Covert and secret key expansion over quantum channels under collective attacks

Mehrdad Tahmasbi and Matthieu R. Bloch

Abstract

We consider an enhanced measure of security for a quantum key distribution protocol, in which we require that the adversary not only obtains no information about the key but also remains unaware that a key generation protocol has been executed. When the adversary applies the same quantum channel independently to each transmitted quantum state, akin to a collective attack in the quantum key distribution literature, we propose a protocol that achieves covert and secret key expansion under mild restrictions. A crucial component of the protocol is a covert estimation stage, which is then combined with universal channel coding for reliability and resolvability in the covert regime.

I. Introduction

The search for alternative cryptographic methods has become relevant because of the existence of algorithms that could be run on a quantum computer and threaten the security of some traditional cryptographic methods, such as some asymmetric cryptographic schemes. In particular, the generation of unconditionally secure key bits over quantum channels, known as Quantum Key Distribution (QKD) [1], has attracted much attention because of its ability to exploit the unique characteristics of quantum channels and enable the detection of an adversary tampering with the transmitted data. Starting from the seminal work of Bennett and Brassard [2], this field has witnessed tremendous advances in the last decade from both theoretical and practical perspectives. In particular, security proofs have evolved to consider several types of attacks such as individual, collective, coherent attacks, and the possibility of an adversary controlling the legitimate users' apparatus [3], and the statistical effects of measuring and processing a finite amount of data [4]. In addition, experimental QKD systems have matured to stably operate over long distances and over long periods of time [5].

Another desirable feature for a key generation protocol is undetectability or covertness, by which the legitimate parties aim at hiding the fact that the key generation is happening from an unwanted party. From a statistical point of view, the adversary should be unable to distinguish between the statistics of its observation with and without key generation. The study of covert communication has been initiated by [6] for classical channels, in which a square-root-law has been established for Additive White Gaussian Noise (AWGN) channels, i.e., at most $O(\sqrt{n})$ of bits can be reliably and covertly transmitted over n uses of an AWGN channel. Covert communication in a classical setting has been extensively investigated, including the study of the exact asymptotics in covert communication

This work was supported by the National Science Foundation under awards 1527074 and 1910859.

[7], [8], covert communication in multi-terminal networks [9], practical code design [10], and state-dependent channel [11]. Furthermore, a few studies have extended these results to the quantum setting [12], [13], [14], [15], [16]; in particular, [12], [13] have established inner- and outer-bounds for the covert capacity of classical-quantum (cq) channels, and [17] proved a no-go result for covert communication over bosonic channels when there is no imperfection on the adversary's observations. An actual demonstration of covert communication has also been demonstrated in the presence of thermal noise on the transmission channel or dark count in the photo-detector of the adversary.

The main purpose of the present paper is to investigate the possibility of guaranteeing the covertness of a QKD protocol. The authors of [15] have suggested the pessimistic result that "*covert QKD consumes more key that it can generate*." In their protocol, legitimate parties *coordinate* transmissions in $\sqrt{n}$ locations out of n, for which $O(\sqrt{n}\log n)$ bits are required. This amount, unfortunately, dominates the amount of generated key bits, which only scales as $O(\sqrt{n})$. In a previous work [18], we have established a framework for covert and secret key expansion over quantum channels, in which the use of public communication is precisely defined, and we have established lower-bounds on the amount of key bits that can be generated when the channels are known. Our achievability results were based on the transmission of independent and identically distributed (i.i.d.) signals over the channel with highly biased distribution instead of coordinating the locations for transmission ahead of time. The main shortcoming of our previous work is the requirement that the channel should be known, which might not be realistic in practice. To address this issue, we consider here a model in which the adversary controls the quantum channel under some conditions. Specifically, we consider an instance of quantum covert and secret key generation in which the quantum channel is fixed but under the control of the adversary and *unknown* to the legitimate users. Under some conditions that limit the power of the adversary, which we precisely characterize, we prove the existence of covert and secret key generation protocols consisting of a channel estimation phase followed by a key-generation phase. The estimation phase is based on a covert quantum tomography protocol, which estimates the required parameters of the channel, and the key generation phase is based on universal results for covert quantum communication. While covertness cannot be unconditionally guaranteed, our protocol offers the legitimate parties with the ability to successfully abort before engaging in key generation. We do not instantiate explicit codes but recent progress in designing codes for covert communications [10] suggests that the protocols described here can be implemented with low complexity.

The remainder of the paper is organized as follows. In Section II, we briefly introduce the notation used throughout the paper. In Section III, we formally describe the problem under investigation and state our main result. We devote Section IV to the proof of our main theorem.

II. Notation

For a finite-dimensional Hilbert space $\mathcal{H}$, $\dim \mathcal{H}$ denotes the dimension of $\mathcal{H}$, and $\mathcal{L}(\mathcal{H})$ denotes the space of all linear operators from $\mathcal{H}$ to $\mathcal{H}$. We denote the adjoint of an operator $X \in \mathcal{L}(\mathcal{H})$ by $X^\dagger$, and call X Hermitian if $X = X^\dagger$. $X \in \mathcal{L}(\mathcal{H})$ is positive (semi-)definite, if it is Hermitian and all of its eigenvalues are positive (non-negative). $\mathcal{D}(\mathcal{H})$ denotes the set of all density operators on $\mathcal{H}$, i.e., all non-negative operators with unit trace. For $X, Y \in \mathcal{L}(\mathcal{H})$, we write $X \succ Y$ ($X \succeq Y$), if $X - Y$ is positive (semi-)definite. For $X \in \mathcal{H}$, let $\sigma_{\min}(X)$ and

$\sigma_{\max}(X)$ denote the minimum and the maximum singular value of X, respectively, and if X is Hermitian, let $\lambda_{\min}(X)$ and $\lambda_{\max}(X)$ denote the minimum and maximum eigenvalue of X. Furthermore, we define norms of $X \in \mathcal{L}(\mathcal{H})$ as $\|X\|_1 \triangleq \mathrm{tr}\left(\sqrt{X^\dagger X}\right)$ and $\|X\|_2 \triangleq \sqrt{\mathrm{tr}\,(X^\dagger X)}$. For a Hermitian operator $X \in \mathcal{L}(\mathcal{H})$ with eigen-decomposition $X = \sum_x x|x\rangle\langle x|$, we define the projection $\{X \succeq 0\} \triangleq \sum_{x \geqslant 0} |x\rangle\langle x|$. A quantum channel $\mathcal{E}_{A\to B}$ is a completely positive and trace preserving linear map from $\mathcal{L}(\mathcal{H}^A)$ to $\mathcal{L}(\mathcal{H}^B)$. An isomorphic extension of $\mathcal{E}_{A\to B}$, $U_{A\to BE}$, satisfies $\mathcal{E}_{A\to B}(\rho^A) = \mathrm{tr}_E(U_{A\to BE}\rho^A U_{A\to BE}^\dagger)$ for all $\rho^A \in \mathcal{D}(\mathcal{H}^A)$. We denote the complementary channel of $\mathcal{E}_{A\to B}$ by $\mathcal{E}^c_{A\to B}(\rho^A) \triangleq \mathcal{E}_{A\to E}(\rho^A) \triangleq \mathrm{tr}_B(U_{A\to BE}\rho^A U_{A\to BE}^\dagger)$, which is well-defined and unique up to a unitary transformation [19, Exercise 5.2.5]. Let $\mathcal{E}_{A\to B} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_B)$ be a quantum channel and $|1\rangle, \cdots, |d\rangle$ be an orthonormal basis for $\mathcal{H}^B$. We let $\widetilde{E}_{d(n-1)+m} \triangleq |n\rangle\langle m|$ so that $\widetilde{E}_1, \cdots, \widetilde{E}_{d^2}$ forms an orthonormal basis for $\mathcal{L}(\mathcal{H})$. By [20], there exists coefficients $\chi_{j,k}$ such that

$$\mathcal{E}(\rho) = \sum_{j,k} \widetilde{E}_j \rho \widetilde{E}_k^\dagger \chi_{j,k}. \tag{1}$$

We call the matrix $\chi = [\chi_{j,k}]_{j,k=1,\cdots,d^2}$ the chi-representation of the channel $\mathcal{E}_{A\to B}$ with respect to the orthonormal basis $|1\rangle, \cdots, |d\rangle$. A cq-channel is a map from an abstract set $\mathcal{X}$ to $\mathcal{D}(\mathcal{H})$, denoted by $x \mapsto \rho_x$.

For $\rho^A \in \mathcal{D}(\mathcal{H}^A)$ we define the von Neumann entropy $H(\rho^A) \triangleq \mathbb{H}(A)_\rho \triangleq -\mathrm{tr}\left(\rho^A \log \rho^A\right)$. For $\rho^{AB} \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$, we define the conditional von Neumann entropy $\mathbb{H}(A|B)_\rho \triangleq H(\rho^{AB}) - H(\rho^B)$ where $\rho^B \triangleq \mathrm{tr}_A(\rho^{AB})$, and the quantum mutual information $\mathbb{I}(A;B)_\rho \triangleq H(\rho^A) + H(\rho^B) - H(\rho^{AB})$. Similarly, we define the conditional quantum mutual information $\mathbb{I}(A;B|C) \triangleq H(\rho^{AC}) + H(\rho^{BC}) - H(\rho^{ABC}) - H(\rho^C)$ for any $\rho^{ABC} \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^B \otimes \mathcal{H}^C)$. If P_X is a distribution on $\mathcal{X}$ and $x \mapsto \rho_x$ is a cq-channel, we denote the Holevo information by

$$I(P_X, \rho_x) \triangleq H\left(\sum_x P_X(x)\rho_x\right) - \sum_x P_X(x) H(\rho_x). \tag{2}$$

For $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, the quantum relative entropy is

$$\mathbb{D}(\rho\|\sigma) \triangleq \begin{cases} \mathrm{tr}\left(\rho\left(\log\rho - \log\sigma\right)\right) & \text{if } \mathrm{supp}(\rho) \subset \mathrm{supp}(\sigma), \\ \infty & \text{otherwise.} \end{cases} \tag{3}$$

We also define $\eta(\rho,\sigma))$ similar to [13, Eq. (12)] as

$$\eta\left(\rho\|\sigma\right) \triangleq \begin{cases} \mathrm{tr}\left(\int_0^\infty \rho(\sigma+s)^{-1}\rho(\sigma+s)^{-1}ds\right) - 1 & \text{if } \mathrm{supp}(\rho) \subset \mathrm{supp}(\sigma), \\ \infty & \text{otherwise.} \end{cases} \tag{4}$$

We denote by $[\![n, m]\!]$ the interval $\{i \in \mathbb{Z} : n \leqslant i \leqslant m\}$. We also define

$$[a, b] = \{x \in \mathbb{R} : a \leqslant x \leqslant b\}, \tag{5}$$

$$]a, b] = \{x \in \mathbb{R} : a < x \leqslant b\}, \tag{6}$$

$$[a, b[= \{x \in \mathbb{R} : a \leqslant x < b\}, \tag{7}$$

$$]a, b[= \{x \in \mathbb{R} : a < x < b\}. \tag{8}$$

Let $\mathrm{wt}(\mathbf{x}) \triangleq \sum_{i=1}^n x_i$ for $\mathbf{x} = (x_1, \cdots, x_n) \in \mathcal{X}^n$.

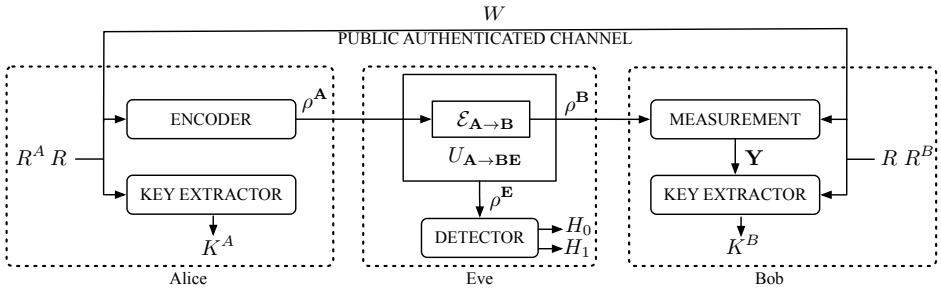

Figure 1. Model of covert and secret key expansion

III. Model for covert and secret key generation under collective attacks and main results

In this section, we formalize key generation protocols, describe our assumption on the eavesdropper, define the metrics used to assess a protocol and state our main result. We finally illustrate our result for a quantum phase-flip channel.

a) Protocol description: As depicted in Fig. 1, two legitimate parties, Alice and Bob, attempt to expand shared secret key using multiple uses of a quantum channel and two-way communication over a public but authenticated classical channel without being detected by an adversary, Eve. Prior to the transmission, Alice and Bob possess local randomness R^A and R^B, respectively, and a shared secret randomness R. In T time steps, Alice prepares a quantum state using her available information, which is sent to Bob through a quantum channel $\mathcal{E}_{A\to B}$. Bob can perform a possibly joint measurement on his received states. Alice and Bob also communicate with each other over the public classical channel after each time step. We denote all communications over the public channel by W. Alice and Bob use their available information at the end of the T^{th} time step to compute two long binary strings s^X and s^Y, respectively, as well as the number of bits ℓ^X and ℓ^Y, respectively, to use as a secret key. The length of s^X and s^Y is public and fixed at the beginning of the protocol. Alice finally sets her key k^X to be the first ℓ^X bits of s^X while Bob sets his key k^Y to be the first ℓ^Y bits of s^Y.

b) Attacker model: We assume that Eve initially chooses the quantum channel $\mathcal{E}_{A\to B}$ under some mild constraints precisely defined in the statement of Theorem 1 but that the channel remains unchanged during the transmission. The channel is a priori unknown to Alice and Bob. When Alice transmits ρ^A, Eve observes $\mathcal{E}^c_{A\to B}(\rho^A)$, which accounts for the maximum amount of information that she can possibly gain, i.e., the state corresponding to a reference system for an isomorphic extension of the channel from Alice to Bob. When no communication happens, Eve expects the "innocent" symbol ρ_0^A to be sent by Alice. She is also fully aware of the details of the protocol used by Alice and Bob for covert and secret key generation.

Similar to [18], we call a protocol an (ϵ, δ, μ)-protocol if it satisfies the following reliability, secrecy, and covertness conditions.

- ϵ-reliability: $P_e \triangleq \mathbb{P}\big(K^X \neq K^Y\big) \leqslant \epsilon$, which implies that $\ell^X = \ell^Y$ with probability at least $1-\epsilon$;
- δ-secrecy: $S \triangleq \mathbb{D}\Big(\rho^{\mathbf{E}WS^X} \| \rho^{\mathbf{E}W} \otimes \rho^{S^X}_{\text{unif}}\Big) \leqslant \delta$, where $\rho^{\mathbf{E}WS^X}$ is the joint density matrix of the eavesdropper's observations, public messages and Alice's random string, and $\rho^{S^X}_{\text{unif}}$ is a mixed state for S^X corresponding to

a uniform distribution;

- μ-covertness: $C \triangleq \mathbb{D}\big(\rho^{\mathbf{E}W} \| \left(\rho_0^{\mathbf{E}}\right) \otimes \rho_{\text{unif}}^{W}\big) \leqslant \mu$, where $\rho_0^{\mathbf{E}} \triangleq (\mathcal{E}_{A\to B}^{c}(\rho_0^A))^{\otimes T}$ is the density matrix of the eavesdropper's observations when no communication takes place and ρ_{unif}^{W} is the density operator corresponding to uniform distribution on W.[1]

Remark 1. *We discuss here the operational meaning of the role of public communication in our covertness constraint. We assume that the adversary always expects a classical communication on the public channel between Alice and Bob (such as a constant traffic in a network to maintain persistence) in the absence of the quantum key generation protocol, but the classical messages on the public channel have pre-specified distribution and are statistically independent of the content of the quantum channel. This assumption might be restrictive in certain scenarios, in which this classical communication does not exist. The results of this paper would not be then applicable and one would still need to formalize the problem and develop protocols according to restricted available resources at Alice and Bob.*

A protocol is *efficient* if it allows key expansion so that the number of key bits created exceeds the number of common randomness bits consumed. Our goal is to analyze under what conditions efficient (ϵ, μ, δ)-protocols might exist.

d) Main result:

Theorem 1. *Let χ be the chi-representation of the channel $\mathcal{E}_{A\to B}$ with respect to some orthonormal basis. Let ρ_1^A be an arbitrary density operator on $\mathcal{H}^A$ and define $\rho_x^B \triangleq \mathcal{E}_{A\to B}(\rho_x^A)$ and $\rho_x^E \triangleq \mathcal{E}_{A\to B}^{c}(\rho_x^A)$ for $x = 0, 1$. Let $\widetilde{\lambda}^{\chi}$, $\widetilde{\lambda}^{B}$ and $\widetilde{\lambda}^{E}$ be fixed in $]0, 1]$. Let $\{\alpha_T\}_{T\geqslant 1}$ be such that*

$$\alpha_T \in \omega\left(\left(\frac{\log T}{T}\right)^{\frac{2}{3}}\right) \cap o\left(\frac{1}{\sqrt{T}}\right). \tag{9}$$

For any $\zeta > 0$, there exists a vanishing sequence $\{\epsilon_T\}_{T\geqslant 1}$ and a sequence of $(\epsilon_T, \epsilon_T, \mu_T)$ covert and secret key generation protocols such that for all quantum channels $\mathcal{E}_{A\to B}$ with

$$\lambda_{\min}(\rho_0^E) \geqslant \widetilde{\lambda}^{E}, \tag{10}$$

we have

$$\mu_T \leqslant (1 + \epsilon_T)\frac{\alpha_T^2 \eta(\rho_1^E \| \rho_0^E) T}{2}. \tag{11}$$

If in addition to Eq. (10)*, it holds that*

$$\lambda_{\min}(\chi) \geqslant \widetilde{\lambda}^{\chi}, \tag{12}$$

$$\lambda_{\min}(\rho_0^B) \geqslant \widetilde{\lambda}^{B}, \tag{13}$$

[1] While the trace distance is the commonly used distance measure in QKD, the relative entropy is more convenient to study covert communication because of its analytical properties. Since Pinsker's inequality allows us to bound the trace distance through the relative entropy, the operational meaning of the bounds remains unchanged.

