ECE, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: matthieu.bloch@ece.gatech.edu).

Digital Object Identifier 10.1109/JSAIT.2021.3056640

of discrimination of finitely many quantum states has been characterized [25], [26] and resembles the optimal error exponent of *classical* states discrimination; this exponent is known as the multiple quantum Chernoff distance. For *quantum channel* discrimination, as the probing signal can be any quantum state and could be arbitrarily entangled with the environment and previously received signals, several intriguing phenomena, specific to the quantum world, could happen. For example, the legitimate user can substantially decrease the probability of estimation error by keeping its environment entangled with the probing signals [23, Example 3.36] or quantum channels that cannot be perfectly distinguished with a single probing can be distinguished with multiple probing with zero probability of error [27].

In a previous conference paper [28], we revisited the idea of covert sensing put forward in [10], [11] from the perspective of active hypothesis testing in a classical setting, which we called *active covert sensing*. Therein, we have characterized the exponent of the probability of detection error subject to covertness constraints for non-adaptive non-sequential strategies and illustrated the benefits of adaptive non-sequential strategies. In the present work, we expand upon these results in the quantum setting, in which the legitimate parties attempt to discriminate quantum channels subject to a covertness constraint. Some of the results developed hereafter supersede those in [28] but, unlike [28], we do not consider the adaptation of the probing signals with respect to the previous outputs of the quantum channel; instead, we explore the potential benefits of using entanglement for covert sensing. Specifically, we offer the following two contributions: i) when the legitimate user's probing signals are classical, but the received state by both the legitimate user and the warden are quantum, we characterize the exact detection error exponent of non-adaptive non-sequential strategies subject to a covertness constraint; ii) when the legitimate user's channel is a unitary depending on the unknown parameter, we show that the legitimate user can estimate the unknown parameter *with zero error* while satisfying a stronger notion of covertness compared to what we could achieve over classical-quantum (cq)-channels. We also prove a converse result showing that the asymptotic scaling of the covertness in our achievability result is optimal.

The remainder of this article is organized as follows. We introduce our notation in Section II and formalize the problem in Section III. We provide our main results for cq-channels and unitary channels in Section IV and prove them in Section V. We defer the most technical parts of the proofs to the appendices to streamline the presentation.

## II. NOTATION

We denote a vector of length  $n$  (e.g.,  $(x_1, \dots, x_n)$ ) in  $\mathcal{X}^n$  by a boldface letter (e.g.,  $\mathbf{x}$ ). We let  $T_{\mathbf{x}}$  denote the type of the vector  $\mathbf{x}$ , which is a Probability Mass Function (PMF) over  $\mathcal{X}$  defined by

$$T_{\mathbf{x}}(a) \triangleq \frac{|\{i \in [1, n] : x_i = a\}|}{n}, \quad (1)$$

where  $[m, n] \triangleq \{i \in \mathbb{Z} : m \leq i \leq n\}$ . We define  $\mathcal{P}_n(\mathcal{X}) \triangleq \{T_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}^n\}$  and  $\mathcal{T}_Q \triangleq \{\mathbf{x} : T_{\mathbf{x}} = Q\}$  for  $Q \in \mathcal{P}_n(\mathcal{X})$ .  $P_X$

denotes a probability distribution over the set  $\mathcal{X}$  and  $P_X \otimes P_Y$  is the product distribution over  $\mathcal{X} \times \mathcal{Y}$  induced by two marginals  $P_X$  and  $P_Y$ .  $P_X^{\otimes n}$  also denotes the  $n$ -fold product distribution of  $P_X$  over  $\mathcal{X}^n$ . We define  $\mathbb{H}_b(x) \triangleq -x \log x - (1-x) \log(1-x)$  for  $x \in [0, 1]$ . Let  $\mathbb{1}\{\cdot\}$  denote the indicator function. We use standard asymptotic notation  $\mathcal{O}(\cdot)$ ,  $o(\cdot)$ ,  $\omega(\cdot)$ , and  $\Theta(\cdot)$ . To emphasize that the constant hidden in  $\mathcal{O}(\cdot)$  could only depend on a parameter  $\theta$ , we write  $\mathcal{O}_{\theta}(\cdot)$ .

A quantum system  $A$  is described by a finite-dimensional Hilbert space, which we also denote by  $A$  with a slight abuse of notation. Let  $\dim A$  be the dimension of  $A$  and  $\mathbf{1}_A$  be the identity map on  $A$ . We denote the tensor product of  $A$  and  $B$  by  $A \otimes B$  or  $AB$ .  $\mathcal{L}(A)$  denotes the set of all linear operators from  $A$  to  $A$  and  $\mathcal{D}(A)$  denotes the set of all density operators acting on  $A$ , which are the possible states of the quantum system  $A$ . Given two density operators  $\rho_A \in \mathcal{D}(A)$  and  $\rho_B \in \mathcal{D}(B)$ , we denote the product state on  $AB$  by  $\rho_A \otimes \rho_B$ . We also define  $\rho_A^{\otimes n}$  as the  $n$ -fold tensor product of  $\rho_A$ . A pure state is of the form  $|\phi\rangle\langle\phi|_A$  for a unit vector  $|\phi\rangle_A \in A$ . We use  $\phi_A$  to denote  $|\phi\rangle\langle\phi|_A$  when there is no confusion. For  $X \in \mathcal{L}(A)$ , the trace norm of  $X$  is  $\|X\|_1 \triangleq \text{tr}(\sqrt{X^\dagger X})$ , and  $\nu(X)$  denotes the number of *distinct* eigenvalues of  $X$ . We also define the support of  $X \in \mathcal{L}(\mathcal{H}_A)$  as the subspace orthogonal to  $\text{Ker}(X)$ , which we denote by  $\text{supp}(X)$ . We denote the adjoint of  $X$  by  $X^\dagger$ . When  $X$  is Hermitian, i.e.,  $X = X^\dagger$ ,  $\lambda_{\min}(X)$  denotes the minimum eigenvalue of  $X$ . The fidelity between two density operators  $\rho$  and  $\sigma$  is defined as  $F(\rho, \sigma) \triangleq \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$ . A quantum channel  $\mathcal{N}_{A \rightarrow B}$  is a linear trace-preserving completely positive map from  $\mathcal{L}(A)$  to  $\mathcal{L}(B)$ . Given two quantum channels  $\mathcal{N}$  and  $\mathcal{M}$ , we denote their tensor product by  $\mathcal{N} \otimes \mathcal{M}$ . Let  $\text{id}_A$  be the identity channel on  $\mathcal{L}(A)$ . For two states  $\rho$  and  $\sigma$  with  $\text{supp}(\rho) \subset \text{supp}(\sigma)$ , we define

$$\chi_2(\rho\|\sigma) \triangleq \text{tr}(\rho^2 \sigma^{-1}) - 1, \quad (2)$$

$$\mathbb{D}(\rho\|\sigma) \triangleq \text{tr}(\rho(\log \rho - \log \sigma)). \quad (3)$$

Additionally, given the spectral decomposition of a state  $\sigma = \sum_i \lambda_i P_i$ , we define

$$\begin{aligned} \eta(\rho\|\sigma) &= \sum_{i \neq j} \frac{\log \lambda_i - \log \lambda_j}{\lambda_i - \lambda_j} \text{tr}((\rho - \sigma)P_i(\rho - \sigma)P_j) \\ &\quad + \sum_i \frac{1}{\lambda_i} \text{tr}((\rho - \sigma)P_i(\rho - \sigma)P_i). \end{aligned} \quad (4)$$

Finally, we use standard notions from differential geometry such as tangent space and derivative of a smooth functions.

## III. PROBLEM FORMULATION

Let  $A$ ,  $B$ , and  $W$  be quantum systems and  $\Theta$  be a finite set of parameters.<sup>1</sup> As illustrated in Fig. 1, let  $\{\mathcal{N}_\theta : \mathcal{L}(A) \rightarrow \mathcal{L}(B)\}_{\theta \in \Theta}$  and  $\{\mathcal{E}_\theta : \mathcal{L}(A) \rightarrow \mathcal{L}(W)\}_{\theta \in \Theta}$  be two families of quantum channels. Alice's estimation strategy consists of the following. Alice prepares a possibly entangled

<sup>1</sup>We only consider classical parameters, estimating an unknown quantum parameter is of interest for certain applications but is outside the scope of this article.

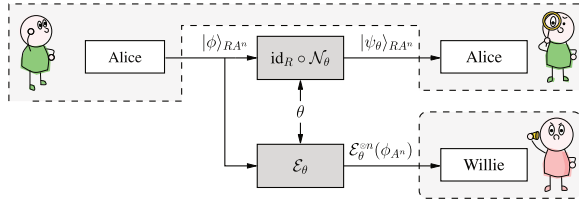


Fig. 1. Model for quantum covert sensing.

state  $|\phi\rangle_{RA^n}$ , where  $R$  is a reference system with  $\dim R = \dim A^n$ , and transmits the subsystem  $A^n$  in the quantum state  $\phi_{A^n}$ . Alice then receives  $\psi_{\theta, RA^n} \triangleq (\text{id}_R \otimes \mathcal{N}_{\theta}^{\otimes n})(|\phi\rangle\langle\phi|_{RA^n})$ , on which Alice performs a POVM  $\{\Gamma_{\theta}\}_{\theta \in \Theta}$  to estimate the unknown parameter  $\theta$ . We measure the estimation error through<sup>2</sup>

$$\max_{\theta \in \Theta} (1 - \text{tr}(\psi_{\theta, RA^n} \Gamma_{\theta})). \quad (5)$$

Let  $\mathcal{S}(n, \epsilon)$  denote the set of all states  $|\phi\rangle_{RA^n}$  such that the estimation error is not greater than  $\epsilon$  for *some* POVM. The strategy is called non-sequential because the parameter  $n$  is fixed, and non-adaptive because the probing signals are not adapted to past observations.

Willie observes what Alice transmits through the memoryless channel  $\mathcal{E}_{\theta}$  when the parameter is  $\theta$ , i.e., Willie receives  $\mathcal{E}_{\theta}^{\otimes n}(\phi_{A^n})$ . One of the input vectors is denoted by  $|0\rangle \in A$  and corresponds to Alice being “inactive,” i.e., Willie expects Alice to transmit  $|0\rangle^{\otimes n}$  when no estimation strategy is run. This allows us measure the inability of Willie to detect probing signals through the covertness metric

$$\max_{\theta \in \Theta} \mathbb{D}(\mathcal{E}_{\theta}^{\otimes n}(\phi_{A^n}) \| \mathcal{E}_{\theta}^{\otimes n}(|0\rangle\langle 0|^{\otimes n})). \quad (6)$$

We refer the reader to [4], [6] for a discussion on how upper-bounding (6) yields a bound on the probability of error of any strategy employed by Willie to detect the presence of an estimation strategy.

We finally define the quantities

$$C(n, \epsilon) \triangleq \min_{|\phi\rangle_{RA^n} \in \mathcal{S}(n, \epsilon)} \max_{\theta \in \Theta} \mathbb{D}(\mathcal{E}_{\theta}^{\otimes n}(\phi_{A^n}) \| \mathcal{E}_{\theta}^{\otimes n}(|0\rangle\langle 0|^{\otimes n})), \quad (7)$$

$$E(n, \delta) \triangleq \inf \{\epsilon \in ]0, 1[ : C(n, \epsilon) \leq \delta\}, \quad (8)$$

which will be useful to express the fundamental limits of Alice’s performance.

*Remark 1:* To ensure that our model is physically realizable, the whole process from Alice to herself and Willie should be described by a completely positive trace-preserving map as required by the laws of quantum mechanics. We also allow the systems received by Alice and Willie have some sub-system in common. This assumption is necessary because we shall study the scenario in which Alice’s channel is unitary and the complementary channel of a unitary channel is constant. If Alice and Willie’s systems were disjoint, Willie would not obtain any useful information. Formally, we state our requirement as follows. There should exist quantum systems  $B'_{\theta}$ ,

$W'_{\theta}$ , and  $C_{\theta}$ , isomorphic isometries  $V_{\theta, B} : B'_{\theta} C_{\theta} \rightarrow B$  and  $V_{\theta, W} : W'_{\theta} C_{\theta} \rightarrow W$  (i.e.,  $B'_{\theta} C_{\theta} \cong B$  and  $W'_{\theta} C_{\theta} \cong W$ ) and a quantum channel  $\mathcal{M}_{\theta} : \mathcal{L}(A) \rightarrow \mathcal{L}(B'W'C)$  such that

$$\mathcal{N}_{\theta}(X) = V_{\theta, B} \text{tr}_{W'_{\theta}}(\mathcal{M}_{\theta}(X)) V_{\theta, B}^{\dagger} \quad \forall X \in \mathcal{L}(A) \quad (9)$$

$$\mathcal{E}_{\theta}(X) = V_{\theta, W} \text{tr}_{B'_{\theta}}(\mathcal{M}_{\theta}(X)) V_{\theta, W}^{\dagger} \quad \forall X \in \mathcal{L}(A). \quad (10)$$

Because some parts of Alice’s and Willie’s output systems could be in common, the order in which Alice and Willie observe their outputs matters. We assume that Willie first observes his outputs and, should he decide to disturb the systems, Alice is notified through another means of communication.

## IV. MAIN RESULTS

### A. Cq-Channels

We first provide the full asymptotic characterization of Alice’s optimal performance for covert sensing over cq-channels. Although the cq channels we consider here have finite dimension, cq channels are good models for those channels that arise in quantum optics, such as bosonic channels in which the input is a classical parameter of the transmitted states, see, e.g., [10]. In particular, we assume that both  $\mathcal{N}_{\theta}$  and  $\mathcal{E}_{\theta}$  are cq for all  $\theta$ . That is, there exist an orthonormal basis  $\{|a_u\rangle_A : u \in \mathcal{U}\}$  for  $A$  and two sets of quantum states  $\{\rho_{B|\theta}^u\}_{u \in \mathcal{U}, \theta \in \Theta}$  and  $\{\rho_{W|\theta}^u\}_{u \in \mathcal{U}, \theta \in \Theta}$  such that

$$\mathcal{N}_{\theta}(|a_u\rangle\langle a_{u'}|) = \mathbb{1}\{u = u'\} \rho_{B|\theta}^u \quad \forall u, u' \in \mathcal{U} \quad (11)$$

$$\mathcal{E}_{\theta}(|a_u\rangle\langle a_{u'}|) = \mathbb{1}\{u = u'\} \rho_{W|\theta}^u \quad \forall u, u' \in \mathcal{U}. \quad (12)$$

Note that  $\mathcal{E}_{\theta}$  and  $\mathcal{N}_{\theta}$  are uniquely characterized by the linearity of quantum channels. We finally assume that  $0 \in \mathcal{U}$  and that  $|a_0\rangle$  is the innocent state  $|0\rangle$ . We also impose some mild restrictions on the cq-channels that are required to make the problem meaningful.

