Abstract—We study quantum soft covering and privacy amplification against quantum side information. The former task aims to approximate a quantum state by sampling from a prior distribution and querying a quantum channel. The latter task aims to extract uniform and independent randomness against quantum adversaries. For both tasks, we use trace distance to measure the closeness between the processed state and the ideal target state. We show that the minimal amount of samples for achieving an $\varepsilon$-covering is given by the $(1 - \varepsilon)$-hypothosis testing information (with additional logarithmic additive terms), while the maximal extractable randomness for an $\varepsilon$-secret extractor is characterized by the conditional $(1 - \varepsilon)$-hypothosis testing entropy. When performing independent and identical repetitions of the tasks, our one-shot characterizations lead to tight asymptotic expansions of the above-mentioned operational quantities. We establish their second-order rates given by the quantum mutual information variance and the quantum conditional information variance, respectively. Moreover, our results extend to the moderate deviation regime, which are the optimal asymptotic rates when the trace distances vanish at sub-exponential speed. Our proof technique is direct analysis of trace distance without smoothing.

Index Terms—Privacy amplification, quantum side information, soft covering, second-order analysis, hypothesis-testing divergence, randomness extraction.

I. INTRODUCTION

QUESTIONING the optimal rates for information-processing tasks is a core problem in classical and quantum information theory [2], [3], [4], [5], [6], [7], [8], [9], [10]. Nowadays, considerable research focus has shifted from the first-order characterization of the optimal rates to the second-order quantities in the asymptotic expansions of the optimal rates in coding blocklengths [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. Such second-order terms quantifying how much extra cost one has to pay in non-asymptotic scenarios are of significant importance both in theory and practice.

Indeed, much progress has been made in deriving the exact second-order rates for numerous quantum information-theoretic protocols. Nevertheless, this problem remains open for certain tasks such as privacy amplification against quantum side information (or called randomness extraction) [16], [22], where the operational quantities usually being used as a security criterion is the trace distance [20], [22], [23], [24], [25], [26], [27], [28]. This manifests the fact that existing analysis on operational quantities in terms of the trace distance as in various quantum information-theoretic tasks [20], [29], [30], [31], [32], [33] still has room for improvement. Hence, this problem will be the main focus of this work. Moreover, we hope the proposed analysis on the trace distance would shed new lights on the one-shot quantum information theory [20].

In this paper, we study two tasks. The first task is privacy amplification against quantum side information [20], [22], [27], [34], [35], [36], [37], [38], [39], [40], [41]. Suppose that a classical source $X$ at Alice’s disposal may be correlated with a quantum adversary Eve, which can be modelled as a joint classical-quantum (c-q) state $\rho_{XE} = \sum_{x \in X} p_X(x) |x\rangle\langle x| \otimes \rho_E$. The goal of Alice is to extract randomness from her original system $X$ to an output system, $Z$, which is also independent of Eve’s system. Namely, Alice wants to operate on her system $X$ such that the ideal output state is $\frac{1}{|Z|} \otimes \rho_E$. Due to operational motivation of composability [20], [22], [23], [25], [35], the trace distance is usually adopted as the security criterion to measure how far the extracted state is from the perfectly uniform and independent randomness, i.e.

$$\Delta(X | E)_\rho := \frac{1}{2} \text{Ent}(\rho_{XE}) - \frac{1}{|Z|} \otimes \rho_E^1,$$

where $\rho^1$ denotes the random hash function applied by Alice, and $\| \cdot \|_1$ is the Schatten 1-norm. A randomness extractor satisfying $\Delta(X | E)_\rho \leq \varepsilon$ is then said to be $\varepsilon$-secret. We define

Optimal Second-Order Rates for Quantum Soft Covering and Privacy Amplification

Yu-Chen Shen, Li Gao, and Hao-Chung Cheng, Member, IEEE
\(\ell^\varepsilon(X | E)_{\rho}\) as the maximal extractable randomness \(|Z|\) for \(\varepsilon\)-secret randomness extractors.

The second task studied in this work is quantum soft covering \([42], [43]\). Consider a c-q state \(p_{X:B} := \sum_{x \in X} p_X(x) \rho_{x} \otimes \rho_B\). The goal of quantum soft covering is to approximate the marginal state \(\rho\) using a random codebook \(C\) with certain size \(|C|\). Here, the codewords in \(C\) are independently sampled from the distribution \(p_X\); through the c-q channel \(x \mapsto \rho_{x,B}\), one may construct an approximation state \(\frac{1}{|C|} \sum_{x \in C} \rho_{x,B}\) to accomplish the goal. Again, the trace distance is used as the figure of merit to measure closeness, i.e.

\[
\Delta(X : B)_{\rho} := \frac{1}{2} \mathbb{E}_C \left\| \frac{1}{|C|} \sum_{x \in C} \rho_{x,B} - \rho_{B} \right\|_1.
\]

We say that the random codebook \(C\) achieves an \(\varepsilon\)-covering if it satisfies \(\Delta(X : B)_{\rho} \leq \varepsilon\). We then define the \(\varepsilon\)-covering number \(M^\varepsilon(X : B)_{\rho}\) as the minimum random codebook size \(|C|\) to realize an \(\varepsilon\)-covering.

Our main result is to provide one-shot characterizations for both the operational quantities \(\ell^\varepsilon(X | E)_{\rho}\) and \(M^\varepsilon(X : B)_{\rho}\). For privacy amplification, we show that for every \(0 < \varepsilon < 1\), the maximal extractable randomness using 2-universal hash functions is characterized by the \((1 - \varepsilon)\)-conditional hypothesis testing entropy \(H^1_{h}-\varepsilon(X | E)_{\rho}\) \([16]\):

\[
\log \ell^\varepsilon(X | E)_{\rho} \approx H^1_{h}-\varepsilon\delta(X | E)_{\rho}.
\]

Here, \(\approx\) means equality up to some logarithmic additive terms, and \(\delta\) is a parameter that be chosen for optimization (see Theorem 1 for the precise statement, and see Section II for detailed definitions). With a similar flavor, we prove that for every \(0 < \varepsilon < 1\), the minimal random codebook size for quantum soft covering is characterized by the \((1 - \varepsilon)\)-hypothesis testing information: \(I^1_{h}-\varepsilon(X : B)_{\rho}\) \([44]\):

\[
\log M^\varepsilon(X : B)_{\rho} \approx I^1_{h}-\varepsilon\delta(X : B)_{\rho}
\]

(see Theorem 2 for the precise statement). Contrary to the previous studies, the established one-shot entropic characterizations established in this work are not smoothed min- and max-entropies \([16], [20], [22], [45], [46], [47], [48]\).

In the scenario that the underlying states are identical and independently distributed, the established one-shot characterizations lead to the following second-order asymptotic expansions of the optimal rates, respectively (Propositions 11 and 14):

\[
\frac{1}{n} \log \ell^\varepsilon(X^n | E^n)_{\rho^n} = n H(X | E)_{\rho} + \sqrt{n V(X | E)_{\rho}} \Phi^{-1}(\varepsilon) + O(\log n);
\]

\[
\frac{1}{n} \log M^\varepsilon(X^n : B^n)_{\rho^n} = n I(X : B)_{\rho} - \sqrt{n V(X : B)_{\rho}} \Phi^{-1}(\varepsilon) + O(\log n).
\]

Here, the first-order terms are the conditional quantum entropy \(H(X | E)_{\rho}\) and the quantum mutual information \(I(X : B)_{\rho}\), whereas the second-order rates that can be expressed as the quantum conditional variance \(V(X | E)_{\rho}\) and the quantum mutual information variance \(V(X : B)_{\rho}\); and \(\Phi^{-1}\) is the inverse of the cumulative normal distribution. Note that in the scenario of privacy amplification but the adversary state is purely classical (which is also called classical side information), such second-order bound had already been solved in References \([37]\) and \([49]\).

Furthermore, our results extend to the moderate deviation regime \([50], [51]\). Namely, we derive both the optimal rates when the trace distances approach zero sub-exponentially (Propositions 12 and 15):

\[
\frac{1}{n} \log \ell^\varepsilon(X^n | E^n)_{\rho^n} = H(X | E)_{\rho} - \sqrt{2 V(X | E)_{\rho}} a_n + o(a_n);
\]

\[
\frac{1}{n} \log M^\varepsilon(X^n : B^n)_{\rho^n} = I(X : B)_{\rho} + \sqrt{2 V(X : B)_{\rho}} a_n + o(a_n).
\]

Here, \((a_n)_{n \in \mathbb{N}}\) is any moderate sequence satisfying \(a_n \to 0\) and \(n a_n^2 \to \infty\); and \(\varepsilon_n := e^{-n a_n^2} \to 0\). We remark that the moderate deviation rates for classical side information has been derived in Ref. \([49]\).

### A. Related Works

For privacy amplification, when the purified distance \(P\) is used as the security criterion, Tomamichel and Hayashi in Ref. \([16], \text{Theorem 8}\) derived a one-shot characterization in terms of the smooth min-entropy which leads to matched second-order asymptotics of the maximal extractable randomness:

\[
\frac{1}{n} \log \ell^p\varepsilon(X^n | E^n)_{\rho^n} = n H(X | E)_{\rho} + \sqrt{n V(X | E)_{\rho}} \Phi^{-1}(\varepsilon^2) + O(\log n).
\]

Note that while this provides the same first-order rate \(H(X | E)_{\rho}\), the two security criteria result in different second-order rates; i.e.

\[
\sqrt{V(X | E)_{\rho}} \Phi^{-1}(\varepsilon)\]

for trace distance (Propositions 11) and

\[
\sqrt{V(X : B)_{\rho}} \Phi^{-1}(\varepsilon^2)\]

for purified distance. We remark that the two results are incomparable; Ref. \([16], \text{Theorem 8}\) exploited the smooth entropy framework, where as the present work directly analyses trace distance.