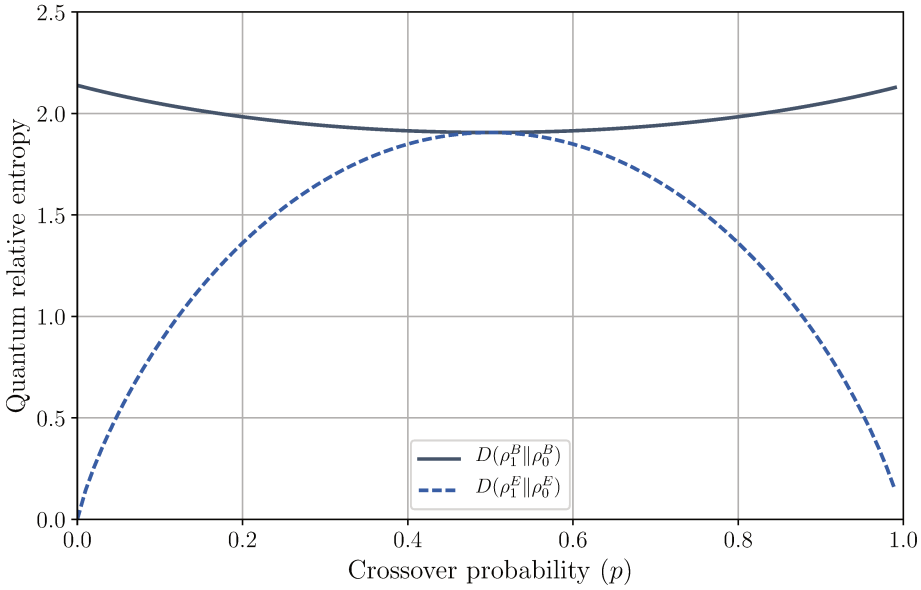

Figure 2. Quantum relative entropy of main and warden channels

then, with probability at least $1-\epsilon_T$*, the length of the generated key is at least*

$$(1-\zeta)\left[\mathbb{D}\big(\rho_1^B\|\rho_0^B\big)-\mathbb{D}\big(\rho_1^E\|\rho_0^E\big)\right]\alpha_T T. \tag{14}$$

We now discuss the meaning and implications of Theorem 1. Firstly, the parameter α_T defined in (9) controls the fraction of the state ρ_1^A transmitted over T channel uses, and should be understood as close to but slightly less than $1/\sqrt{T}$. For this choice, the covertness parameter μ_T vanishes with T while the number of secret key bits covertly generated scales almost as $\Omega(\sqrt{T})$. Secondly, the condition (10) states that the channel to the adversary should be noisy enough to allow covert operation a priori. This condition should not be surprising, as we know that covert communication is impossible in a general setting [17]. The conditions (12) and (13) are technical conditions to avoid extreme cases, which are necessary to ascertain the reliability of our estimation protocol and can be set to the technological limits on the quantum channel $\mathcal{E}_{A\to B}$. As apparent in the proof, these three conditions only affect the sequence $\{\epsilon_T\}_{T\geqslant 1}$ but not the asymptotic covert and secret key rate guaranteed in (14). Finally, Alice and Bob can test whether $\lambda_{\min}(\rho_0^E)\geqslant\widetilde{\lambda}^E$ with high probability at the end of the estimation phase and abort the protocol if the condition is violated. In that case, the covertness constraint cannot be guaranteed as defined before because too many non-innocent symbols have been sent in the estimation phase; however, aborting the protocol and transmitting innocent symbols in the remaining channel uses still minimizes $\mathbb{D}\big(\rho^{\mathbf{E}}\|\rho_0^{\mathbf{E}}\big)$.

We illustrate the result of Theorem 1 for a specific quantum channel, $\mathcal{E}_{A\to B}$, from Alice to Bob. We assume that $\mathcal{E}_{A\to B}$ is a phase flip channel with flipping probability p, namely, $\mathcal{E}_{A\to B}(\rho^A)=(1-p)\rho+p\sigma_z\rho\sigma_z$, and the

matrix representation of Alice's transmitted states in the computational basis is

$$\rho_0^A = \begin{bmatrix} 0.95 & 0 \\ 0 & 0.05 \end{bmatrix} \tag{15}$$

$$\rho_1^A = \begin{bmatrix} 0.2 & 0.3 \\ 0.3 & 0.8 \end{bmatrix}. \tag{16}$$

For $x \in \{0,1\}$, we define $\rho_x^B \triangleq \mathcal{E}_{A\to B}(\rho_x^A)$ and $\rho_x^E \triangleq \mathcal{E}^c_{A\to B}(\rho_x^A)$, and in Fig. 2, we illustrate the behavior of $\mathbb{D}\big(\rho_1^B \| \rho_0^B\big)$ and $\mathbb{D}\big(\rho_1^E \| \rho_0^E\big)$ as a function of p. By Theorem 1, the number of generated covert and secret key bits is on the order of $(\mathbb{D}\big(\rho_1^B \| \rho_0^B\big) - \mathbb{D}\big(\rho_1^E \| \rho_0^E\big))\alpha_T T$, which scales as $O(\alpha_T T)$ except for $p = 0.5$.

IV. Proof of Theorem 1

For clarity, we have divided the proof of Theorem 1 in three steps.

1) In Section IV-A, we show the existence of a sequence of universal codes for reliable, secure, and covert transmission of a message upon the existence of bounds on some parameters of the effective wiretap cq-channel $x \mapsto \rho_x^{BE}$ from Alice to Bob and Eve (Theorem 2). Our proof follows a standard random wiretap code construction, which is summarized next. Let M be the number of messages and M' be a large integer to be determined later. To transmit message $w \in [\![1, M]\!]$, Alice chooses a codeword uniformly at random from the set $\{\mathbf{x}_{w,1}, \cdots, \mathbf{x}_{w,M'}\}$ and transmits the codeword over n uses of the cq-channel. Each codeword $\mathbf{x}_{w,i}$ is independently generated according to $P_X^{\otimes n}$, where P_X is a Bernoulli distribution with parameter α_n vanishing with n. To analyze the protocol, we first quantize the set of possible channels to obtain a finite set that approximates the entire set of channels with desired parameters. The reliability analysis of the protocol relies on the universal properties of a decoder based on Schur-Weyl duality [21]. The secrecy and covertness analysis of the protocol relies on channel resolvability results. Specifically, we use super-exponential bounds to analyze the output statistics generated by the coding scheme.
2) In Section IV-B, we show how quantum tomography can be covertly implemented to derive the bounds on the channel parameters that were assumed to be known in the first step of the proof. Specifically, Alice transmits non-zero states in a sparse set of positions that is shared with Bob using a shared key. Bob performs a measurement in those positions to estimate the channel and compute the desired parameters.
3) Finally, in Section IV-C. we complete the proof of Theorem 1 by combining the first two steps. Specifically, i) Alice and Bob first run the tomography protocol described in Section IV-B; ii) Bob then computes the required parameters and sends them over the public channel one-time-padded with a pre-shared key; and, finally, iii) Alice and Bob implement the universal protocol described in Section IV-A. The required pre-shared key is the one needed for one-time-padding the estimated parameters of the channel and sharing the position of the state sent for tomography. The size of this key is shown to be smaller than the size of the generated key.

A. Universal covert communication

Let ρ_x^A, ρ_x^B, ρ_x^E be defined as in the statement of Theorem 1 for $x = 0, 1$. The following theorem shows that knowing bounds on $\lambda_{\min}(\rho_0^B)$, $\lambda_{\min}(\rho_0^E)$, $\mathbb{D}\big(\rho_1^B \| \rho_0^B\big)$ and $\mathbb{D}\big(\rho_1^E \| \rho_0^E\big)$ is all that is required to covertly generate a secret key.

Theorem 2. *Let D^B, D^E, $\widetilde{\lambda}^B$, and $\widetilde{\lambda}^E$ be fixed numbers and $\{\alpha_T\}_{T \geqslant 1}$ be as in* (9)*. For any $\zeta > 0$, there exists a sequence of codes $\{\mathcal{C}_T\}_{T \geqslant 1}$ such that for all cq-channels $x \mapsto \rho_x^B$ and $x \mapsto \rho_x^E$ satisfying*

$$\mathbb{D}\big(\rho_1^B \| \rho_0^B\big) \geqslant D^B, \tag{17}$$

$$\mathbb{D}\big(\rho_1^E \| \rho_0^E\big) \leqslant D^E, \tag{18}$$

$$\lambda_{\min}(\rho_0^B) \geqslant \widetilde{\lambda}^B, \tag{19}$$

$$\lambda_{\min}(\rho_0^E) \geqslant \widetilde{\lambda}^E, \tag{20}$$

we have

$$P_e \leqslant 2T^{-5}, \tag{21}$$

$$S \leqslant L_1 T^{-4}, \tag{22}$$

$$C \leqslant \frac{\alpha_T^2 \eta(\rho_1^E \| \rho_0^E)}{2} T + L_1 T^{-4} + 2\sqrt{L_1} \log \frac{2}{\widetilde{\lambda}^E} T^{-1} + L_2 \alpha_T^3 T,$$

$$\log M = (1 - 2\zeta)(D^B - D^E)\alpha_T T, \tag{23}$$

where P_e, S, C denote the probability of error, secrecy, and covertness, respectively, as defined in Section III $L_1, L_2 > 0$ depend on the $\dim \mathcal{H}^E$ and $\widetilde{\lambda}^E$.

The remainder of this section is dedicated to the proof of the above result. We first adapt a result from [22], which shows that for any class of cq-channels, there exists a *finite* class of cq-channels that approximates the main class with high precision.

Lemma 1. *Consider a compound cq-channel $x \mapsto \rho_x^B(\theta)$ where $x \in \mathcal{X}$, $\rho_x^B \in \mathcal{D}(\mathcal{H})$, $\mathcal{H}$ is a d-dimensional Hilbert space, and $\theta \in \Theta$ is an arbitrary index set. There exists a constant $K > 0$ that depends only on d such that for all $T \in \mathbb{N}$, there exists another compound cq-channel $x \mapsto \rho_x^B(\widetilde{\theta})$ with $x \in \mathcal{X}$, $\rho_x^B \in \mathcal{D}(\mathcal{H})$, and $\widetilde{\theta} \in \widetilde{\Theta}$ such that*

1) the set $\widetilde{\Theta}$ is finite, i.e.,

$$|\widetilde{\Theta}| \leqslant K^{|\mathcal{X}|} T^{6|\mathcal{X}|d^2}; \tag{24}$$

2) for all $\theta \in \Theta$, there exists a $\widetilde{\theta} \in \widetilde{\Theta}$ such that for all $\mathbf{x} \in \mathcal{X}^T$, we have

$$\|\rho_{\mathbf{x}}^{\mathbf{B}}(\theta) - \rho_{\mathbf{x}}^{\mathbf{B}}(\widetilde{\theta})\|_1 \leqslant T^{-5}; \tag{25}$$

3) for all Probability Mass Functions (PMFs) P_X over $\mathcal{X}$, we have

$$\min_{\widetilde{\theta} \in \widetilde{\Theta}} I(P_X, \rho_x^B(\widetilde{\theta})) \geqslant \inf_{\theta \in \Theta} I(P_X, \rho_x^B(\theta)) - 2T^{-6} \log\left(T^6 d\right). \tag{26}$$

Proof. We modify the proof provided in [22] to derive a tighter upper-bound on the approximation error of the new compound channel at the expense of increasing its size. By [22, Theorem 5.5], for all $\kappa > 0$, there exists a partition of all cq-channels from $\mathcal{X}$ to $\mathcal{D}(\mathcal{H})$ denoted by $\Pi = \{\pi_1, \cdots, \pi_n\}$ such that $n \leqslant K^{|\mathcal{X}|}\kappa^{-|\mathcal{X}|d^2}$, where K only depends on the dimension of $\mathcal{H}$, d, and the diameter of Π is at most κ, i.e., for all $i \in [\![1, n]\!]$, for any two channels $x \mapsto \rho_x^B$ and $x \mapsto \widetilde{\rho}_x^B$ in π_i, for any $x \in \mathcal{X}$, we have $\|\rho_x^B - \widetilde{\rho}_x^B\|_1 \leqslant \kappa$. Setting $\kappa = T^{-6}$, this implies that there exists a partition of size at most $K^{|\mathcal{X}|}T^{6|\mathcal{X}|d^2}$ and diameter at most T^{-6}. We construct the new compound cq-channel $x \mapsto \rho_x^B(\widetilde{\theta})$ by selecting an arbitrary channel from each π_i whose intersection with $\{x \mapsto \rho_x^B(\theta) : \theta \in \Theta\}$ is non-empty. We now show that this compound channel satisfies the conditions mentioned in the statement of the lemma. Since we select at most one channel from each π_i, $|\widetilde{\Theta}| \leqslant n \leqslant K^{|\mathcal{X}|}T^{6|\mathcal{X}|d^2}$, and thus, we have (24). To prove (25), consider any $\theta \in \Theta$. By our construction, there should be a $\widetilde{\theta} \in \widetilde{\Theta}$ such that $x \mapsto \rho_x^B(\widetilde{\theta})$ and $x \mapsto \rho_x^B(\theta)$ belong to the same π_i. Therefore, for any $\mathbf{x} \in \mathcal{X}^T$, we have

$$\|\rho_{\mathbf{x}}^{\mathbf{B}}(\theta) - \rho_{\mathbf{x}}^{\mathbf{B}}(\widetilde{\theta})\|_1 = \|\rho_{x_1}^B(\theta) \otimes \cdots \otimes \rho_{x_T}^B(\theta) - \rho_{x_1}^B(\widetilde{\theta}) \otimes \cdots \otimes \rho_{x_T}^B(\widetilde{\theta})\|_1 \tag{27}$$

$$\leqslant \sum_{t=1}^{T} \|\rho_{x_t}^B(\theta) - \rho_{x_t}^B(\widetilde{\theta})\|_1 \tag{28}$$

$$\overset{(a)}{\leqslant} T^{-5}, \tag{29}$$

where (a) follows since $x \mapsto \rho_x^B(\theta)$ and $x \mapsto \rho_x^B(\widetilde{\theta})$ belong to the same π_i, and the diameter of the partition is less than T^{-6}. Finally, let P_X be any PMF over $\mathcal{X}$; to lower-bound $\min_{\widetilde{\theta} \in \widetilde{\Theta}} I(P_X, \rho_x^B(\widetilde{\theta}))$ as in (26), take any $\widetilde{\theta} \in \widetilde{\Theta}$ and consider θ such that $x \mapsto \rho_x^B(\theta)$ and $x \mapsto \rho_x^B(\widetilde{\theta})$ belong to the same π_i. To complete the lemma, it is enough to show that

$$I(P_X, \rho_x^B(\widetilde{\theta})) \geqslant I(P_X, \rho_x^B(\theta)) - 2T^{-6} \log\left(T^6 d\right). \tag{30}$$

To this end, we have

$$I(P_X, \rho_x^B(\widetilde{\theta})) = H\left(\sum_x P_X(x)\rho_x^B(\widetilde{\theta})\right) - \sum_x P_X(x) H\left(\rho_x^B(\widetilde{\theta})\right) \tag{31}$$

$$\overset{(a)}{\geqslant} H\left(\sum_x P_X(x)\rho_x^B(\theta)\right) - \sum_x P_X(x) H\left(\rho_x^B(\theta)\right) \tag{32}$$

$$- 2T^{-6} \log\left(T^6 d\right) \tag{33}$$

$$= I(P_X, \rho_x^B(\widetilde{\theta})) - 2T^{-6} \log\left(T^6 d\right), \tag{34}$$

where (a) follows from Fannes's inequality which states that for any two ρ and σ in $\mathcal{D}(\mathcal{H})$, if $\|\rho - \sigma\|_1 \leqslant \delta \leqslant e^{-1}$, we have $|H(\rho) - H(\sigma)| \leqslant \delta \log(d\delta^{-1})$. □

1) Universal reliability result: We next prove a universal reliability result suitable for covert communications. Note that we cannot use the result of [22] directly since the input distribution used to analyze covert communications changes with the block-length. Indeed, our inspection of the proof of [22] suggests that the technique cannot be adapted for the covert case since the the penalty arising from the approximation of a class of channels dominates the number of bits that one can transmit covertly, which scales as $O\left(\sqrt{T}\right)$. Therefore, we use a different approach

based on the quantum universal decoder introduced by Hayashi in [21]. We first state the following lemma from [22] which is a general achievability result for cq-channels.

Lemma 2 ([22, Theorem 5.4])**.** *Let $x \mapsto \rho_x^B$ be any cq-channel with input set $\mathcal{X}$, and M be a positive integer. For all x, let Γ_x be an operator on $\mathcal{H}^B$ with $0 \leqslant \Gamma_x \leqslant I$, and P_X be a probability distribution over $\mathcal{X}$. If $F : [\![1, M]\!] \to \mathcal{X}$ is a random encoder whose codewords are iid according to P_X, there exists a "universal" decoder corresponding to a POVM $\{\Lambda_w\}_{w=1}^M$ depending on the operators Γ_x and the encoder F (not on the channel) such that the average probability of error satisfies*

$$\mathbb{E}_F\left(\sum_{w=1}^{M}\left(1-\text{tr}\left(\rho_{F(w)}^B\Lambda_w\right)\right)\right) \leqslant 2\sum_x P_X(x)\text{tr}\left(\rho_x^B\Gamma_x\right) + 4M\sum_x P_X(x)\text{tr}\left(\rho^B\Gamma_x\right), \tag{35}$$

where $\rho^B \triangleq \sum_x P_X(x)\rho_x^B$.

We next consider a stationary memoryless cq-channel $x \mapsto \rho_x^B$ with T channel uses and for each codeword $\mathbf{x} \in \mathcal{X}^T$, we aim to construct the operator $\Gamma_{\mathbf{x}}$ *independent* of the channel such that we would be able to upper-bound the right hand side of (35). We shall follow the approach in [21], which is based on the following result from representation theory.

Theorem 3 (Schur-Weyl Duality)**.** *Let H be a d-dimensional Hilbert space over $\mathbb{C}$. For any $T \geqslant 1$, we have the decomposition*

$$H^{\otimes T} = \bigoplus_{\mathbf{t}\in Y_T^d} \mathcal{U}_{\mathbf{t}} \otimes \mathcal{V}_{\mathbf{t}}, \tag{36}$$

where $Y_T^d \triangleq \{(t_1, \cdots, t_d) \in \mathbb{Z}^d : t_1 \geqslant \cdots \geqslant t_d \geqslant 0, \sum_{i=1}^d t_i = T\}$, $\mathcal{U}_{\mathbf{t}}$ is an irreducible representation of SU(d), and $\mathcal{V}_{\mathbf{t}}$ is an irreducible representation of the T^{th} order symmetric group.

In [21], for all $\mathbf{t} \in Y_T^d$ and all T, the author introduced several quantum states that satisfy universal matrix inequalities for all density matrices and all cq-channels. Since those quantum states are a substantial ingredient of the construction of our universal decoder, we state here their definition and properties from [21].

Definition 1. *For $\mathbf{t} \in Y_Y^d$, let $I_{\mathbf{t}}$ be the projection onto the subspace $\mathcal{U}_{\mathbf{t}} \otimes \mathcal{V}_{\mathbf{t}}$. Define*

$$\sigma_{\mathbf{t}} \triangleq \frac{1}{\dim(\mathcal{U}_{\mathbf{t}} \otimes \mathcal{V}_{\mathbf{t}})} I_{\mathbf{t}} \tag{37}$$

$$\sigma_{U,T} \triangleq \sum_{\mathbf{t}\in Y_T^d} \frac{1}{|Y_T^d|}\sigma_{\mathbf{t}}. \tag{38}$$

Moreover, for $\mathbf{x}' = (0, \cdots, 0, 1, \cdots, 1) \in \mathcal{X}^T$ with $\text{wt}(\mathbf{x}') = m$, we define $\sigma_{\mathbf{x}'} \triangleq \sigma_{U,T-m} \otimes \sigma_{U,m}$. For any $\mathbf{x} \in \mathcal{X}^T$ with $\text{wt}(\mathbf{x}) = m$, let π be a permutation of T elements such that $\mathbf{x} = \pi\mathbf{x}'$. We then define $\sigma_{\mathbf{x}} \triangleq U_\pi \sigma_{\mathbf{x}'} U_\pi^\dagger$ where U_π is the unitary representation of π.