- 1)  $\tilde{\Theta} \triangleq \{\theta \in \Theta : \exists \theta' \in \Theta \setminus \{\theta\} : \rho_{B|\theta}^0 = \rho_{B|\theta'}^0\} \neq \emptyset$ . This ensures that Alice cannot distinguish all parameters by sending only 0, which would result in perfect covertness.
- 2) There exists  $\theta \in \Theta$  such that no distribution  $P$  over  $\mathcal{U} \setminus \{0\}$  is such that  $\sum_u P(u) \rho_{W|\theta}^u = \rho_{W|\theta}^0$ . This ensures that Alice cannot simulate sending 0 by a random selection of other inputs.
- 3) For all  $\theta \in \Theta$  and for all  $u \in \mathcal{U}$   $\text{supp}(\rho_{W|\theta}^u) \subset \text{supp}(\rho_{W|\theta}^0)$ . This ensures that Willie cannot detect the estimation with non-vanishing probability when a state with support not included in  $\text{supp}(\rho_{W|\theta}^0)$  is transmitted.

We introduce the notion of conditional Chernoff information to state our main result.

*Definition 1:* Let  $\theta$  and  $\theta'$  be two parameters in  $\Theta$  and  $P$  be a distribution over  $\mathcal{U}$ . The conditional Chernoff information is

$$\mathbb{C}(\theta \| \theta' | P) \triangleq \sup_{s \in [0, 1]} - \sum_u P(u) \log \left( \text{tr} \left( \left( \rho_{B|\theta}^u \right)^s \left( \rho_{B|\theta'}^u \right)^{1-s} \right) \right). \quad (13)$$

*Theorem 1:* Under the assumption on the cq-channel discussed above, we have  $-\log E(n, \delta_n) = \Theta(\sqrt{n \delta_n})$  for any

<sup>2</sup>Note that this definition of estimation error is only meaningful for finite parameter sets.

sequence  $\{\delta_n\}_{n \geq 1} = \mathcal{O}(1) \cap \omega(\frac{\log n}{n})$ , and in particular,

$$\begin{aligned} & \lim_{n \rightarrow \infty} -\frac{\log E(n, \delta_n)}{\sqrt{n\delta_n}} \\ &= \sup_P \frac{\sqrt{2} \min_{\theta \neq \theta': \rho_{B|\theta}^u = \rho_{B|\theta'}^0} \mathbb{C}(\theta \| \theta' | P)}{\sqrt{\max_{\theta} \eta \left( \sum_u P(u) \rho_{W|\theta}^u \| \rho_{W|\theta}^0 \right)}}, \end{aligned} \quad (14)$$

where the supremum is taken over all probability distributions  $P$  over  $\mathcal{U} \setminus \{0\}$ .

*Remark 2* [28, Th. 1]: can be obtained as a special case of Theorem 1, corresponding to the situation in which all operators in  $\{\rho_{B|\theta}^u\}_{u \in \mathcal{U}, \theta \in \Theta}$  and all operators in  $\{\rho_{W|\theta}^u\}_{u \in \mathcal{U}, \theta \in \Theta}$  mutually commute.

### B. Unitary Channels: The Power of Entanglement

We now consider a situation in which  $\{U_{\theta}\}_{\theta \in \Theta}$  is a family of unitaries acting on  $A$  such that  $\mathcal{N}_{\theta}(\rho) = U_{\theta} \rho U_{\theta}^{\dagger}$  for all  $\theta \in \Theta$ . This corresponds to an ideal situation in which Alice is able to receive all transmitted signals without any loss. We also assume that  $\mathcal{E}_{\theta}$  is independent of  $\theta$ , i.e., there exists a quantum channel  $\mathcal{E} : \mathcal{L}(A) \rightarrow \mathcal{L}(W)$  such that  $\mathcal{E}_{\theta} = \mathcal{E}$  for all  $\theta \in \Theta$ . The latter assumption helps us simplify the expression of our results but our proof does not exploit this assumption. Also note that this assumption does not trivialize the problem since we still require the sensing to be covert.

1) *Achievability*: We require again mild assumptions in our achievability result to make the problem meaningful.

- 1)  $U_{\theta} \neq U_{\theta'}$  for all  $\theta \neq \theta'$ . Without this assumption Alice would be unable to distinguish at least two parameters.
- 2)  $\text{supp}(\mathcal{E}(\rho)) \subset \text{supp}(\mathcal{E}(|0\rangle\langle 0|))$  for all  $\rho \in \mathcal{D}(A)$ . Without this assumption, the transmission of such  $\rho$  would allow Willie to systematically detect Alice.

*Theorem 2*: Under the above assumptions, there exists a positive integer  $N$  depending on  $\{U_{\theta}\}_{\theta \in \Theta}$  such that for all  $n \geq N$

$$C(n, 0) \leq \mathcal{O}_{\{U_{\theta}\}, \mathcal{E}}\left(\frac{1}{n}\right). \quad (15)$$

Note that there is a significant difference between the optimal performance of unitary channels and cq-channels. Indeed, according to Theorem 1, we have  $C(n, \exp(-\mathcal{O}(\sqrt{n}\delta))) \geq \delta$  for all  $\delta > 0$  and for all cq-channels, while  $C(n, 0) \leq \mathcal{O}(1/n)$  when Alice's channel is a unitary for all parameters  $\theta$ . As we show next, the rate of decay of the covertness metric with  $n$  is optimal under mild assumptions.

2) *Converse*: Our converse result holds under the following mildly restrictive assumptions.

- 1)  $U_{\theta}|0\rangle = U_{\theta'}|0\rangle$  for some  $\theta \neq \theta'$ , i.e., Alice cannot distinguish all parameters by always sending  $|0\rangle$  and trivially ensuring covertness.
- 2)  $\mathcal{E}(\rho) \neq \mathcal{E}(|0\rangle\langle 0|)$  for all  $\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}$ , i.e., Alice cannot simulate sending  $|0\rangle$  using other quantum states.
- 3) There exists no sequence  $\{\rho_n\}_{n \geq 1} \subset \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}$  such that

$$\lim_{n \rightarrow \infty} \frac{\|\rho_n - |0\rangle\langle 0|\|_1}{\|\mathcal{E}(\rho_n) - \mathcal{E}(|0\rangle\langle 0|)\|_1} = \infty. \quad (16)$$

This last assumption prevents Alice to send states whose image under  $\mathcal{E}$  is close to  $\mathcal{E}(|0\rangle\langle 0|)$ . We show that testing this assumption is possible by providing a *computable* equivalent form in Lemma 1.

*Theorem 3*: Under the above assumptions, for all  $\epsilon \in [0, 1]$ , we have

$$C(n, \epsilon) \geq \mathcal{O}_{\mathcal{E}}\left(\frac{(1 - \epsilon)^4}{n}\right). \quad (17)$$

*Remark 3*: Note that Theorem 2 and Theorem 3 require different assumptions because we need to restrict Willie's power for achievability and Alice's power for converse.

We now provide a computable equivalent form of our last assumption in the converse result. We first need the following definition, which introduces a map from  $\mathcal{L}(A)$  to a Euclidean space.

*Definition 2*: Let  $d \triangleq \dim A$  and  $\{|a_1\rangle, \dots, |a_d\rangle\}$  be an orthonormal basis for  $A$  such that  $|a_1\rangle = |0\rangle$ . We define a function  $f : \mathcal{L}(A) \rightarrow \mathbb{R}^{2d^2}$  which maps  $X \in \mathcal{L}(A)$  to the vector  $(\text{Re}(\langle a_i | X | a_j \rangle), \text{Im}(\langle a_i | X | a_j \rangle))_{i,j \in [1,d]}$ .

We also define  $2d - 2$  vectors  $a_1, \dots, a_{2d-2} \in \mathbb{R}^{2d^2}$  such that the  $j$ th component of  $a_i$  is  $\mathbb{1}\{i = j\} + \mathbb{1}\{j = [i/2]d + 1\}$ .

We show in Appendix E that the vectors  $a_1, \dots, a_{2d-2} \in \mathbb{R}^{2d^2}$  form a basis for the tangent space of

$$f(\{|\phi\rangle\langle \phi| : \phi \in A, \|\phi\| = 1\} \setminus |0\rangle\langle 0|) \quad (18)$$

at the origin.

*Lemma 1*: Let  $\mathcal{E} : \mathcal{L}(A) \rightarrow \mathcal{L}(B)$  be a quantum channel and  $|0\rangle \in A$  be a unit vector. Suppose that  $\mathcal{E}(\rho) \neq \mathcal{E}(|0\rangle\langle 0|)$  for all  $\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}$ . We then have

$$\sup_{\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}} \frac{\|\rho - |0\rangle\langle 0|\|_1}{\|\mathcal{E}(\rho) - \mathcal{E}(|0\rangle\langle 0|)\|_1} < \infty \quad (19)$$

if and only if  $f(\text{Ker}(\mathcal{E})) \cap \text{span}(a_1, \dots, a_{2d-2}) = \{0\}$  where the function  $f$  and the vectors  $a_1, \dots, a_{2d-2}$  are defined in Definition 2.

*Proof*: See Appendix E. ■

*Remark 4*: If  $\mathcal{E} : \mathcal{L}(A) \rightarrow \mathcal{L}(E)$  is an invertible map (as a linear map), there exists no sequence satisfying the above conditions, which is consistent with our result as  $\text{Ker}(\mathcal{E}) = \{0\}$ . However, there might be some quantum channels  $\mathcal{E}$  that are not invertible, but for which we still have  $f(\text{Ker}(\mathcal{E})) \cap \text{span}\{a_1, \dots, a_{2d-2}\} = \{0\}$ .

## V. PROOFS

### A. Achievability Proof of Theorem 1

We first derive a general bound on the reliability of a strategy when the input is generated according to  $P_U$ .

*Lemma 2*: Let  $P_U$  be any distribution over  $\mathcal{U}^n$ . There exists an  $(n, \epsilon, \delta)$  non-adaptive strategy with

$$\begin{aligned} \log \epsilon &= \max_{\theta \neq \theta'} \log \left( \sum_{Q \in \mathcal{P}_n(\mathcal{U})} P_U(\mathcal{T}_Q) \exp(-nC(\theta \| \theta' | Q)) \right) \\ &\quad + \mathcal{O}_{\dim B, |\mathcal{U}|, |\Theta|}(\log n) \end{aligned} \quad (20)$$



and

$$\delta = \max_{\theta \in \Theta} \mathbb{D} \left( \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \rho_{\mathbf{W}|\theta}^{\mathbf{u}} \| \rho_{\mathbf{W}|\theta}^0 \right). \quad (21)$$

*Proof:* See Appendix B. ■

Deterministic strategies, for which  $P_{\mathbf{U}}$  is positive only on one input sequence, cannot achieve any positive exponent as shown next. Let  $\tilde{\Theta}$  be as defined in Section IV-A and  $\theta \in \tilde{\Theta}$  be such that no distribution  $P$  over  $\mathcal{U} \setminus \{0\}$  is such that  $\sum_u P(u) \rho_{W|\theta}^u = \rho_{W|\theta}^0$ . If Alice transmits a fixed sequence  $\mathbf{u}$ , we have

$$\delta \geq \mathbb{D}(\rho_{\mathbf{W}|\theta}^{\mathbf{u}} \| \rho_{\mathbf{W}|\theta}^0) \quad (22)$$

$$\geq |\{i \in [1, n] : u_i \neq 0\}| \min_{u \in \mathcal{U} \setminus \{0\}} \mathbb{D}(\rho_{W|\theta}^u \| \rho_{W|\theta}^0). \quad (23)$$

By our assumption on  $\theta$ ,  $\min_{u \in \mathcal{U} \setminus \{0\}} \mathbb{D}(\rho_{W|\theta}^u \| \rho_{W|\theta}^0)$  is positive. Therefore, the number of non-zero elements of  $\mathbf{u}$  is uniformly bounded. By definition of  $\tilde{\Theta}$ , there also exists  $\theta' \in \Theta \setminus \{\theta\}$  such that  $\rho_{B|\theta}^0 = \rho_{B|\theta'}^0$ . Thus, even when restricting the parameter set to  $\{\theta, \theta'\}$ , the estimation error cannot vanish. Hence, no positive exponent is achievable.

Furthermore, independent and identically distributed (i.i.d.) actions cannot achieve the optimal exponent since, with exponentially small probability, the type of the input sequence largely deviates from the typical input type and affects the achievable exponent.

We now introduce an input probability distribution  $P_{\mathbf{U}}$  that circumvents the challenges discussed above. Intuitively,  $P_{\mathbf{U}}$  should be such that 1) we can control the type of the sequences in its support and 2) we can ensure covertness. Let  $P$  be any distribution over  $\mathcal{U}$  and define  $\alpha \triangleq 1 - P(0)$ . We set for  $\zeta > 0$ ,

$$\mathcal{Q} \triangleq \{Q \in \mathcal{P}_n(\mathcal{U}) : |Q(u) - P(u)| \leq \alpha \zeta \ \forall u \in \mathcal{U} \setminus \{0\}\}, \quad (24)$$

$$\mathcal{A} \triangleq \cup_{Q \in \mathcal{Q}} \mathcal{T}_Q, \quad (25)$$

$$P_{\mathbf{U}}(\mathbf{u}) \triangleq \begin{cases} \frac{P^{\otimes n}(\mathbf{u})}{P^{\otimes n}(\mathcal{A})} & \mathbf{u} \in \mathcal{A}, \\ 0 & \mathbf{u} \notin \mathcal{A}. \end{cases} \quad (26)$$

Intuitively, the parameter  $\alpha$  allows us to finely control the type of sequences in  $\mathcal{A}$  with  $\alpha$  possibly depending on  $n$ . In the following lemma, we provide bounds on (20) and (21) for this specific choice of  $P_{\mathbf{U}}$ .