In the classical case (when \(\{\rho_{x,B}\}_{x \in X}\) mutually commute), the direct part of Proposition 11 (i.e. lower bound) has been derived by Hayashi \([37]\) using 2-universal hash functions or dual 2-universal hash functions. The converse part (i.e. upper bound) has been obtained by Hayashi and Watanabe \([49]\) for general hash functions. Moreover, our one-shot characterization for the maximal extractable randomness (Theorem 1) can be seen as a quantum generalization of \([52], \text{Theorem 1}\), wherein the classical side information was considered. Anshu et al. in \([26], \text{Theorem 14}\) also established a one-shot characterization in terms of the so-called partially smoothed min-entropy, which also yields the optimal second-order rate for classical side information. We remark that the relation of the partially smoothed min-entropy and the conditional information-spectrum entropy (as well as the conditional hypothesis-testing entropy) has been proved for
the classical case in [26, Theorem 9]. The result remains open for the quantum case. Recently, Hayashi et al. studied an incoherent randomness extraction protocol [53], and the one-shot characterization and the second-order asymptotics were obtained. Before applying hash functions, this setting involves a quantum operation that outputs only some incoherent states followed by a completely dephasing channel. Hence, it can be seen as a generalization of Ref. [16] using purified distance too.

As for soft covering, [54, Equation (85)] provided an error exponent when relative entropy between the approximate state and the true marginal state is used as the error criterion. Parts of the authors established an error exponent for trace distance too. [43]. The first-order rate established in Proposition 14 is in agree with the existing results.

For positive semi-definite operator $L$ and positive operator $K$ we define the noncommutative quotient.

\[
\frac{K}{L} := L^{-\frac{1}{2}} K L^{-\frac{1}{2}}. \tag{1}
\]

We use $\mathbb{E}_{x \sim p_X}$ to stand for taking expectation where the underlying random variable is $x$ with probability distribution $p_X$ (with finite support), e.g. $\rho_{XB} = \sum_{x \in X} p_X(x) |x\rangle \langle x| \otimes \rho_B^x = \mathbb{E}_{x \sim p_X} (|x\rangle \langle x| \otimes \rho_B^x)$. For density operators $\rho$ and $\sigma$, we define the $\varepsilon$-information spectrum divergence [10], [58] as

\[
D^\varepsilon(\rho \parallel \sigma) := \sup_{\varepsilon \in \mathbb{R}} \left\{ -\log \text{Tr}[\sigma T] : \text{Tr}[\rho T] \leq 1 - \varepsilon \right\}. \tag{2}
\]

By a classical-quantum state $\rho_{XB} = \sum_{x \in X} p_X(x) |x\rangle \langle x| \otimes \rho_B^x$, we mean $p_X$ is a probability mass function on $X$ and each $\rho_B^x$ is a density operator on system $B$. We define the $\varepsilon$-hypothesis testing divergence [15], [16], [59] as

\[
D^\varepsilon_2(\rho \parallel \sigma) := \sup_{0 \leq T \leq 1} \left\{ -\log \text{Tr}[\sigma T] : \text{Tr}[\rho T] \geq 1 - \varepsilon \right\}. \tag{3}
\]

For positive semi-definite operator $\rho$ and positive definite operator $\sigma$, we define the collision divergence [22] as

\[
D^\varepsilon_2(\rho \parallel \sigma) := \log \text{Tr} \left[ \left( \frac{\sigma - \frac{1}{2} \rho \sigma - \frac{1}{2} \rho^*}{\sqrt{\sigma}} \right)^2 \right]. \tag{4}
\]

The quantum relative entropy [60], [61] and quantum relative entropy variance [15], [16] for density operator $\rho$ and positive definite operator $\sigma$ are defined as

\[
D(\rho \parallel \sigma) := \text{Tr} [\rho (\log \rho - \log \sigma)];
\]

\[
V(\rho \parallel \sigma) := \text{Tr} \left[ \rho (\log \rho - \log \sigma)^2 \right] - (D(\rho \parallel \sigma))^2.
\]

II. NOTATION AND AUXILIARY LEMMAS

For an integer $M \in \mathbb{N}$, we denote $[M] := \{1, \ldots, M\}$. We use ‘∧’ to indicate ‘minimum value’ between two scalars or the conjunction ‘and’ between two statements. We use $\mathbb{I}_A$ to denote the indicator function for a condition $A$. The density operators considered in this paper are positive semi-definite operators with unit trace. For a trace-class operator $H$, the trace class norm (also called Schatten-1 norm) is defined by

\[
\|H\|_1 := \text{Tr} \left[ \sqrt{H^* H} \right].
\]

For positive semi-definite operator $K$ and positive operator $L$, we use the following short notation for the noncommutative quotient.

\[
K \perp L := L^{-\frac{1}{2}} K L^{-\frac{1}{2}}. \tag{1}
\]
Lemma 1 (A Cauchy–Schwarz-Type Inequality [31, Exercise 6.9]): For any operator $X$ and a density operator $\rho$, it holds that
\[
\|X\|_1 \leq \sqrt{\text{Tr}[X^\dagger X \rho^{-1}]}.
\]

For any self-adjoint operator $H$ with eigenvalue decomposition $H = \sum \lambda_i |e_i\rangle\langle e_i|$, we define the set $\text{spec}(H) := \{\lambda_i\}_i$ to be the eigenvalues of $H$, and $|\text{spec}(H)|$ to be the number of distinct eigenvalues of $H$. We define the pinching map with respect to $H$ as
\[
\mathcal{P}_H[L] : L \rightarrow \sum_i |e_i\rangle\langle e_i|L|e_i\rangle/\langle e_i|.
\]

Lemma 2 (Pinching inequality [62]): For every $d$-dimensional self-adjoint operator $H$ and positive semi-definite operator $L$,
\[
\mathcal{P}_H[L] \geq \frac{1}{|\text{spec}(H)|} L.
\]
Moreover, it holds that for every $n \in \mathbb{N}$,
\[
|\text{spec}(H^\otimes n)| \leq (n + 1)d^n - 1.
\]

We have the following relation between divergences that will be used in our proofs.

Lemma 3 (Relation Between Divergences [16, Lemma 12, Proposition 13, Theorem 14]): For every density operator $\rho$, positive semi-definite operator $\sigma$, $0 < \varepsilon < 1$, and $0 < \delta < 1 - \varepsilon$, we have
\[
D^\varepsilon_\delta(\rho^\sigma) \geq D^{\varepsilon+\delta}_\delta(\rho^\sigma) + \log |\text{spec}(\sigma)| - 2\log \delta,
\]
\[
D^\varepsilon_\delta(\rho^\sigma) \leq D^{\varepsilon+\delta}_\delta(\rho^\sigma) - \log \delta.
\]

Lemma 4 (Lower Bound on the Collision Divergence [63, Theorem 3]): For every $0 < \eta < 1$ and $\lambda_1, \lambda_2 > 0$, density operator $\rho$, and positive semi-definite $\sigma$, we have
\[
\exp D^\eta_\delta(\rho^\sigma) \geq \frac{1 - \eta}{\lambda_1 + \lambda_2} \cdot e^{-D^\delta(\rho^\sigma)}.
\]

Lemma 5 (Joint Convexity [64, Proposition 3], [17, [65]]): The map
\[(\rho, \sigma) \mapsto \exp D^\eta_\delta(\rho^\sigma)\]
is jointly convex on all positive semi-definite operators $\rho$ and $\sigma$.

Lemma 6 (Variational Formula of the Trace Distance [66, 67, 29, 89]): For density operators $\rho$ and $\sigma$,
\[
\frac{1}{2} \|\rho - \sigma\|_1 = \sup_{0 \leq \Pi \leq 1} \text{Tr}[\Pi(\rho - \sigma)].
\]

We list some basic properties of the noncommutative quotient introduced in (1) as follows. Since they follow straightforwardly from basic matrix theory, we will use them in our derivations without proofs.

Lemma 7 (Properties of the noncommutative quotient): The noncommutative quotient defined in (1) satisfies the following:

(a) $A/B \geq 0$ for all $A \geq 0$ and $B > 0$;
(b) $A/B = A/C + B/C$ for all $A, B \geq 0$ and $C > 0$;
(c) $A + B \leq 1$ for all $A \geq 0$ and $B > 0$;
(d) $\text{Tr}[A/B] = \text{Tr}[B/A]$ for all $A, B \geq 0$ and $C > 0$.

Lemma 8 (Jensen’s Inequalities [68, 69]): Let $\Phi$ be a unital positive linear map between two spaces of bounded operators (possibly with different dimensions), and let $f$ be an operator concave function. Then for every self-adjoint operator $H$, one has
\[
\Phi(f(H)) \leq f(\Phi(H)).
\]

For examples, the unital positive linear map $\Phi$ considered in this paper include (i) taking expectation $\mathbb{E}$ on (possibly matrix-valued) random variables; (ii) the pinching map $\mathcal{P}$; and (iii) the functional $H \mapsto \text{Tr}[\rho H]$ for some density operator $\rho$.

Lemma 9 (Second-Order Expansion [15, 16]): For every density operator $\rho$, positive definite operator $\sigma$, $0 < \varepsilon < 1$, and $\delta = O(1/\sqrt{n})$, we have the following expansion:
\[
D^\varepsilon_\delta(\rho^\sigma) = n D(\rho^\sigma) + \sqrt{n} V(\rho^\sigma) \Phi^{-1}(\varepsilon) + O(\log n),
\]
where $\Phi$ is the cumulative normal distribution $\Phi(u) := \int_{-\infty}^{u} \frac{e^{-t^2}}{\sqrt{2\pi}} dt$, and its inverse $\Phi^{-1}(\varepsilon) := \sup\{u \mid \Phi(u) \leq \varepsilon\}$.

Lemma 10 (Moderate deviations [51, Theorem 1]): Let $(a_n)_{n \in \mathbb{N}}$ be a moderate sequence satisfying
\[
\lim_{n \rightarrow \infty} a_n = 0, \quad \lim_{n \rightarrow \infty} na_n^2 = \infty,
\]
and let $\varepsilon_n := e^{-na_n^2}$. For any density operator $\rho$ and positive definite operator $\sigma$, the following asymptotic expansions hold:
\[
\frac{1}{n} D^\varepsilon_\delta(\rho^\sigma) = D(\rho^\sigma) + \varepsilon_n a_n + o(a_n).
\]

III. PRIVACY AMPLIFICATION AGAINST QUANTUM SIDE INFORMATION

Let $\mathcal{H}_E$ be a finite dimensional Hilbert space. Consider a classical-quantum state $\rho_{XE} = \sum_{x \in \mathcal{X}} p_X(x) |x\rangle\langle x| \otimes \rho^E_x$. Without loss of generality, we assume that the marginal density $\rho^E$ is invertible.