Lemma 3. *For any density matrix ρ on $\mathcal{H}$ and any cq-channel $x \mapsto \rho_x^B$, we have*

$$T^{\frac{d(d-1)}{2}}|Y_T^d|\sigma_{U,T} \succeq \rho^{\otimes T}, \tag{39}$$

$$T^{|\mathcal{X}|\frac{d(d-1)}{2}}|Y_T^d|\sigma_{\mathbf{x}} \succeq \rho_{\mathbf{x}}^{\mathbf{B}}. \tag{40}$$

Proof. See [21, Equation (6) and (7)]. □

Lemma 4. *Fix ζ and $\widetilde{\lambda}$ in $]0,1[$. Let $x \mapsto \rho_x^B(\theta)$ be a compound cq-channel with $\theta \in \Theta$ and $x \in \mathcal{X} = \{0,1\}$ such that $\lambda_{\min}(\rho_0^B) \geqslant \widetilde{\lambda}$ for all $\theta \in \Theta$. For a fixed T, let*

$$\log M \triangleq \lfloor (1-\zeta)\alpha_T \inf_{\theta\in\Theta} \mathbb{D}\big(\rho_1^B(\theta)\|\rho_0^B(\theta)\big)T \rfloor, \tag{41}$$

and $F : [\![1, M]\!] \to \mathcal{X}^T$ be a random encoder such that $F(1), \cdots, F(M)$ are iid according to $P_X^{\otimes T}$ with $P_X =$ Bernoulli(α_T) and α_T as in (9)*.*

Then, there exists T_0 that depends only on $\dim \mathcal{H}$, ζ, and $\widetilde{\lambda}$ such that for all $T \geqslant T_0$,

$$\mathbb{P}_F\big(\forall\theta \in \Theta, P_e(\theta) \leqslant 2T^{-5}\big) \geqslant \frac{2}{3}. \tag{42}$$

Proof. We first consider the compound cq-channel $x \mapsto \rho_x^B(\widetilde{\theta})$ obtained by applying Lemma 1 to the compound cq-channel $x \mapsto \rho_x^B(\theta)$. By Lemma 2, for each $\widetilde{\theta} \in \widetilde{\Theta}$, the expectation of the probability of error with respect to random coding is upper-bounded by

$$2\sum_{\mathbf{x}} P_X^{\otimes T}(\mathbf{x})\mathrm{tr}\left(\rho_{\mathbf{x}}\Gamma_{\mathbf{x}}\right) + 4M\sum_{\mathbf{x}} P_X^{\otimes T}(\mathbf{x})\mathrm{tr}\left(\left(\rho^B\right)^{\otimes T}\Gamma_{\mathbf{x}}\right), \tag{43}$$

where $\Gamma_{\mathbf{x}} \triangleq \{\sigma_{\mathbf{x}} - \gamma\sigma_{U,T} \succeq 0\}$ [21]. To upper-bound the first term in (43), we split the summation into three parts based on the weight of the codeword $\mathbf{x}$. In particular, for two thresholds $w_\ell < T\alpha_T < w_u \leqslant 2T\alpha_T$, we obtain with a Chernoff bound

$$\sum_{\mathbf{x}:\mathrm{wt}(\mathbf{x})<w_\ell} P_X^{\otimes T}(\mathbf{x})\mathrm{tr}\left(\rho_{\mathbf{x}}\Gamma_{\mathbf{x}}\right) \leqslant \sum_{\mathbf{x}:\mathrm{wt}(\mathbf{x})<w_\ell} P_X^{\otimes T}(\mathbf{x}) \tag{44}$$

$$= \mathbb{P}_{P_X^{\otimes T}}(\mathrm{wt}(\mathbf{X}) \leqslant w_\ell) \tag{45}$$

$$\leqslant e^{-\frac{1}{2}\left(1-\frac{w_\ell}{T\alpha_T}\right)^2 T\alpha_T}, \tag{46}$$

and analogously

$$\sum_{\mathbf{x}:\mathrm{wt}(\mathbf{x})>w_u} P_X^{\otimes T}(\mathbf{x})\mathrm{tr}\left(\rho_{\mathbf{x}}\Gamma_{\mathbf{x}}\right) \leqslant e^{-\frac{1}{3}\left(\frac{w_u}{T\alpha_T}-1\right)^2 T\alpha_T}. \tag{47}$$

To upper-bound the remaining terms, for $Q_X \sim$ Bernoulli(p), let us define

$$\phi(s,p) \triangleq -(1-s)\log\left(\mathrm{tr}\left(\left(\sum_x Q_X(x)\left(\rho_x^B(\widetilde{\theta})\right)^{1-s}\right)^{\frac{1}{1-s}}\right)\right). \tag{48}$$

Then, by [21, Equation (18)], we have

$$\sum_{\mathbf{x}:w_\ell\leqslant\mathrm{wt}(\mathbf{x})\leqslant w_u} P_X^{\otimes T}(\mathbf{x})\mathrm{tr}\left(\rho_{\mathbf{x}}\Gamma_{\mathbf{x}}\right) \leqslant \sum_{\mathbf{x}:w_\ell\leqslant\mathrm{wt}(\mathbf{x})\leqslant w_u} P_X^{\otimes T}(\mathbf{x}) \min_{s\in[0,1]} (T+1)^{d+sd(d-1)}|Y_T^d|^{2s}\gamma^s e^{-T\phi\left(s,\frac{\mathrm{wt}(\mathbf{x})}{T}\right)} \tag{49}$$

$$\leqslant (T+1)^{d^2}|Y_T^d|^2 \max_{w\in[\![w_\ell,w_u]\!]} \min_{s\in[0,1]} \gamma^s e^{-T\phi\left(s,\frac{w}{T}\right)}. \tag{50}$$

We introduce a result bounding $\phi(s,p)$ for small s and p.

Lemma 5. *For all $\widetilde{\lambda}, \widetilde{s}, \widetilde{p} \in [0,1]$, there exists a universal constant $B > 0$ such that for all cq-channels $x \mapsto \rho_x^B$ with $\lambda_{\min}(\rho_0^B) \geqslant \widetilde{\lambda}$ and for all $s \leqslant \widetilde{s}$ and $p \leqslant \widetilde{p}$, we have*

$$\phi(s,p) \geqslant sI(p) - B(ps^2 + s^3), \tag{51}$$

where $I(p) \triangleq I(Q_X, \rho_x)$ with $Q_X \sim$ Bernoulli(p). Furthermore, for small enough p, we have

$$I(p) \geqslant p\mathbb{D}\big(\rho_1^B \| \rho_0^B\big) - Bp^2. \tag{52}$$

Proof. See Appendix A. □

Applying Lemma 5 to (50), we obtain for all s small enough,

$$(T+1)^{d^2}|Y_T^d|^2 \max_{w \in [\![w_\ell, w_u]\!]} \gamma^s e^{-T\phi\left(s, \frac{w}{T}\right)} \leqslant (T+1)^{d^2}|Y_T^d|^2 \max_{w \in [\![w_\ell, w_u]\!]} \gamma^s e^{-T\left(sI\left(\frac{w}{T}\right) - B\left(\frac{w}{T}s^2 + s^3\right)\right)} \tag{53}$$

$$\leqslant (T+1)^{d^2}|Y_T^d|^2 \max_{w \in [\![w_\ell, w_u]\!]} \gamma^s e^{-T\left(\frac{w}{T}s\mathbb{D}\left(\rho_1^B \| \rho_0^B\right) - B\left(\frac{w^2}{T^2} + \frac{w}{T}s^2 + s^3\right)\right)} \tag{54}$$

$$\leqslant (T+1)^{d^2}|Y_T^d|^2 \gamma^s e^{-T\left(\frac{w_\ell}{T}s\mathbb{D}\left(\rho_1^B(\widetilde{\theta}) \| \rho_0^B(\widetilde{\theta})\right) - B\left(\frac{w_u^2}{T^2} + \frac{w_u}{T}s^2 + s^3\right)\right)}. \tag{55}$$

To upper-bound the second term in (43), we use the operator inequality $A\{A \succeq 0\} \succeq 0$ for any Hermitian operator A. Hence, we have for all $\mathbf{x}$

$$(\sigma_{\mathbf{x}} - \gamma\sigma_{U,T})\Gamma_{\mathbf{x}} \succeq 0. \tag{56}$$

This implies that

$$\left(\sigma_{\mathbf{x}} - \gamma\sigma_{U,T} + \frac{\gamma}{T^{d(d-1)}|Y_T^d|}\left(\rho^B\right)^{\otimes T} - \frac{\gamma}{T^{d(d-1)}|Y_T^d|}\left(\rho^B\right)^{\otimes T}\right)\Gamma_{\mathbf{x}} \succeq 0. \tag{57}$$

Thus, we have

$$\mathrm{tr}\left(\left(\sigma_{\mathbf{x}} - \frac{\gamma}{T^{d(d-1)}|Y_T^d|}\left(\rho^B\right)^{\otimes T}\right)\Gamma_{\mathbf{x}}\right) \geqslant \mathrm{tr}\left(\left(\gamma\sigma_{U,T} - \frac{\gamma}{T^{d(d-1)}|Y_T^d|}\left(\rho^B\right)^{\otimes T}\right)\Gamma_{\mathbf{x}}\right) \overset{(a)}{\geqslant} 0, \tag{58}$$

where (a) follows since by Lemma 3, $\left(\gamma\sigma_{U,T} - \frac{\gamma}{T^{d(d-1)}|Y_T^d|}\left(\rho^B\right)^{\otimes T}\right) \succeq 0$. Accordingly, we conclude that

$$\sum_{\mathbf{x}} P_X^{\otimes T}(\mathbf{x})\mathrm{tr}\left(\left(\rho^B\right)^{\otimes T}\Gamma_{\mathbf{x}}\right) \leqslant \frac{T^{d(d-1)}|Y_T^d|}{\gamma}. \tag{59}$$

Substituting the derived upper-bounds in (43), we obtain

$$\mathbb{E}_F\left(P_e(\widetilde{\theta})\right) \leqslant 2\left(e^{-\frac{1}{2}\left(1-\frac{w_\ell}{T\alpha_T}\right)^2 T\alpha_T} + e^{-\frac{1}{3}\left(\frac{w_u}{T\alpha_T}-1\right)^2 T\alpha_T} + (T+1)^{d^2}|Y_T^d|^2\gamma^s \times e^{-T\left(\frac{w_\ell}{T}s\mathbb{D}\left(\rho_1^B(\widetilde{\theta}) \| \rho_0^B(\widetilde{\theta})\right) - B\left(\frac{w_u^2}{T^2} + \frac{w_u}{T}s^2 + s^3\right)\right)}\right) + 4M\frac{T^{d(d-1)}|Y_T^d|}{\gamma}. \tag{60}$$

By choosing

$$w_\ell = T\alpha_T - (T\alpha_T)^{\frac{2}{3}}, \tag{61}$$

$$w_u = T\alpha_T + (T\alpha_T)^{\frac{2}{3}}, \tag{62}$$

$$\gamma = \left\lfloor \left(1 - \frac{\zeta}{2}\right) \alpha_T \inf_{\theta\in\Theta} \mathbb{D}\big(\rho_1^B(\theta)\|\rho_0^B(\theta)\big)T \right\rfloor, \tag{63}$$

$$s = o(\sqrt{\alpha_T}) \cap \omega(\frac{\log T}{T\alpha_T}), \tag{64}$$

we obtain

$$\mathbb{E}_F\Big(P_e(\widetilde{\theta})\Big) \leqslant 2^{-\omega(\log T)}, \tag{65}$$

where the term $-\omega(\log T)$ depends on $\widetilde{\lambda}$, ζ, and $\dim \mathcal{H}$. By Markov's inequality and the union bound, we have

$$\mathbb{P}_F\Big(\forall\widetilde{\theta} \in \widetilde{\Theta}, P_e(\widetilde{\theta}) \leqslant 3|\widetilde{\Theta}|\mathbb{E}_F\Big(P_e(\widetilde{\theta})\Big)\Big) \geqslant \frac{2}{3}. \tag{66}$$

By Lemma 1, $|\widetilde{\Theta}|$ is upper-bounded by a polynomial in T. This together with (65) implies that $3|\widetilde{\Theta}|\mathbb{E}_F\Big(P_e(\widetilde{\theta})\Big) = 2^{-\omega(\log T)}$. Finally, by Lemma 1, for all $\theta \in \Theta$, there exists $\widetilde{\theta} \in \widetilde{\Theta}$ such that $P_e(\theta) \leqslant P_e(\widetilde{\theta}) + T^{-5}$. Thus, for large enough T, we have

$$\mathbb{P}_F\big(\forall\theta \in \Theta, P_e(\theta) \leqslant 2T^{-5}\big) \geqslant \frac{2}{3}. \tag{67}$$

□

2) Universal resolvability result: We next prove an asymptotic resolvability result for covert distributions.

Lemma 6. *Fix $\widetilde{\lambda}$ and ζ in $]0,1[$. Consider a cq-channel $x \mapsto \rho_x^E$ with $x \in \mathcal{X} = \{0,1\}$ such that $\lambda_{\min}(\rho_0^E) \geqslant \widetilde{\lambda}$. Define $\rho_{\mathbf{x}}^{\mathbf{E}} \triangleq \rho_{x_1}^E \otimes \cdots \otimes \rho_{x_T}^E$ for all $\mathbf{x} = (x_1, \cdots, x_T) \in \mathcal{X}^T$. Let P_X be Bernoulli(α_T) where α_T is defined as in* (9)*, M' be an integer satisfying*

$$M' \geqslant \lceil (1+\zeta)\alpha_T \mathbb{D}\big(\rho_1^E\|\rho_0^E\big)T \rceil, \tag{68}$$

and $F : [\![1, M']\!] \to \mathcal{X}^T$ be a random encoder such that all codewords are distributed according to $P_X^{\otimes T}$ independently. Then, we have

$$\mathbb{E}_F\Big(\Big\|\widehat{\rho}^{\mathbf{E}} - \big(\rho^E\big)^{\otimes T}\Big\|_1\Big) \leqslant 2^{-\omega(\log T)}, \tag{69}$$

where the constant hidden in $\omega(\log T)$ depends only on ζ, $\widetilde{\lambda}$, and $\dim \mathcal{H}$, $\widehat{\rho}^{\mathbf{E}} \triangleq \frac{1}{M'}\sum_{i=1}^{M'} \rho_{F(i)}^{\mathbf{E}}$ and $\rho^E \triangleq \sum_x P_X(x)\rho_x^E$.

Proof. By [23, Lemma 9.2], we have

$$\mathbb{E}_F\Big(\|\widehat{\rho}^{\mathbf{E}} - \big(\rho^E\big)^{\otimes T}\|_1\Big) \leqslant \sqrt{2^{\gamma s + T\phi(s,\alpha_T)}} + \sqrt{\frac{2^\gamma \nu}{M'}}, \tag{70}$$

where ν is the number of distinct eigenvalues of $\big(\rho^E\big)^{\otimes T}$, and for $Q_X \sim$ Bernoulli(p), we define

$$\phi(s,p) \triangleq \log\left(\sum_x P_X(x)\mathrm{tr}\left(\big(\rho_x^E\big)^{1-s}\big(\rho^E\big)^s\right)\right). \tag{71}$$

For $\gamma = \alpha_T \mathbb{D}\big(\rho_1^E \| \rho_0^E\big) T + \frac{\zeta}{2} \alpha_T T$, we have

$$\sqrt{2^{\gamma s + T\phi(s,\alpha_T)}} + \sqrt{\frac{2^{\gamma}\nu}{M'}} \leqslant \sqrt{2^{s\alpha_T T\left(\mathbb{D}\left(\rho_1^E \| \rho_0^E\right) + \frac{\zeta}{2} + \frac{\phi(s,\alpha_T)}{s\alpha_T}\right)}} + \sqrt{2^{-\frac{\zeta}{2}\alpha_T T}\nu} \tag{72}$$

$$\overset{(a)}{\leqslant} \sqrt{2^{s\alpha_T T\left(\mathbb{D}\left(\rho_1^E \| \rho_0^E\right) + \frac{\zeta}{2} + \frac{\phi(s,\alpha_T)}{s\alpha_T}\right)}} + \sqrt{2^{-\frac{\zeta}{2}\alpha_T T}(T+1)^{\dim \mathcal{H}^E}} \tag{73}$$

$$\leqslant \sqrt{2^{s\alpha_T T\left(\mathbb{D}\left(\rho_1^E \| \rho_0^E\right) + \frac{\zeta}{2} + \frac{\phi(s,\alpha_T)}{s\alpha_T}\right)}} + \frac{1}{2} 2^{-\xi \alpha_T T}, \tag{74}$$

where (a) follows from [23, Lemma 3.7] and ξ is small positive number.

Lemma 7. *Fix $\widetilde{s} < 0$, $\widetilde{p} \in [0,1]$, and $\widetilde{\lambda} \in [0,1]$. There exists a universal constant $B > 0$ such that for all cq-channels $x \mapsto \rho_x^E$, $p \in [0, \widetilde{p}]$, and $s \in [\widetilde{s}, 0]$, we have*

$$\phi(s,p) > -I(p)s - B(ps^2 - s^3), \tag{75}$$

where $I(p) \triangleq I(P_X, \rho_x^E)$.

Proof. See Appendix A. □

Applying Lemma 7 to (74), we obtain

$$\sqrt{2^{s\alpha_T T\left(\mathbb{D}\left(\rho_1^E \| \rho_0^E\right) + \frac{\zeta}{2} + \frac{\phi(s,\alpha_T)}{s\alpha_T}\right)}} \leqslant \sqrt{2^{s\alpha_T T\left(\mathbb{D}\left(\rho_1^E \| \rho_0^E\right) + \frac{\zeta}{2} + \frac{-\alpha_T \mathbb{D}\left(\rho_1^E \| \rho_0^E\right) s - B(\alpha_T^2 + \alpha_T s^2 - s^3)}{s\alpha_T}\right)}} \tag{76}$$

$$= \sqrt{2^{s\alpha_T T\left(\frac{\zeta}{2} + \frac{B(\alpha_T^2 + \alpha_T s - s^2)}{\alpha_T}\right)}} \tag{77}$$

By choosing $s = o(\sqrt{\alpha_T}) \cap \omega(\frac{\log T}{T\alpha_T})$,[2] the above expression goes to zero faster than any polynomial. □

Lemma 8. *Fix ζ and $\widetilde{\lambda}$ in $]0,1[$. Let $x \mapsto \rho_x^E(\theta)$ be a compound cq-channel with $x \in \mathcal{X} = \{0,1\}$ and $\theta \in \Theta$ such that for all $\theta \in \Theta$, $\lambda_{\min}(\rho_0^E) \geqslant \widetilde{\lambda}$. Let P_X be as in Lemma 6. Let M' be an integer satisfying*

$$M' \geqslant \lceil (1+\zeta)\alpha_T \sup_{\theta \in \Theta} \mathbb{D}\big(\rho_1^E(\theta) \| \rho_0^E(\theta)\big) T \rceil, \tag{78}$$

and $F : [\![1, M]\!] \times [\![1, M']\!] \to \mathcal{X}^T$ be a random encoder such that all codewords are independently distributed according to $P_X^{\otimes T}$. Then, there exists T_0 depending only on $\dim \mathcal{H}$, ζ, and $\widetilde{\lambda}$ such that for all $T \geqslant T_0$, we have

$$\mathbb{P}_F\left(\forall \theta \in \Theta, \frac{1}{M}\sum_{w=1}^{M} \left\| \widehat{\rho}_w^{\mathbf{E}} - \big(\rho^E(\theta)\big)^{\otimes T} \right\|_1 \leqslant 2T^{-5}\right) \geqslant \frac{2}{3} \tag{79}$$

where $\widehat{\rho}_w^{\mathbf{E}} \triangleq \frac{1}{M'}\sum_{i=1}^{M'} \rho_{F(w,i)}^{\mathbf{E}}$ and $\rho^E(\theta) \triangleq \sum_x P_X(x) \rho_x^E(\theta)$.