*Lemma 3:* Let  $\theta$  and  $\theta'$  be two distinct elements of  $\Theta$ . We have

$$\begin{aligned} & \log \left( \sum_{Q \in \mathcal{P}_n(\mathcal{U})} P_{\mathbf{U}}(\mathcal{T}_Q) \exp(-n\mathbb{C}(\theta \| \theta' | Q)) \right) \\ & \leq \min \left( -n\mathbb{C}(\theta \| \theta' | P) - \mathcal{O}_{\{\rho_{B|\theta}^u\}}(n\alpha\zeta|\mathcal{U}|), \right. \\ & \quad \left. n \left( 1 - \alpha(1 + \zeta|\mathcal{U}|) \right) \right. \\ & \quad \left. \times \inf_{s \in [0, 1]} \log \left( \text{tr} \left( \left( \rho_{B|\theta}^0 \right)^s \left( \rho_{B|\theta'}^0 \right)^{1-s} \right) \right) \right). \end{aligned} \quad (27)$$

In addition, we have

$$\begin{aligned} \mathbb{D} \left( \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \rho_{\mathbf{W}|\theta}^{\mathbf{u}} \| \rho_{\mathbf{W}|\theta}^0 \right) & \leq n \mathbb{D} \left( \sum_u P(u) \rho_{W|\theta}^u \| \rho_{W|\theta}^0 \right) \\ & \quad + 2|\mathcal{U}| \exp \left( -\frac{\alpha n \zeta^2}{3} \right) \\ & \quad \times \log \left( \frac{\dim W}{\lambda_{\min}(\rho_{W|\theta}^0)} \right)^n \\ & \quad + \mathbb{H}_b \left( 2|\mathcal{U}| \exp \left( -\frac{\alpha n \zeta^2}{3} \right) \right). \end{aligned} \quad (28)$$

*Proof:* See Appendix C. ■

We are now ready to prove the achievability of the exponent in (14). Let  $\bar{P}$  be any distribution over  $\mathcal{U} \setminus \{0\}$  (not depending on  $n$ ) and  $\{\lambda_n\}_{n \geq 0}$  be a vanishing sequence specified later. We define

$$\alpha_n \triangleq \sqrt{\frac{2\delta_n(1 - \lambda_n)}{n \max_{\theta} \eta \left( \sum_u \bar{P}(u) \rho_{W|\theta}^u \| \rho_{W|\theta}^0 \right)}}, \quad (29)$$

$$P(u) \triangleq \begin{cases} 1 - \alpha_n & u = 0, \\ \alpha_n \bar{P}(u) & u \neq 0. \end{cases} \quad (30)$$

We then choose  $P_{\mathbf{U}}$  according to (26), for which we have

$$\mathbb{D} \left( \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \rho_{\mathbf{W}|\theta}^{\mathbf{u}} \| \rho_{\mathbf{W}|\theta}^0 \right) \quad (31)$$

$$\stackrel{(a)}{\leq} n \mathbb{D} \left( \sum_u P(u) \rho_{W|\theta}^u \| \rho_{W|\theta}^0 \right) + 2|\mathcal{U}| \exp \left( -\frac{\alpha_n n \zeta^2}{3} \right) \\ \times \log \frac{\dim W}{\lambda_{\min}(\rho_{W|\theta}^0)} n + \mathbb{H}_b \left( 2|\mathcal{U}| \exp \left( -\frac{\alpha_n n \zeta^2}{3} \right) \right) \quad (32)$$

$$\stackrel{(b)}{=} n \mathbb{D} \left( \sum_u P(u) \rho_{W|\theta}^u \| \rho_{W|\theta}^0 \right) + \exp(-\omega(\log n)) \quad (33)$$

$$\stackrel{(c)}{=} n \left( \frac{\alpha_n^2}{2} \eta \left( \sum_u \bar{P}(u) \rho_{W|\theta}^u \| \rho_{W|\theta}^0 \right) + \mathcal{O}(\alpha_n^3) \right) \\ + \exp(-\omega(\log n)) \quad (34)$$

$$\stackrel{(d)}{\leq} (1 - \lambda_n) \delta_n + \mathcal{O}(\alpha_n^3 n) + \exp(-\omega(\log n)) \quad (35)$$

$$= (1 - \lambda_n) \delta_n + \mathcal{O} \left( \frac{\delta_n^{\frac{3}{2}}}{\sqrt{n}} \right) + \exp(-\omega(\log n)) \quad (36)$$

where (a) follows from (28), (b) follows since we are choosing  $\delta_n = \omega(\frac{\log n}{n})$ , (c) follows from Lemma 5 in Appendix A and (d) follows from (29). We set

$$\lambda_n = \mathcal{O} \left( \frac{\delta_n^{\frac{1}{2}}}{\sqrt{n}} \right) + \frac{1}{\delta_n} \exp(-\omega(\log n)), \quad (37)$$

which is vanishing and ensures that  $\mathbb{D}(\sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \rho_{\mathbf{W}|\theta}^{\mathbf{u}} \| \rho_{\mathbf{W}|\theta}^0) \leq \delta_n$ . Therefore, by Lemma 2,

there exists an  $(n, \epsilon_n, \delta_n)$  with

$$\log \epsilon_n = \max_{\theta \neq \theta'} \log \left( \sum_{Q \in \mathcal{P}_n(\mathcal{U})} P_U(\mathcal{T}_Q) \exp(-n\mathbb{C}(\theta \| \theta' | Q)) \right) + \mathcal{O}_{\dim B, |\mathcal{U}|, |\Theta|}(\log n). \quad (38)$$

To upper-bound  $\epsilon_n$ , we consider two cases for  $\theta$  and  $\theta'$ . If  $\rho_{B|\theta}^0 = \rho_{B|\theta'}^0$ , then (27) yields that

$$-\log \left( \sum_{Q \in \mathcal{P}_n(\mathcal{U})} P_U(\mathcal{T}_Q) \exp(-n\mathbb{C}(\theta \| \theta' | Q)) \right) \quad (39)$$

$$\geq n\mathbb{C}(\theta \| \theta' | P) + \mathcal{O}_{\{\rho_{W|\theta}^u\}}(n\alpha_n \zeta |\mathcal{U}|) \quad (40)$$

$$\stackrel{(a)}{\geq} n\alpha_n \left( 1 + \zeta |\mathcal{U}| \mathcal{O}_{\{\rho_{W|\theta}^u\}}(1) \right) \mathbb{C}(\theta \| \theta' | \bar{P}), \quad (41)$$

where (a) follows from the definition of  $\bar{P}$  and since all terms in the definition of  $\mathbb{C}(\theta \| \theta' | P)$  are non-negative. If  $\rho_{B|\theta}^0 \neq \rho_{B|\theta'}^0$ , we have

$$\begin{aligned} & -\log(P_U(\mathcal{T}_Q) \exp(-n\mathbb{C}(\theta \| \theta' | P))) \\ & \geq -(1 - \alpha_n(1 + \zeta |\mathcal{U}|)) \\ & \quad \times n \inf_{s \in [0, 1]} \log \left( \text{tr} \left( (\rho_{B|\theta}^0)^s (\rho_{B|\theta'}^0)^{1-s} \right) \right) \stackrel{(a)}{=} \Theta(n), \end{aligned} \quad (42)$$

where (a) follows since  $\inf_{s \in [0, 1]} \log(\text{tr}((\rho_{B|\theta}^0)^s (\rho_{B|\theta'}^0)^{1-s})) < 0$  when  $\rho_{B|\theta}^0 \neq \rho_{B|\theta'}^0$ . Therefore, we can exclude all pairs  $(\theta, \theta')$  with  $\rho_{B|\theta}^0 \neq \rho_{B|\theta'}^0$  from the maximization in (38) for large enough  $n$ . Thus, using Lemma 2 and (41), we have

$$\begin{aligned} -\log \epsilon_n & \geq n\alpha_n \left( 1 + \zeta |\mathcal{U}| \mathcal{O}_{\{\rho_{W|\theta}^u\}}(1) \right) \\ & \quad \times \min_{\theta, \theta': \rho_{B|\theta}^0 = \rho_{B|\theta'}^0} \mathbb{C}(\theta \| \theta' | \bar{P}) \\ & \quad + \mathcal{O}_{\dim B, |\mathcal{U}|, |\Theta|}(\log n). \end{aligned} \quad (43)$$

Using the definition of  $\alpha_n$  in (29), we obtain

$$\liminf_{n \rightarrow \infty} -\frac{\log \epsilon_n}{\sqrt{\delta_n n}} \geq \frac{\sqrt{2} \min_{\theta \neq \theta': \rho_{B|\theta}^0 = \rho_{B|\theta'}^0} \mathbb{C}(\theta \| \theta' | \bar{P})}{\sqrt{\max_{\theta} \eta \left( \sum_u \bar{P}(u) \rho_{W|\theta}^u \| \rho_{W|\theta}^0 \right)}}. \quad (44)$$

### B. Converse Proof of Theorem 1

Let us consider a sequence of  $(n, \epsilon_n, \delta_n)$  non-adaptive strategies, for which the input is generated according to a PMF  $P_U$  over  $\mathcal{U}^n$  in the  $n^{\text{th}}$  strategy. We define

$$\bar{P} \triangleq \frac{1}{n} \sum_{i=1}^n P_{U_i}, \quad (45)$$

$$\alpha_n \triangleq 1 - \bar{P}(0), \quad (46)$$

$$\tilde{P}(u) \triangleq \begin{cases} \frac{\bar{P}(u)}{\alpha_n} & u \neq 0, \\ 0 & u = 0. \end{cases} \quad (47)$$

*Lemma 4:* We have

$$\begin{aligned} -\log \epsilon_n & \leq \min_{\theta \neq \theta': \rho_{B|\theta}^0 = \rho_{B|\theta'}^0} n\alpha_n \mathbb{C}(\theta \| \theta' | \tilde{P}) \\ & \quad - \mathcal{O}_{\dim B, |\mathcal{U}|, |\Theta|} \left( \frac{\log n}{n} \right). \end{aligned} \quad (48)$$

As  $n$  goes to infinity,  $\alpha_n$  tends to zero and we have

$$\frac{\delta_n}{n} \geq \frac{\alpha_n^2}{2} \max_{\theta} \eta \left( \sum_u \tilde{P}(u) \rho_{W|\theta}^u \| \rho_{W|\theta}^0 \right) + \mathcal{O}_{\{\rho_{W|\theta}^0\}}(\alpha_n^3). \quad (49)$$

*Proof:* See Appendix D. ■

We therefore have Eq. (52) on the bottom of the next page, where (a) follows from (48), and (b) follows from (49) and the constraint  $\delta_n = \omega(\log n/n)$ . Taking the limit as  $n$  goes to infinity, we obtain the desired converse bound.

### C. Proof of Theorem 2

We first recall from [27] that given a unitary  $U \neq \mathbf{1}_A$  acting on  $A$ , one can find a positive integer  $m$  and a pure state  $|\nu\rangle_{A^m}$  such that  $\langle \nu | U^{\otimes m} | \nu \rangle = 0$ . Applying this result to the unitary  $U_{\theta}^{\dagger} U_{\theta'} \neq \mathbf{1}_A$ , there exist a positive integer  $m_{\theta, \theta'}$  and pure state  $|\nu_{\theta, \theta'}\rangle_{A^{m_{\theta, \theta'}}}$  in  $A^{m_{\theta, \theta'}}$  such that  $\langle \nu_{\theta, \theta'} | (U_{\theta}^{\dagger} U_{\theta'})^{\otimes m_{\theta, \theta'}} | \nu_{\theta, \theta'} \rangle = 0$ . Let  $m \triangleq \sum_{\theta \neq \theta'} m_{\theta, \theta'}$  and  $|\nu\rangle_{A^m}$  be a pure state in  $A^m$  defined as the tensor product of all  $|\nu_{\theta, \theta'}\rangle_{A^{m_{\theta, \theta'}}}$  in an arbitrary order. Let  $\ell \triangleq \lfloor n/m \rfloor$ . Alice decomposes the first  $m\ell$  channel uses into  $\ell$  sub-blocks of length  $m$ , selects one sub-block at random, transmits  $|\nu\rangle_{A^m}$  on that sub-block, and transmits  $|0\rangle$  for any other channel use. By transmitting  $|\nu\rangle_{A^m}$ , Alice can estimate  $\theta$  without error.

We now analyze the covertness. Let us denote the state transmitted through the channels by

$$\phi_{A^n} \triangleq \frac{1}{\ell} \sum_{i=1}^{\ell} (|0\rangle\langle 0|)^{\otimes (i-1)m} \otimes |\nu\rangle\langle \nu|_{A^m} \otimes (|0\rangle\langle 0|)^{n-im}. \quad (53)$$

Note that

$$\mathbb{D}(\mathcal{E}^{\otimes n}(\phi_{A^n}) \| \mathcal{E}^{\otimes n}(|0\rangle\langle 0|^{\otimes n})) \quad (54)$$

$$\stackrel{(a)}{\leq} \frac{\chi_2(\mathcal{E}^{\otimes m}(|\nu\rangle\langle \nu|_{A^m}) \| \mathcal{E}^{\otimes m}(|0\rangle\langle 0|^{\otimes m}))}{\ell} \quad (55)$$

$$\leq \frac{1}{\ell \lambda_{\min}(\mathcal{E}(|0\rangle\langle 0|))^m} \quad (56)$$

$$\stackrel{(b)}{\leq} \frac{m}{(n-m) \lambda_{\min}(\mathcal{E}(|0\rangle\langle 0|))^m}, \quad (57)$$

where (a) follows from [29, eq. (B144)], and (b) follows since  $\ell > (n-m)/m$ . Since  $m$  is a constant independent of  $n$ , we obtain the desired bound on the covertness.