We note that Lemmas 9 and 10 were originally stated for normalized $\sigma$ [15, 16, 51]. They hold for positive definite operator $\sigma$ as well by observing that $D^\varepsilon_\delta(\rho^\sigma) = D^\varepsilon_\delta(\rho^{\sigma\sigma})$.

We note that the $\varepsilon$-hypothesis testing divergence $D^\varepsilon_\delta$ used in [51] has an additional log-convex term than our definition in (3). Nonetheless, this does not affect the moderate deviation results since the additional terms are $\frac{1}{2} \log(1 - \varepsilon_n) = o(a_n)$ and $\frac{1}{2} \log \varepsilon_n = -a_n^2 = o(a_n)$ for any moderate sequence $(a_n)_{n \in \mathbb{N}}$.

---

4Ref. [63, Thm. 3] is stated for $\lambda_1 = \lambda \in (0, 1)$ and $\lambda_2 = 1 - \lambda$. We remark that the result applies to the case $\lambda_1, \lambda_2 > 0$ by following the same proof.
In this work, we consider a \(2\)-universal random hash function \(h : X \to Z\), a random function satisfying that for all \(x, x' \in X\) with \(x \neq x'\),
\[
\Pr_{h} \{ h(x) = h(x') \} = \frac{1}{|Z|}.
\]

Alice applies the linear operation \(R_{X\to Z}^{h}\) on her system \(X\) by the following:
\[
R_{X\to Z}^{h}(ρ_{XE}) := \sum_{x \in X} ρ_{X}(x)|h(x)⟩⟨h(x)| ⊗ ρ_{E}^{a}.
\]

A perfectly randomizing channel \(U_{X\to Z}\) from \(X\) to \(Z\) is defined as
\[
U(θ_{X}) = \frac{1}{|Z|} \left( \sum_{x} θ_{X}(x) \right).
\]

We define the maximal extractable randomness for an \(ε\)-secret extractor [20, §7] as
\[
ℓ^{ε}(X | E)_{ρ} := \sup \left\{ ℓ \in \mathbb{N} : |Z| \geq ℓ \land \frac{1}{2} E_{h} \| R_{X\to Z}^{h}(ρ_{XE}) - U(ρ_{XE}) \| _{1} \leq ε \right\}.
\]
(7)

We remark that the supremum is over all \(2\)-universal hash function \(h\).

The main result of this section is to prove the following one-shot characterization of the operational quantity \(ℓ^{ε}(X | E)_{ρ}\).

**Theorem 1:** Let \(ρ_{XE}\) be a classical-quantum state, and let \(h : X \to Z\) be a \(2\)-universal hash function. Then, for every \(0 < ε < 1\) and \(0 < c < δ < \frac{3}{4} \land \frac{1−ε}{2}\), we have
\[
H_{h}^{1−ε+3δ}(X | E)_{ρ} − \log \frac{ε^{2}}{δ} \leq \log ℓ^{ε}(X | E)_{ρ}
\]
\[
≤ H_{h}^{1−ε−2δ}(X | E)_{ρ} + \log \left( \frac{1+c}{δ−c} \right) + \log \left( \frac{ε+c}{δ} \right).
\]

Here, \(H_{h}^{ε}\) is the conditional \(ε\)-hypothesis testing entropy defined in (5) and \(ρ = |\text{spec}(ρ_{E})|\).

In the i.i.d setting, the one-shot characterization from Theorem 1 combined with the second-order expansion of \(H_{h}^{ε}\) leads to the following second-order asymptotics of \(\log ℓ^{ε}(X^{n} | E^{n})_{ρ^{⊗n}}\).

**Proposition 11:** Let \(ε \in (0, 1)\). For any \(2\)-universal hash function \(h^{n} : X^{n} \to Z^{n}\) and
\[
\frac{1}{2} E_{h^{n}} \| (R_{X^{n}\to Z^{n}}^{h^{n}} - U^{⊗n})(ρ_{XE}^{⊗n}) - U(ρ_{XE}^{⊗n}) \| _{1} \leq ε,
\]
we have the asymptotic expansion:
\[
\log ℓ^{ε}(X^{n} | E^{n})_{ρ^{⊗n}} = n H(X | E)_{ρ} + \sqrt{n V(X | E)_{ρ}Φ^{-1}(ε)} + O (\log n).
\]

**Proof:** Since \(|\text{spec}(ρ_{E})| \leq (n + 1)^{|\mathcal{H}_{E}|−1}\) for \(|\mathcal{H}_{E}|\) being the rank of \(ρ_{E}\), the additive terms grow of order \(O(\log n)\). We apply the established one-shot characterization, Theorem 1, with \(δ = n^{−(1/2)}\) and \(c = \frac{1}{2} δ\), together with the second-order expansion of the conditional hypothesis testing entropy, Lemma 9, to arrive at the claim. □

**Remark 1:** In the classical case (when \(\{ρ_{E}^{a}\}_{a∈X}\) mutually commute), the direct part of Proposition 11 (i.e. lower bound) has been derived by Hayashi [37] using \(2\)-universal hash functions or dual \(2\)-universal hash functions. Moreover, the converse part (i.e. upper bound) has been obtained by Hayashi and Watanabe [49] for general hash functions.

**Remark 2:** When the purified distance \(\mathcal{P}\) is used as security criterion, [16, Theorem 8] derived a one-shot characterization using smooth min-entropy. \(\log ℓ^{ε}(X | E)_{ρ} \approx H_{min}^{1−ε+δ}(X | E)_{ρ}\). Their result leads to matched second-order asymptotic expansions of optimal rates:
\[
\log ℓ^{ε}(X^{n} | E^{n})_{ρ^{⊗n}} = n H(X | E)_{ρ} + \sqrt{n V(X | E)_{ρ}Φ^{-1}(ε^{2})} + O (\log n).
\]

According to relation between trace distance and purified distance, our result and the above agree in the first-order rate, while the second-order rates are different.

**Remark 3:** Our converse bound is an “ensemble converse”, which characterizes the optimality of the ensemble of all \(2\)-universal hash functions instead of arbitrary hash functions. We focus on \(2\)-universal hash functions because it is one of the mostly used extractors in the achievability proof of privacy amplification (e.g. [16], [27], [52], [70]). The lower bound in Theorem 1 provides an achievable one-shot bound, while the upper bound shows the tightness with respect to the \(2\)-universal hash functions. The second-order converse analysis for arbitrary hash functions is left for future work.

Moreover, the one-shot characterization can be extended to the moderate deviation regime [50], [51]. Namely, we derive the optimal rate of the maximal required output dimension when the error approaches zero moderately quickly.

**Proposition 12 (Moderate Deviations for Privacy Amplification):** For every classical-quantum state \(ρ_{XE}\) and any moderate sequence \((a_{n})_{n∈\mathbb{N}}\) satisfying (6) and \(ε_{n} := e^{−na_{n}}\), we have
\[
\left\{ \frac{1}{n} \log ℓ^{ε_{n}}(X^{n} | E^{n})_{ρ^{⊗n}} \right\} = H(X | E)_{ρ} − \sqrt{2V(X | E)_{ρ}a_{n} + o(a_{n})};
\]
\[
\frac{1}{n} \log ℓ^{1−ε_{n}}(X^{n} | E^{n})_{ρ^{⊗n}} = H(X | E)_{ρ} + \sqrt{2V(X | E)_{ρ}a_{n} + o(a_{n})}.
\]

**Proof:** We prove the expansion for \(ℓ^{ε_{n}}(X^{n} | E^{n})\). For every moderate deviation sequence \((a_{n})_{n∈\mathbb{N}}\), we let \(δ_{n} = \frac{1}{4} e^{−na_{n}}\) satisfying (6) and let \(c_{n} = \frac{1}{2} δ_{n}\). Then, \(ε_{n} + 2δ_{n}\) or \(ε_{n}−3δ_{n}\) can be viewed as another \(e^{−nδ_{n}}\) for another moderate deviation sequence \((b_{n})_{n∈\mathbb{N}}\); we have \(b_{n} = a_{n} + (b_{n}−a_{n}) = o(a_{n})\) (see e.g. [71, §7.2]). \(\frac{1}{n} \log δ_{n} = −a_{n}^{2} = o(a_{n})\) and \(\frac{1}{n} \log |\text{spec}(ρ_{E}^{⊗n})| = O (\frac{\log n}{n}) = o(a_{n})\). Then, applying Theorem 1 with Lemma 10 leads to our first claim of the moderate derivation for privacy amplification. The second line follows similarly. □

**Remark 4:** We note that the moderate deviation analysis for the optimal errors (in terms of the trace distance) when rate approaching the first-order term has been studied in the authors’ previous work [28, §6] (see also the results by Hayashi and Watanabe [49, Theorem 31] for classical side information with moderate deviation sequence \(a_{n} = n^{−t}\) for any \(t ∈ (0, 1/2)\)). Here, the moderate deviation analysis is taken from the different perspective—we derive the optimal
rates of privacy amplification when the errors converge to 0 or 1 at sub-exponential speed.

We prove the lower bound and upper bound of \( \log \mathcal{E}(X : E) \) in Section III-A and Section III-B, respectively.

### A. Direct Bound

We prove the direct (lower) bound on \( \log \mathcal{E}(X | E) \) here. **Proof:** [Proof of achievability in Theorem 1]

We first claim the following achievability bound that for any \( c > 0 \) and 2-universal hash function \( h : X \to Z \),

\[
\frac{1}{2} E_h \left\| \mathcal{R}^h(\rho_{XE}) - \mathcal{U}(\rho_{XE}) \right\|_1 \\
\leq \text{Tr} \left[ \rho_{XE} \{ \rho_{XE} [X > c1_X \otimes \rho_E] \} \right] \\
+ \sqrt{c} \| \text{spec}(\rho_E) \|_Z.
\]  

(8)

Let \( \delta \in (0, c) \) and let

\[
c = \exp \left\{ D_1^{1-\varepsilon} \right\}
\]

\[
\text{for some small } \xi > 0. \text{ Then, by definition of the information spectrum divergence, (2), we have}
\]

\[
\text{Tr} \left[ \rho_{XE} \{ \rho_{XE} [X > c1_X \otimes \rho_E] \} \right] < \varepsilon - \delta.
\]  

(9)

Choose \( |Z| = \left\lfloor \frac{\varepsilon^2}{c \| \text{spec}(\rho_E) \|_Z} \right\rfloor \). Then, by (8) and (9), the \( \varepsilon \)-secret criterion is satisfied, i.e.

\[
\frac{1}{2} E_h \left\| \mathcal{R}^h(\rho_{XE}) - \mathcal{U}(\rho_{XE}) \right\|_1 \leq \varepsilon.
\]

By the definition of (7), we have the following lower bound on \( \mathcal{E}(X | E)_\rho \):

\[
\log \mathcal{E}(X | E)_\rho \\
\geq -D_1^{1-\varepsilon} \left\{ \rho_{XE} [X > c1_X \otimes \rho_E] \right\} - \xi - \log \| \text{spec}(\rho_E) \|_1 \\
+ 2 \log \delta \\
\geq \mathcal{H}_1^{1-\xi} (X | E)_\rho - \xi - 2 \log \| \text{spec}(\rho_E) \|_1 + 4 \log \delta,
\]

where we have used Lemma 3 in the last inequality. Since \( \xi > 0 \) is arbitrary, taking \( \xi \to 0 \) gives our claim of lower bound in Theorem 1.