Proof. We again consider the compound cq-channel $x \mapsto \rho_x^E(\widetilde{\theta})$ from Lemma 1. By Lemma 6, for all $\widetilde{\theta} \in \widetilde{\Theta}$, we have

$$\mathbb{E}_F\left(\frac{1}{M}\sum_{w=1}^{M} \left\| \widehat{\rho}_w^{\mathbf{E}} - \left(\rho^E(\widetilde{\theta})\right)^{\otimes T} \right\|_1\right) = \frac{1}{M}\sum_{w=1}^{M} \mathbb{E}_F\left(\left\| \widehat{\rho}_w^{\mathbf{E}} - \left(\rho^E(\widetilde{\theta})\right)^{\otimes T} \right\|_1\right) \tag{80}$$

$$\leqslant 2^{-\omega(\log T)}. \tag{81}$$

[2]To find such s, it is required that $\sqrt{\alpha_T} = \omega(\frac{\log T}{T\alpha_T})$ or equivalently $\alpha_T = \omega\left(\left(\frac{\log T}{T}\right)^{\frac{2}{3}}\right)$

By Markov's inequality and the union bound, we have

$$\mathbb{P}_F\left(\forall\widetilde{\theta}\in\widetilde{\Theta},\frac{1}{M}\sum_{w=1}^{M}\left\|\widehat{\rho}_w^{\mathbf{E}}-\left(\rho^E(\widetilde{\theta})\right)^{\otimes T}\right\|_1\leqslant 3|\widetilde{\Theta}|\mathbb{E}_F\left(\frac{1}{M}\sum_{w=1}^{M}\left\|\widehat{\rho}_w^{\mathbf{E}}-\left(\rho^E(\widetilde{\theta})\right)^{\otimes T}\right\|_1\right)\right)\geqslant\frac{2}{3}. \tag{82}$$

Since $|\widetilde{\Theta}|$ is upper-bounded by a polynomial in T, we have

$$3|\widetilde{\Theta}|\mathbb{E}_F\left(\frac{1}{M}\sum_{w=1}^{M}\left\|\widehat{\rho}_w^{\mathbf{E}}-\left(\rho^E(\widetilde{\theta})\right)^{\otimes T}\right\|_1\right)=2^{-\omega(\log T)}. \tag{83}$$

Finally, by Lemma 1, for all $\theta\in\Theta$, there exists $\widetilde{\theta}\in\widetilde{\Theta}$ such that

$$\frac{1}{M}\sum_{w=1}^{M}\left\|\widehat{\rho}_w^{\mathbf{E}}-\left(\rho^E(\theta)\right)^{\otimes T}\right\|_1\leqslant\frac{1}{M}\sum_{w=1}^{M}\left\|\widehat{\rho}_w^{\mathbf{E}}-\left(\rho^E(\widetilde{\theta})\right)^{\otimes T}\right\|_1+T^{-5}. \tag{84}$$

Thus, for large enough T, we have

$$\mathbb{P}_F\left(\forall\theta\in\Theta,\frac{1}{M}\sum_{w=1}^{M}\left\|\widehat{\rho}_w^{\mathbf{E}}-\left(\rho^E(\theta)\right)^{\otimes T}\right\|_1\leqslant 2T^{-5}\right)\geqslant\frac{2}{3}. \tag{85}$$

□

3) Proof of Theorem 2: We are now ready to provide the proof of the main result of this section. Our code construction is similar to [8], which uses wiretap coding to ensure the security of a covert message. Fix ζ, $\widetilde{\lambda}^B$, $\widetilde{\lambda}^E$, D^B, and D^E, and let Θ be an arbitrary indexing of all cq-channels $x\mapsto\rho_x^{BE}$ satisfying (17)-(20) for which the corresponding cq-channel to $\theta\in\Theta$ is $x\mapsto\rho_x^{BE}(\theta)$. Considering the sequence $\{\alpha_T\}_{T\geqslant 1}$ as in (9), for a fixed large enough T, let P_X be Bernoulli(α_T); let $F:[\![1,M]\!]\times[\![1,M']\!]\to\mathcal{X}^T$ be a random encoder whose codewords are iid according to $P_X^{\otimes T}$ that encodes two messages W and W' uniformly distributed over $[\![1,M]\!]$ and $[\![1,M']\!]$, respectively, to a codeword $\mathbf{X}$. By Lemma 4, for

$$\log M+\log M'=\lfloor(1-\zeta)\alpha_T\inf_{\theta\in\Theta}\mathbb{D}\big(\rho_1^B\|\rho_0^B\big)T\rfloor \tag{86}$$

$$\geqslant\lfloor(1-\zeta)\alpha_T D^B T\rfloor, \tag{87}$$

we have

$$\mathbb{P}_F\big(\forall\theta\in\Theta,P_e(\theta)\leqslant 2T^{-5}\big)\geqslant\frac{2}{3}, \tag{88}$$

where $P_e(\theta)$ is the probability that at least one of the messages W and W' is not decoded correctly at the receiver when the cq-channel corresponding to index θ is used. Moreover, by Lemma 8, for

$$\log M'=\lceil(1+\zeta)\alpha_T\sup_{\theta\in\Theta}\mathbb{D}\big(\rho_1^E(\theta)\|\rho_0^E(\theta)\big)T\rceil \tag{89}$$

$$\leqslant\lceil(1+\zeta)\alpha_T D^E T\rceil, \tag{90}$$

we have

$$\mathbb{P}_F\left(\forall\theta\in\Theta,\frac{1}{M}\sum_{w=1}^{M}\mathbb{D}\Big(\widehat{\rho}_w^{\mathbf{E}}\|\left(\rho^E(\theta)\right)^{\otimes T}\Big)\leqslant 2T^{-5}\right)\geqslant\frac{2}{3}, \tag{91}$$

where $\widehat{\rho}_w^{\mathbf{E}}$ and $\rho^E(\theta)$ are defined in the statement of Lemma 8. Inequalities (88) and (91) imply that there exists a realization f of F such that for all $\theta \in \Theta$,

$$P_e(\theta) \leqslant 2T^{-5}, \tag{92}$$

$$\frac{1}{M}\sum_{w=1}^{M} \left\| \widehat{\rho}_w^{\mathbf{E}} - \left(\rho^E(\theta)\right)^{\otimes T} \right\|_1 \leqslant 2T^{-5}. \tag{93}$$

Hence, by Lemma 18, we upper-bound the quantum relative entropy between the induced quantum states and $\left(\rho^E(\theta)\right)^{\otimes T}$ as

$$\frac{1}{M}\sum_{w=1}^{M} \mathbb{D}\Big(\widehat{\rho}_w^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big) \leqslant 2T^{-5} \log \frac{d^T}{\left(\lambda_{\min}(\rho^E(\theta))\right)^T 2T^{-5}} \tag{94}$$

$$= 2T^{-4}\left(\log \frac{d}{\lambda_{\min}(\rho^E(\theta))} + \frac{5\log T - \log 2}{T}\right). \tag{95}$$

To lower-bound the minimum eigenvalue of $\rho^E(\theta)$, we use Lemma 13 in Appendix B to obtain for large T,

$$\lambda_{\min}(\rho^E(\theta)) = \lambda_{\min}(\alpha_T \rho_1^E(\theta) + (1-\alpha_T)\rho_0^E(\theta)) \tag{96}$$

$$\geqslant \lambda_{\min}((1-\alpha_T)\rho_0^E(\theta)) - \|\alpha_T \rho_1^E(\theta)\|_1 \tag{97}$$

$$\geqslant (1-\alpha_T)\lambda_{\min}(\rho_0^E(\theta)) - \alpha_T \tag{98}$$

$$\geqslant \frac{\widetilde{\lambda}^E}{2}. \tag{99}$$

Therefore, for some constant $L_1 > 0$ depending on d and $\widetilde{\lambda}^E$, we have

$$\frac{1}{M}\sum_{w=1}^{M} \mathbb{D}\Big(\widehat{\rho}_w^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big) \leqslant L_1 T^{-4}. \tag{100}$$

To analyze the secrecy of the protocol, since there is no public communication and W is the key extracted at Alice's end, the information leakage to the adversary is

$$\mathbb{D}\big(\rho^{\mathbf{E}W} \| \rho^{\mathbf{E}} \otimes \rho_{\text{unif}}^W\big) \overset{(a)}{=} \mathbb{D}\big(\rho^{\mathbf{E}W} \| \rho^{\mathbf{E}} \otimes \rho_{\text{unif}}^W\big) \tag{101}$$

$$\leqslant \mathbb{D}\Big(\rho^{\mathbf{E}W} \| \left(\rho^E(\theta)\right)^{\otimes T} \otimes \rho_{\text{unif}}^W\Big) \tag{102}$$

$$= \frac{1}{M}\sum_{w=1}^{M} \mathbb{D}\Big(\widehat{\rho}_w^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big) \tag{103}$$

$$\leqslant L_1 T^{-4}, \tag{104}$$

where (a) follows since there is no public communication. For the covertness, first note that by convexity of quantum relative entropy, we have

$$\mathbb{D}\Big(\rho^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big) \leqslant \frac{1}{M}\sum_{w=1}^{M} \mathbb{D}\Big(\widehat{\rho}_w^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big) \leqslant L_1 T^{-4}. \tag{105}$$

We can subsequently bound $\mathbb{D}\Big(\rho^{\mathbf{E}} \| \left(\rho_0^E(\theta)\right)^{\otimes T}\Big)$ as

$$\mathbb{D}\Big(\rho^{\mathbf{E}} \| \left(\rho_0^E(\theta)\right)^{\otimes T}\Big) = \mathbb{D}\Big(\rho^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big) + \mathbb{D}\Big(\left(\rho^E(\theta)\right)^{\otimes T} \| \left(\rho_0^E(\theta)\right)^{\otimes T}\Big) + \text{tr}\left(\left(\rho^{\mathbf{E}} - \left(\rho^E(\theta)\right)^{\otimes T}\right) \times \left(\log\left(\rho^E(\theta)\right)^{\otimes T} - \log\left(\rho_0^E(\theta)\right)^{\otimes T}\right)\right) \tag{106}$$

$$\leqslant \mathbb{D}\Big(\rho^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big) + \mathbb{D}\Big(\left(\rho^E(\theta)\right)^{\otimes T} \| \left(\rho_0^E(\theta)\right)^{\otimes T}\Big)$$

$$+ \left\| \rho^{\mathbf{E}} - \left(\rho^E(\theta)\right)^{\otimes T} \right\|_1 T \left(\log \frac{1}{\lambda_{\min}(\rho_0^E(\theta))\lambda_{\min}(\theta)} \right) \quad (107)$$

$$\overset{(a)}{\leqslant} \mathbb{D}\Big(\rho^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big) + \mathbb{D}\Big(\left(\rho^E(\theta)\right)^{\otimes T} \| \left(\rho_0^E(\theta)\right)^{\otimes T}\Big) + \sqrt{\mathbb{D}\Big(\rho^{\mathbf{E}} \| \left(\rho^E(\theta)\right)^{\otimes T}\Big)} T \times \left(\log \frac{1}{\lambda_{\min}(\rho_0^E(\theta))\lambda_{\min}(\rho^E(\theta))} \right) \quad (108)$$

$$\leqslant L_1 T^{-4} + \mathbb{D}\Big(\left(\rho^E(\theta)\right)^{\otimes T} \| \left(\rho_0^E(\theta)\right)^{\otimes T}\Big) + \sqrt{L_1 T^{-4}} T \log \frac{1}{\lambda_{\min}(\rho_0^E(\theta))\lambda_{\min}(\rho^E(\theta))} \quad (109)$$

$$= \mathbb{D}\Big(\left(\rho^E(\theta)\right)^{\otimes T} \| \left(\rho_0^E(\theta)\right)^{\otimes T}\Big) + L_1 T^{-4} + \sqrt{L_1} \log \frac{1}{\lambda_{\min}(\rho_0^E(\theta))(\rho^E(\theta))} T^{-1}, \quad (110)$$

$$\overset{(b)}{\leqslant} \frac{\alpha_T^2 \eta(\rho_1^E(\theta) \| \rho_0^E(\theta))}{2} T + L_2 \alpha_T^3 T + L_1 T^{-4} + \sqrt{L_1} \log \frac{1}{\lambda_{\min}(\rho_0^E)\lambda_{\min}(\rho^E(\theta))} T^{-1} \quad (111)$$

$$\leqslant \frac{\alpha_T^2 \eta(\rho_1^E(\theta) \| \rho_0^E(\theta))}{2} T + L_2 \alpha_T^3 T + L_1 T^{-4} + 2\sqrt{L_1} \log \frac{2}{\widetilde{\lambda}^E} T^{-1}, \quad (112)$$

where (a) follows from Pinsker inequality, and (b) follows from [13, Lemma 1].

B. Covert Quantum Tomography

1) Instantiation of a covert estimation protocol: We now detail how Alice and Bob can covertly form estimates of $\mathbb{D}\big(\rho_1^B \| \rho_0^B\big)$ and $\mathbb{D}\big(\rho_1^W \| \rho_0^W\big)$. If the channel from Alice to Bob is $\mathcal{E}_{A\to B}$, the goal of the estimation phase would be to first verify the conditions (12) and (13), and if they hold, to estimate $D^B(\mathcal{E}) \triangleq \mathbb{D}\big(\mathcal{E}_{A\to B}(\rho_1^A) \| \mathcal{E}_{A\to B}(\rho_0^A)\big)$ and $D^E(\mathcal{E}) \triangleq \mathbb{D}\big(\mathcal{E}^c_{A\to B}(\rho_1^A) \| \mathcal{E}^c_{A\to B}(\rho_0^A)\big)$. The protocol will be aborted otherwise. We shall use standard quantum tomography [20] and adapt it to be covert. We start the description of the estimation phase by formally defining an estimation protocol. Suppose Alice and Bob have access to private randomness R distributed according to P_R over $\mathcal{R}$ and use T' channel uses for the estimation phase. The estimation protocol consists of an encoder function $f : \mathcal{R} \to \mathcal{D}(\mathcal{H})^{T'}$ for Alice, a POVM $\mathbf{M}_r = \{M_r^j\}_{j\in\mathcal{J}}$ for each $r \in \mathcal{R}$ applied by Bob to his received state $\rho^{\mathbf{B}}$ when $R = r$ and results in an output j in $\mathcal{J}$ the set of all possible outputs of the measurement, one function $H : \mathcal{J} \to \{0, 1\}$ used by Bob to verify that (12) and (13) hold, and two estimators $\widehat{D}^B : \mathcal{J} \to \mathbb{R}$ and $\widehat{D}^W : \mathcal{J} \to \mathbb{R}$ used by Bob to form estimations of $\mathbb{D}\big(\mathcal{E}_{A\to B}(\rho_1^A) \| \mathcal{E}_{A\to B}(\rho_0^A)\big)$ and $\mathbb{D}\big(\mathcal{E}^c_{A\to B}(\rho_1^A) \| \mathcal{E}^c_{A\to B}(\rho_0^A)\big)$, respectively.

We now explicitly instantiate a covert estimation protocol. Consider any number of channel uses T' and any quantum channel $\mathcal{E} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ where $\mathcal{H}$ is a d-dimensional Hilbert space. Let $\widetilde{E}_1, \cdots, \widetilde{E}_{d^2}$ be defined as $\widetilde{E}_{d(n-1)+m} \triangleq |n\rangle\langle m|$ for an orthonormal basis $|1\rangle, \cdots, |d\rangle$. Our goal is to estimate $\mathcal{E}(\widetilde{E}_n)$ for all $n \in [\![1, d^2]\!]$ from which we would have a complete characterization of the quantum channel $\mathcal{E}$. To do so, the main idea is that Alice would send some states through Pulse-Position Modulation (PPM) to Bob for which Bob performs quantum state tomography. More concretely, Alice and Bob first agree on two integers q and ℓ such that $q\ell \leqslant T'$ and sample an iid sequence $U_1, \cdots, U_\ell$ from their private randomness, where each U_i is uniformly distributed over $[\![1, q]\!]$. Alice then transmits the innocent state ρ_0^A on the i^{th} channel uses unless

$$i \in \mathcal{I} \triangleq \{U_1, U_2 + q, \cdots, U_\ell + q(\ell - 1)\}. \quad (113)$$

To determine the state that should be sent by Alice on the positions in $\mathcal{I}$, let us define the vectors

$$|n,m\rangle_+ \triangleq \frac{|n\rangle + |m\rangle}{\sqrt{2}} \tag{114}$$

$$|n,m\rangle_- \triangleq \frac{|n\rangle + i|m\rangle}{\sqrt{2}} \tag{115}$$

and consider pure states

$$\mathcal{S} \triangleq \{|n,m\rangle_+\langle n,m|_+ : n \neq m\} \cup \{|n,m\rangle_-\langle n,m|_- : n \neq m\} \cup \{|n\rangle\langle n| : n \in [\![1,d]\!]\}, \tag{116}$$

where $|\mathcal{S}| = 2d^2 - d$. On the positions in $\mathcal{I}$, in an arbitrary but known order, Alice transmits each state in $\mathcal{S}$ $\lfloor \ell / |S| \rfloor$ times. Then, for each state $\rho \in \mathcal{S}$, Bob receives $\lfloor \ell / |S| \rfloor$ independent copies of $\mathcal{E}(\rho)$, and performs a POVM defined by $\{\widetilde{\rho}, I - \widetilde{\rho}\}$ for each operator $\widetilde{\rho} \in \mathcal{S}$, $\widetilde{\ell} \triangleq \lfloor \lfloor \ell / |S| \rfloor / |\mathcal{S}| \rfloor$ times. Let $\widehat{N}(\rho, \widetilde{\rho})$ be the number of times the result of the measurement $\{\widetilde{\rho}, I - \widetilde{\rho}\}$ on $\mathcal{E}(\rho)$ corresponds to $\widetilde{\rho}$ and let $\widehat{f}(\rho, \widetilde{\rho}) \triangleq \widehat{N}(\rho, \widetilde{\rho}) / \widetilde{\ell}$. Bob subsequently estimates $\mathcal{E}(\rho)$ for each $\rho \in \mathcal{S}$ as

$$\widehat{\mathcal{E}}(\rho) \triangleq \sum_{n \neq m} |n\rangle\langle m| \left(\widehat{f}(\rho, |n,m\rangle_+\langle n,m|_+) - i\widehat{f}(\rho, |n,m\rangle_-\langle n,m|_-) - \frac{1-i}{2}\widehat{f}(\rho, |n\rangle\langle n|) \right.$$
$$\left. - \frac{1-i}{2}\widehat{f}(\rho, |m\rangle\langle m|) \right) + \sum_n |n\rangle\langle n| \widehat{f}(\rho, |n\rangle\langle n|). \tag{117}$$

Since $\{\widetilde{E}_j : j \in [\![1, d^2]\!]\}$ is an orthonormal basis for $\mathcal{L}(\mathcal{H})$, we can write $\widehat{\mathcal{E}}(\rho) \triangleq \sum_j \widetilde{E}_j \widehat{\lambda}_{\rho,j}$ for some unique $\widehat{\lambda}_{\rho,j}$. Then, for $n, m \in [\![1,d]\!]$, we define

$$\widehat{\mathcal{E}}(\widetilde{E}_{d(n-1)+m}) \triangleq \begin{cases} \widehat{\mathcal{E}}(|n,m\rangle_+\langle n,m|_+) + i\widehat{\mathcal{E}}(|n,m\rangle_-\langle n,m|_-) - \frac{1+i}{2}\widehat{\mathcal{E}}(|n\rangle\langle n|) - \frac{1+i}{2}\widehat{\mathcal{E}}(|m\rangle\langle m|) & n \neq m, \\ \widehat{\mathcal{E}}(|n\rangle\langle n|) & n = m, \end{cases} \tag{118}$$

which is enough to characterize a quantum channel. We can similarly write $\mathcal{E}(\widetilde{E}_{d(n-1)+m}) = \sum_j \widetilde{E}_j \lambda_{d(n-1)+m,j}$ for some unique $\lambda_{d(n-1)+m,j}$. We next attempt to form an estimation of the chi-representation of the channel $\mathcal{E}$, $\{\chi_{j,k}\}$. By [20], for some fixed $\kappa_{j,k}^{j',k'}$,

$$\chi_{j,k} = \sum_{j',k'} \kappa_{j,k}^{j',k'} \lambda_{j',k'}. \tag{119}$$

We thus define $\widehat{\chi}_{j,k} \triangleq \sum_{j',k'} \kappa_{j,k}^{j',k'} \widehat{\lambda}_{j',k'}$. Finally, for some $\tau > 0$, we define

$$H \triangleq \mathbb{1}\Big\{ \lambda_{\min}(\widehat{\chi}) \geqslant \widetilde{\lambda}^{\chi} - \tau \text{ and } \lambda_{\min}(\widehat{\mathcal{E}}(\rho_0^A)) \geqslant \widetilde{\lambda}^{B} - \tau \Big\}, \tag{120}$$

$$\widehat{D}^B \triangleq \mathbb{D}\Big(\widehat{\mathcal{E}}(\rho_1^A) \| \widehat{\mathcal{E}}(\rho_0^A) \Big) - \tau, \tag{121}$$

$$\widehat{D}^E \triangleq \mathbb{D}\Big(\widehat{\mathcal{E}}^c(\rho_1^A) \| \widehat{\mathcal{E}}^c(\rho_0^A) \Big) + \tau. \tag{122}$$

The next theorem establishes bounds on the performance of the described covert estimation protocol.