### D. Proof of Theorem 3

We consider a general strategy, in which Alice initially prepares  $|\phi\rangle_{RA^n}$  and, after receiving  $|\psi_{\theta}\rangle_{RA^n} \triangleq (\mathbf{1}_R \otimes U_{\theta}^{\otimes n})$  for an unknown parameter  $\theta$ , performs a POVM to estimate  $\theta$ . We assume that the estimation error as defined in (5) is  $\epsilon$  and the covertness as defined in (6) is  $\delta$ . We desire to prove that  $\delta \geq \mathcal{O}_{\mathcal{E}}((1-\epsilon)^4/n)$ . We show this result in three steps sketched as follows.

- 1) We first use the assumption that  $U_\theta|0\rangle = U_{\theta'}|0\rangle$  for some  $\theta \neq \theta'$  to show that  $\epsilon \geq 1 - 2\|\phi_{A^n} - (|0\rangle\langle 0|)^{\otimes n}\|_1$ .
- 2) We upper-bound  $\|\phi_{A^n} - (|0\rangle\langle 0|)^{\otimes n}\|_1$  by  $\mathcal{O}_{\mathcal{E}}((n \sum_{i=1}^n \mathbb{D}(\mathcal{E}(\phi_{A_i})\|\mathcal{E}(|0\rangle\langle 0|)))^{\frac{1}{4}})$ . The proof of this step relies on both our assumptions on  $\mathcal{E}$ , i.e.,  $\mathcal{E}(\rho) \neq \mathcal{E}(|0\rangle\langle 0|)$  for all  $\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}$  and

$$\sup_{\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}} \frac{\|\rho - |0\rangle\langle 0|\|_1}{\|\mathcal{E}(\rho) - \mathcal{E}(|0\rangle\langle 0|)\|_1} < \infty. \quad (58)$$

- 3) We use standard converse argument for covert communication to show that

$$\sum_{i=1}^n \mathbb{D}(\mathcal{E}(\phi_{A_i})\|\mathcal{E}(|0\rangle\langle 0|)) \leq \delta. \quad (59)$$

Combining these three steps yields that  $\epsilon \geq 1 - \mathcal{O}_{\mathcal{E}}((n\delta)^{\frac{1}{4}})$ , which is equivalent to  $\delta \geq \mathcal{O}_{\mathcal{E}}((1-\epsilon)^4/n)$  as desired. We now prove each step.

*a) Proof of step 1:* The estimation error,  $\epsilon$ , is lower-bounded by

$$\max_{\theta \neq \theta'} |\langle \psi_\theta | \psi_{\theta'} \rangle|^2 \quad (60)$$

$$= \max_{\theta \neq \theta'} \left| \langle \phi | \left( \mathbf{1}_R \otimes (U_\theta^\dagger U_{\theta'})^{\otimes n} \right) | \phi \rangle \right|^2 \quad (61)$$

$$= \max_{\theta \neq \theta'} \left| \text{tr} \left( \phi_{A^n} (U_\theta^\dagger U_{\theta'})^{\otimes n} \right) \right|^2 \quad (62)$$

$$= \max_{\theta \neq \theta'} \left| \text{tr} \left( (|0\rangle\langle 0|)^{\otimes n} - \phi_{A^n} \right) (U_\theta^\dagger U_{\theta'})^{\otimes n} \right| - \text{tr} \left( (|0\rangle\langle 0|)^{\otimes n} (U_\theta^\dagger U_{\theta'})^{\otimes n} \right) \quad (63)$$

$$\stackrel{(a)}{\geq} \min_{\theta \neq \theta'} \left| \text{tr} \left( (\phi_{A^n} - (|0\rangle\langle 0|)^{\otimes n}) (U_\theta^\dagger U_{\theta'})^{\otimes n} \right) - 1 \right|^2 \quad (64)$$

$$\stackrel{(b)}{\geq} \min_{\theta \neq \theta'} 1 - 2 \left| \text{tr} \left( (\phi_{A^n} - (|0\rangle\langle 0|)^{\otimes n}) (U_\theta^\dagger U_{\theta'})^{\otimes n} \right) \right| \quad (65)$$

$$\stackrel{(c)}{\geq} 1 - 2 \|\phi_{A^n} - (|0\rangle\langle 0|)^{\otimes n}\|_1 \quad (66)$$

$$(67)$$

where (a) follows because  $U_\theta|0\rangle = U_{\theta'}|0\rangle$  for some  $\theta \neq \theta'$ , (b) follows from  $|1-z|^2 = 1 - 2\text{Re}(z) + |z|^2 \geq 1 - 2|z|$  for any complex number  $z$ , and (c) follows from  $|\text{tr}(XY)| \leq \|X\| \|Y\|_1$  for all  $X, Y \in \mathcal{L}(A^n)$  and  $\|(U_\theta^\dagger U_{\theta'})^{\otimes n}\| = 1$ .

*b) Proof of step 2:* We have

$$\|\phi_{A^n} - (|0\rangle\langle 0|)^{\otimes n}\|_1 \stackrel{(a)}{\leq} \sqrt{1 - F(\phi_{A^n}, (|0\rangle\langle 0|)^{\otimes n})} \quad (68)$$

$$= \sqrt{1 - \langle 0|^{\otimes n} \phi_{A^n} | 0 \rangle^{\otimes n}} \quad (69)$$

$$\stackrel{(b)}{\leq} \sqrt{\sum_{i=1}^n (1 - \langle 0 | \phi_{A_i} | 0 \rangle)} \quad (70)$$

$$\stackrel{(c)}{\leq} \sqrt{\sum_{i=1}^n \| |0\rangle\langle 0| - \phi_{A_i} \|_1}, \quad (71)$$

where (a) follows from [30, Th. 1], (b) follows from (classical) union bound, and (c) follows since  $1 - F(\rho, \sigma) \leq \|\rho - \sigma\|_1$  when  $\rho$  is pure. We now state a lemma that allows us to bound  $\|\rho - |0\rangle\langle 0|\|_1$  using  $\|\mathcal{E}(\rho - |0\rangle\langle 0|)\|_1$ . By our assumption on  $\mathcal{E}$  there exists  $B > 0$  such that for all  $i \in [1, n]$ ,

$$\| |0\rangle\langle 0| - \phi_{A_i} \|_1 \leq B \|\mathcal{E}(|0\rangle\langle 0|) - \mathcal{E}(\phi_{A_i})\|_1 \quad (72)$$

$$\leq B \sqrt{\mathbb{D}(\mathcal{E}(\phi_{A_i})\|\mathcal{E}(|0\rangle\langle 0|))}. \quad (73)$$

This implies that

$$\|\phi_{A^n} - (|0\rangle\langle 0|)^{\otimes n}\|_1 \quad (74)$$

$$\leq \sqrt{B} \sqrt{\sum_{i=1}^n \sqrt{\mathbb{D}(\mathcal{E}(\phi_{A_i})\|\mathcal{E}(|0\rangle\langle 0|))}} \quad (75)$$

$$\leq \sqrt{B} \sqrt{n} \sqrt{\sum_{i=1}^n \mathbb{D}(\mathcal{E}(\phi_{A_i})\|\mathcal{E}(|0\rangle\langle 0|))} \quad (76)$$

*c) Proof of step 3:* We have

$$\mathbb{D}(\mathcal{E}^{\otimes n}(\phi_{A^n})\|\mathcal{E}^{\otimes n}(|0\rangle\langle 0|)^{\otimes n}) \quad (77)$$

$$= -H(\mathcal{E}^{\otimes n}(\phi_{A^n})) + \text{tr}(\mathcal{E}^{\otimes n}(\phi_{A^n}) \log((\mathcal{E}(|0\rangle\langle 0|))^{\otimes n})) \quad (78)$$

$$= -H(\mathcal{E}^{\otimes n}(\phi_{A^n})) + \sum_{i=1}^n \text{tr}(\mathcal{E}(\phi_{A_i}) \log(\mathcal{E}(|0\rangle\langle 0|))) \quad (79)$$

$$\geq -\sum_{i=1}^n H(\mathcal{E}(\phi_{A_i})) + \sum_{i=1}^n \text{tr}(\mathcal{E}(\phi_{A_i}) \log(\mathcal{E}(|0\rangle\langle 0|))) \quad (80)$$

$$= \sum_{i=1}^n \mathbb{D}(\mathcal{E}(\phi_{A_i})\|\mathcal{E}(|0\rangle\langle 0|)). \quad (81)$$

$$-\frac{\log \epsilon_n}{\sqrt{\delta_n n}} \stackrel{(a)}{\leq} \frac{n\alpha_n \min_{\theta \neq \theta': \rho_{B|\theta}^0 = \rho_{B|\theta'}^0} \mathbb{C}(\theta\|\theta'|\tilde{P}) - \mathcal{O}_{\dim B, |\mathcal{U}|, |\Theta|} \left( \frac{\log n}{n} \right)}{\sqrt{\delta_n n}} \quad (50)$$

$$\stackrel{(b)}{\leq} \frac{\sqrt{\frac{2n\delta_n(1+o(1))}{\max_{\theta} \eta \left( \sum_u \tilde{P}(u) \rho_{W|\theta}^u \|\rho_{W|\theta}^0 \right)}} \min_{\theta \neq \theta': \rho_{B|\theta}^0 = \rho_{B|\theta'}^0} \mathbb{C}(\theta\|\theta'|\tilde{P})}{\sqrt{\delta_n n}} \quad (51)$$

$$= \sqrt{\frac{2(1+o(1))}{\max_{\theta} \eta \left( \sum_u \tilde{P}(u) \rho_{W|\theta}^u \|\rho_{W|\theta}^0 \right)}} \min_{\theta \neq \theta': \rho_{B|\theta}^0 = \rho_{B|\theta'}^0} \mathbb{C}(\theta\|\theta'|\tilde{P}) \quad (52)$$

## APPENDIX A

## APPROXIMATION OF QUANTUM RELATIVE ENTROPY

We characterize in the next lemma the expansion of  $\mathbb{D}(\alpha\rho_1 + (1-\alpha)\rho_0\|\rho_0)$  in  $\alpha$  around zero.

*Lemma 5:* Let  $\rho_1$  and  $\rho_0$  be two density operators on  $A$  such that  $\rho_0$  is invertible. We have for small  $\alpha > 0$

$$\mathbb{D}(\alpha\rho_1 + (1-\alpha)\rho_0\|\rho_0) = \frac{1}{2}\alpha^2\eta(\rho_1\|\rho_0) + \mathcal{O}_{\rho_0}(\alpha^3). \quad (82)$$

*Remark 5:* This result is similar to [8, Lemma 1], but the expression for  $\eta(\rho_1\|\rho_0)$  in [8] is

$$\text{tr}\left(\int_0^\infty \rho_1(\rho_0 + s)^{-1} \rho_1(\rho_0 + s)^{-1} ds\right) - 1, \quad (83)$$

which involves an integration. In addition, in Lemma 5, the constant behind the higher order term is independent of  $\rho_1$ , which is not shown in [8, Lemma 1] and which is crucial in our converse argument.

We first recall two results from functional calculus before proving Lemma 5.

*Lemma 6:* Let  $X$  be a positive operator in  $\mathcal{L}(\mathcal{H})$  with eigen-decomposition  $X = \sum_{i=1}^q \lambda_i P_i$ , where  $\lambda_1, \dots, \lambda_q$  are distinct eigenvalue of  $X$  and  $P_i$  is the projection onto the eigensubspace corresponding to  $\lambda_i$ . There exists  $\epsilon > 0$  such that for all  $Y$  with  $\|X - Y\| \leq \epsilon$ ,  $X + Y$  is positive and

$$\log(X + Y) = \log(X) + \sum_{i,j} D_{i,j} P_i Y P_j + \mathcal{O}_X(\|Y\|^2), \quad (84)$$

where

$$D_{i,j} = \begin{cases} \frac{\log \lambda_i - \log \lambda_j}{\lambda_i - \lambda_j} & i \neq j \\ \frac{1}{\lambda_i} & i = j. \end{cases} \quad (85)$$

*Proof:* It follows from applying [31, Th. 4.2] to the function  $\log(\cdot)$ . ■

*Lemma 7:* Let  $I \subset \mathbb{R}$  be an open interval in  $\mathbb{R}$  and  $f : I \rightarrow \mathbb{R}$  be a smooth function. Let  $A$  and  $B$  be two self-adjoint operator in  $\mathcal{L}(\mathcal{H})$ . We define  $g(t) \triangleq \text{tr}(f(A + tB))$  for all  $t$  such that all eigenvalues of  $A + tB$  are in  $I$ . Then, the domain of  $g$  is open and for each  $t$  in the domain of  $g$ ,

$$g'(t) = \text{tr}(f'(A + tB)B). \quad (86)$$

*Proof:* See [32, eq. (11.176)]. ■

We now prove Lemma 5. Let  $\rho_0$  has eigen-decomposition  $\sum_i \lambda_i P_i$  and define

$$D_{i,j} = \begin{cases} \frac{\log \lambda_i - \log \lambda_j}{\lambda_i - \lambda_j} & i \neq j \\ \frac{1}{\lambda_i} & i = j. \end{cases} \quad (87)$$

We also define  $\Delta \triangleq \rho_1 - \rho_0$  and

$$g(\alpha) \triangleq \mathbb{D}(\alpha\rho_1 + (1-\alpha)\rho_0\|\rho_0) \quad (88)$$

$$= \mathbb{D}(\rho_0 + \alpha\Delta\|\rho_0) \quad (89)$$

$$= \text{tr}((\rho_0 + \alpha\Delta) \log(\rho_0 + \alpha\Delta)) - \text{tr}((\rho_0 + \alpha\Delta) \log(\rho_0)). \quad (90)$$