Now we move on to prove (8). We first consider case where \( \rho^c_E \) is invertible for each \( x \in X \) and later argue that the general case follows from approximation. Shorthand \( p \equiv p_X \) and \( \rho_x \equiv \rho^c_x \). For every \( x \in X \), we take the projection

\[
\Pi_x = \{ \rho_{XE} [p(x) x \rho_x] \leq c \rho_x \}; \\
\Pi_x^c := 1_E - \Pi_x,
\]

or in the joint space \( XE \):

\[
\Pi := \{ \rho_{XE} [X \rho_x] \leq c1_X \otimes \rho_E \} \\
\Pi^c := 1_X \otimes 1_E - \sum_{x \in X} |x> <x| \otimes \Pi_x.
\]

We then use the fact that the Schatten 1-norm \( \| \cdot \|_1 \) is additive for direct sums to calculate

\[
\frac{1}{2} E_h \left\| \mathcal{R}^h(\rho_{XE}) - \mathcal{U}(\rho_{XE}) \right\|_1 \\
= \frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\rho_{XE} (\Pi + \Pi^c)) \right\|_1
\]

\[
\leq \frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\rho_{XE} \Pi^c) \right\|_1 + \frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\rho_{XE} \Pi) \right\|_1
\]

\[
\leq \frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\Pi^c \rho_{XE} \Pi^c) \right\|_1 \\
+ \frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi) \right\|_1
\]

\[
\leq \frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi) \right\|_1 + \frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi) \right\|_1,
\]

(10)

where (a) follows from the triangle inequality of the Schatten 1-norm \( \| \cdot \|_1 \), and in (b) we use triangle inequality again to decompose the first term of RHS of (a).

We bound the three terms of (10) as follows. The first term is bounded by

\[
\frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi^c) \right\|_1 \\
= \frac{1}{2} E_h \left\| (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi^c) \right\|_1 \\
\leq \frac{1}{2} E_h \left[ \text{Tr} \left[ (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi^c) \right] \right] \\
= \frac{1}{2} E_h \left[ \text{Tr} \left[ (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi^c) \right] \right]
\]

The second term of (10) can be bounded by

\[
\frac{1}{2} E_h \left( \text{Tr} \left[ (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi) \right] \right)
\]

\[
\leq \frac{1}{2} \left( E_h \text{Tr} \left[ (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi) \right] \right)
\]

where (a) follows from Lemma 1 by setting \( X = (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi) \) and \( \rho = \frac{1}{2 \rho} \otimes \rho_E \), and (b) follows from the concavity of the square root.

In (11), the term in the square root can be separated by

\[
\text{Tr} \left[ (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi) \right] \left[ (\mathcal{R}^h - \mathcal{U})(\Pi \rho_{XE} \Pi) \right]
\]

\[
\leq \frac{1}{2} \left( E_h \text{Tr} \left[ \mathcal{R}^h(\Pi \rho_{XE} \Pi) \right] \right) \left[ \mathcal{R}^h(\Pi \rho_{XE} \Pi) \right] \\
= \frac{1}{2} \left( E_h \text{Tr} \left[ \mathcal{R}^h(\Pi \rho_{XE} \Pi) \right] \right) \left[ \mathcal{R}^h(\Pi \rho_{XE} \Pi) \right]
\]

(12)
The last term of (12) is equal to

\[ E_{th} \text{Tr} \left[ |R^h(\Pi^c \rho_{XE} \Pi)|^2 \right] \]

The first term of (12) is equal to

\[ E_{th} \text{Tr} \left[ \sum_{x} \| \Pi_x(p(x) \rho_x) \|_{\infty} \| \Pi_x(p(x) \rho_x) \Pi_x^c \|_{\infty} \right] \]

The second term of (12) is equal to

\[ E_{th} \text{Tr} \left[ \partial \left( \Pi^c \rho_{XE} \Pi \right) \right] \]

\[ E_{th} \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

\[ = E_{th} \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

\[ = E_{th} \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

\[ = \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

Using (12) and the above four equations (13) (14) (15) (16), (11) then becomes

\[ \frac{1}{2} \left( E_{th} \text{Tr} \left[ \partial \left( \Pi^c \rho_{XE} \Pi \right) \right] \right)^{(1/2)} \]

The third term of (12) is indeed the same of the second term. Thus,

\[ E_{th} \text{Tr} \left[ |U(\Pi^c \rho_{XE} \Pi)|^2 \right] \]

\[ = E_{th} \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

\[ = \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

The last term of (12) is equal to

\[ E_{th} \text{Tr} \left[ |U(\Pi^c \rho_{XE} \Pi)|^2 \right] \]

\[ = E_{th} \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

\[ = \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

\[ = \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

\[ \leq \text{Tr} \left[ \sum_{x} \Pi_x(p(x) \rho_x) \Pi_x(p(x) \rho_x) \Pi_x^c \right] \]

In (a) we use that,

\[ \Pi_x \rho_x^2 \Pi_x \leq c \Pi_x \rho_x (p(x) \rho_x)^{-1} \Pi_x \]

\[ \leq c \rho_x (p(x) \rho_x)^{-1} \rho_x \]

\[ \leq c \text{spec}(\rho_x) \rho_x \]

where the last inequality follows from Lemma 2. The third term of (10) can be bounded similarly to the second term.

\[ \frac{1}{2} E_{th} \| (R^h - U)(\Pi \rho_{XE}) \|_1 \leq \frac{1}{2} \sqrt{|Z| c \text{spec}(\rho_x)} \]

Combining all three terms of (10) and we get the desired result (8).

For the general non-invertible states \{\rho_x\}_x we define the approximation

\[ \rho_x^c := (1 - \epsilon) \rho_x + \epsilon \omega_x \]

where \epsilon \in [0, 1]; each \omega_x is full rank and it will be determined later. Let \rho_{XE}^c be denoted as follows,

\[ \rho_{XE}^c = \sum_{x \in X} p(x) |x\rangle \langle x| \otimes \rho_x^c \]

It is clear that

\[ \lim_{\epsilon \to 0} \frac{1}{2} E_{th} \| R^h(\rho_{XE}^c) - U(\rho_{XE}^c) \|_1 \]

\[ = \frac{1}{2} E_{th} \| R^h(\rho_{XE}) - U(\rho_{XE}) \|_1 \]

On bounding the term \[ \frac{1}{2} E_{th} \| R^h(\rho_{XE}^c) - U(\rho_{XE}^c) \|_1 \], we follow similar process of our proof where \rho_x is invertible, but with the projection operator being

\[ \Pi_x = \{ \mathcal{P}_{\rho_x} [p(x) \rho_x^c] \leq c \rho_x \} ; \]

\[ \Pi = \sum_{x \in X} \langle x| \otimes \Pi_x = \{ \mathcal{P}_{\rho_x} [p(x) \rho_x^c] \leq c \mathbb{I}_X \otimes \rho_x \} \].

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Following the same steps as in (10) leads to,
\[
\begin{align*}
\frac{1}{2} E_h \left\| R^h(\rho_{XE}) - U(\rho_{XE}) \right\|_1 &
\leq \frac{1}{2} E_h \left\| (R^h - U)(\Pi^\rho_{XE} \Pi^\rho_{XE}) \right\|_1 \\
&+ \frac{1}{2} E_h \left\| (R^h - U)(\Pi' \rho_{XE} \Pi') \right\|_1 \\
&+ \frac{1}{2} \left\| (R^h - U)(\rho_{XE} \Pi) \right\|_1.
\end{align*}
\] (18)

For bounding the first term on the right-hand side of (18), suppose \( \rho_E = \sum_{e \in E} \rho_e |e\rangle \langle e | \) and let \( \omega_x = \sum_{e \in E} q_{e}^x |e\rangle \langle e | \). Then,
\[
\begin{align*}
\lim_{\epsilon \to 0} \left\{ \{ \rho_{pE} | \rho_{XE}^\epsilon \} > c \mathbb{I}_X \otimes \rho_E \right\} \\
= \lim_{\epsilon \to 0} \left\{ \sum_{x \in \mathcal{X}} |x\rangle \langle x| \otimes \{ \rho_{pE} | p(x) \rho_x^\epsilon > c \rho_E \} \right\} \\
= \lim_{\epsilon \to 0} \sum_{x \in \mathcal{X}} |x\rangle \langle x| \\
\otimes \{ (1 - \epsilon) p(x) \rho_{pE} | p(x)_\rho_x > (1 - \epsilon) c \rho_E + \epsilon c \rho_E \} \\
= \sum_{x \in \mathcal{X}} |x\rangle \langle x| \otimes \{ p(x) \rho_{pE} | p(x) \rho_x > c \rho_E \} \\
\otimes \{ (p(x) \rho_{pE} | p(x) \rho_x = c \rho_E) \cap (p(x) \rho_x > c \rho_E) \}. \quad (19)
\end{align*}
\]

Here, (19) comes from that in the projection \( \{ (1 - \epsilon) p(x) \rho_{pE} | p(x) \rho_x > (1 - \epsilon) c \rho_E + \epsilon c \rho_E \} \), as \( \epsilon \to 0 \), the comparison between \( (1 - \epsilon) p(x) \rho_{pE} | p(x) \rho_x \) and \( (1 - \epsilon) c \rho_E \) dominates the other terms. On the other hand, in the eigenspace where the above two terms are equal, then comparing the two perturbation terms \( c \rho(x) \omega_x \) and \( c \epsilon \rho_E \) determines the projection. Note also that each operator in the projection is mutually diagonalized in the basis \{ |e\rangle |e\rangle \}, and we are allowed to take intersection of commuting projections.