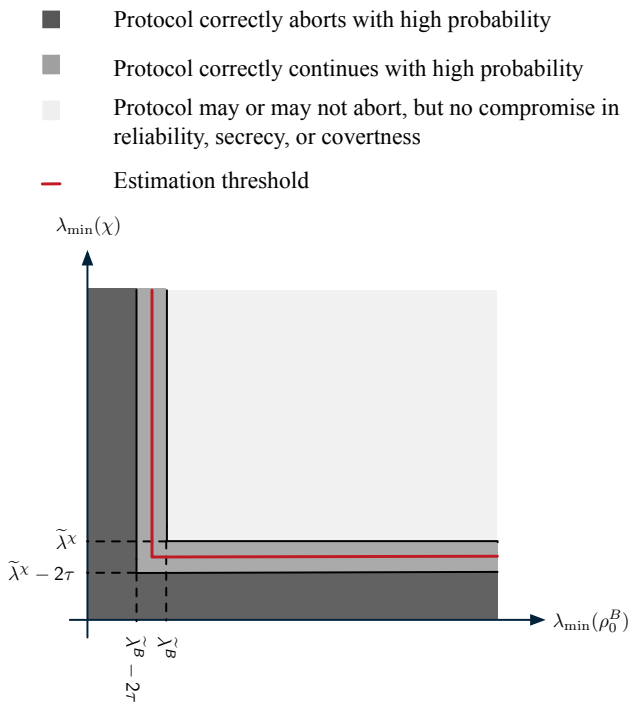

Figure 3. Testing the conditions (12) and (13)

Theorem 4. *There exist $\xi > 0$ that depends on τ, d, $\widetilde{\lambda}^{\chi}$, $\widetilde{\lambda}^{B}$, and $\widetilde{\lambda}^{E}$ such that*

$$\mathbb{D}\big(\rho^{\mathbf{E}W} \| (\rho_0^{\mathbf{E}}) \otimes \rho_{unif}^{W}\big) \leqslant \frac{\ell}{q}\left(\frac{\dim \mathcal{H}^E}{\widetilde{\lambda}^E} - 1\right), \tag{123}$$

$$\mathbb{P}\Big(H = 0 | \lambda_{\min}(\chi) \geqslant \widetilde{\lambda}^{\chi}, \lambda_{\min}(\mathcal{E}(\rho_0^A)) \geqslant \widetilde{\lambda}^{B}\Big) \leqslant 2^{-\xi n}, \tag{124}$$

$$\mathbb{P}\left(H = 1 | \lambda_{\min}(\chi) \leqslant \widetilde{\lambda}^{\chi} - 2\tau \text{ or } \lambda_{\min}(\mathcal{E}(\rho_0^A)) \leqslant \widetilde{\lambda}^{B} - 2\tau\right) \leqslant 2^{-\xi n}, \tag{125}$$

and

$$\mathbb{P}\Big(D^B(\mathcal{E}) - 2\tau \leqslant \widehat{D}^B \leqslant D^B(\mathcal{E}), D^E(\mathcal{E}) \leqslant \widehat{D}^E \leqslant D^E(\mathcal{E}) + 2\tau \,\Big| \, \lambda_{\min}(\chi) \geqslant \widetilde{\lambda}^{\chi} - 2\tau, \lambda_{\min}(\mathcal{E}(\rho_0^A)) \geqslant \widetilde{\lambda}^{B} - 2\tau\Big) \geqslant 1 - 2^{-\xi \ell}. \tag{126}$$

We shall prove Theorem 4 in Section IV-B2. Note that (123) characterizes the covertness of the estimation protocol by bounding the relative entropy between the state induced by the estimation protocol and the state in which there is no communication. (124) and (125) characterize the robustness of estimation since (124) bounds the probability that the channel satisfies the required condition (12) and (13) but Alice and Bob abort the protocol while (125) bounds the probability that Alice and Bob run the key generation phase but the channel does not satisfy the required conditions. Finally, (126) characterizes the accuracy of the estimation by bounding the probability that the estimated parameters of the channel are close to their true values.

As depicted in Fig. 3, there is a technical subtlety in verifying (12) and (13) because the channel estimation error obtained in finite number of channel uses prevents us from testing with absolute certainty that (12) and (13) hold.

In other words, there could exist a set of channels for which, based on the estimation error, Alice and Bob may or may not abort the protocol; regardless, the protocol ensures that if the key generation phase is executed, it is reliable, secure, and covert.

We conclude this section by analyzing the performance of a covert key generation protocol obtained by combining the covert estimation protocol with the universal code introduced in Section IV-A. More precisely, Alice and Bob first perform the described estimation protocol $\mathcal{P}$ over T' channel uses. Using $O(\log T')$ channel uses and $O(\log T')$ bits of private common randomness, Bob transmits the one-time-padded H, $\widehat{D}^B$, and $\widehat{D}^W$ over the public channel. If $H=0$, Alice abort the protocol. If $H=1$, after obtaining $\widehat{D}^B$ and $\widehat{D}^E$, Alice and Bob run the universal code $\mathcal{C}_T$ introduced in Theorem 2 for $D^B=\widehat{D}^B$, $D^E=\widehat{D}^E$, and the lower-bounds on the minimum eigenvalue of ρ_0^B and ρ_0^E, $\widetilde{\lambda}^B-2\tau$ and $\widetilde{\lambda}^E$, respectively. The rationale behind the conservative choice for the minimum eigenvalue of ρ_0^B is that, because to the estimation error, Alice and Bob might accept the channels for which $\lambda_{\min}(\rho_0^B)$ is slightly less than $\widetilde{\lambda}_B$. We characterize the reliability, secrecy, and covertness of the overall protocol in the next lemma and provide the poof in Section IV-B2.

Lemma 9. *For all channels $\mathcal{E}_{A\to B}$ if we only know $\lambda_{\min}(\mathcal{E}^c_{A\to B}(\rho_0^A))\geqslant\widetilde{\lambda}^E$, we have*

$$P_e\leqslant\mathbb{P}(H=1), \tag{127}$$

$$S\leqslant\mathbb{P}(H=1)\left(T\log\frac{1}{\widetilde{\lambda}^E}+\ell^{\max}\right), \tag{128}$$

$$C\leqslant\mathbb{P}(H=1)T\log\frac{1}{\widetilde{\lambda}^E}+\delta, \tag{129}$$

where $L_1,L_2>0$ depend on $\dim\mathcal{H}^E$ and $\widetilde{\lambda}^E$, and $\ell^{\max}$ is the maximum length of the key.

In addition, for a quantum channel $\mathcal{E}_{A\to B}$ with $\lambda_{\min}(\mathcal{E}^c_{A\to B}(\rho_0^A))\geqslant\widetilde{\lambda}^E$ and $\lambda_{\min}(\mathcal{E}_{A\to B}(\rho_0^A))\geqslant\widetilde{\lambda}^B-2\tau$, and an estimation protocol $\mathcal{P}$, define $\epsilon\triangleq\mathbb{P}\Big(H=1$ and $(D^B(\mathcal{E})\leqslant\widehat{D}^B$ or $D^E(\mathcal{E})\geqslant\widehat{D}^E)\Big)$ and $\delta\triangleq\mathbb{D}\big(\rho^{\mathbf{E}W}\|(\rho_0^{\mathbf{E}})\otimes\rho^W_{\rm unif}\big)$. For the protocol described above, we have

$$P_e\leqslant 2T^{-5}+\epsilon, \tag{130}$$

$$S\leqslant L_1T^{-4}+\epsilon\left(T\log\frac{1}{\widetilde{\lambda}^E}+\ell^{\max}\right), \tag{131}$$

$$C\leqslant\frac{\alpha_T^2\eta(\rho_1^E(\theta)\|\rho_0^E(\theta))}{2}T+L_2\alpha_T^3T+L_1T^{-4}+2\sqrt{L_1}\log\frac{2}{\widetilde{\lambda}^E}T^{-1}+\epsilon T\log\frac{1}{\widetilde{\lambda}^E}+\delta. \tag{132}$$

2) Proof of Theorem 4 and Lemma 9: To show that the desired parameters of the channel are approximated properly by their associated estimators, we first show that the estimated channel $\widehat{\mathcal{E}}$ defined in (118) is close to the true channel $\mathcal{E}$, i.e., for all j and k, with high probability, $\widehat{\lambda}_{j,k}$ is close to $\lambda_{j,k}\triangleq\text{tr}\left(\widetilde{E}_k^\dagger\mathcal{E}(\widetilde{E}_j)\right)$.

Lemma 10. *For all $\gamma>0$ and $\kappa_{\max}\triangleq\max_{j,k,j',k'}|\kappa_{j,k}^{j',k'}|$, we have*

$$\mathbb{P}\Big(\exists j,k:|\lambda_{j,k}-\widehat{\lambda}_{j,k}|\geqslant\gamma\Big)\leqslant 16d^4e^{-\frac{1}{256}\widetilde{\ell}\gamma^2}, \tag{133}$$

and

$$\mathbb{P}\big(\exists j,k:|\chi_{j,k}-\widehat{\chi}_{j,k}|\geqslant d^2\kappa_{\max}\gamma\big)\leqslant 16d^4e^{-\frac{1}{256}\widetilde{\ell}\gamma^2}. \tag{134}$$

Proof. We only prove (133) as (134) then follows from the definition of $\widehat{\chi}_{j,k}$. Notice first that, by our construction, the distribution of $\widehat{N}(\rho,\widetilde{\rho}) = \widetilde{\ell}\widehat{f}(\rho,\widetilde{\rho})$ is Binomial$(\mathrm{tr}\left(\widetilde{\rho}\mathcal{E}(\rho)\right),\widetilde{\ell})$ for all $\rho,\widetilde{\rho}\in\mathcal{S}$. Therefore, Hoeffding's inequality yields for all $\gamma>0$ that

$$\mathbb{P}\Big(|\widehat{f}(\rho,\widetilde{\rho}) - \mathrm{tr}\left(\widetilde{\rho}\mathcal{E}(\rho)\right)| \geqslant \gamma\Big) \leqslant 2\exp -2\widetilde{\ell}\gamma^2. \tag{135}$$

For all n, m, n', and m', using the equality

$$\widetilde{E}_{d(n-1)+m} = |n,m\rangle_+\langle n,m|_+ + i|n,m\rangle_-\langle n,m|_- - \frac{1+i}{2}|n\rangle\langle n| - \frac{1+i}{2}|m\rangle\langle m|, \tag{136}$$

we expand $\lambda_{j,k}$ and $\widehat{\lambda}_{j,k}$ in terms of $\mathrm{tr}\left(\widetilde{\rho}\mathcal{E}(\rho)\right)$ and $\widehat{f}(\rho,\widetilde{\rho})$, respectively, and apply (135). More precisely, by definition of $\lambda_{d(n-1)+m,d(n'-1)+m'}$, we have

$$\begin{aligned}\lambda_{d(n-1)+m,d(n'-1)+m'} &= \mathrm{tr}\left(\widetilde{E}^\dagger_{d(n'-1)+m'}\mathcal{E}(\widetilde{E}_{d(n-1)+m})\right) &(137)\\ &= \mathrm{tr}\left(\widetilde{E}^\dagger_{d(n'-1)+m'}\mathcal{E}(|n,m\rangle_+\langle n,m|_+)\right) + i\mathrm{tr}\left(\widetilde{E}^\dagger_{d(n'-1)+m'}\mathcal{E}(|n,m\rangle_-\langle n,m|_-)\right)\\ &\quad - \frac{1+i}{2}\mathrm{tr}\left(\widetilde{E}^\dagger_{d(n'-1)+m'}\mathcal{E}(|n\rangle\langle n|)\right) - \frac{1+i}{2}\mathrm{tr}\left(\widetilde{E}^\dagger_{d(n'-1)+m'}\mathcal{E}(|m\rangle\langle m|)\right). &(138)\end{aligned}$$

We now fix $n',m'\in[\![1,d]\!]$ and $\rho\in\mathcal{S}$ and for simplicity, let $j\triangleq d(n'-1)+m'$, $|a\rangle\triangleq|n',m'\rangle_+$, $|b\rangle\triangleq|n',m'\rangle_-$. Then, by (136),

$$\mathrm{tr}\left(\widetilde{E}^\dagger_j\mathcal{E}(\rho)\right) = \mathrm{tr}\left(|a\rangle\langle a|\mathcal{E}(\rho)\right) - i\mathrm{tr}\left(|b\rangle\langle b|\mathcal{E}(\rho)\right) - \frac{1-i}{2}\mathrm{tr}\left(|n\rangle\langle n|\mathcal{E}(\rho)\right) - \frac{1-i}{2}\mathrm{tr}\left(|m\rangle\langle m|\mathcal{E}(\rho)\right). \tag{139}$$

Therefore, we obtain the upper-bound in (140).

$$\begin{aligned}&\mathbb{P}\Big(|\widehat{\lambda}_{\rho,j} - \mathrm{tr}\left(\widetilde{E}_j\mathcal{E}(\rho)\right)| \geqslant \gamma\Big)\\ &= \mathbb{P}\left(|\widehat{f}(\rho,|a\rangle\langle a|) - i\widehat{f}(\rho,|b\rangle\langle b|) - \frac{1-i}{2}\widehat{f}(\rho,|n'\rangle\langle n'|) - \frac{1-i}{2}\widehat{f}(\rho,|m'\rangle\langle m'|) - \mathrm{tr}\left(\widetilde{E}_j\mathcal{E}(\rho)\right)| \geqslant \gamma\right)\\ &\leqslant \mathbb{P}\Big(|\widehat{f}(\rho,|a\rangle\langle a|) - \mathrm{tr}\left(|a\rangle\langle a|\mathcal{E}(\rho)\right)| \geqslant \frac{\gamma}{4}\Big) + \mathbb{P}\Big(|\widehat{f}(\rho,|b\rangle\langle b|) - \mathrm{tr}\left(|b\rangle\langle b|\mathcal{E}(\rho)\right)| \geqslant \frac{\gamma}{4}\Big) +\\ &\quad \mathbb{P}\left(|\widehat{f}(\rho,|n'\rangle\langle n'|) - \mathrm{tr}\left(|n'\rangle\langle n'|\mathcal{E}(\rho)\right)| \geqslant \frac{\gamma}{2\sqrt{2}}\right) + \mathbb{P}\left(|\widehat{f}(\rho,|m'\rangle\langle m'|) - \mathrm{tr}\left(|m'\rangle\langle m'|\mathcal{E}(\rho)\right)| \geqslant \frac{\gamma}{2\sqrt{2}}\right)\\ &\leqslant 4e^{-\frac{1}{8}\widetilde{\ell}\gamma^2}.\end{aligned} \tag{140}$$

Similarly, to analyze the second term in the right hand side of (117), we have

$$\mathbb{P}\Big(|\widehat{f}(\rho,|n'\rangle\langle n'|) - \mathrm{tr}\left(|n'\rangle\langle n'|\mathcal{E}(\rho)\right)| \geqslant \gamma\Big) \leqslant e^{-2\widetilde{\ell}\gamma^2}. \tag{141}$$

Thus, the union bound implies that

$$\mathbb{P}\Big(\exists j : |\widehat{\lambda}_{\rho,j} - \mathrm{tr}\left(\widetilde{E}_j\mathcal{E}(\rho)\right)| \geqslant \gamma\Big) \leqslant d(d-1)4e^{-\frac{1}{8}\widetilde{\ell}\gamma^2} + de^{-2\widetilde{\ell}\gamma^2} \leqslant 4d^2e^{-\frac{1}{8}\widetilde{\ell}\gamma^2}. \tag{142}$$

Moreover, because we have

$$\widehat{\lambda}_{j,k} = \widehat{\lambda}_{|a\rangle\langle a|,k} + i\widehat{\lambda}_{|b\rangle\langle b|,k} - \frac{1+i}{2}\widehat{\lambda}_{|n'\rangle\langle n'|,k} - \frac{1+i}{2}\widehat{\lambda}_{|m'\rangle\langle m'|,k}, \tag{143}$$

we obtain

$$\mathbb{P}\Big(\exists j,k : |\lambda_{j,k} - \widehat{\lambda}_{j,k}| \geqslant \gamma\Big) \leqslant 16d^4e^{-\frac{1}{256}\widetilde{\ell}\gamma^2}. \tag{144}$$

□

Lemma 11. *For any $\rho \in \mathcal{D}(\mathcal{H})$ and $0 < \gamma < \frac{\lambda_{\min}(\chi)}{d^5 \kappa_{\max}}$, we have*

$$\mathbb{P}\Big(\|\mathcal{E}(\rho) - \widehat{\mathcal{E}}(\rho)\|_1 \geqslant d^6 \kappa_{\max} \gamma \Big) \leqslant 16 d^4 e^{-\frac{1}{256}\widetilde{\ell}\gamma^2}, \tag{145}$$

$$\mathbb{P}\left(\|\mathcal{E}^c(\rho) - \widehat{\mathcal{E}}^c(\rho)\|_1 \geqslant \frac{d^{18}\kappa_{\max}^2 \lambda_{\max}(\widetilde{\rho}) \sqrt{\lambda_{\max}(\chi)} \gamma^2}{2\left(\lambda_{\min}(\chi) - d^5 \kappa_{\max}\gamma\right)} \right) \leqslant 16 d^4 e^{-\frac{1}{256}\widetilde{\ell}\gamma^2}, \tag{146}$$

where $\widetilde{\rho} \triangleq \sum_{j,k} \operatorname{tr}\left(\widetilde{E}_j \rho \widetilde{E}_k^\dagger\right) |j\rangle\langle k|$.