Note that

$$g'(\alpha) \quad (91)$$

$$\stackrel{(a)}{=} \text{tr}(\Delta(\log(\rho_0 + \alpha\Delta) + \mathbf{1})) - \text{tr}(\Delta \log \rho_0) \quad (92)$$

$$= \text{tr}(\Delta(\log(\rho_0 + \alpha\Delta) - \log(\rho_0))) \quad (93)$$

$$\stackrel{(b)}{=} \text{tr}\left(\Delta\left(\sum_{i,j} D_{i,j} P_i(\alpha\Delta) P_j + \mathcal{O}_{\rho_0}(\|\alpha\Delta\|^2)\right)\right) \quad (94)$$

$$= \text{tr}\left(\Delta\left(\sum_{i,j} D_{i,j} P_i(\alpha\Delta) P_j + \mathcal{O}_{\rho_0}(\|\alpha\Delta\|^2)\right)\right) \quad (95)$$

$$= \alpha \sum_{i,j} D_{i,j} \text{tr}(\Delta P_i \Delta P_j) + \text{tr}(\Delta \mathcal{O}_{\rho_0}(\|\alpha\Delta\|^2)) \quad (96)$$

$$\stackrel{(c)}{=} \alpha \sum_{i,j} D_{i,j} \text{tr}(\Delta P_i \Delta P_j) + \mathcal{O}_{\rho_0}(\alpha^2) \quad (97)$$

$$= \alpha \eta(\rho_1\|\rho_0) + \mathcal{O}_{\rho_0}(\alpha^2) \quad (98)$$

where (a) follows from Lemma 7, (b) follows from Lemma 6, and (c) follows since the norm of  $\Delta = \rho_1 - \rho_0$  is bounded for all density operators  $\rho_0$  and  $\rho_1$ . We then have

$$\left|g(\alpha) - \frac{1}{2}\alpha^2\eta(\rho_1\|\rho_0)\right| \stackrel{(a)}{=} \left|\int_0^\alpha (g'(\beta) - \beta\eta(\rho_1\|\rho_0))d\beta\right| \quad (99)$$

$$\leq \int_0^\alpha |g'(\beta) - \beta\eta(\rho_1\|\rho_0)|d\beta \quad (100)$$

$$\stackrel{(b)}{=} \int_0^\alpha \mathcal{O}_{\rho_0}(\beta^2)d\beta \quad (101)$$

$$= \mathcal{O}_{\rho_0}(\alpha^3) \quad (102)$$

where (a) follows from the fundamental theorem of calculus and (b) follows from Eq. (98).

## APPENDIX B

## PROOF OF LEMMA 2

We first recall a result from [26] on the optimal performance of discriminating multiple quantum states.

*Lemma 8:* Let  $\{\rho_\theta\}_{\theta \in \Theta}$  be a finite family of density operators acting on a finite dimensional space. There exists a Positive Operator Valued Measurement (POVM)  $\{\Gamma_\theta\}_{\theta \in \Theta}$  such that

$$\begin{aligned} \max_{\theta \in \Theta} \text{tr}(\rho_\theta(\mathbf{1} - \Gamma_\theta)) &\leq 10(|\Theta| - 1)^2 \\ &\times \max_{\theta \in \Theta} \nu(\rho_\theta) \sum_{\theta' \neq \theta} \inf_{s \in [0,1]} \text{tr}(\rho_\theta^s \rho_{\theta'}^{1-s}). \end{aligned} \quad (103)$$

*Proof:* It follows from combining [26, Th. 2] and [26, eq. (35)]. ■

Alice samples the input sequence  $\mathbf{u}$  according to  $P_{\mathbf{U}}$  and receives  $\rho_{\mathbf{B}|\theta}^{\mathbf{u}} \triangleq \rho_{B_1|\theta}^{u_1} \otimes \dots \otimes \rho_{B_n|\theta}^{u_n}$ . Alice then performs the POVM  $\{\Gamma_\theta^{\mathbf{u}} : \theta \in \Theta\}$  given by Lemma 8 for the states  $\{\rho_{\mathbf{B}|\theta}^{\mathbf{u}}\}_{\theta \in \Theta}$ , resulting in estimation error

$$\begin{aligned} &10(|\Theta| - 1)^2 \max_{\theta \in \Theta} \nu(\rho_{\mathbf{B}|\theta}^{\mathbf{u}}) \\ &\times \sum_{\theta' \neq \theta} \inf_{s \in [0,1]} \text{tr}\left((\rho_{\mathbf{B}|\theta}^{\mathbf{u}})^s (\rho_{\mathbf{B}|\theta'}^{\mathbf{u}})^{1-s}\right). \end{aligned} \quad (104)$$



Note that  $\max_{\theta \in \Theta} \nu(\rho_{\mathbf{B}|\theta}^{\mathbf{u}}) \leq (n+1)^{\dim B|\mathcal{U}|}$  and

$$\inf_{s \in [0,1]} \text{tr} \left( \left( \rho_{\mathbf{B}|\theta}^{\mathbf{u}} \right)^s \left( \rho_{\mathbf{B}|\theta'}^{\mathbf{u}} \right)^{1-s} \right) \quad (105)$$

$$= \inf_{s \in [0,1]} \prod_{i=1}^n \text{tr} \left( \left( \rho_{B|\theta}^{u_i} \right)^s \left( \rho_{B|\theta'}^{u_i} \right)^{1-s} \right) \quad (106)$$

$$= \exp \left( - \sup_{s \in [0,1]} - \sum_{i=1}^n \log \left( \text{tr} \left( \left( \rho_{B|\theta}^{u_i} \right)^s \left( \rho_{B|\theta'}^{u_i} \right)^{1-s} \right) \right) \right) \quad (107)$$

$$= \exp \left( -n \sup_{s \in [0,1]} - \sum_{u \in \mathcal{U}} T_{\mathbf{u}}(u) \log \left( \text{tr} \left( \left( \rho_{B|\theta}^u \right)^s \left( \rho_{B|\theta'}^u \right)^{1-s} \right) \right) \right) \quad (108)$$

$$= \exp(-n\mathbb{C}(\theta\|\theta'|T_{\mathbf{u}})). \quad (109)$$

This concludes the proof.

#### APPENDIX C PROOF OF LEMMA 3

We introduce a notation that simplifies our expressions. Let us define for  $u \in \mathcal{U}$  and  $s \in [0,1]$ ,

$$f(s; u) \triangleq \log \left( \text{tr} \left( \left( \rho_{W|\theta}^u \right)^s \left( \rho_{W|\theta'}^u \right)^{1-s} \right) \right), \quad (110)$$

which is always non-positive and

$$L_f \triangleq \min_{u' \in \mathcal{U} \setminus \{0\}} \min_{s' \in [0,1]} f(s'; u') > -\infty. \quad (111)$$

We then have

$$\log \left( \sum_{Q \in \mathcal{P}_n(\mathcal{U})} P_{\mathbf{U}}(\mathcal{T}_Q) \exp(-n\mathbb{C}(\theta\|\theta'|Q)) \right) \quad (112)$$

$$\leq \max_{Q \in \mathcal{P}_n(\mathcal{U}): P_{\mathbf{U}}(\mathcal{T}_Q) > 0} [-n\mathbb{C}(\theta\|\theta'|Q)] \quad (113)$$

$$= \max_{Q \in \mathcal{P}_n(\mathcal{U}): P_{\mathbf{U}}(\mathcal{T}_Q) > 0} \left[ -n \sup_{s \in [0,1]} - \sum_{u \in \mathcal{U}} Q(u) f(s; u) \right] \quad (114)$$

$$\stackrel{(a)}{\leq} \max_{Q \in \mathcal{P}_n(\mathcal{U}): P_{\mathbf{U}}(\mathcal{T}_Q) > 0} \left[ -n \sup_{s \in [0,1]} - \sum_{u \in \mathcal{U} \setminus \{0\}} Q(u) f(s; u) \right] \quad (115)$$

$$\stackrel{(b)}{\leq} -n \sup_{s \in [0,1]} - \sum_{u \in \mathcal{U} \setminus \{0\}} (P(u) - \zeta \alpha) f(s; u) \quad (116)$$

$$\leq \mathbb{C}(\theta\|\theta'|P) - n\alpha\zeta|\mathcal{U}|L_f, \quad (117)$$

where (a) follows since  $f(s; 0) \leq 0$ , and (b) follows since  $P_{\mathbf{U}}(\mathcal{T}_Q) > 0$  if and only if  $|P(u) - Q(u)| \leq \alpha\zeta$  for all  $u \in \mathcal{U} \setminus \{0\}$  (see Eq. (24)-(26)). Note also that  $Q(0) = 1 - \sum_{u \in \mathcal{U} \setminus \{0\}} Q(u) \geq 1 - \sum_{u \in \mathcal{U} \setminus \{0\}} (P(u) + \alpha\zeta) \geq P(0) - \alpha\zeta|\mathcal{U}| = 1 - \alpha(1 + \zeta|\mathcal{U}|)$  for all  $Q \in \mathcal{Q}$ . The same line of reasoning as

in Eq. (112)-(117) then provides that

$$\log \left( \sum_{Q \in \mathcal{P}_n(\mathcal{U})} P_{\mathbf{U}}(\mathcal{T}_Q) \exp(-n\mathbb{C}(\theta\|\theta'|Q)) \right) \leq -n(1 - \alpha(1 + \zeta|\mathcal{U}|)) \inf_{s \in [0,1]} f(s, 0), \quad (118)$$

which yields Eq. (27) together with Eq. (117).

Let  $\mathbf{U}$  be distributed according to  $P^{\otimes n}$  and  $T_{\mathbf{U}}$  denote its type, which is a random element of  $\mathcal{P}_n(\mathcal{U})$ . We have

$$1 - P^{\otimes n}(\mathcal{A}) \stackrel{(a)}{\leq} \sum_{u \in \mathcal{U} \setminus \{0\}} \mathbb{P}(|T_{\mathbf{U}}(u) - P(u)| \geq \zeta \alpha) \quad (119)$$

$$\stackrel{(b)}{=} \sum_{u \in \mathcal{U} \setminus \{0\}: P(u) > 0} \mathbb{P}(|T_{\mathbf{U}}(u) - P(u)| \geq \zeta \alpha) \quad (120)$$

$$\stackrel{(c)}{\leq} \sum_{u \in \mathcal{U} \setminus \{0\}: P(u) > 0} 2 \exp \left( -\frac{\alpha^2 n \zeta^2}{3P(u)} \right) \quad (121)$$

$$\stackrel{(d)}{\leq} 2|\mathcal{U}| \exp \left( -\frac{\alpha n \zeta^2}{3} \right), \quad (122)$$

where (a) follows from the union bound, (b) follows since  $T_{\mathbf{U}}(u) = 0$  with probability one when  $P(u) = 0$ , (c) follows from a Chernoff bound, and (d) follows since  $P(u) \leq 1 - P(0) = \alpha$  for all  $u \in \mathcal{U} \setminus \{0\}$ . Note that  $\frac{1}{2} \|P_{\mathbf{U}} - P^{\otimes n}\|_1 = 1 - P^{\otimes n}(\mathcal{A}) \leq 2|\mathcal{U}| \exp(-\frac{\alpha n \zeta^2}{3})$  by the definition of  $P_{\mathbf{U}}$ . Hence, the data processing inequality implies that

$$\begin{aligned} & \frac{1}{2} \left\| \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \rho_{\mathbf{W}|\theta}^{\mathbf{u}} - \sum_{\mathbf{u}} P^{\otimes n}(\mathbf{u}) \rho_{\mathbf{W}|\theta}^{\mathbf{u}} \right\|_1 \\ & \leq 2|\mathcal{U}| \exp \left( -\frac{\alpha n \zeta^2}{3} \right). \end{aligned} \quad (123)$$

Finally, the following continuity result for the relative entropy completes the proof of (28).

*Lemma 9:* Let  $\rho_{\mathbf{B}}$  and  $\sigma_{\mathbf{B}}$  be two density operators over  $B^n$  such that  $\frac{1}{2} \|\rho_{\mathbf{B}} - \sigma_{\mathbf{B}}\|_1 \leq \epsilon$ . We then have

$$\begin{aligned} & \left| \mathbb{D}(\rho_{\mathbf{B}} \| \rho_{\mathbf{W}|\theta}^0) - \mathbb{D}(\sigma_{\mathbf{B}} \| \rho_{\mathbf{W}|\theta}^0) \right| \\ & \leq \epsilon \log \left( \frac{\dim B}{(\lambda_{\min}(\rho_{\mathbf{W}|\theta}^0))^2} \right) n + \mathbb{H}_b(\epsilon). \end{aligned} \quad (124)$$

*Proof:* Note that

$$\left| \mathbb{D}(\rho_{\mathbf{B}} \| \rho_{\mathbf{W}|\theta}^0) - \mathbb{D}(\sigma_{\mathbf{B}} \| \rho_{\mathbf{W}|\theta}^0) \right| \quad (125)$$

$$= \left| H(\rho_{\mathbf{B}}) - H(\sigma_{\mathbf{B}}) + \text{tr} \left( (\rho_{\mathbf{B}} - \sigma_{\mathbf{B}}) \log(\rho_{\mathbf{W}|\theta}^0) \right) \right| \quad (126)$$

$$\leq |H(\rho_{\mathbf{B}}) - H(\sigma_{\mathbf{B}})| + \left| \text{tr} \left( (\rho_{\mathbf{B}} - \sigma_{\mathbf{B}}) \log(\rho_{\mathbf{W}|\theta}^0) \right) \right| \quad (127)$$

$$\stackrel{(a)}{\leq} \epsilon \log(\dim B) n + \mathbb{H}_b(\epsilon) + \left| \text{tr} \left( (\rho_{\mathbf{B}} - \sigma_{\mathbf{B}}) \log(\rho_{\mathbf{W}|\theta}^0) \right) \right| \quad (128)$$

$$\leq \epsilon \log(\dim B) n + \mathbb{H}_b(\epsilon) + \|\rho_{\mathbf{B}} - \sigma_{\mathbf{B}}\|_1 \left\| \log(\rho_{\mathbf{W}|\theta}^0) \right\| \quad (129)$$

$$\leq \epsilon \log(\dim B)n + \mathbb{H}_b(\epsilon) + 2\epsilon \log\left(\frac{1}{\lambda_{\min}(\rho_{W|\theta}^0)}\right)n \quad (130)$$

where (a) follows from Fannes' inequality. ■

#### APPENDIX D PROOF OF LEMMA 4

Eq. (49) follows from the same argument used to obtain [8, eq. (39)], except using Lemma 5 instead of [8, Lemma 1]. We prove Eq. (48) in four steps as summarized below.