For every \( x \), we now choose every \( q_{e}^x \) as follows: for all \( x \in \mathcal{X} \) and \( e \in E \), if \( p(x) |e\rangle |\rho_{pE} | p(x)_\rho_x > (1 - \epsilon) c \rho_E + \epsilon c \rho_E \), we avoid the situation \( p(x) q_{e}^x > c \rho_E \) while keeping \( q_{e}^x \) non-zero. This can always be done for every \( q_{e}^x \) since (1) if there exists \( e' \in E \) such that \( p(x) |e'\rangle |\rho_{pE} | p(x)_\rho_x \neq c \rho_E \), we can take such \( q_{e}^x \) large enough so that for any other \( e \neq e' \), \( q_{e}^x \) can be small enough to avoid \( p(x) q_{e}^x > c \rho_E \). (2) If for every \( e \in E \), \( p(x) |e\rangle |\rho_{pE} | p(x)_\rho_x = c \rho_E \), \( c \) can only be \( p(x) \) in this case. Then we can set \( q_{e}^x = p(x) \) for all \( e \).

For any \( x \in \mathcal{X} \), by this choice of \( \omega_x \),
\[
\{ p(x) \rho_{pE} | p(x) \rho_x = c \rho_E \} \cap (p(x) \rho_x > c \rho_E) = 0.
\]
Thus,
\[
\begin{align*}
\lim_{\epsilon \to 0} \left\{ \rho_{pE} | \rho_{XE}^\epsilon > c \mathbb{I}_X \otimes \rho_E \right\} \\
\leq \sum_{x \in \mathcal{X}} |x\rangle \langle x| \otimes \{ \rho_{pE} | p(x) \rho_x > c \rho_E \} \\
= \{ \rho_{pE} | \rho_{XE} > c \mathbb{I}_X \otimes \rho_E \}.
\end{align*}
\] (20)

In the limit \( \epsilon \to 0 \), the first term of (18) becomes
\[
\begin{align*}
\lim_{\epsilon \to 0} \frac{1}{2} E_h \left\| (R^h - U)(\Pi^\rho_{XE} \Pi^\rho_{XE}) \right\|_1 \\
\leq \lim_{\epsilon \to 0} \left\{ \{ \rho_{pE} | \rho_{XE}^\epsilon > c \mathbb{I}_X \otimes \rho_E \} \right\} \\
\leq \lim_{\epsilon \to 0} \left\{ \{ \rho_{pE} | \rho_{XE} > c \mathbb{I}_X \otimes \rho_E \} \right\} \\
= \left\{ \rho_{pE} | \rho_{XE} > c \mathbb{I}_X \otimes \rho_E \right\}.
\end{align*}
\] (21)

where in (a) we use (20). On the other hand, the sum of the second and third term of (18) can be bounded similarly as before,
\[
\begin{align*}
\frac{1}{2} E_h \left\| (R^h - U)(\Pi^\rho_{XE} \Pi') \right\|_1 + \frac{1}{2} \left\| (R^h - U)(\rho_{XE} \Pi) \right\|_1 \\
\leq \sqrt{c} \text{spec}(\rho_E) \| Z \|.
\end{align*}
\] (22)

Combining (17), (18), (21), and (22), we obtain our desired (8).

Remark 5: To obtain the one-shot characterization in Theorem 1, it suffices to have (8) for densely many \( c \). This can be argued by a simpler approximation \( \rho_{XE} = (1 - \epsilon) \rho_{XE}^\epsilon + \epsilon \rho_X \otimes \frac{1}{|E|} \). Nevertheless, in the above proof, we managed to prove (8) for all \( c \) by the careful choices of \( \omega_x \) instead of \( |E| \).

B. Converse Bound

In this section, we prove the upper bound of \( \log \ell^p(X | E) \).

Lemma 13: For positive semi-definite operators \( K \) and \( L \),
\[
\sup_{0 \leq \Pi \leq 1} \text{Tr} [K \Pi] + \text{Tr} [(1 - \Pi) L] = \text{Tr} \left[ \frac{K + L + |K - L|}{2} \right].
\]

Therefore,
\[
\frac{1}{2} \left\| K - L \right\|_1 = \sup_{0 \leq \Pi \leq 1} \text{Tr} ((K - L) \Pi) - \text{Tr} \left[ \frac{K - L}{2} \right].
\]

Proof: [Proof of converse of Theorem 1]

We denote \( p = p_X \) and \( \rho_x = \rho_x^\epsilon \). We introduce the notation \( \rho_{h(x)E} := p(x) |h(x)\rangle \langle h(x)| \otimes \rho_x \).

Note that
\[
\begin{align*}
R^h(\rho_{XE}) &= \sum_{x \in \mathcal{X}} \rho_{h(x)E}, \\
\text{and} \\
U(\rho_{XE}) &= 1_Z \otimes \frac{\rho_{XE}}{|Z|}.
\end{align*}
\]

For arbitrary set of matrices \( \{ \Pi_z | 0 \leq z \leq 1 \wedge z \in Z \} \),
\[
\epsilon \geq E_{h, Z} \frac{1}{2} \left\| (R^h - U)(\rho_{XE}) \right\|_1 \\
= E_{h, Z} \sum_{z: h(x) = z} \rho_{h(x)E} - |z\rangle \langle z| \otimes \frac{\rho_{XE}}{|Z|} \\
= E_{h, Z} \sum_{z: h(x) = z} \rho_{x} - \rho_{XE} |Z| \\
\geq E_{h, Z} \sum_{z: h(x) = z} \text{Tr} \left[ \left( \sum_{x: h(x) = z} p(x) \rho_x - \rho_{XE} |Z| \right) \Pi_z \right] \\
= E_{h, Z} \sum_{z: h(x) = z} \text{Tr} \left[ \left( \sum_{x: h(x) = z} p(x) \rho_x - \rho_{XE} |Z| \right) \Pi_z \right] \\
= E_{h, Z} \sum_{z: h(x) = z} \text{Tr} \left[ \sum_{x: h(x) = z} (p(x) \rho_x) \Pi_z \right].
\]

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where in (a) we use Lemma \ref{lem:bound}, and (b) follows since the last
term of (a) is cancelled after summation over \( z \).

For every \( 0 < c < \delta < \frac{1}{|Z|} \), and for any realization of the
random hash function \( h \), we choose \( \Pi_z \) by:

\[
\Pi_z = \frac{\sum_{x : h(x) = z} (p(x) \rho_x)}{\sum_{x : h(x) = z} (p(x) \rho_x) + c^{-1} \frac{\rho_E}{|Z|}},
\]

The first term of (23) can be bounded by

\[
\mathcal{E}_h \sum_{z \in Z} \operatorname{Tr} \left[ \frac{\rho_E}{|Z|} \Pi_z \right] = \mathcal{E}_h \sum_{x \in \mathcal{X}} \operatorname{Tr} \left[ \frac{p(x) \rho_x}{|Z|} \sum_{x : h(x) = z} (p(x') \rho_{x'}) + c^{-1} \frac{\rho_E}{|Z|} \right]
\]

\[
\geq \mathcal{E}_h \sum_{x \in \mathcal{X}} \operatorname{Tr} \left[ \frac{p(x) \rho_x}{|Z|} \sum_{x : h(x) = z} (p(x') \rho_{x'}) + c^{-1} \frac{\rho_E}{|Z|} \right] = \mathcal{E}_h \sum_{x \in \mathcal{X}} \operatorname{Tr} \left[ |x\rangle \langle x| \otimes \frac{p(x) \rho_x}{|Z|} \sum_{x : h(x) = z} (p(x') \rho_{x'}) + c^{-1} \frac{\rho_E}{|Z|} \right]
\]

\[
\geq \sum_{x} \operatorname{Tr} \left[ \frac{p(x) \rho_x}{|Z|} \sum_{x : h(x) = z} (p(x') \rho_{x'}) + c^{-1} \frac{\rho_E}{|Z|} \right] = \sum_{x} \operatorname{Tr} \left[ |x\rangle \langle x| \otimes \frac{p(x) \rho_x}{|Z|} \sum_{x : h(x) = z} (p(x') \rho_{x'}) + c^{-1} \frac{\rho_E}{|Z|} \right]
\]

\[
= \sum_{x} \operatorname{Tr} \left[ |x\rangle \langle x| \otimes p(x) \rho_x \sum_{x : h(x) = z} (p(x') \rho_{x'}) + c^{-1} \frac{\rho_E}{|Z|} \right] = \sum_{x} \operatorname{Tr} \left[ |x\rangle \langle x| \right]
\]

\[
\leq |x| \langle x| \otimes p(x) \rho_x \sum_{x : h(x) = z} (p(x') \rho_{x'}) + c^{-1} \frac{\rho_E}{|Z|} \]

\[
= \mathcal{E}_h \sum_{x \in \mathcal{X}} \operatorname{Tr} \left[ \sum_{x : h(x) = z} (p(x') \rho_{x'}) + c^{-1} \frac{\rho_E}{|Z|} \right] = c.
\]

Combining (23), (24), and (25) gives

\[
\varepsilon \geq (\delta + \varepsilon) \left( 1 - \frac{1}{|Z|} + \frac{1}{|Z|} e^{-D_1^{1-\varepsilon-\delta}(\rho_{XE} \| 1_X \otimes \rho_E)} \right)^{-1} \]

\[
- c,
\]

which can be translated to

\[
\log |Z| \leq -D_1^{1-\varepsilon-\delta}(\rho_{XE} \| 1_X \otimes \rho_E) + \log(1 + c^{-1})
\]

\[
- \log \left( \frac{\delta - c}{\varepsilon + c} + \frac{1}{|Z|} \right)
\]

\[
\leq -D_1^{1-\varepsilon-\delta}(\rho_{XE} \| 1_X \otimes \rho_E) + \log(1 + c^{-1}) - \log \left( \frac{\delta - c}{\varepsilon + c} \right) - \log \delta
\]

\[
\leq H_1^{1-\varepsilon-\delta}(X | E) + \log \left( \frac{1 + c}{c \delta} \right) + \log \left( \frac{\varepsilon + c}{\delta - c} \right),
\]

where we applied Lemma \ref{lem:convex_slab} in (a). That completes the proof.