Proof. Using the triangle inequality, we obtain

$$\|\mathcal{E}(\rho) - \widehat{\mathcal{E}}(\rho)\|_1 = \left\| \sum_{j,k} \widetilde{E}_j \rho \widetilde{E}_k^\dagger \chi_{j,k} - \sum_{j,k} \widetilde{E}_j \rho \widetilde{E}_k^\dagger \widehat{\chi}_{j,k} \right\|_1 \tag{147}$$

$$\leqslant \sum_{j,k} \left\| \widetilde{E}_j \rho \widetilde{E}_k^\dagger \right\|_1 |\chi_{j,k} - \widehat{\chi}_{j,k}| \tag{148}$$

$$\leqslant \sum_{j,k} |\chi_{j,k} - \widehat{\chi}_{j,k}|. \tag{149}$$

Furthermore,

$$\|\mathcal{E}^c(\rho) - \widehat{\mathcal{E}}^c(\rho)\|_1 = \|\sqrt{\chi}^* \widetilde{\rho} \sqrt{\chi}^* - \sqrt{\widehat{\chi}}^* \widetilde{\rho} \sqrt{\widehat{\chi}}^* \|_1 \tag{150}$$

$$= \|\sqrt{\chi}^* \widetilde{\rho} (\sqrt{\chi}^* - \sqrt{\widehat{\chi}}) - (\sqrt{\widehat{\chi}}^* - \sqrt{\chi}^*) \widetilde{\rho} \sqrt{\widehat{\chi}}^* \|_1 \tag{151}$$

$$\leqslant \|\sqrt{\chi}^* \widetilde{\rho} (\sqrt{\chi}^* - \sqrt{\widehat{\chi}})\|_1 + \|(\sqrt{\widehat{\chi}}^* - \sqrt{\chi}^*) \widetilde{\rho} \sqrt{\widehat{\chi}}^* \|_1 \tag{152}$$

$$\overset{(a)}{\leqslant} \sigma_{\max}(\sqrt{\chi}^* \widetilde{\rho}) \|\sqrt{\chi}^* - \sqrt{\widehat{\chi}}\|_1 + \sigma_{\max}(\widetilde{\rho} \sqrt{\widehat{\chi}}^*) \|\sqrt{\widehat{\chi}}^* - \sqrt{\chi}^*\|_1 \tag{153}$$

$$\leqslant \sigma_{\max}(\widetilde{\rho}) \left(\lambda_{\max}(\sqrt{\chi}) + \lambda_{\max}(\sqrt{\widehat{\chi}}) \right) \|\sqrt{\chi} - \sqrt{\widehat{\chi}}\|_1 \tag{154}$$

$$\overset{(b)}{\leqslant} \sigma_{\max}(\widetilde{\rho}) \left(2\sigma_{\max}(\sqrt{\chi})) + \|\sqrt{\chi} - \sqrt{\widehat{\chi}}\|_1 \right) \|\sqrt{\chi} - \sqrt{\widehat{\chi}}\|_1, \tag{155}$$

where (a) follows from Lemma 16 in Appendix B, and (b) follows from Lemma 13 in Appendix B. To upper-bound $\|\sqrt{\chi} - \sqrt{\widehat{\chi}}\|_1$, let us define $F(x) \triangleq \sqrt{\chi + x(\widehat{\chi} - \chi)}$; then, we have

$$\|\sqrt{\chi} - \sqrt{\widehat{\chi}}\|_1 = \|F(0) - F(1)\|_1 \tag{156}$$

$$\leqslant \sup_{x \in [0,1]} \|F'(x)\|_1. \tag{157}$$

Applying Lemma 17 in Appendix B for $f(\mu) = \sqrt{\mu}$ and $A(x) = \chi + x(\widehat{\chi} - \chi)$, we obtain

$$\|F'(x)\|_1 = \left\| \frac{d}{dx} f(A(x)) \right\|_1 \tag{158}$$

$$\leqslant \sup_{\mu \in [\lambda_{\min}(A(x)), \lambda_{\max}(A(x))]} d^4 |f'(\mu)| \|\chi - \widehat{\chi}\|_1 \tag{159}$$

$$= \sup_{\mu \in [\lambda_{\min}(A(x)), \lambda_{\max}(A(x))]} d^4 \left| \frac{1}{2\sqrt{\mu}} \right| \|\chi - \widehat{\chi}\|_1 \tag{160}$$

$$= \frac{d^4 \|\chi - \widehat{\chi}\|_1}{2\sqrt{\lambda_{\min}(A(x))}}. \tag{161}$$

Moreover, by Lemma 13, we know that

$$\lambda_{\min}(A(x)) = \lambda_{\min}(\chi + x(\widehat{\chi} - \chi)) \tag{162}$$

$$\geqslant \lambda_{\min}(\chi) - \|\widehat{\chi} - \chi\|_1. \tag{163}$$

Hence, we have

$$\|\sqrt{\chi} - \sqrt{\widehat{\chi}}\|_1 \leqslant \frac{d^4\|\chi - \widehat{\chi}\|_1}{2\sqrt{\lambda_{\min}(\chi) - \|\widehat{\chi} - \chi\|_1}} \tag{164}$$

If for all j, k, we have $|\chi_{jk} - \widehat{\chi}_{jk}| \leqslant d^2\kappa_{\max}\gamma$, then $\|\chi - \widehat{\chi}\|_1 \leqslant d^5\kappa_{\max}\gamma$. Thus, (134) yields the upper-bound

$$\mathbb{P}\left(\|\mathcal{E}^c(\rho) - \widehat{\mathcal{E}}^c(\rho)\|_1 \geqslant \frac{d^9\kappa_{\max}\gamma\lambda_{\max}(\widetilde{\rho})}{2\sqrt{\lambda_{\min}(\chi) - d^5\kappa_{\max}\gamma}}\left(2\sqrt{\lambda_{\max}(\chi)}\frac{d^9\kappa_{\max}\gamma}{2\sqrt{\lambda_{\min}(\chi) - d^5\kappa_{\max}\gamma}}\right)\right) \leqslant 16d^4e^{-\frac{1}{256}\widetilde{\ell}\gamma^2}. \tag{165}$$

□

Proof of Theorem 4. **Covertness analysis**: Let $\rho_i^{\mathbf{E}}$ denote Eve's state during the channel uses from $q(i-1)+1$ to qi. Since, $U_1, \cdots, U_\ell$ are independent, then

$$\mathbb{D}\Big(\rho^{\mathbf{E}} \| \left(\rho_0^E\right)^{\otimes T'}\Big) = \sum_{i=1}^{\ell} \mathbb{D}\Big(\rho_i^{\mathbf{E}} \| \left(\rho_0^E\right)^{\otimes q}\Big). \tag{166}$$

We now focus on the block from channel use $q(i-1)+1$ to qi. Define $\overline{\rho}$ as the state sent by Alice on the position $q(i-1)+U_i$ and $\rho(j) \triangleq \left(\rho_0^E\right)^{\otimes(j-1)} \otimes \overline{\rho} \otimes \left(\rho_0^E\right)^{\otimes(q-j)}$. One can check that $\rho_i^{\mathbf{E}} = \frac{1}{q}\sum_{j=1}^{q}\rho(j)$. Thus, we have

$$\mathbb{D}\Big(\rho_i^{\mathbf{E}} \| \left(\rho_0^E\right)^{\otimes q}\Big) \overset{(a)}{\leqslant} \operatorname{tr}\left(\left(\rho_i^{\mathbf{E}}\right)^2\left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right) - 1 \tag{167}$$

$$= \operatorname{tr}\left(\left(\frac{1}{q}\sum_{j=1}^{q}\rho(j)\right)^2\left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right) - 1 \tag{168}$$

$$= \frac{1}{q^2}\sum_{j=1}^{q}\sum_{\widetilde{j}=1}^{q}\operatorname{tr}\left(\rho(j)\rho(\widetilde{j})\left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right), \tag{169}$$

where (a) follows from [24]. Note that for $j < \widetilde{j}$, we have (173).

$$\operatorname{tr}\left(\rho(j)\rho(\widetilde{j})\left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right) = \operatorname{tr}\left(\left(\left(\rho_0^E\right)^{\otimes(j-1)} \otimes \overline{\rho} \otimes \left(\rho_0^E\right)^{\otimes(q-j)}\right)\left(\left(\rho_0^E\right)^{\otimes(\widetilde{j}-1)} \otimes \overline{\rho} \otimes \left(\rho_0^E\right)^{\otimes(q-\widetilde{j})}\right)\left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right) \tag{170}$$

$$= \operatorname{tr}\left(\left(\rho_0^E\right)^{\otimes(j-1)} \otimes \left(\overline{\rho}\rho_0^E\left(\rho_0^E\right)^{-1}\right) \otimes \left(\rho_0^E\right)^{\otimes(\widetilde{j}-j-1)} \otimes \left(\rho_0^E\overline{\rho}\left(\rho_0^E\right)^{-1}\right) \otimes \left(\rho_0^E\right)^{\otimes(q-\widetilde{j})}\right) \tag{171}$$

$$= \operatorname{tr}(\overline{\rho})\operatorname{tr}\left(\rho_0^E\overline{\rho}\left(\rho_0^E\right)^{-1}\right)\left(\operatorname{tr}\left(\rho_0^E\right)\right)^{q-2} \tag{172}$$

$$= 1. \tag{173}$$

Similarly, one can show that for $j > \widetilde{j}$, we have $\mathrm{tr}\left(\rho(j)\rho(\widetilde{j})\left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right) = 1$. Furthermore, when $j = \widetilde{j}$, we have

$$\mathrm{tr}\left(\rho(j)\rho(\widetilde{j})\left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right) = \mathrm{tr}\left(\left(\left(\rho_0^E\right)^{\otimes(j-1)} \otimes \overline{\rho} \otimes \left(\rho_0^E\right)^{\otimes(q-j)}\right)^2 \left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right) \tag{174}$$

$$= \mathrm{tr}\left(\left(\rho_0^E\right)^{\otimes(j-1)} \otimes \left(\overline{\rho}^2\left(\rho_0^E\right)^{-1}\right) \otimes \left(\rho_0^E\right)^{\otimes(q-j)}\right) \tag{175}$$

$$= \mathrm{tr}\left(\left(\overline{\rho}^2\left(\rho_0^E\right)^{-1}\right)\right)\mathrm{tr}\left(\rho_0^E\right)^{q-1} \tag{176}$$

$$= \mathrm{tr}\left(\left(\overline{\rho}^2\left(\rho_0^E\right)^{-1}\right)\right). \tag{177}$$

Therefore, we obtain

$$\frac{1}{q^2}\sum_{j=1}^{q}\sum_{\widetilde{j}=1}^{q}\mathrm{tr}\left(\rho(j)\rho(\widetilde{j})\left(\left(\rho_0^E\right)^{\otimes q}\right)^{-1}\right) - 1 = \frac{1}{q^2}\left(q(q-1) + q\mathrm{tr}\left(\left(\overline{\rho}^2\left(\rho_0^E\right)^{-1}\right)\right)\right) - 1 \tag{178}$$

$$= \frac{1}{q}\left(\mathrm{tr}\left(\left(\overline{\rho}^2\left(\rho_0^E\right)^{-1}\right)\right) - 1\right) \tag{179}$$

$$\leqslant \frac{1}{q}\left(\frac{\dim \mathcal{H}^E}{\widetilde{\lambda}^E} - 1\right) \tag{180}$$

Error analysis: To prove (124) and (125), it is enough to show that

$$\mathbb{P}\left(|\lambda_{\min}(\chi) - \lambda_{\min}(\widehat{\chi})| \leqslant \tau \text{ and } |\lambda_{\min}(\widehat{\mathcal{E}}(\rho_0^A)) - \lambda_{\min}(\mathcal{E}(\rho_0^A))| \leqslant \tau\right) \geqslant 1 - 2^{-\xi\ell}. \tag{181}$$

To this end, note that

$$\mathbb{P}(|\lambda_{\min}(\chi) - \lambda_{\min}(\widehat{\chi})| \leqslant \tau) \overset{(a)}{\geqslant} \mathbb{P}\left(\|\chi - \widehat{\chi}\|_1 \leqslant \tau\right) \tag{182}$$

$$\overset{(b)}{\geqslant} \mathbb{P}\left(\sum_{j,k}|\chi_{j,k} - \widehat{\chi}_{j,k}| \leqslant \tau\right), \tag{183}$$

where (a) follows from [25, Lemma 11.1], and (b) follows from the triangle inequality. By (149), we also have

$$\mathbb{P}\Big(|\lambda_{\min}(\widehat{\mathcal{E}}(\rho_0^A)) - \lambda_{\min}(\mathcal{E}(\rho_0^A))| \leqslant \tau\Big) \leqslant \mathbb{P}\left(\sum_{j,k}|\chi_{j,k} - \widehat{\chi}_{j,k}| \leqslant \tau\right). \tag{184}$$

Using (134), we thus obtain

$$\mathbb{P}\left(|\lambda_{\min}(\chi) - \lambda_{\min}(\widehat{\chi})| \leqslant \tau \text{ and } |\lambda_{\min}(\widehat{\mathcal{E}}(\rho_0^A)) - \lambda_{\min}(\mathcal{E}(\rho_0^A))| \leqslant \tau\right) \geqslant 1 - 16d^4 e^{-\frac{1}{256d^1\kappa_{\max}^2}\widetilde{\ell}\tau^2}. \tag{185}$$

We now establish bounds on the accuracy of the estimates $\widehat{D}^B$ and $\widehat{D}^E$ when $\lambda_{\min}(\chi) \geqslant \widetilde{\lambda}^\chi - 2\tau, \lambda_{\min}(\mathcal{E}(\rho_0^A)) \geqslant \widetilde{\lambda}^B - 2\tau$. We choose $\epsilon > 0$ small enough such that

$$\epsilon\left(\frac{\log(d-1)}{2} + d\log\frac{1}{\min(\widetilde{\lambda}^B - 2\tau, \widetilde{\lambda}^E)} + \frac{d^2}{\min(\widetilde{\lambda}^B - 2\tau, \widetilde{\lambda}^E) - \epsilon}\right) + \mathbb{H}_b\left(\frac{\epsilon}{2}\right) \leqslant \tau. \tag{186}$$

By Lemma 14, we can choose $\gamma > 0$ independent of $\lambda_{\max}(\chi)$ such that

$$d^6\kappa_{\max}\gamma \leqslant \epsilon, \tag{187}$$

$$\frac{d^{18}\kappa_{\max}^2\lambda_{\max}(\widetilde{\rho})\sqrt{\lambda_{\max}(\chi)}\gamma^2}{2\left(\lambda_{\min}(\chi) - 2\tau - d^5\kappa_{\max}\gamma\right)} \leqslant \epsilon. \tag{188}$$

By Lemma 11 and Lemma 19, we have

$$\mathbb{P}\left(D^B(\mathcal{E})-2\tau\leqslant\widehat{D}^B\leqslant D^B(\mathcal{E}),D^E(\mathcal{E})\leqslant\widehat{D}^E\leqslant D^E(\mathcal{E})+2\tau\right)\leqslant 32d^4e^{-\frac{1}{256}\widetilde{\ell}\gamma^2}. \tag{189}$$

Since $\widetilde{\ell}\geqslant\frac{\frac{\ell}{2d^2}-1}{2d^2}-1$, we can choose $\xi>0$ small enough such that the above upper-bound is less than $2^{-\xi\ell}$. □

Proof of Lemma 9. We only prove the second part of the lemma and the proof of the second part can be obtained by the exact same approach. Let $P_e(D^B,D^E)$, $S(D^B,D^E)$, $C(D^B,D^E)$ indicate the probability of error, secrecy, and covertness of the protocol discussed in the proof of Theorem 2, respectively, when we use the parameters D^B and D^E. By the law of total probability, the probability of error of the overall protocol is

$$\mathbb{E}_{\widehat{D}^B\widehat{D}^E}\Big(P_e(\widehat{D}^B,\widehat{D}^E)\Big)=\mathbb{E}\Big(P_e(\widehat{D}^B,\widehat{D}^E)|\mathcal{A}\Big)\mathbb{P}(\mathcal{A})+\mathbb{E}\Big(P_e(\widehat{D}^B,\widehat{D}^E)|\mathcal{A}^c\Big)\mathbb{P}(\mathcal{A}^c) \tag{190}$$

$$\overset{(a)}{\leqslant} 2T^{-5}+\mathbb{E}\Big(P_e(\widehat{D}^B,\widehat{D}^E)|\mathcal{A}^c\Big)\mathbb{P}(\mathcal{A}^c) \tag{191}$$

$$\leqslant 2T^{-5}+\epsilon, \tag{192}$$

where $\mathcal{A}\triangleq\{\widehat{D}^B\leqslant D^B(\mathcal{E}),\widehat{D}^E\geqslant D^E(\mathcal{E})\}\cup\{H=0\}$, (a) follows from Theorem 2. For the secrecy, first note that the estimation phase does not leak any information about the key. Furthermore, by convexity of the quantum relative entropy, we have

$$S\leqslant\mathbb{E}_{\widehat{D}^B\widehat{D}^E}\Big(S(\widehat{D}^B,\widehat{D}^E)\Big) \tag{193}$$

$$=\mathbb{E}\Big(S(\widehat{D}^B,\widehat{D}^E)|\mathcal{A}\Big)\mathbb{P}(\mathcal{A})+\mathbb{E}\Big(S(\widehat{D}^B,\widehat{D}^E)|\mathcal{A}^c\Big)\mathbb{P}(\mathcal{A}^c) \tag{194}$$

$$\overset{(a)}{\leqslant} L_1T^{-4}+\mathbb{E}\Big(S(\widehat{D}^B,\widehat{D}^E)|\mathcal{A}^c\Big)\mathbb{P}(\mathcal{A}^c) \tag{195}$$

$$\overset{(b)}{\leqslant} L_1T^{-4}+\left(T\log\frac{1}{\widetilde{\lambda}^E}+\ell^{\max}\right)\epsilon, \tag{196}$$

where (a) follows from Theorem 2, and (b) follows from the upper-bound $S\leqslant T\log\frac{1}{\widetilde{\lambda}^E}+\ell^{\max}$. Finally, for covertness, since the estimation and transmission phases are independent, we have

$$C\leqslant\delta+\mathbb{E}_{\widehat{D}^B\widehat{D}^E}\Big(C(\widehat{D}^B,\widehat{D}^E)\Big). \tag{197}$$