- *Step 1:* We lower-bound the estimation error of this strategy by

$$\begin{aligned} & \max_{\theta \in \Theta} \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \text{tr}(\Gamma_{\theta}^{\mathbf{u}} \rho_{\mathbf{B}|\theta}^{\mathbf{u}}) \\ & \geq \frac{1}{|\Theta|} \max_{\theta \neq \theta'} \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \left(1 - \frac{1}{2} \|\rho_{\mathbf{B}|\theta}^{\mathbf{u}} - \rho_{\mathbf{B}|\theta'}^{\mathbf{u}}\|_1\right). \end{aligned} \quad (131)$$

- *Step 2:* Let us now consider the spectral decomposition of  $\rho_{\mathbf{B}|\theta}^{\mathbf{u}} = \sum_{y \in \mathcal{Y}} p_{\theta}^{\mathbf{u}}(y) |e_{\theta}^{\mathbf{u}}(y)\rangle \langle e_{\theta}^{\mathbf{u}}(y)|$ , where  $\mathcal{Y}$  is a set of size  $\dim B$ ,  $p_{\theta}^{\mathbf{u}}$  is a PMF over  $\mathcal{Y}$ , and  $\{|e_{\theta}^{\mathbf{u}}(y)\rangle : y \in \mathcal{Y}\}$  forms an orthonormal basis for  $B$ . We also define  $q_{\theta, \theta'}^{\mathbf{u}}(y, y') \triangleq p_{\theta}^{\mathbf{u}}(y) | \langle e_{\theta}^{\mathbf{u}}(y) | e_{\theta'}^{\mathbf{u}}(y') \rangle |^2$ , which is a PMF over  $\mathcal{Y} \times \mathcal{Y}$ , and  $q_{\theta}^{\mathbf{u}} \triangleq q_{\theta}^{\mathbf{u}_1} \otimes \cdots \otimes q_{\theta}^{\mathbf{u}_n}$ , which is a PMF over  $(\mathcal{Y} \times \mathcal{Y})^n$ . We shall show that

$$1 - \frac{1}{2} \|\rho_{\mathbf{B}|\theta}^{\mathbf{u}} - \rho_{\mathbf{B}|\theta'}^{\mathbf{u}}\|_1 \geq \frac{1}{2} \left(1 - \frac{1}{2} \|q_{\theta, \theta'}^{\mathbf{u}} - q_{\theta, \theta}^{\mathbf{u}}\|_1\right). \quad (132)$$

- *Step 3:* Let  $P_{V|U}$  and  $P_{\tilde{V}|U}$  be two conditional distributions,  $\mathbf{u} \in \mathcal{U}^n$  be a sequence with type  $T_U$ , and  $\mathbf{V}$  and  $\tilde{\mathbf{V}}$  be distributed according to  $P_{\mathbf{V}} = P_{V|U=u_1} \otimes \cdots \otimes P_{V|U=u_n}$  and  $P_{\tilde{\mathbf{V}}} = P_{\tilde{V}|U=u_1} \otimes \cdots \otimes P_{\tilde{V}|U=u_n}$ , respectively. We shall show that

$$\sum_{\mathbf{v}} \min(P_{\mathbf{V}}(\mathbf{v}), P_{\tilde{\mathbf{V}}}(\mathbf{v}))$$

$$\begin{aligned} & \geq \exp\left(-n \sup_{s \in [0,1]} \log\right. \\ & \quad \times \left(\sum_{\mathbf{u}} T_U(\mathbf{u}) \sum_{\mathbf{v}} P_{V|U}(\mathbf{v}|\mathbf{u})^s P_{\tilde{V}|U}(\mathbf{v}|\mathbf{u})^{1-s}\right) \\ & \quad \left. + \mathcal{O}\left(\frac{\log n}{n}\right)\right). \end{aligned} \quad (133)$$

- *Step 4:* We show that

$$\begin{aligned} & 1 - \frac{1}{2} \|\rho_{\mathbf{B}|\theta}^{\mathbf{u}} - \rho_{\mathbf{B}|\theta'}^{\mathbf{u}}\|_1 \\ & = \exp\left(-n \mathbb{C}(\theta \|\theta | T_U) + \mathcal{O}\left(\frac{\log n}{n}\right)\right), \end{aligned} \quad (134)$$

which concludes the proof together with Jensen's inequality and the convexity of the exponential function. We now provide the detailed proof of each step.

- a) *Proof of step 1:* Note that

$$\max_{\theta \in \Theta} \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \text{tr}(\Gamma_{\theta}^{\mathbf{u}} \rho_{\mathbf{B}|\theta}^{\mathbf{u}}) \quad (135)$$

$$\geq \frac{1}{|\Theta|} \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \sum_{\theta} \text{tr}(\Gamma_{\theta}^{\mathbf{u}} \rho_{\mathbf{B}|\theta}^{\mathbf{u}}) \quad (136)$$

$$\stackrel{(a)}{\geq} \frac{1}{|\Theta|} \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \max_{\theta \neq \theta'} \left(1 - \frac{1}{2} \|\rho_{\mathbf{B}|\theta}^{\mathbf{u}} - \rho_{\mathbf{B}|\theta'}^{\mathbf{u}}\|_1\right) \quad (137)$$

$$\geq \frac{1}{|\Theta|} \max_{\theta \neq \theta'} \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \left(1 - \frac{1}{2} \|\rho_{\mathbf{B}|\theta}^{\mathbf{u}} - \rho_{\mathbf{B}|\theta'}^{\mathbf{u}}\|_1\right), \quad (138)$$

where (a) follows from the variational characterization of the trace norm  $\frac{1}{2} \|\rho - \sigma\|_1 = \max_{0 \leq \Gamma \leq \mathbf{I}} \text{tr}(\Gamma(\rho - \sigma))$ .

b) *Proof of step 2:* The proof is in [25], but we provide the proof for completeness. We first define  $p_{\theta}^{\mathbf{u}} \triangleq p_{\theta}^{\mathbf{u}_1} \otimes \cdots \otimes p_{\theta}^{\mathbf{u}_n}$  and  $|e_{\theta}^{\mathbf{u}}(y)\rangle \triangleq |e_{\theta}^{\mathbf{u}_1}(y_1)\rangle \otimes \cdots \otimes |e_{\theta}^{\mathbf{u}_n}(y_n)\rangle$ . We have Eq. (144) on the bottom of this page, where (a) follows since  $(1-\Gamma) \succ (1-\Gamma)^2$  for all  $0 \leq \Gamma \leq \mathbf{I}$ , and (b) follows from  $|x|^2 + |y|^2 \geq |x+y|^2/2$  for any two complex numbers  $x$  and  $y$ .

$$1 - \frac{1}{2} \|\rho_{\mathbf{B}|\theta}^{\mathbf{u}} - \rho_{\mathbf{B}|\theta'}^{\mathbf{u}}\|_1 = \inf_{0 \leq \Gamma \leq \mathbf{I}} \left[ \text{tr}(\Gamma \rho_{\mathbf{B}|\theta}^{\mathbf{u}}) + \text{tr}((\mathbf{I} - \Gamma) \rho_{\mathbf{B}|\theta'}^{\mathbf{u}}) \right] \quad (139)$$

$$= \inf_{0 \leq \Gamma \leq \mathbf{I}} \left[ \sum_{\mathbf{y}} p_{\theta}^{\mathbf{u}}(\mathbf{y}) \langle e_{\theta}^{\mathbf{u}}(\mathbf{y}) | \Gamma | e_{\theta}^{\mathbf{u}}(\mathbf{y}) \rangle + \sum_{\mathbf{y}} p_{\theta'}^{\mathbf{u}}(\mathbf{y}) \langle e_{\theta'}^{\mathbf{u}}(\mathbf{y}) | (\mathbf{I} - \Gamma) | e_{\theta'}^{\mathbf{u}}(\mathbf{y}) \rangle \right] \quad (140)$$

$$\stackrel{(a)}{\geq} \inf_{0 \leq \Gamma \leq \mathbf{I}} \left[ \sum_{\mathbf{y}} p_{\theta}^{\mathbf{u}}(\mathbf{y}) \|\Gamma | e_{\theta}^{\mathbf{u}}(\mathbf{y}) \rangle\|_2^2 + \sum_{\mathbf{y}} p_{\theta'}^{\mathbf{u}}(\mathbf{y}) \|(\mathbf{I} - \Gamma) | e_{\theta'}^{\mathbf{u}}(\mathbf{y}) \rangle\|_2^2 \right] \quad (141)$$

$$= \inf_{0 \leq \Gamma \leq \mathbf{I}} \left[ \sum_{\mathbf{y}, \mathbf{y}'} p_{\theta}^{\mathbf{u}}(\mathbf{y}) | \langle e_{\theta'}^{\mathbf{u}}(\mathbf{y}') | \Gamma | e_{\theta}^{\mathbf{u}}(\mathbf{y}) \rangle |^2 + \sum_{\mathbf{y}, \mathbf{y}'} p_{\theta'}^{\mathbf{u}}(\mathbf{y}) | \langle e_{\theta}^{\mathbf{u}}(\mathbf{y}) | (\mathbf{I} - \Gamma) | e_{\theta'}^{\mathbf{u}}(\mathbf{y}') \rangle |^2 \right] \quad (142)$$

$$\geq \inf_{0 \leq \Gamma \leq \mathbf{I}} \sum_{\mathbf{y}, \mathbf{y}'} \min(p_{\theta}^{\mathbf{u}}(\mathbf{y}), p_{\theta'}^{\mathbf{u}}(\mathbf{y}')) \left( | \langle e_{\theta'}^{\mathbf{u}}(\mathbf{y}') | \Gamma | e_{\theta}^{\mathbf{u}}(\mathbf{y}) \rangle |^2 + | \langle e_{\theta}^{\mathbf{u}}(\mathbf{y}) | (\mathbf{I} - \Gamma) | e_{\theta'}^{\mathbf{u}}(\mathbf{y}') \rangle |^2 \right) \quad (143)$$

$$\stackrel{(b)}{\geq} \frac{1}{2} \sum_{\mathbf{y}, \mathbf{y}'} \min(p_{\theta}^{\mathbf{u}}(\mathbf{y}), p_{\theta'}^{\mathbf{u}}(\mathbf{y}')) | \langle e_{\theta}^{\mathbf{u}}(\mathbf{y}) | e_{\theta'}^{\mathbf{u}}(\mathbf{y}') \rangle |^2 \quad (144)$$

c) *Proof of step 3:* Deploying standard method of type arguments, we have

$$\sum_{\mathbf{v}} \min(P_{\mathbf{V}}(\mathbf{v}), P_{\tilde{\mathbf{V}}}(\mathbf{v})) \quad (145)$$

$$\geq \max_{T_{V|U} \in \mathcal{P}_n(\mathcal{V}|\mathbf{u})} \sum_{\mathbf{v} \in \mathcal{T}_{T_{V|U}}(\mathbf{u})} \min(P_{\mathbf{V}}(\mathbf{v}), P_{\tilde{\mathbf{V}}}(\mathbf{v})) \quad (146)$$

$$\stackrel{(a)}{\geq} \max_{T_{V|U} \in \mathcal{P}_n(\mathcal{V}|\mathbf{u})} (n+1)^{-|\mathcal{V}||\mathcal{U}|} \times \exp\left(-n \max \left( \mathbb{D}(T_{V|U} \| P_{V|U} | T_U), \mathbb{D}(T_{V|U} \| P_{\tilde{V}|U} | T_U) \right) \right) \quad (147)$$

$$= (n+1)^{-|\mathcal{V}||\mathcal{U}|} \times \exp\left(-n \min_{T_{V|U} \in \mathcal{P}_n(\mathcal{V}|\mathbf{u})} \max \left( \mathbb{D}(T_{V|U} \| P_{V|U} | T_U), \mathbb{D}(T_{V|U} \| P_{\tilde{V}|U} | T_U) \right) \right), \quad (148)$$

where (a) follows from [33, eq. (2.8)]. Next note that for an arbitrary conditional distribution  $Q_{V|U}$ , there exists  $T_{V|U} \in \mathcal{P}_n(\mathcal{V}|\mathbf{u})$  such that  $\Delta_u \triangleq \frac{1}{2} \|T_{V|U=u} - Q_{V|U=u}\|_1 \leq \frac{|\mathcal{V}|}{nT_U(u)}$  for all  $u \in \text{supp}(T_U)$ . Thus, for such a  $T_{V|U}$ , we have Eq. (155) on the bottom of this page, where (a) follows from Fannes' inequality, (b) follows since  $\mathbb{H}_b(x) \leq x \log \frac{e}{x}$ , and (c) follows since  $\Delta_u \leq \frac{|\mathcal{V}|}{nT_U(u)}$  by our choice of  $T_{V|U}$ .