\section{IV. Quantum Soft Covering}

In this section, we consider a classical-quantum state 
\( \rho_{XB} = \sum_{x \in \mathcal{X}} p(x) |x\rangle \langle x| \otimes \rho_B^x \). We assume that \( \rho_B \) is invertible and the Hilbert space \( \mathcal{H}_B \) is finite dimensional. Let \( C \) be a random codebook where each codeword \( x \in \mathcal{X} \) is drawn pairwise independently according to distribution \( p_X \). The goal of quantum soft covering is to approximate the state \( \rho_B^x \) using the (random) codebook-induced state \( \frac{1}{|Z|} \sum_{z \in Z} \rho_B^z \). We define the minimal random codebook size for an \( \varepsilon \)-covering as

\[
M^\varepsilon(X : B) := \inf \left\{ M \in \mathbb{N} : |C| \leq M \right\}
\]

\[
\wedge \frac{1}{2} \mathcal{E}_C \rho_{p_X} \left\| \frac{1}{|C|} \sum_{x \in C} \rho_B^x - \rho_B \right\|_1 \leq \varepsilon .
\]

We remark that the definition \( M^\varepsilon \) is for the random codebook sampled from a fixed \( p_X \). If one allows to choose an arbitrary codebook, the setting then becomes the channel resolvability. In this paper, we focus on the setting of soft covering.

The main result of this section is to prove the following one-shot characterization of the operational quantity \( M^\varepsilon(X : B) \).

\textbf{Theorem 2} (One-Shot Characterization for Quantum Soft Covering): Given a classical-quantum state \( \rho_{XB} \), for every \( 0 < \varepsilon < 1 \) and \( 0 < c < \delta < \frac{\varepsilon}{3} \wedge \frac{1-\varepsilon}{2} \), we have

\[
H_1^{1-\varepsilon-2\delta}(X : B) - \log \frac{1 + c}{c \delta} - \log \frac{\varepsilon + c}{\delta - c} \leq \log M^\varepsilon(X : B).
\]
\[
\leq I_1^{1-\varepsilon + 3\delta} (X : B)_\rho + \log \frac{\nu^2}{n}.
\]

Here, \(I_h^\varepsilon\) is the \(\varepsilon\)-hypothesis testing information defined in (5) and \(\nu = |\text{spec}(\rho_B)|\).

**Remark 6:** In Theorem 2, we express the operational quantity \(\log M^\varepsilon(X : B)_\rho\) in terms of the \((1-\varepsilon)\)-hypothesis testing information. However, the lower bound can be improved to \(D_1^{1-\varepsilon-\delta}(\rho_{XB} \| \rho_X \otimes \rho_B)\) and the upper bound can be improved to \(D_1^{1-\varepsilon+\delta}(\rho_{PB} \| \rho_{XB} \otimes \rho_B)\) (both with additional additive logarithmic terms).

In the scenario where the underlying state is identical and independently prepared, i.e. \(\rho_{XB}^\otimes\), the established one-shot characterization, Theorem 2, gives the following second-order asymptotics of the logarithmic random codebook size, \(\log M^\varepsilon(X^n : B^n)_{\rho_{\otimes n}}\), as a function of blocklength \(n\), in which the optimal second-order rate is obtained.

**Proposition 14 (Second-Order Rate for Quantum Soft Covering):** For every classical-quantum state \(\rho_{XB}\), and \(0 < \varepsilon < 1\), we have

\[
\log M^\varepsilon(X^n : B^n)_{\rho_{\otimes n}} = n I(X : B)_\rho - \sqrt{n V(X : B)_\rho} \Phi^{-1}(\varepsilon) + O(\log n),
\]

**Proof:** Since \(|\text{spec}(\rho_B^\otimes)| \leq (n + 1)|H_B|^{-1}\) for \(|H_B|\) being the rank of \(\rho_B\), the additive terms grow order \(O(\log n)\). Then applying Theorem 2, with \(\delta_n = n^{-1/2}\) and \(c_n = \frac{1}{2} n^{-1/2}\), together with the second-order expansion of the hypothesis testing information, Lemma 9, proves our claim.

Moreover, the one-shot characterization can be extended to the moderate deviation regime \([50, 51]\); namely, we derive the optimal rates of the minimal required random codebook size when the error approaches 0 or 1 moderately.

**Proposition 15 (Moderate Deviations for Quantum Soft Covering):** For every classical-quantum state \(\rho_{XB}\) and every moderate sequence \((a_n)_{n \in \mathbb{N}}\) satisfying (6) and \(\varepsilon_n := e^{-n a_n}\), we have

\[
\frac{1}{n} \log M^\varepsilon(X^n : B^n)_{\rho_{\otimes n}} = I(X : B)_\rho + \sqrt{V(X : B)_\rho} a_n + o(a_n);
\]

\[
\frac{1}{n} \log M^1 - \varepsilon(X^n : B^n)_{\rho_{\otimes n}} = I(X : B)_\rho - \sqrt{V(X : B)_\rho} a_n + o(a_n).
\]

**Proof:** We prove the first assertion. For every moderate deviation sequence \((a_n)_{n \in \mathbb{N}}\), we let \(\delta_n = \frac{1}{4} e^{-n a_n/2}\) satisfying (6) and let \(c_n = \frac{1}{2} \delta_n\). Then, \(\varepsilon_n - 2\delta_n\) or \(\varepsilon + 3\delta_n\) can be viewed as \(e^{-n b_n}\) for another moderate deviation sequence \((b_n)_{n \in \mathbb{N}}\) with \(b_n = a_n + (b_n - a_n) = o(a_n)\) (see e.g. [71, §7.2]). \(\frac{1}{n} \log \delta_n = -\frac{1}{2} a_n = o(a_n)\) and \(\frac{1}{n} \log |\text{spec}(\rho_B^\otimes)| = O\left(\frac{\log n}{n}\right) = o(a_n)\). Then, applying Theorem 2 together with Lemma 10 leads to our first claim of the moderate derivation for quantum soft covering. The second line follows similarly.

**Remark 7:** The proof of achievability, i.e. the upper bound of Theorem 2, is inspired by a result of classical soft covering by Hayashi [72, Lemma 2], and hence it can be viewed as a quantum generalization of [72, Lemma 2]. The idea of [72, Lemma 2] was partly extended to the quantum setting in [31, Lemma 9.2]. In particular, the derivation of (9.67) in the proof of [31, Lemma 9.2] is similar to the former part of the proof of Theorem 2. That is, the evaluation of the second term of (9.67) is similar to the first term of (28) of the present paper. However, the next step of (9.68) in the proof of [31, Lemma 9.2] could be not tight (because it involves an inverse of operator). Hence, our contribution is to strengthen the derivations of the second and third term of (28). We remark that while this step is straightforward in the commuting case, in the noncommutative setting we exploit certain properties of the matrix-valued variance to overcome the technical challenges. Such an improvement also appears in the second and third terms of (10) of the present paper for deriving the achievability for privacy amplification with quantum side information.

The proofs of the one-shot achievability (i.e. upper bound) and converse (i.e. lower bound) of Theorem 2 are presented in Section IV-A and IV-B, respectively.

**A. Direct Bound**

We prove the upper bound on \(\log M^\varepsilon(X : B)_\rho\) here.

**Proof:** [Proof of achievability of Theorem 2] Throughout the proof, we use the short notation: \(\rho_x \equiv \rho_B^\otimes\) and \(M \equiv |\mathcal{C}|.\) For every \(x \in \mathcal{C}\), we define a projection \(\Pi_x := \{\rho_{PB} | \rho_{PB} \leq c \rho_B\}\) and its complement \(\Pi_x^c := 1_B - \Pi_x\).

We claim that for any random codebook \(C\) with its codeword independently drawn according to distribution \(p_X\) and for any \(c > 0\), we have

\[
\frac{1}{2} E C \left| \frac{1}{|\mathcal{C}|} \sum_{x \in \mathcal{C}} \rho_{PB}^x - \rho_B \right|_1 \leq \text{Tr} \left[ \rho_{PB} \{ \rho_{PB} | \rho_{PB} > c \rho_B \} \right] + \sqrt{\frac{|\text{spec}(\rho_B)| c}{|\mathcal{C}|}}.
\]

Then, let \(\delta \in (0, \varepsilon)\) and choose

\[
c = \exp \left\{ D_1^{1-\varepsilon-\delta} (\rho_{PB} | \rho_X \otimes \rho_B) + \xi \right\}
\]

for some small \(\xi > 0\). By definition of the \(\varepsilon\)-information spectrum divergence (2),

\[
\text{Tr} \left[ \rho_{XB} \{ \rho_{PB} | \rho_{XB} > c \rho_X \otimes \rho_B \} \right] = \text{Tr} \left[ \rho_{PB} \{ \rho_{PB} | \rho_{PB} > c \rho_X \otimes \rho_B \} \right] < \varepsilon - \delta.
\]

Letting \(|\mathcal{C}| = |\text{spec}(\rho_B)| c \delta^{-2}\), we obtain the \(\varepsilon\)-covering, i.e.

\[
\frac{1}{2} E C \left| \frac{1}{|\mathcal{C}|} \sum_{x \in \mathcal{C}} \rho_{PB}^x - \rho_B \right|_1 \leq \varepsilon,
\]

and then together (27) we have the following upper bound:

\[
\log M^\varepsilon(X : B)_\rho \leq D_1^{1-\varepsilon+\delta} (\rho_{PB} | \rho_X \otimes \rho_B) + \xi + \log |\text{spec}(\rho_B)| - 2 \log \delta
\]

\[
\leq I_1^{1-\varepsilon+3\delta} (X : B)_\rho + \xi + 2 \log |\text{spec}(\rho_B)| - 4 \log \delta.
\]
where we have used Lemma 3 in the last inequality. Since $\xi > 0$ is arbitrary, we take $\xi \to 0$ to obtain the upper bound in (26).