Similar to secrecy, we also have

$$\mathbb{E}_{\widehat{D}^B\widehat{D}^E}\Big(C(\widehat{D}^B,\widehat{D}^E)\Big)\leqslant\frac{\alpha_T^2\eta(\rho_1^E(\theta)\|\rho_0^E(\theta))}{2}T+L_2\alpha_T^3T+L_1T^{-4}+2\sqrt{L_1}\log\frac{2}{\widetilde{\lambda}^E}T^{-1}+\epsilon T\log\frac{1}{\widetilde{\lambda}^E}. \tag{198}$$

□

C. Proof of Theorem 1

We describe a protocol running over $\widetilde{T}>0$ channel uses. Let $T'=\lfloor\sqrt{\widetilde{T}}\rfloor$ and $T=\widetilde{T}-T'-O(\log T')$. Alice and Bob use the first $T'+O(\log T')$ channel uses for the estimation protocol described in Section IV-B1 for parameters q and ℓ to obtain H as well as estimates $D^B(\mathcal{E})$ and $D^E(\mathcal{E})$. If $H=0$ the protocol is aborted and if $H=1$, the rest of T channel uses will be used for transmission using the universal protocol as described before for $\widehat{D}^B$, $\widehat{D}^E$,

$\widetilde{\lambda}^B - 2\tau$ and $\widetilde{\lambda}^W$. For a channel satisfying $\lambda_{\min}(\chi) \geqslant \widetilde{\lambda}^\chi - 2\tau, \lambda_{\min}(\mathcal{E}(\rho_0^A)) \geqslant \widetilde{\lambda}^B - 2\tau$, by applying the second part of Lemma 9 and Theorem 4, for some $\xi > 0$, we have

$$P_e \leqslant 2T^{-5} + 2^{-\xi\ell}, \tag{199}$$

$$S \leqslant L_1 T^{-4} + 2^{-\xi\ell}\left(T \log \frac{1}{\widetilde{\lambda}^E} + \ell^{\max}\right), \tag{200}$$

$$C \leqslant \frac{\alpha_T^2 \eta(\rho_1^E(\theta)\|\rho_0^E(\theta))}{2} T + L_2 \alpha_T^3 T + L_1 T^{-4} + 2\sqrt{L_1} \log \frac{2}{\widetilde{\lambda}^E} T^{-1} + 2^{-\xi\ell} T \log \frac{1}{\widetilde{\lambda}^E} + \frac{\ell}{q}\left(\frac{\dim \mathcal{H}^E}{\widetilde{\lambda}^E} - 1\right). \tag{201}$$

One can check that if $\ell \in \omega(\log T) \cap o\left(\alpha_T T^{-\frac{3}{4}}\right)$, which is non-empty by definition of α_T, we can always find the sequence $\epsilon_{\widetilde{T}}$ satisfying the conditions in Theorem 1. If the channel satisfies $\lambda_{\min}(\chi) \geqslant \widetilde{\lambda}^\chi, \lambda_{\min}(\mathcal{E}(\rho_0^A)) \geqslant \widetilde{\lambda}^B$, by (124), with probability $2^{-\xi\ell}$, the number of transmitted bits is lower-bounded by

$$(1 - 2\zeta)(D^B(\mathcal{E}) - D^E(\mathcal{E}) - 2\tau)\alpha_T T. \tag{202}$$

If the channel does not satisfy $\lambda_{\min}(\chi) \geqslant \widetilde{\lambda}^\chi, \lambda_{\min}(\mathcal{E}(\rho_0^A)) \geqslant \widetilde{\lambda}^B$, by (125) and the first part of Lemma 9, we have

$$P_e \leqslant 2^{-\xi\ell}, \tag{203}$$

$$S \leqslant 2^{-\xi\ell}\left(T \log \frac{1}{\widetilde{\lambda}^E} + \ell^{\max}\right), \tag{204}$$

$$C \leqslant 2^{-\xi\ell} T \log \frac{1}{\widetilde{\lambda}^E} + \frac{\ell}{q}\left(\frac{\dim \mathcal{H}^E}{\widetilde{\lambda}^E} - 1\right), \tag{205}$$

but no key is generated.

CONCLUSION

We prove the existence of covert secret key expansion protocols that achieve the square root law for a wide range of quantum channels. Our security measure is more stringent than that of traditional QKD as we require that the adversary be unable to detect the execution of the protocol in addition to obtaining negligible information about the key. Our result is obtained by combining an undetectable covert tomography protocol and universal covert communication codes over cq-channels.

APPENDIX A
ERROR EXPONENT CALCULATIONS

Proof of Lemma 5. Consider any cq-channel $x \mapsto \rho_x^B$ with $\lambda_{\min}(\rho_0^B) = \lambda_{\min} > 0$. We first show that the corresponding function ϕ is smooth enough to use Taylor theorem. Let us define

$$A(s,p) \triangleq \left((1-p)\left(\rho_0^B\right)^{1-s} + p\left(\rho_0^B\right)^{1-s}, s\right) \tag{206}$$

$$g(M,s) \triangleq \left(\text{tr}\left(M^{\frac{1}{1-s}}\right), s\right) \tag{207}$$

$$\psi(x,s) \triangleq -(1-s)\log(x). \tag{208}$$

By definition, we have $\phi(s,p) = (\psi \circ g \circ A)(s,p)$. Additionally, all these three functions are from a subset of a Banach space to a Banach space, which means that we can consider their Fréchet derivative. In the following lemma, we show that they are infinitely many times differentiable.

Lemma 12. *The functions A, g, and ψ are infinitely many times differentiable on*

$$[0,1[\times[0,1[, \tag{209}$$

$$\{M \in \mathcal{L}(\mathcal{H}) : M \textit{ is Hermitian}, M \succ 0\} \times [0,1[, \tag{210}$$

$$[0,1[\times[0,\infty[, \tag{211}$$

respectively.[3]

Proof. We investigate each function separately.

- **Differentiability of** A: It is enough to check the differentiability of $A_1(s,p) \triangleq (1-p)\left(\rho_0^B\right)^{1-s} + p\left(\rho_1^B\right)^{1-s}$. We shall provide explicit expressions for all partial derivatives of A_1 to any order. For any Hermitian operator $\rho \in \mathcal{L}(\mathcal{H})$ with $\rho \succeq 0$ and $\rho \neq 0$, let $\rho = \sum_e \lambda_e |e\rangle\langle e|$ be an eigen-decomposition for ρ. We define $\log \rho \triangleq \sum_{e:\lambda_e \neq 0} \log(\lambda_e)|e\rangle\langle e|$, which is different from the usual definition since we disregard the zero eigenvalues. With this definition, one can check that for any $i \geqslant 1$, we have

$$\frac{d^i}{ds^i}(\rho^{1-s}) = \rho^{1-s}\left(-\log\rho\right)^i. \tag{212}$$

  Hence, using the linearity of the Fréchet derivative, if we take i partial derivatives with respect to s and j partial derivatives with respect to p at any order, the result is

$$\begin{cases} (1-p)\left(\rho_0^B\right)^{1-s}\left(-\log\rho_0^B\right)^i + p\left(\rho_1^B\right)^{1-s}\left(-\log\rho_1^B\right)^i & j=0, \\ -\left(\rho_0^B\right)^{1-s}\left(-\log\rho_0^B\right)^i + \left(\rho_1^B\right)^{1-s}\left(-\log\rho_1^B\right)^i & j=1, \\ 0 & j \geqslant 2. \end{cases} \tag{213}$$

  This also means that all partial derivative are differentiable and therefore continuous. Accordingly, A_1 is infinitely many times Fréchet differentiable.

- **Differentiability of** g: Again we only check the differentiability of $g_1(M,s) \triangleq \mathrm{tr}\left(M^{\frac{1}{1-s}}\right)$. In this case, it is more challenging to obtain a closed-form expression for partial derivatives. However, we will prove that any partial derivative is a multilinear form mapping $(K_1, \cdots, K_m) \in \mathcal{L}(\mathcal{H})^m$ to $\mathbb{R}$ and is a summation of terms of the form

$$\frac{p(s)}{(1-s)^i}\mathrm{tr}\left(K_1 \cdots K_m M^{\frac{q(s)}{(1-s)^j}}(\log M)^k\right), \tag{214}$$

  where q and p are polynomial in s, and i, j, and k are non-negative integers. Using induction on the total number of partial derivative taken and linearity of the derivative, it is enough to show that if we take the

[3]For the boundary points we consider the one-sided derivative.

derivative of (214) with respect to s or M, we would have an expression that is a summation of term of the same form. Applying the rules of differentiation, one can check that

$$\frac{\partial}{\partial s}\left(\frac{p(s)}{(1-s)^i}\mathrm{tr}\left(K_1\cdots K_m M^{\frac{q(s)}{(1-s)^j}}(\log M)^k\right)\right)=\frac{p(s)\left(jq(s)+(1-s)q'(s)\right)}{(1-s)^{i+j+1}}$$
$$\times\,\mathrm{tr}\left(K_1\cdots K_m M^{\frac{q(s)}{(1-s)^j}}(\log M)^{k+1}\right)+\frac{ip(s)+(1-s)p'(s)}{(1-s)^{i+1}}\mathrm{tr}\left(K_1\cdots K_m M^{\frac{q(s)}{(1-s)^j}}(\log M)^k\right), \quad (215)$$

and

$$\frac{\partial}{\partial M}\left(\frac{p(s)}{(1-s)^i}\mathrm{tr}\left(K_1\cdots K_m M^{\frac{q(s)}{(1-s)^j}}(\log M)^k\right)\right)=K\mapsto\frac{p(s)}{(1-s)^i}\frac{q(s)}{(1-s)^j}$$
$$\times\,\mathrm{tr}\left(KK_1\cdots K_m\left(\frac{q(s)}{(1-s)^j}M^{\frac{q(s)}{(1-s)^j}-1}(\log M)^k+kM^{\frac{q(s)}{(1-s)^j}-1}(\log M)^{k-1}\right)\right). \quad (216)$$

Therefore, g_1 has partial derivatives of any order. Using the same argument that we used for A_1, we conclude that g_1 is infinitely many times Fréchet differentiable.

- **Differentiability of** ψ: ψ is product of two smooth functions $(x,s)\mapsto -(1-s)$ and $(x,s)\mapsto \log x$, and therefore, it is smooth on its domain.

□

We next check that $A(s,p)$ lies in the set $\{M\in\mathcal{L}(\mathcal{H}): M \text{ is Hermitian}, M\succ 0\}$ where g is differentiable. By our assumption that $\lambda_{\min}>0$, ρ_0^B is positive semi-definite, and so is $\left(\rho_0^B\right)^{1-s}$ for $s\in[0,1[$. Furthermore, since $\rho_1^B\succeq 0$, we have $A(s,p)\succ 0$ for all $(s,p)\in[0,1[\times[0,1[$. Thus, by the chain rule, ϕ is a smooth function on $[0,1[\times[0,1[$. Apply Taylor's theorem, we have

$$\phi(s,p)=\phi(0,p)+\frac{\partial\phi(0,p)}{\partial s}s+\frac{1}{2}\frac{\partial^2\phi(0,p)}{\partial^2 s}s^2+\frac{1}{6}\frac{\partial^3\phi(\eta,p)}{\partial^3 s}s^3, \quad (217)$$

for some $\eta\in[0,s]$ that can depend on s. Similarly, we have

$$\frac{\partial^2\phi(0,p)}{\partial^2 s}=\frac{\partial^2\phi(0,0)}{\partial^2 s}+\frac{\partial^3\phi(0,\tau)}{\partial^2 s\partial p}p, \quad (218)$$

for some $\tau\in[0,p]$. Additionally, one can check that $A(s,p)$ and all its derivatives depend continuously on ρ_0^B and ρ_1^B. Since any continuous function achieves its maximum on a compact domain, we have

$$\sup_{\tau\in[0,\widetilde{p}],\rho_0^B\in\mathcal{D}(\mathcal{H}),\rho_1^B\in\mathcal{D}(\mathcal{H}):\lambda_{\min}(\rho_0^B)\geqslant\widetilde{\lambda}}\left|\frac{\partial^3\phi(0,\tau)}{\partial^2 s\partial p}\right|<\infty, \quad (219)$$

$$\sup_{\eta\in[0,\widetilde{s}],p\in[0,\widetilde{p}],\rho_0^B\in\mathcal{D}(\mathcal{H}),\rho_1^B\in\mathcal{D}(\mathcal{H}):\lambda_{\min}(\rho_0^B)\geqslant\widetilde{\lambda}}\left|\frac{\partial^3\phi(\eta,p)}{\partial^3 s}\right|<\infty. \quad (220)$$

Moreover, from the definition, one can check that $\phi(0,p)=0$, $\frac{\partial^2\phi(0,0)}{\partial^2 s}=0$, and by [21], $\frac{\partial\phi(0,p)}{\partial s}=I(p)$. This implies that there exists $B>0$, such that for all cq-channels $x\mapsto\rho_x^B$ with $\lambda_{\min}(\rho_0^B)\geqslant\widetilde{\lambda}$, we have

$$\phi(s,p)\geqslant I(p)s-B\left(ps^2+s^3\right). \quad (221)$$

Furthermore, using the same approach, we can prove $I(p)\geqslant p\mathbb{D}\big(\rho_1^B\|\rho_0^B\big)-Bp^2$. □

Proof of Lemma 7. If we define

$$A(s,p) \triangleq \left((1-p)\left(\rho_0^E\right)^{1-s} p \left(\rho_1^E\right)^{1-s}\right)\left((1-p)\rho_0^E + p\rho_1^E\right)^s \tag{222}$$

$$g(M) \triangleq \mathrm{tr}\,(M) \tag{223}$$

$$\psi(x) \triangleq \log(x), \tag{224}$$

similar to the proof of Lemma 5, one can check that all these functions are infinitely many times Fréchet differentiable. Since, $\phi = \psi \circ g \circ A$, the rest of proof is exactly similar to that of Lemma 5. □

APPENDIX B
TECHNICAL LEMMAS

Lemma 13. *Suppose A and B are Hermitian in $\mathcal{L}(\mathcal{H})$. Then, we have*

$$\lambda_{\min}(A) \geqslant \lambda_{\min}(B) - \|A-B\|_2 \geqslant \lambda_{\min}(B) - \|A-B\|_1 \tag{225}$$

$$\lambda_{\max}(A) \leqslant \lambda_{\max}(B) + \|A-B\|_2 \leqslant \lambda_{\max}(B) + \|A-B\|_1 \tag{226}$$

Proof. If $\lambda_{\min}(A) \triangleq \lambda_1 \leqslant \cdots \leqslant \lambda_d \triangleq \lambda_{\max}(A)$ and $\lambda_{\min}(B) \triangleq \gamma_1 \leqslant \cdots \leqslant \gamma_d \triangleq \lambda_{\max}(B)$ are the eigenvalues of A and B, respectively, then by [26, Corollary 6.3.8], we have $\|A-B\|_1^2 \geqslant \|A-B\|_2^2 \geqslant \sum_{i=1}^d (\lambda_i - \gamma_i)^2$ which results in the desired bounds. □

Lemma 14. *For any quantum channel $\mathcal{E} : \mathcal{L}(\mathcal{H}^A) \to \mathcal{L}(\mathcal{H}^A)$ with chi-representation matrix χ, we have $\lambda_{\max}(\chi) \leqslant \sqrt{d}$, where $d \triangleq \dim(\mathcal{H}^A)$.*

Proof. Since χ is Hermitian, it admits an eigen-decomposition representation, i.e., for some unitary matrix U and real values $\sigma_1, \cdots, \sigma_{d^2}$, we have $\chi_{i,j} = \sum_{k=1}^{d^2} d_i U_{i,k} U_{j,k}^*$. By [20, Eq. (8.168)], $\mathcal{E}$ has a Kraus representation $\mathcal{E}(\rho) = \sum_{i=1}^{d^2} E_i \rho E_i^\dagger$ for $E_i = \sqrt{\sigma_i}\sum_{j=1}^{d^2} U_{j,i}\widetilde{E}_j$. We hence have

$$\|E_i\|_2 = \sqrt{\sigma_i}\left\|\sum_{j=1}^{d^2} U_{j,i}\widetilde{E}_j\right\|_2 \tag{227}$$

$$= \sqrt{\sigma_i}\sqrt{\mathrm{tr}\left(\left(\sum_{j=1}^{d^2} U_{j,i}^*\widetilde{E}_j^\dagger\right)\left(\sum_{j'=1}^{d^2} U_{j',i}\widetilde{E}_{j'}\right)\right)} \tag{228}$$

$$= \sqrt{\sigma_i}\sqrt{\sum_{j=1}^{d^2}\sum_{j'=1}^{d^2} U_{j,i}^* U_{j',i}\mathrm{tr}\left(E_j^\dagger E_{j'}\right)} \tag{229}$$

$$= \sqrt{\sigma_i}\sqrt{\sum_{j=1}^{d^2} U_{j,i}^* U_{j,i}} \tag{230}$$

$$\overset{(a)}{=} \sqrt{\sigma_i}, \tag{231}$$

where (a) follows since U is unitary. Because $\mathcal{E}$ is a quantum channel, we have $\sum_{i=1}^{d^2} E_i^\dagger E_i = I$. Taking the trace from this equality, we obtain that

$$d = \mathrm{tr}\,(I) = \mathrm{tr}\left(\sum_{i=1}^{d^2} E_i^\dagger E_i\right) = \sum_{i=1}^{d^2} \|E_i\|_2^2. \tag{232}$$

Using (231) and (232), we conclude that

$$\lambda_{\max}(\chi) = \max_{i\in[\![1,d^2]\!]} \Lambda_i \leqslant \|E_i\|_2 \leqslant \sqrt{d}. \tag{233}$$

□

Lemma 15. *Consider any quantum channel $\mathcal{E} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ with $\dim \mathcal{H} = d$ and characterized by $\mathcal{E}(\rho) = \sum_{i,j} \widetilde{E}_i \rho \widetilde{E}_j^\dagger \chi_{ij}$. Define another Hilbert space $\mathcal{H}^c$ spanned by an orthonormal basis $\{|j\rangle : j \in [\![1, d^2]\!]\}$. Then, up to a unitary transformation, the complementary channel $\mathcal{E}^\dagger : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}^c)$ would be*

$$\mathcal{E}^c(\rho) = \sqrt{\chi}^* \widetilde{\rho} \sqrt{\chi}^*, \tag{234}$$

where

$$\chi \triangleq \sum_{j,k} |j\rangle\langle k| \chi_{jk} \tag{235}$$

$$\widetilde{\rho} \triangleq \sum_{j,k} |j\rangle\langle k| \mathrm{tr}\left(\widetilde{E}_j \rho \widetilde{E}_k^\dagger\right). \tag{236}$$