Hence,

$$\begin{aligned} & \min_{T_{V|U} \in \mathcal{P}_n(\mathcal{V}|\mathbf{u})} \max \left( \mathbb{D}(T_{V|U} \| P_{V|U} | T_U), \mathbb{D}(T_{V|U} \| P_{\tilde{V}|U} | T_U) \right) \\ & \geq \min_{Q_{V|U} \in \mathcal{P}(\mathcal{V}|\mathcal{U})} \max \left( \mathbb{D}(Q_{V|U} \| P_{V|U} | T_U), \mathbb{D}(Q_{V|U} \| P_{\tilde{V}|U} | T_U) \right) - \mathcal{O}\left(\frac{\log n}{n}\right). \end{aligned} \quad (156)$$

Finally, [16, eq. (39)] implies that

$$\begin{aligned} & \min_{Q_{V|U} \in \mathcal{P}(\mathcal{V}|\mathcal{U})} \max \left( \mathbb{D}(Q_{V|U} \| P_{V|U} | T_U), \mathbb{D}(Q_{V|U} \| P_{\tilde{V}|U} | T_U) \right) \\ & = - \sup_{s \in [0,1]} \log \left( \sum_u T_U(u) \sum_v P_{V|U}(v|u)^s P_{\tilde{V}|U}(v|u)^{1-s} \right). \end{aligned} \quad (157)$$

d) *Proof of step 4:* Combining the result of step one and two, we have

$$\begin{aligned} & 1 - \frac{1}{2} \left\| \rho_{\mathbf{B}|\theta}^{\mathbf{u}} - \rho_{\mathbf{B}|\theta'}^{\mathbf{u}} \right\|_1 \\ & \geq \exp\left(-n \sup_{s \in [0,1]} \log \left( \sum_u T_U(u) \times \sum_{y,y'} \left( p_{\theta}^u(y) | \langle e_{\theta}^u(y) | e_{\theta'}^u(y') \rangle | \right)^2 \right)^s \right. \\ & \quad \times \left. \left( p_{\theta'}^u(y') | \langle e_{\theta}^u(y) | e_{\theta'}^u(y') \rangle | \right)^{1-s} \right) \\ & \quad + \mathcal{O}\left(\frac{\log n}{n}\right). \end{aligned} \quad (158)$$

Note that

$$\sum_{y,y'} \left( p_{\theta}^u(y) | \langle e_{\theta}^u(y) | e_{\theta'}^u(y') \rangle | \right)^2 \left( p_{\theta'}^u(y') | \langle e_{\theta}^u(y) | e_{\theta'}^u(y') \rangle | \right)^{1-s} \quad (159)$$

$$= \sum_{y,y'} p_{\theta}^u(y)^s p_{\theta'}^u(y')^{1-s} | \langle e_{\theta}^u(y) | e_{\theta'}^u(y') \rangle |^2 \quad (160)$$

$$\begin{aligned} & = \text{tr} \left( \left( \sum_y p_{\theta}^u(y)^s | e_{\theta}^u(y) \rangle \langle e_{\theta}^u(y) | \right) \right. \\ & \quad \times \left. \left( \sum_{y'} p_{\theta'}^u(y')^{1-s} | e_{\theta'}^u(y') \rangle \langle e_{\theta'}^u(y') | \right) \right) \end{aligned} \quad (161)$$

$$= \text{tr} \left( \left( \rho_{B|\theta}^{\theta} \right)^s \left( \rho_{B|\theta'}^{\theta'} \right)^{1-s} \right). \quad (162)$$

$$\left| \mathbb{D}(T_{V|U} \| P_{\tilde{V}|U} | T_U) - \mathbb{D}(Q_{V|U} \| P_{\tilde{V}|U} | T_U) \right| \quad (149)$$

$$\leq \sum_u T_U(u) \left| \mathbb{D}(T_{V|U=u} \| P_{\tilde{V}|U=u}) - \mathbb{D}(Q_{V|U=u} \| P_{\tilde{V}|U=u}) \right| \quad (150)$$

$$\leq \sum_u T_U(u) \left( |H(T_{V|U=u}) - H(Q_{V|U=u})| + \sum_v |T_{V|U}(v|u) - Q_{V|U}(v|u)| \log \frac{1}{P_{V|U}(v|u)} \right) \quad (151)$$

$$\stackrel{(a)}{\leq} \sum_u T_U(u) \left( \Delta_u \log |\mathcal{V}| + \mathbb{H}_b(\Delta_u) + \max_v \log \frac{1}{P_{V|U}(v|u)} \Delta_u \right) \quad (152)$$

$$\stackrel{(b)}{\leq} \sum_u T_U(u) \left( \Delta_u \log |\mathcal{V}| + \Delta_u \log \frac{e}{\Delta_u} + \max_v \log \frac{1}{P_{V|U}(v|u)} \Delta_u \right) \quad (153)$$

$$\stackrel{(c)}{\leq} \sum_u T_U(u) \left( \frac{|\mathcal{V}|}{nT_U(u)} \log |\mathcal{V}| + \frac{|\mathcal{V}|}{nT_U(u)} \log \frac{nT_U(u)e}{|\mathcal{V}|} + \max_v \log \frac{1}{P_{V|U}(v|u)} \frac{|\mathcal{V}|}{nT_U(u)} \right) \quad (154)$$

$$\leq \frac{|\mathcal{V}||\mathcal{U}| \log |\mathcal{V}|}{n} + \frac{|\mathcal{V}||\mathcal{U}| \log n}{n} + \max_{v,n} \log \frac{1}{P_{V|U}} \frac{|\mathcal{V}|}{n} = \mathcal{O}\left(\frac{\log n}{n}\right) \quad (155)$$

Substituting Eq. (162) into Eq. (158), we have

$$1 - \frac{1}{2} \|\rho_{\mathbf{B}|\theta}^{\mathbf{u}} - \rho_{\mathbf{B}|\theta'}^{\mathbf{u}}\|_1 \quad (163)$$

$$\geq \exp\left(-n \sup_{s \in [0,1]} \log\left(\sum_u T_U(u) \times \text{tr}\left(\left(\rho_{\mathbf{B}|\theta}^{\theta}\right)^s \left(\rho_{\mathbf{B}|\theta'}^{\theta'}\right)^{1-s}\right)\right) + \mathcal{O}\left(\frac{\log n}{n}\right)\right) \quad (164)$$

$$= \exp\left(-n\mathbb{C}(\theta|\theta|T_U) + \mathcal{O}\left(\frac{\log n}{n}\right)\right). \quad (165)$$

## APPENDIX E PROOF OF LEMMA 1

We divide the proof into four steps.

a) *Step 1:* By the observation  $\|X\|_2 \leq \|X\|_1 \leq \sqrt{\dim A} \|X\|_2$  for all  $X \in \mathcal{L}(A)$ , it holds that

$$\sup_{\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}} \frac{\|\rho - |0\rangle\langle 0|\|_1}{\|\mathcal{E}(\rho) - \mathcal{E}(|0\rangle\langle 0|\|_1)} < \infty \quad (166)$$

if and only if

$$\sup_{\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}} \frac{\|\rho - |0\rangle\langle 0|\|_2}{\|\mathcal{E}(\rho) - \mathcal{E}(|0\rangle\langle 0|\|_2)} < \infty. \quad (167)$$

b) *Step 2:* We state a result that relates the norm of the output of a linear operator to the norm of the output of projection onto the kernel of the linear operator. This implies that one only needs to know  $\text{Ker}(\mathcal{E})$  to verify (167).

*Proposition 1:* Let  $V$  and  $W$  be Hilbert spaces and  $A: V \rightarrow W$  be a non-zero linear map. Let  $P$  be the projection onto  $\text{Ker}(A)$ . There exist positive constants  $B_1$  and  $B_2$  such that for all  $v \in V$ ,

$$B_1 \|Av\| \leq \|(\mathbf{1}_V - P)v\| \leq B_2 \|Av\|. \quad (168)$$

*Proof:* See Appendix E-A

By Proposition 1, we have

$$\sup_{\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}} \frac{\|\rho - |0\rangle\langle 0|\|_2}{\|\mathcal{E}(\rho) - \mathcal{E}(|0\rangle\langle 0|\|_2)} < \infty \quad (169)$$

if and only if

$$\sup_{\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}} \frac{\|\rho - |0\rangle\langle 0|\|_2}{\|(\text{id}_A - P)(\rho - |0\rangle\langle 0|\|_2)} < \infty. \quad (170)$$

where  $P$  is the projection onto  $\text{Ker}(\mathcal{E})$ .

a) *Step 3:* It will be more convenient in the sequel to consider linear operators acting on  $A$  as points in  $\mathbb{R}^{2d^2}$ . We use the function  $f$  defined in Definition 2, for which we list here some useful properties.

*Proposition 2:* The function  $f$  defined in Definition 2 satisfies the following properties.

- 1)  $f$  is bijective
- 2)  $f(aX + bY) = af(X) + bf(Y)$  for all  $X, Y \in \mathcal{L}(A)$  and for all  $a, b \in \mathbb{R}$
- 3)  $\|f(X)\|_2 = \|X\|_2$  for all  $X \in \mathcal{A}$
- 4) If  $Q$  is a projection onto a linear subspace  $E \subset \mathcal{L}(A)$ , then  $f(E)$  is also a linear subspace of  $\mathbb{R}^{2d^2}$  and  $f(Q(X)) = Q'(f(X))$  where  $Q'$  denotes the projection onto  $f(E)$ .

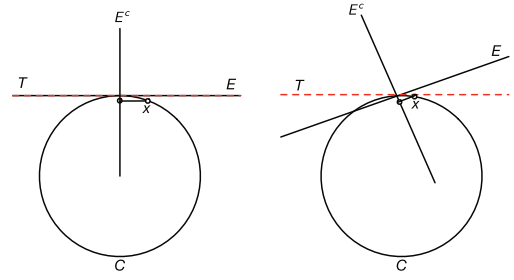


Fig. 2. Illustration of Proposition 3: On the left,  $\|x\|_2$  cannot be uniformly bounded by  $\|(\mathbf{1} - P)x\|_2$  when  $x$  is close to the origin, while on the right,  $\|x\|_2$  can be uniformly bounded by  $\|(\mathbf{1} - P)x\|_2$  when  $x$  is close to the origin.

- 5) If  $X$  is a compact convex subset of  $\mathcal{L}(A)$ , then  $f(X)$  is a compact convex subset of  $\mathbb{R}^{2d^2}$  and  $\partial f(X) = f(\partial X)$  where  $\partial f(X)$  and  $\partial X$  denote the boundaries of  $f(X)$  and  $X$ , respectively.

*Proof:* We only prove item 4 and the other items are straightforward consequence of the definition of  $f$ . We have  $f(Q(X)) = f(\arg\min_{Y \in E} \|Y - X\|_2) = f(\arg\min_{Y \in E} \|f(Y) - f(X)\|_2) = \arg\min_{Y' \in f(E)} \|Y' - f(X)\|_2 = Q'(f(X))$ . ■

Proposition 2 implies that

$$\sup_{\rho \in \mathcal{D}(A) \setminus \{|0\rangle\langle 0|\}} \frac{\|\rho - |0\rangle\langle 0|\|_2}{\|(\text{id}_A - P)(\rho - |0\rangle\langle 0|\|_2)} < \infty \quad (171)$$

if and only if

$$\sup_{x \in f(\mathcal{D}(A) - |0\rangle\langle 0|) \setminus \{0\}} \frac{\|x\|_2}{\|(\mathbf{1} - P')(x)\|_2} < \infty, \quad (172)$$

where  $P'$  is the projection onto  $f(\text{Ker}(\mathcal{E}))$ .

a) *Step 4:* We now provide a geometric characterization for Eq. (172) (See Fig. 2).

*Proposition 3:* Let  $C$  be a compact convex subset of  $\mathbb{R}^k$  containing the origin on its boundary and  $E$  be a linear subspace of  $\mathbb{R}^k$  such that  $C \cap E = \{0\}$ . We assume that the boundary of  $C$ ,  $\partial C$ , is a smooth manifold embedded in  $\mathbb{R}^k$  and the tangent space of  $\partial C$  at the origin is  $T$ . Then, upon denoting the projection onto  $E$  by  $P$ ,

$$\sup_{x \in C \setminus \{0\}} \frac{\|x\|_2}{\|(\mathbf{1} - P)x\|_2} < \infty \quad (173)$$

if and only if  $T \cap E = \{0\}$ .

*Proof:* See Appendix E-B. ■

a) *Step 5:* We show here that the tangent space at the origin of the boundary of  $f(\mathcal{D}(A) - |0\rangle\langle 0|)$  is  $\text{span}(a_1, \dots, a_{2d-2})$ , where  $\{a_i\}_{i \in [1, 2d-2]}$  is defined in Definition 2. First, note that the boundary of  $f(\mathcal{D}(A) \setminus |0\rangle\langle 0|)$  is  $f(\{|\phi\rangle\langle\phi| : \|\phi\| = 1\} \setminus |0\rangle\langle 0|)$  because the boundary of  $\mathcal{D}(A)$  are pure states and because of item 5 of Proposition 2. We define two map  $g : |\phi\rangle \mapsto |\phi\rangle\langle\phi|$  and  $h : \mathbb{R}^{2d-1} \rightarrow A$  as

$$(x_1, x_2, y_2, x_3, y_3, \dots, x_d, y_d) \mapsto x_1|e_1\rangle + \sum_{j=2}^d (x_j + iy_j)|e_j\rangle. \quad (174)$$

Note that the coefficient of  $|e_1\rangle$  is always real for all vectors in the range of  $h$  as we have freedom to choose the phase of



a quantum state. Following our definition of  $f$ ,  $g$ , and  $h$ , the  $2((j-1) \times d + k) - 1$  and  $2((j-1) \times d + k)$  components of  $(f \circ g \circ h)(x_1, x_2, y_2, x_3, y_3, \dots, x_d, y_d)$  are  $x_j x_k - y_j y_l$  and  $y_j x_k + x_j y_l$ , respectively. Thus,  $f \circ g \circ h$  is a smooth function. We also calculate the derivative of  $f \circ g \circ h$  at  $(1, 0, \dots, 0)$ , which is represented by the matrix  $[a_0 | a_1 | \dots | a_{2d-2}]$ , where  $a_0$  is a vector and  $a_1, \dots, a_{2d-2}$  are as in Definition 2.

Let  $\mathcal{S} \triangleq \{x \in \mathbb{R}^{2d-1} : \|x\|_2 = 1\}$  be the unit sphere in  $\mathbb{R}^{2d-1}$ . The restriction of  $f \circ g \circ h$  is also a smooth function. The tangent space of  $\mathcal{S}$  at  $(1, 0, \dots, 0)$  is the span of  $(b_2, \dots, b_{2d-1})$ , where  $b_1, \dots, b_{2d-1}$  form the standard basis for  $\mathbb{R}^{2d-1}$ . Therefore, the image of the derivative of  $f \circ g \circ h$  restricted to the tangent space of  $\mathcal{S}$  at  $(1, 0, \dots, 0)$  is the space of  $(a_1, \dots, a_{2d-2})$ . Since  $a_1, \dots, a_{2d-2}$  are linearly independent and the dimension of  $f(\{|\phi\rangle\langle\phi| : \|\phi\| = 1\} \setminus |0\rangle\langle 0|)$  is  $2d - 2$ , the whole tangent space of  $f(\{|\phi\rangle\langle\phi| : \|\phi\| = 1\} \setminus |0\rangle\langle 0|)$  at the origin should be the span of  $a_1, \dots, a_{2d-2}$ .