Now, we move on to prove (27). We first prove the case where all $\{\rho_x\}_x$ are invertible. Use triangle inequality of the norm $\| \cdot \|_1$, we obtain

$$
\frac{1}{2} E \left\| \frac{1}{|C|} \sum_{x \in C} \rho_x - \rho_B \right\|_1
= \frac{1}{2} E \left\| \frac{1}{|C|} \sum_{x \in C} \rho_x ( \Pi_x^c + \Pi_x ) \right\|_1
- E \left[ \frac{1}{|C|} \sum_{x \in C} \rho_x ( \Pi_x^c + \Pi_x ) \right] \right\|_1
\leq \frac{1}{2} E \left\| \frac{1}{|C|} \sum_{x \in C} \rho_x \Pi_x^c \right\|_1
+ \frac{1}{2} E \left\| \frac{1}{|C|} \sum_{x \in C} \rho_x \Pi_x \right\|_1
\leq \frac{1}{2} E \left\| \frac{1}{|C|} \sum_{x \in C} \rho_x \Pi_x \right\|_1
= E \left( \frac{1}{|C|} \left[ \sum_{x \in C} \rho_x \Pi_x \right] \right).
$$

(28)

The first term in (28) can be further bounded using triangle inequality again as:

$$
\frac{1}{2} E \left\| \frac{1}{|C|} \sum_{x \in C} \Pi_x^c \rho_x \Pi_x^c \right\|_1
\leq \frac{1}{2} E \left\| \frac{1}{|C|} \sum_{x \in C} \Pi_x^c \rho_x \Pi_x^c \right\|_1
\leq E \left[ \frac{1}{|C|} \left[ \sum_{x \in C} \Pi_x^c \rho_x \Pi_x^c \right] \right].
$$

(29)

To further upper bound this term, we use Jensen’s inequality, Lemma 8, with the pinching map $\mathcal{P}_{\rho_B}$ and the operator concavity of square-root to have

$$
\begin{aligned}
&\text{Tr} \sqrt{E_{x \sim p_X} \left[ \Pi_x \rho_B^2 \Pi_x \right]} \\
&= \text{Tr} \mathcal{P}_{\rho_B} \sqrt{E_{x \sim p_X} \left[ \Pi_x \rho_B^2 \Pi_x \right]} \\
&\leq \text{Tr} \mathcal{P}_{\rho_B} \left[ E_{x \sim p_X} \left[ \Pi_x \rho_B^2 \Pi_x \right] \right] \rho_B^{-2} \\
&\leq \frac{1}{\text{Tr} \left[ E_{x \sim p_X} \left[ \Pi_x \rho_B^2 \Pi_x \right] \right]} \rho_B^{-2},
\end{aligned}
$$

(30)

where (a) follows from Jensen’s inequality, Lemma 8, with the functional $\text{Tr}[\rho_B( \cdot )]$ and the operator concavity of square-root.

Now, since

$$
\Pi_x = \left\{ \mathcal{P}_{\rho_B} [\rho_x] \leq c \rho_B \right\} = \left\{ \rho_B^1 \leq c (\mathcal{P}_{\rho_B} [\rho_x])^{-1} \right\},
$$

we obtain

$$
\Pi_x \rho_B^{-1} \Pi_x \leq c \Pi_x \left( \mathcal{P}_{\rho_B} [\rho_x] \right)^{-1} \Pi_x = c (\mathcal{P}_{\rho_B} [\rho_x])^{-1/2} \Pi_x (\mathcal{P}_{\rho_B} [\rho_x])^{-1/2},
$$

where we used the fact that $\Pi_x$ commutes with $\mathcal{P}_{\rho_B} [\rho_x]$. Then for each $x$,

$$
\text{Tr} \left[ \rho_x^2 \Pi_x \rho_B^{-1} \Pi_x \right]
$$
where in (a) we used the pinching inequality (Lemma 2), i.e.
\[ \rho \leq |\text{spec}(\rho_B)| |\rho_B|, \]
and the operator monotonicity of inversion. Combining (28), (29), (30), and (31) arrives at the desired (27). For the general non-invertible state \( \rho_x \), we define the approximation
\[ \rho^*_x := (1 - \epsilon)\rho_x + \epsilon \omega_B, \]
where \( \omega_B \) is a full rank quantum state and will be determined later. Let \( \rho^*_x B \) be
\[ \rho^*_x B = \sum_{x \in A} p(x)|x\rangle\langle x| \otimes \rho^*_x. \]
It is clear that
\[ \lim_{\epsilon \to 0} \frac{1}{2} E_C \left\| \frac{1}{|C|} \sum_{x \in C} \rho^*_x - \rho_B \right\|_1 \]
\[ = \frac{1}{2} E_C \left\| \frac{1}{|C|} \sum_{x \in C} \rho_x - \rho_B \right\|_1. \]
On bounding the term \( \frac{1}{2} E_C \left\| \frac{1}{|C|} \sum_{x \in C} \rho^*_x - \rho^*_B \right\|_1 \), we follow similar process of our proof where \( \rho_x \) is invertible, but with the projection operator being
\[ \Pi_x = \{ \rho_B | \rho^*_x \leq \epsilon \rho_B \}. \]
Following the same steps leading to (28),
\[ \frac{1}{2} E_C \left\| \frac{1}{|C|} \sum_{x \in C} \rho^*_x - \rho^*_B \right\|_1 \]
\[ \leq \frac{1}{2} E_C \left\| \frac{1}{|C|} \sum_{x \in C} \Pi_x \rho^*_x \Pi_x^* - \rho^*_B \right\|_1 \]
\[ + \frac{1}{2} E_C \left\| \frac{1}{|C|} \sum_{x \in C} \Pi_x \rho_x \Pi_x^* - \rho_B \right\|_1 \]
\[ + \frac{1}{2} E_C \left\| \frac{1}{|C|} \sum_{x \in C} \Pi_x \rho^*_x \Pi_x^* - \rho^*_B \right\|_1. \]
For bounding the first term of (33), suppose \( \rho_B = \sum_{b \in B} p_B(b)|b\rangle\langle b| \) and let \( \omega_x = \sum_{x \in X} q^*_b(b)|b\rangle\langle b|. \) Then,
\[ \lim_{\epsilon \to 0} \{ \rho_B | |\rho^*_x| > \epsilon \rho_B \} \]
\[ = \lim_{\epsilon \to 0} \{(1 - \epsilon)\rho_B(\rho_x) + \epsilon \omega_x > (1 - \epsilon)\rho_B + \epsilon \omega_B \}
\[ = \{ \rho_B(\rho_x) > \rho_B \} + \{ \rho_B = \rho_B \} \cap (\omega_x > \epsilon \rho_B) \}, \]
where the last equality holds for a similar reason as in (19).
For every \( x \) in \( X \), we now choose every \( q^*_b \) as follows: for \( b \in B \), if \( \{ \rho_B| |\rho_B(\rho_x)| > \epsilon \rho_B \} \), then avoid the situation \( q^*_b > \epsilon \rho_B \) while keeping \( q^*_b \) non-zero. We argue that for every \( x \) in \( X \), this can always be done for every \( q^*_b \) because (1) if there exists a \( b \) in \( B \) such that \( \{ b| |\rho_B(\rho_x)| > \epsilon \rho_B \} \), we can take such \( q^*_b \) large enough so that for any other \( b \neq b' \), \( q^*_b \) can be small enough to observe \( q^*_b \leq \epsilon \rho_B \). (2) If for every \( b \) in \( B \), \( \{ b| |\rho_B(\rho_x)| > \epsilon \rho_B \} \), then \( \epsilon = 1 \) and we can set \( q^*_b = \epsilon \rho_B \) for all \( b \). With this choice, \( \{ \rho_B(\rho_x) = \rho_B \} \cap (\omega_x > \epsilon \rho_B) \} = 0 \), and thus
\[ \lim_{\epsilon \to 0} \{ |\rho^*_x| > \epsilon \rho_B \} \leq \{ |\rho_B| > \epsilon \rho_B \}. \]
In the limit of \( \epsilon \to 0 \), the first term of (33) then becomes
\[ \lim_{\epsilon \to 0} \frac{1}{2} E_C \left\| \frac{1}{|C|} \sum_{x \in C} \Pi_x \rho^*_x \Pi_x^* - \rho_B \right\|_1 \]
\[ \leq \frac{1}{2} E_{x \sim \rho_B} \left\| \frac{1}{|C|} \sum_{x \in C} \Pi_x \rho^*_x \Pi_x^* - \rho_B \right\|_1 \]
\[ = \frac{1}{2} E_{x \sim \rho_B} \left\| \frac{1}{|C|} \sum_{x \in C} \rho^*_x - \rho_B \right\|_1 \]
\[ \leq \frac{1}{|\text{spec}(\rho_B)|} \epsilon. \]
The combination of equations (32), (33), (35), and (36) proves (27) by approximation. \( \square \)

B. Converse Bound

We prove the lower bound on \( \log M^{\epsilon}(X:B) \) here. \textit{Proof:} [Proof of converse of Theorem 2] Throughout this proof, we write \( \rho_x \equiv \rho^*_B \) and \( M := |C| \). For every \( 0 < c < 3/4 \) and for any realization of the random codebook \( \mathcal{C} \), we choose the noncommutative quotient
\[ \Pi = \frac{1}{M} \sum_{x \in X} \rho_x. \]
Lemma 6 then implies that
\[ \frac{1}{2} \left\| \frac{1}{M} \sum_{x \in C} \rho^*_x - \rho_B \right\|_1 \]
\[ \geq \text{Tr} \left[ \frac{1}{M} \sum_{x \in C} \rho^*_x \Pi \right] - \text{Tr} [\rho_B \Pi]. \]
Using Lemma 7-d), the second term in (37) can be lower bounded as
\[ - \text{Tr} [\rho_B \Pi] \]
\[ = -c \text{Tr} \left[ \frac{1}{M} \sum_{x \in C} \rho_x \cdot \frac{1}{M} \sum_{x \in C} \rho_x + c^{-1} \rho_B \right] \]
\[ \geq -c \text{Tr} \left[ \frac{1}{M} \sum_{x \in C} \rho_x \right]. \]
\[
\text{(37)} \quad \text{where we apply Lemma 4 in (a) with } \eta = 1 - \varepsilon - \delta, \lambda_1 = M^{-1}, \text{ and } \lambda_2 = 1 + c^{-1} - M^{-1}. \text{ In other words, we get an lower bound on } \log M \text{ as }
\]
\[
\log M \geq D_{\varepsilon} \log (1 + c^{-1} - M^{-1}) + \log \frac{\delta - c}{\varepsilon + c}
\]
\[
\geq D_{\varepsilon} \log (1 + c^{-1}) + \log \frac{\delta - c}{\varepsilon + c}
\]
\[
= D_{\varepsilon} \log (1 + c^{-1}) + \log \frac{\delta - c}{\varepsilon + c}
\]
\[
= D_{\varepsilon} \log (1 + c^{-1}) + \log \frac{\delta - c}{\varepsilon + c}
\]
\[
= I_{\varepsilon} \log (1 + c^{-1}) + \log \frac{\delta - c}{\varepsilon + c}
\]
\[
= I_{\varepsilon} \log (1 + c^{-1}) + \log \frac{\delta - c}{\varepsilon + c}
\]

\[
\text{completing the proof.}
\]

V. Conclusion

The large deviation analysis [17], [71], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84] of privacy amplification against quantum side information and quantum soft covering has been investigated in previous literature [22], [27], [28], [43], [70], and [83], wherein one fixes the rate or the size of \(|Z|\) and \(|C|\) and studies the optimal errors in terms of the trace distance. Also, some moderate deviation analysis [50], [51] were studied for characterizing the minimal trace distance while the rates approach the first-order limits with certain speed [28], [43]. In this paper, we took another perspective—what are the optimal rates when the trace distances are upper bounded by a constant \(\varepsilon \in (0, 1)\). This corresponds to the so-called small error regime [11], [12], [13] or the non-vanishing error regime [14]. We establish the second-order rates for fixed \(\varepsilon \in (0, 1)\) and establish the optimal rates when trace distances vanishes no faster than \(O((1/\sqrt{m}))\).