Proof. By [20], without loss of generality we can assume that χ is Hermitian. Therefore, let $\chi = \sum_j d_j |u_j\rangle\langle u_j|$ be an eigen-decomposition of χ. For $E_j \triangleq \sum_k \sqrt{d_j}\langle k|u_j\rangle \widetilde{E}_k$, we have

$$\sum_j E_j \rho E_j^\dagger = \sum_j \left(\sum_k \sqrt{d_j}\langle k|u_j\rangle \widetilde{E}_k\right) \rho \left(\sum_{k'} \sqrt{d_j}\langle u_j|k'\rangle \widetilde{E}_{k'}^\dagger\right) \tag{237}$$

$$= \sum_k \sum_{k'} \sum_j \widetilde{E}_k \rho \widetilde{E}_{k'}^\dagger d_j \langle k|u_j\rangle\langle u_j|k'\rangle \tag{238}$$

$$= \sum_k \sum_{k'} \widetilde{E}_k \rho \widetilde{E}_{k'}^\dagger \langle k| \left(\sum_j d_j |u_j\rangle\langle u_j|\right) |k'\rangle \tag{239}$$

$$= \sum_k \sum_{k'} \widetilde{E}_k \rho \widetilde{E}_{k'}^\dagger \langle k|\chi|k'\rangle \tag{240}$$

$$= \sum_k \sum_{k'} \widetilde{E}_k \rho \widetilde{E}_{k'}^\dagger \chi_{kk'} \tag{241}$$

$$= \mathcal{E}(\rho). \tag{242}$$

This implies that $\sum_j E_j \rho E_j^\dagger$ is a Kraus representation for $\mathcal{E}$, and therefore, by [19], a representation for the complementary channel is

$$\widetilde{\mathcal{E}}^c(\rho) = \sum_{j,k} \mathrm{tr}\left(E_j \rho E_k^\dagger\right) |j\rangle\langle k|. \tag{243}$$

Hence, it is enough to show that for some unitary operator U onto $\mathcal{H}^c$, we have

$$\sqrt{\chi}^* \widetilde{\rho} \sqrt{\chi}^* = U \widetilde{\mathcal{E}}^c(\rho) U^\dagger. \tag{244}$$

Let $U \triangleq \sum_j |\widetilde{u}_j\rangle\langle j|$ where $|\widetilde{u}_j\rangle \triangleq \sum_i \langle u_j|i\rangle|i\rangle$. One can check that it is a unitary operator, and we have

$$U\widetilde{\mathcal{E}}^{\mathrm{c}}(\rho)U^\dagger = \left(\sum_j |\widetilde{u}_j\rangle\langle j|\right)\left(\sum_{k,k'} \mathrm{tr}\left(E_k\rho E_{k'}^\dagger\right)|k\rangle\langle k'|\right)\left(\sum_{j'}|j'\rangle\langle\widetilde{u}_{j'}|\right) \tag{245}$$

$$= \sum_{jj'kk'} \mathrm{tr}\left(E_k\rho E_{k'}^\dagger\right)|\widetilde{u}_j\rangle\langle j||k\rangle\langle k'||j'\rangle\langle\widetilde{u}_{j'}| \tag{246}$$

$$= \sum_{kk'} \mathrm{tr}\left(E_k\rho E_{k'}^\dagger\right)|\widetilde{u}_k\rangle\langle\widetilde{u}_{k'}| \tag{247}$$

$$= \sum_{kk'} \mathrm{tr}\left(\left(\sum_j \sqrt{d_k}\langle j|u_k\rangle\widetilde{E}_j\right)\rho\left(\sum_{j'}\sqrt{d_{k'}}\langle u_{k'}|j'\rangle\widetilde{E}_{j'}^\dagger\right)\right)|\widetilde{u}_k\rangle\langle\widetilde{u}_{k'}| \tag{248}$$

$$= \sum_{jj'kk'} \sqrt{d_k}\sqrt{d_{k'}}\mathrm{tr}\left(\langle j|u_k\rangle\langle u_{k'}|j'\rangle\widetilde{E}_j\rho\widetilde{E}_{j'}^\dagger\right)|\widetilde{u}_k\rangle\langle\widetilde{u}_{k'}| \tag{249}$$

$$= \sum_{jj'} \mathrm{tr}\left(\widetilde{E}_j\rho\widetilde{E}_{j'}^\dagger\right)\sum_{kk'}\sqrt{d_k}\sqrt{d_{k'}}\langle j|u_k\rangle\langle u_{k'}|j'\rangle|\widetilde{u}_k\rangle\langle\widetilde{u}_{k'}| \tag{250}$$

$$= \sum_{jj'} \mathrm{tr}\left(\widetilde{E}_j\rho\widetilde{E}_{j'}^\dagger\right)\left(\sum_k\sqrt{d_k}\langle j|u_k\rangle|\widetilde{u}_k\rangle\right)\left(\sum_{k'}\sqrt{d_{k'}}\langle u_{k'}|j'\rangle\langle\widetilde{u}_{k'}|\right) \tag{251}$$

$$= \sum_{jj'} \mathrm{tr}\left(\widetilde{E}_j\rho\widetilde{E}_{j'}^\dagger\right)\left(\sum_k\sqrt{d_k}\langle\widetilde{u}_k|j\rangle|\widetilde{u}_k\rangle\right)\left(\sum_{k'}\sqrt{d_{k'}}\langle j'|\widetilde{u}_{k'}\rangle\langle\widetilde{u}_{k'}|\right) \tag{252}$$

$$= \left(\sum_k\sqrt{d_k}|\widetilde{u}_k\rangle\langle\widetilde{u}_k|\right)\left(\sum_{jj'}\mathrm{tr}\left(\widetilde{E}_j\rho\widetilde{E}_{j'}^\dagger\right)|j\rangle\langle j'|\right)\left(\sum_{k'}\sqrt{d_{k'}}|\widetilde{u}_{k'}\rangle\langle\widetilde{u}_{k'}|\right) \tag{253}$$

$$= \sqrt{\chi}^*\widetilde{\rho}\sqrt{\chi}^*. \tag{254}$$

□

Lemma 16. *Let $A, B \in \mathcal{L}(\mathcal{H})$ and B be Hermitian. Then,*

$$\|AB\|_1 \leqslant \sigma_{\max}(A)\|B\|_1, \tag{255}$$

where $\sigma_{\max}(A)$ is the maximum singular value of the A.

Proof. Consider an eigen-decomposition of B, i.e., $B = \sum_b b|b\rangle\langle b|$. Then,

$$\|AB\|_1 = \left\|A\left(\sum_b b|b\rangle\langle b|\right)\right\|_1 \tag{256}$$

$$\leqslant \sum_b |b|\||A|b\rangle\langle b|\|_1 \tag{257}$$

$$\leqslant \sum_b |b|\mathrm{tr}\left(\sqrt{|b\rangle\langle b|A^\dagger A|b\rangle\langle b|}\right) \tag{258}$$

$$= \sum_b |b|\sqrt{\langle b|A^\dagger A|b\rangle} \tag{259}$$

$$= \sum_b |b| \| A|b\rangle \|_2 \tag{260}$$

$$\leqslant \sigma_{\max}(A) \left(\sum_b |b| \right) \tag{261}$$

$$= \sigma_{\max}(A) \|B\|_1. \tag{262}$$

□

Lemma 17. *Let $\mathcal{I} \subset \mathbb{R}$ be an interval and $f : \mathcal{I} \to \mathbb{R}$ and $A(x) : \mathbb{R} \to \mathcal{L}(\mathcal{H})$ be differentiable functions such $A(x)$ is Hermitian and its spectrum is included in $\mathcal{I}$ for all x. For any operator norm $\| \cdot \|$ satisfying $\max(\|PA\|, \|AP\|) \leqslant \|A\|$ where A is an arbitrary operator and P is a projection, we have*

$$\left\| \frac{d}{dx'} f(A(x')) \Big|_{x'=x} \right\| \leqslant d^2 \sup_{\mu \in [\lambda_{\min}(A'(x)), \lambda_{\max}(A'(x))]} |f'(\mu)| \|A'(x)\|. \tag{263}$$

Proof. We use a formula in [19] for the derivative of an operator-valued function. Let $f : \mathbb{R} \to \mathbb{R}$ and $A(x) : \mathbb{R} \to \mathcal{L}(\mathcal{H})$ be a differentiable functions. Then,

$$\frac{d}{dx'} f(A(x')) \Big|_{x'=x} = \sum_{\nu,\eta} f^{[1]}(\nu, \eta) P_{A(x)}(\nu) A'(x) P_{A(x)}(\eta), \tag{264}$$

where the summation is taken over all eigenvalues of $A(x)$, $P_{A(x)}(\nu)$ is the projector onto the subspace of all eigenvectors corresponding to ν, and

$$f^{[1]}(\nu, \eta) = \begin{cases} \frac{f(\nu) - f(\eta)}{\nu - \eta} & \nu \neq \eta \\ f'(\nu) & \nu = \eta \end{cases}. \tag{265}$$

We can now upper-bound the norm of $\frac{d}{dx} f(A(x))$ by

$$\left\| \frac{d}{dx'} f(A(x')) \Big|_{x'=x} \right\| = \left\| \sum_{\nu,\eta} f^{[1]}(\nu, \eta) P_{A(x)}(\nu) A'(x) P_{A(x)}(\eta) \right\| \tag{266}$$

$$\leqslant \sum_{\nu,\eta} |f^{[1]}(\nu, \eta)| \left\| P_{A(x)}(\nu) A'(x) P_{A(x)}(\eta) \right\| \tag{267}$$

$$\overset{(a)}{\leqslant} \sum_{\nu,\eta} |f^{[1]}(\nu, \eta)| \left\| A'(x) \right\|, \tag{268}$$

where (a) follows from our assumption that $\max(\|PA\|, \|AP\|) \leqslant \|A\|$. By the mean value theorem, we also have that $f^{[1]}(\nu, \eta) = f'(\mu)$ for some μ between ν and η. Thus,

$$\sum_{\nu,\eta} |f^{[1]}(\nu, \eta)| \left\| A'(x) \right\| \leqslant d^2 \sup_{\mu \in [\lambda_{\min}(A'(x)), \lambda_{\max}(A'(x))]} |f'(\mu)| \|A'(x)\|. \tag{269}$$

□

Lemma 18. *Suppose ρ and σ are two density matrices on Hilbert space $\mathcal{H}$ with $\dim \mathcal{H} = d$ such that $\mathrm{supp}\rho \subset \mathrm{supp}\sigma$ and $\|\rho - \sigma\|_1 \leqslant \epsilon \leqslant e^{-1}$. Then,*

$$\mathbb{D}(\rho \| \sigma) \leqslant \epsilon \log \frac{d}{\lambda_{\min}(\sigma) \epsilon}. \tag{270}$$

Proof. Since $\text{supp}(\rho) \subset \text{supp}(\sigma)$, we have

$$\mathbb{D}(\rho\|\sigma) = \text{tr}\left(\rho(\log\rho - \log\sigma)\right) \tag{271}$$

$$= -H(\rho) + H(\sigma) - \text{tr}\left((\rho-\sigma)\log\sigma\right) \tag{272}$$

$$\overset{(a)}{\leqslant} \epsilon \log\frac{d}{\epsilon} - \text{tr}\left((\rho-\sigma)\log\sigma\right) \tag{273}$$

$$\leqslant \epsilon\log\frac{d}{\epsilon} + \epsilon\log\frac{1}{\lambda_{\min}(\sigma)}, \tag{274}$$

where (a) follows from Fannes inequality. □

Lemma 19. *Suppose $\rho, \rho', \sigma, \sigma' \in \mathcal{D}(\mathcal{H})$ with $\dim\mathcal{H} = d$, $\text{supp}(\rho) \subset \text{supp}(\sigma)$, and $\text{supp}(\rho') \subset \text{supp}(\sigma')$. Let $\|\rho-\rho'\|_1 \leqslant \epsilon$, $\|\sigma-\sigma'\|_1 \leqslant \epsilon$, and $\lambda_{\min}(\sigma)$ be the minimum eigenvalue of σ with $\lambda_{\min}(\sigma) \geqslant \epsilon$). Then,*

$$|\mathbb{D}(\rho\|\sigma) - \mathbb{D}(\rho'\|\sigma')| \leqslant \epsilon\left(\frac{\log(d-1)}{2} + d\log\frac{1}{\lambda_{\min}(\sigma))} + \frac{d^2}{\lambda_{\min}(\sigma)-\epsilon}\right) + \mathbb{H}_b\left(\frac{\epsilon}{2}\right). \tag{275}$$

Proof. By definition, we have

$$|\mathbb{D}(\rho\|\sigma) - \mathbb{D}(\rho'\|\sigma')| = |-\mathbb{H}(\rho) + \mathbb{H}(\rho') - \text{tr}\left(\rho\log\sigma\right) + \text{tr}\left(\rho'\log\sigma'\right)| \tag{276}$$

$$\leqslant |-\mathbb{H}(\rho) + \mathbb{H}(\rho')| + |\text{tr}\left((\rho-\rho')\log\sigma\right)| + |\text{tr}\left(\rho'(\log\sigma' - \log\sigma)\right)|. \tag{277}$$

By Fannes inequality, we have

$$|-\mathbb{H}(\rho) + \mathbb{H}(\rho')| \leqslant \frac{1}{2}\|\rho-\rho'\|_1\log(d-1) + \mathbb{H}_b\left(\frac{1}{2}\|\rho-\rho'\|_1\right). \tag{278}$$

Furthermore, Cauchy-Schwartz inequality for Hilbert-Schmidt inner-products implies that

$$|\text{tr}\left((\rho-\rho')\log\sigma\right)| \leqslant \|\rho-\rho'\|_2\|\log\sigma\|_2 \tag{279}$$

$$\leqslant \|\rho-\rho'\|_1\|\log\sigma\|_2 \tag{280}$$

$$\leqslant \|\rho-\rho'\|_1 d\log\frac{1}{\lambda_{\min}(\sigma)}. \tag{281}$$

Using Cauchy-Schwartz again, we obtain

$$|\text{tr}\left(\rho'(\log\sigma' - \log\sigma)\right)| \leqslant \|\rho'\|_2\|\log\sigma' - \log\sigma\|_2 \tag{282}$$

$$\leqslant \|\log\sigma' - \log\sigma\|_2. \tag{283}$$

To upper-bound $\|\log\sigma' - \log\sigma\|_2$, let us define $F(x) \triangleq \log(\sigma + x(\sigma'-\sigma))$ for $t \in [0,1]$. Then,

$$\|\log\sigma' - \log\sigma\|_2 \ = \|F(1) - F(0)\|_2 \tag{284}$$

$$\overset{(a)}{\leqslant} \sup_{x\in[0,1]} \|F'(x)\|_2. \tag{285}$$

where (a) follows from mean value theorem of multi-variable functions. Applying Lemma 17 for $f \triangleq \log$ and $A(x) = \sigma + x(\sigma' - \sigma)$, we obtain

$$\|F'(x)\|_2 \leqslant d^2 \sup_{\mu \in [a,b]} |f'(\mu)| \|A'(x)\|_2 \tag{286}$$

$$\leqslant d^2 \frac{1}{\lambda_{\min}(\sigma + x(\sigma' - \sigma))} \|\sigma' - \sigma\|_2 \tag{287}$$

$$\leqslant d^2 \frac{1}{\lambda_{\min}(\sigma + x(\sigma' - \sigma))} \|\sigma' - \sigma\|_1. \tag{288}$$

Finally, for $x \in [0, 1]$, we have

$$\lambda_{\min}(\sigma + x(\sigma' - \sigma)) \leqslant \lambda_{\min}(\sigma) - \|x(\sigma' - \sigma)\|_2 \tag{289}$$

$$\leqslant \lambda_{\min}(\sigma) - \|\sigma' - \sigma\|_1. \tag{290}$$

□

REFERENCES

[1] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, pp. 145–195, Mar 2002.

[2] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *Theoretical Computer Science*, vol. 560, p. 7–11, Dec 2014.

[3] U. Vazirani and T. Vidick, "Fully device-independent quantum key distribution," *Phys. Rev. Lett.*, vol. 113, p. 140501, Sep 2014.

[4] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, "Tight finite-key analysis for quantum cryptography," *Nature Communications*, vol. 3, p. 634, Jan. 2012.

[5] H.-L. Yin, T.-Y. Chen, Z.-W. Yu, H. Liu, L.-X. You, Y.-H. Zhou, S.-J. Chen, Y. Mao, M.-Q. Huang, W.-J. Zhang, H. Chen, M. J. Li, D. Nolan, F. Zhou, X. Jiang, Z. Wang, Q. Zhang, X.-B. Wang, and J.-W. Pan, "Measurement-device-independent quantum key distribution over a 404 km optical fiber," *Phys. Rev. Lett.*, vol. 117, p. 190501, Nov 2016.

[6] B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE Journal of Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, September 2013.

[7] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Info. Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[8] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Info. Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[9] K. S. K. Arumugam and M. R. Bloch, "Keyless covert communication over multiple-access channels," in *Proc. of IEEE International Symposium on Information Theory*, Barcelona, Spain, July 2016, pp. 2229–2233.

[10] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, "Multilevel-coded pulse position modulation for covert communications," in *Proc. of IEEE International Symposium on Information Theory*, Vail, CO, Jun. 2018, pp. 1864–1868.

[11] Q. E. Zhang, M. Bakshi, and S. Jaggi, "Covert communication over adversarially jammed channels," in *2018 IEEE Information Theory Workshop (ITW)*. IEEE, Nov 2018, p. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/8613405/

[12] A. Sheikholeslami, B. A. Bash, D. Towsley, D. Goeckel, and S. Guha, "Covert communication over classical-quantum channels," in *Proc. of IEEE International Symposium on Information Theory*, Barcelona, Spain, July 2016, pp. 2064–2068.

[13] L. Wang, "Optimal throughput for covert communication over a classical-quantum channel," in *Proc. of IEEE Information Theory Workshop*, Cambridge, UK, September 2016, pp. 364–368.

[14] Y. Liu, J. M. Arrazola, W.-Z. Liu, W. Zhang, I. W. Primaatmaja, H. Li, L. You, Z. Wang, V. Scarani, Q. Zhang, and et al., "Experimental unconditionally secure covert communication in dense wavelength-division multiplexing networks," Sep 2017, arXiv: 1709.06755. [Online]. Available: http://arxiv.org/abs/1709.06755

[15] J. M. Arrazola and V. Scarani, "Covert quantum communication," *Phys. Rev. Lett.*, vol. 117, p. 250503, Dec 2016.

[16] J. M. Arrazola and R. Amiri, "Secret-key expansion from covert communication," *Phys. Rev. A*, vol. 97, p. 022325, Feb 2018.

[17] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, "Quantum-secure covert communication on bosonic channels," *Nature Communications*, vol. 6, no. 1, p. 8626, Dec 2015.

[18] M. Tahmasbi and M. R. Bloch, "A framework for covert and secret key expansion over quantum channels," submitted to Physical Review A, Nov. 2018.

[19] M. M. Wilde, *Quantum information theory*. Cambridge University Press, 2013.

[20] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, 10th ed. Cambridge University Press, 2010.

[21] M. Hayashi, "Universal coding for classical-quantum channel," *Communications in Mathematical Physics*, vol. 289, no. 3, pp. 1087–1098, 2009.

[22] I. Bjelakovic and H. Boche, "Classical capacities of compound and averaged quantum channels," *IEEE Transactions on Information theory*, vol. 55, no. 7, pp. 3360–3374, 2009.

[23] M. Hayashi, *Quantum information*. Springer, 2006.

[24] M. Ruskai and F. H. Stillinger, "Convexity inequalities for estimating free energy and relative entropy," *Journal of Physics A: Mathematical and General*, vol. 23, no. 12, p. 2421, 1990.

[25] D. Petz, *Quantum information theory and quantum statistics*, 1st ed., ser. Theoretical and Mathematical Physics, 2008, vol. 2008, no. 1, pp. 1–209.

[26] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.