### A. Proof of Proposition 1

Let  $V/\text{Ker}(V)$  be the quotient space and  $\pi : V \rightarrow V/\text{Ker}(A)$  be the quotient map. We can define a norm on  $V/\text{Ker}(A)$  by  $\|\pi(v)\| = \inf_{x \in \text{Ker}(A)} \|v - x\| = \|v - Pv\| = \|(\mathbf{1}_V - P)v\|$ . By the first isomorphism theorem of linear algebra, there exists a linear isomorphism  $\tilde{A} : V/\text{Ker}(A) \rightarrow W$  such that  $A\tilde{A} = (\tilde{A} \circ \pi)v$ . Since any linear operator from a finite dimensional space is bounded, we have

$$\|A\tilde{A}\| = \|\tilde{A}(\pi(v))\| \leq \|\tilde{A}\| \|\pi(v)\| = \|\tilde{A}\| \|(\mathbf{1}_V - P)v\|$$

and  $\|(\mathbf{1}_V - P)v\| \leq \|\tilde{A}^{-1}\| \|A\tilde{A}\|$ . The result therefore holds for  $B_1 = 1/\|\tilde{A}\|$  and  $B_2 = \|\tilde{A}^{-1}\|$ .

### B. Proof of Proposition 3

a) *Step 1:* We first show that

$$\sup_{x \in C \setminus \{0\}} \frac{\|x\|_2}{\|(\mathbf{1} - P)x\|_2} < \infty \quad (175)$$

if and only if

$$\sup_{x \in \partial C \setminus \{0\}} \frac{\|x\|_2}{\|(\mathbf{1} - P)x\|_2} < \infty, \quad (176)$$

If  $Px = 0$  for  $x \in C \setminus \{0\}$ , we have  $\frac{\|x\|_2}{\|(\mathbf{1} - P)x\|_2} = 1$ , which is bounded. Let  $x \in C \setminus \{0\}$  such that  $Px \neq 0$  and define  $\phi : [0, 1] \rightarrow \mathbb{R}^k$  by  $\phi(t) \triangleq (1 - t)x + tPx$ . We know that  $\phi^{-1}(C)$  is closed and connected, because  $C$  is closed and connected (as a convex set) and  $\phi$  is continuous. The only closed and connected subsets of  $[0, 1]$  are of closed intervals and as  $\phi(0) = x \in C$ , we have  $\phi^{-1}(C) = [0, a]$  for some  $a \in [0, 1]$ . Since  $Px \neq 0$  and  $E \cap C = \{0\}$ , we have  $Px \notin C$  and therefore  $a < 1$ .  $\phi(a)$  is on the boundary of  $C$  because  $a$  is on the boundary of  $\phi^{-1}(C) = [0, a]$  and  $\phi$  is continuous. Now note that

$$P\phi(a) = P((1 - a)x + aPx) \quad (177)$$

$$= (1 - a)Px + aP^2x = Px. \quad (178)$$

Hence,

$$\begin{aligned} \|(1 - P)\phi(a)\|_2 &= \|(1 - a)x + aPx - Px\|_2 \\ &= (1 - a)\|(1 - P)x\|_2 \leq \|(1 - P)x\|_2 \end{aligned}$$

Therefore,

$$\frac{\|x\|_2}{\|(\mathbf{1} - P)x\|_2} = \frac{\sqrt{\|(1 - P)x\|_2^2 + \|Px\|_2^2}}{\|(1 - P)x\|_2} \quad (179)$$

$$\stackrel{(a)}{\leq} \frac{\sqrt{\|(1 - P)\phi(a)\|_2^2 + \|P\phi(a)\|_2^2}}{\|(1 - P)\phi(a)\|_2} \quad (180)$$

$$= \frac{\|\phi(a)\|_2}{\|(1 - P)\phi(a)\|_2}. \quad (181)$$

This completes the proof of the first step.

b) *Step 2:* We now show that

$$\sup_{x \in \partial C \setminus \{0\}} \frac{\|x\|_2}{\|(\mathbf{1} - P)x\|_2} < \infty, \quad (182)$$

if and only if  $E \cap T = \{0\}$ .

First suppose that  $v \in E \cap T$  is non-zero. We will find  $x \in \partial C \setminus \{0\}$  such that  $\|x\|/\|(\mathbf{1} - P)x\| \geq K$  for a given  $K > 0$ . By definition of tangent space, there exists a smooth curve  $\gamma : (-1, 1) \rightarrow \partial C$  such that  $\gamma(0) = 0$  and  $\gamma'(0) = v$ , i.e.,  $\lim_{t \rightarrow 0} \gamma(t)/t = v$ . There exists some  $t_0 > 0$  such that  $\|\gamma(t)/t - v\| \leq \|v\|/2$  for all  $0 < |t| < t_0$ . We therefore have  $\|\gamma(t)\| \geq t\|v\|/2$ . Additionally,  $\lim_{t \rightarrow 0} (\mathbf{1} - P)\gamma(t) = 0$  because  $\mathbf{1} - P$  is continuous and  $(\mathbf{1} - P)v = v - Pv = 0$ . Thus, there exists  $t_1 > 0$  such that  $\|(\mathbf{1} - P)\gamma(t)\| \leq 2/(K\|v\|)$ . For any  $t$  such that  $0 < t < \min(t_0, t_1)$ , we have

$$\frac{\|\gamma(t)\|}{\|(\mathbf{1} - P)\gamma(t)\|} \geq \frac{t\|v\|/2}{2/(K\|v\|)} = K, \quad (183)$$

as claimed.

We now prove the other direction. Let  $\mathcal{B}(\epsilon) \triangleq \{x \in \mathbb{R}^k : \|x\| < \epsilon\}$  denote the open ball of radius  $\epsilon$  at the origin. To show (182), it is enough to check for arbitrary small  $\epsilon > 0$

$$\sup_{x \in \partial C \cap \mathcal{B}(\epsilon) \setminus \{0\}} \frac{\|x\|_2}{\|(\mathbf{1} - P)x\|_2} < \infty \quad (184)$$

because  $\partial C \setminus \mathcal{B}(\epsilon)$  is a compact set, on which the distance from  $E$  is non-zero and varies continuously. Let  $Q$  denote the projection onto  $T$  the tangent space of  $\partial C$  at the origin. We know that for a point  $x$  on  $\partial C$  close to origin we have  $\|x\| = \|Qx\| + o(\|x\|)$ . We can hence find an  $\epsilon > 0$  such that for  $x \in \partial C \cap \mathcal{B}(\epsilon) \setminus \{0\}$  we have  $\|x\| \leq 2\|Qx\|$ . Furthermore, considering the linear map  $A : T \rightarrow E^\perp$  defined by  $x \mapsto (\mathbf{1} - P)x$ , it is injective. Therefore, for some constant  $B > 0$ , we have  $\|Ax\| = \|(\mathbf{1} - P)x\| \geq B\|x\|$  for all  $x \in T$ . We can also choose  $\epsilon > 0$  such that  $\|(1 - Q)x\| \leq B/2\|Qx\|$ . Therefore, we have for all  $x \in \partial C \cap \mathcal{B}(\epsilon) \setminus \{0\}$

$$\begin{aligned} \frac{\|x\|}{\|(\mathbf{1} - P)x\|} &\leq \frac{2\|Qx\|}{\|(\mathbf{1} - P)x\|} \\ &\leq \frac{2\|Qx\|}{\|(\mathbf{1} - P)Qx\| - \|(\mathbf{1} - P)(\mathbf{1} - Q)x\|} \\ &\leq \frac{2\|Qx\|}{\|(\mathbf{1} - P)Qx\| - \|(\mathbf{1} - P)(\mathbf{1} - Q)x\|} \\ &\leq \frac{2\|Qx\|}{C\|Qx\| - C\|Qx\|/2} \leq 4/C. \end{aligned}$$

## REFERENCES

- [1] D. Lentner and G. Kramer, "Stealth communication with vanishing power over binary symmetric channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Los Angeles, CA, USA, Jun. 2020, pp. 822–827.
- [2] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 601–605.
- [3] B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.
- [4] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.
- [5] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.
- [6] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.
- [7] A. Sheikholeslami, B. A. Bash, D. Towsley, D. Goeckel, and S. Guha, "Covert communication over classical-quantum channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 2064–2068.
- [8] L. Wang, "Optimal throughput for covert communication over a classical-quantum channel," in *Proc. IEEE Inf. Theory Workshop*, Cambridge, U.K., Sep. 2016, pp. 364–368.
- [9] C. N. Gagatsos, M. S. Bullock, and B. A. Bash, "Covert capacity of bosonic channels," 2020. [Online]. Available: arXiv: 2002.06733.
- [10] B. A. Bash, C. N. Gagatsos, A. Datta, and S. Guha, "Fundamental limits of quantum-secure covert optical sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 3210–3214.
- [11] C. N. Gagatsos, B. A. Bash, A. Datta, Z. Zhang, and S. Guha, "Covert sensing using floodlight illumination," *Phys. Rev. A*, vol. 99, Jun. 2019, Art. no. 062321.
- [12] D. Goeckel, B. A. Bash, A. Sheikholeslami, S. Guha, and D. Towsley, "Covert active sensing of linear systems," in *Proc. 51st Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2017, pp. 1692–1696.
- [13] R. F. Schaefer, H. Boche, and H. V. Poor, "Secure communication under channel uncertainty and adversarial attacks," *Proc. IEEE*, vol. 103, no. 10, pp. 1796–1813, Oct. 2015.
- [14] M. Naghshvar and T. Javidi, "Sequentiality and adaptivity gains in active hypothesis testing," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 5, pp. 768–782, Oct. 2013.
- [15] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *Ann. Stat.*, vol. 41, no. 6, pp. 2703–2738, Dec. 2013.
- [16] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2451–2464, Oct. 2013.
- [17] R. Blahut, "An hypothesis-testing approach to information theory (Ph.D. thesis abstr.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, p. 253, Mar. 1973.
- [18] H. Chernoff, "Sequential design of experiments," *Ann. Math. Stat.*, vol. 30, no. 3, pp. 755–770, Sep. 1959.
- [19] M. Hayashi, "Discrimination of two channels by adaptive methods and its application to quantum system," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3807–3820, Aug. 2009.
- [20] M. Naghshvar and T. Javidi, "Extrinsic Jensen-Shannon divergence with application in active hypothesis testing," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 2191–2195.
- [21] M. Franceschetti, S. Marano, and V. Matta, "Chernoff test for strong-or-weak radar models," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 289–302, Jan. 2017.
- [22] S.-E. Chiu, N. Ronquillo, and T. Javidi, "Active learning and CSI acquisition for mmwave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2474–2489, Nov. 2019.
- [23] J. Watrous, *The Theory of Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2018.
- [24] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, "Fundamental limits to quantum channel discrimination," *NPJ Quantum Inf.*, vol. 5, no. 1, p. 50, Dec. 2019.
- [25] M. Nussbaum and A. Szkola, "The chernoff lower bound for symmetric quantum hypothesis testing," *Ann. Stat.*, vol. 37, no. 2, pp. 1040–1057, Apr. 2009.
- [26] K. Li, "Discriminating quantum states: The multiple chernoff distance," *Ann. Stat.*, vol. 44, no. 4, pp. 1661–1679, Aug. 2016.
- [27] A. Acín, "Statistical distinguishability between unitary operations," *Phys. Rev. Lett.*, vol. 87, no. 17, Oct. 2001, Art. no. 177901.
- [28] M. Tahmasbi and M. R. Bloch, "Active covert sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, Mar. 2020, pp. 840–845.
- [29] M. Tahmasbi and M. R. Bloch, "A framework for covert and secret key expansion over quantum channels," 2018. [Online]. Available: arXiv:1811.05626.
- [30] C. A. Fuchs and J. van de Graaf, "Cryptographic distinguishability measures for quantum-mechanical states," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1216–1227, May 1999.
- [31] A. Shapiro, *On Differentiability of Symmetric Matrix Valued Functions*, Dept. School Ind. Syst. Eng., Georgia Inst. Technol., Atlanta, GA, USA, 2002.
- [32] M. M. Wilde, *Quantum Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2017. [Online]. Available: https://doi.org/10.1017/9781316809976
- [33] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

**Mehrdad Tahmasbi** received two B.Sc. degrees in electrical engineering and pure mathematics from the Sharif University of Technology, Tehran, Iran, and two M.S. degrees in mathematics and electrical engineering and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, USA. He is a Postdoctoral Fellow with the University of Amsterdam and QuSoft. His research interests include quantum cryptography and quantum computing. He was a recipient of the Silver Medal at the 22nd International Olympiad in Informatics in 2010.

**Matthieu R. Bloch** (Senior Member, IEEE) received the engineering degree from Supélec, Gif-sur-Yvette, France, the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 2003, the Ph.D. degree in engineering science from the Université de Franche-Comté, Besançon, France, in 2006, and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology in 2008. He is an Associate Professor with the School of Electrical and Computer Engineering, Georgia Tech. From 2008 to 2009, he was a Postdoctoral Research Associate with the University of Notre Dame, Notre Dame, IN, USA. Since July 2009, he has been on the faculty of the School of Electrical and Computer Engineering, Georgia Institute of Technology. His research interests are in the areas of information theory, error-control coding, wireless communications, and cryptography. He is the co-recipient of the IEEE Communications Society and IEEE Information Theory Society 2011 Joint Paper Award and the coauthor of the textbook *Physical-Layer Security: From Information Theory to Security Engineering* published by Cambridge University Press. He was the Chair of the Online Committee of the IEEE Information Theory Society from 2011 to 2014, an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2016 to 2019, and he has been on the Board of Governors of the IEEE Information Theory Society since 2016. He has been an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY since 2016.