In light of the duality between smooth min- and max-entropies, the purified distance has been recognized as an appropriate distance measure [16], [26], [47], [48], [70]. Our work suggests that if one considers the trace distance as the performance benchmark without going into the smooth entropy framework [16], [26], [35], [37], [49], [86], the conditional hypothesis testing entropy and the hypothesis testing information\(^7\) are the natural one-shot characterizations.\(^8\) An interesting open question is comparison between the conditional hypothesis testing entropy with the partially trace-distance-smoothed min-entropy [26].

\(^7\)In Ref. [16], it was shown that up to second-order terms, \(D_{\varepsilon}^{1-\varepsilon}\) scales as the relative entropy version of the smooth min-entropy \(D_{\varepsilon}^{1-\varepsilon}\). Although the two quantities are asymptotically equivalent, they arise in very different proof methodologies.

\(^8\)In Ref. [27], Dupuis raised a question: is it possible to use the Rényi-type entropies or information to characterize operational quantities in one-shot information theory? We would like to point out that although there are essentially no differences between the three deviation regimes in the one-shot setting, there are at least two different types of operational quantities of interest; their characterizations might be different. As observed in [27] and our previous works [28], [43], indeed, the Rényi-type quantities are more favorable in characterizing the optimal error given a fixed size or cardinality such as \(|Z|\) and \(|C|\) considered in this paper. On the other hand, if one considers the size or cardinality given a fixed error, the hypothesis-testing-type quantities or the information-spectrum-type quantities might be more direct for characterizations.
We remark that the second-order converses established in the present paper are a kind of ensemble converse. The converses for arbitrary codes (e.g. as the scenario of channel resolvability in [57]) in the quantum case remains open.

ACKNOWLEDGMENT

The authors thank the anonymous referees for many helpful comments, which greatly improves this article. Hao-Chung Cheng would like to thank Kai-Min Chung for his insightful discussions and also thank Marco Tomamichel for helpful comments and pointing out relevant references.

REFERENCES

[1] Y.-C. Shen, L. Gao, and H.-C. Cheng, “Strong converse for privacy amplification against quantum side information,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2022, pp. 1880–1885.

[2] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, Jul. 1948.

[3] J. Wolfowitz, “The coding of messages subject to chance errors,” Illinois J. Math., vol. 1, no. 4, pp. 591–606, Dec. 1957. [Online]. Available: http://projecteuclid.org/euclid.ijm/1255380682

[4] A. Holevo, “Bounds for the quantity of information transmitted by a quantum communication channel,” Problems Inf. Transm., vol. 9, no. 3, pp. 177–183, 1973.

[5] B. Schumacher and M. D. Westmoreland, “Sending classical information via noisy quantum channels,” Phys. Rev. A, Gen. Phys., vol. 56, pp. 131–138, Jul. 1997.

[6] A. S. Holevo, “The capacity of the quantum channel with general signal states,” IEEE Trans. Inf. Theory, vol. 44, no. 1, pp. 269–273, 1998.

[7] T. Ogawa and H. Nagaoka, “Strong converse to the quantum channel coding theorem,” IEEE Trans. Inf. Theory, vol. 45, no. 7, pp. 2486–2489, 1999.

[8] A. Winter, “Coding theorems of quantum information theory,” Ph.D. thesis, Dept. Math., Univ. Bielefeld, 1999.

[9] A. Winter, “Coding theorem and strong converse for quantum channels,” IEEE Trans. Inf. Theory, vol. 45, no. 7, pp. 2481–2485, Nov. 1999.

[10] M. Hayashi and H. Nagaoka, “General formulas for capacity of classical-quantum channels,” IEEE Trans. Inf. Theory, vol. 49, no. 7, pp. 1753–1768, Jul. 2003.

[11] V. Strassen, “Asymptotische abschatzungen in Shannon’s informationstheorie,” in Proc. Third Prague Conf. Inf. Theory, 1962, pp. 689–723.

[12] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” IEEE Trans. Inf. Theory, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[13] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” IEEE Trans. Inf. Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.

[14] V. Y. F. Tan, “Asymptotic estimates in information theory with non-vanishing error probabilities,” Found. Trends Commun. Inf. Theory, vol. 10, no. 4, pp. 1–184, 2014.

[15] K. Li, “Second–order asymptotics for quantum hypothesis testing,” Ann. Statist., vol. 42, no. 1, pp. 171–189, Feb. 2014.

[16] M. Tomamichel and M. Hayashi, “A hierarchy of information quantities for finite block length analysis of quantum tasks,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7693–7710, Nov. 2013.

[17] M. M. Wilde, A. Winter, and D. Yang. “Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy,” Commun. Math. Phys., vol. 331, no. 2, pp. 593–622, Jul. 2014.

[18] M. Tomamichel and V. Y. F. Tan, “Second–order asymptotics for the classical capacity of image-additive quantum channels,” Commun. Math. Phys., vol. 338, no. 1, pp. 103–137, May 2015.

[19] N. Datta, M. Tomamichel, and M. M. Wilde, “On the second–order asymptotics for entanglement-assisted communication,” Quantum Inf. Process., vol. 15, no. 6, pp. 2569–2591, Mar. 2016.

[20] M. Tomamichel, Quantum Information Processing with Finite Resources. Berlin, Germany: Springer, 2016.

[21] M. Tomamichel, M. Berta, and J. M. Renes, “Quantum coding with finite resources,” Nature Commun., vol. 7, no. 1, p. 11419, May 2016.

[22] H. Yuen, “Security of quantum key distribution,” 2016, arXiv:1602.07602.

[23] R. Canetti, “Universally composable security: A new paradigm for cryptographic protocols,” in Proc. 42nd IEEE Symp. Found. Comput. Sci., Oct. 2001, pp. 136–145.

[24] R. Renner and R. König, “Universally composable privacy amplification against quantum adversaries,” in Theory of Cryptography, J. Kilian, Ed. Berlin, Germany: Springer, 2005, pp. 407–425.

[25] D. Unruh, “Universally composable quantum multi-party computation,” in Advances in Cryptology—EUROCRYPT, Berlin, Germany: Springer, 2010, pp. 486–505.

[26] A. Anshu, M. Berta, R. Jain, and M. Tomamichel, “Partially smoothed information measures,” IEEE Trans. Inf. Theory, vol. 66, no. 8, pp. 5022–5036, Aug. 2020.

[27] F. Dupuis, “Privacy amplification and decoupling without smoothing,” 2021, arXiv:2105.05342.

[28] Y.-C. Shen, L. Gao, and H.-C. Cheng, “Strong converse for privacy amplification against quantum side information,” 2022, arXiv:2202.10263.

[29] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[30] M. M. Wilde, Quantum Information Theory. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[31] M. Hayashi, Quantum Information Processing, 2017.

[32] J. Watrous, “The Theory of Quantum Information,” Cambridge, U.K.: Cambridge Univ. Press, Apr. 2018.

[33] S. Khatri and M. M. Wilde, “Principles of quantum communication theory: A modern approach,” 2020, arXiv:2011.04672.

[34] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, “Leftover hashing against quantum side information,” IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 5524–5535, Aug. 2011.

[35] M. Hayashi, “Tight exponential analysis of universally composable privacy amplification and its applications,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7728–7746, Nov. 2013.

[36] C. Portmann and R. Renner, “Cryptographic security of quantum key distribution,” 2014, arXiv:1409.522.

[37] M. Hayashi, “Security analysis of e-almost dual universal hash functions: Smoothing of min entropy versus smoothing of Rényi entropy of order 2,” IEEE Trans. Inf. Theory, vol. 62, no. 6, pp. 3451–3476, Jun. 2016.

[38] M. Hayashi and T. Tsurumaru, “More efficient privacy amplification with less random seeds via dual universal hash function,” IEEE Trans. Inf. Theory, vol. 62, no. 4, pp. 2213–2232, Apr. 2016.

[39] C. H. Bennett, G. Brassard, and J.-M. Robert, “Privacy amplification by public discussion,” SIAM J. Comput., vol. 17, no. 2, pp. 210–229, Apr. 1988.

[40] R. Salzmann and N. Datta, “Total insecurity of communication via strong converse for quantum privacy amplification,” 2022, arXiv:2202.11090.

[41] K. Li, Y. Yao, and M. Hayashi, “Tight exponential analysis for smoothing the max-relative entropy and for quantum privacy amplification,” IEEE Trans. Inf. Theory, vol. 60, no. 3, pp. 1680–1694, Mar. 2014.

[42] R. Ahlswede and A. Winter, “Strong converse for identification via quantum channels,” IEEE Trans. Inf. Theory, vol. 48, no. 3, pp. 569–579, Mar. 2002.

[43] H.-C. Cheng and L. Gao, “Error exponent and strong converse for quantum soft covering,” 2022, arXiv:2202.10995.

[44] W. Matthews and S. Wehner, “Finite blocklength converse bounds for quantum channels,” IEEE Trans. Inf. Theory, vol. 60, no. 11, pp. 7317–7329, Nov. 2014.

[45] R. König, R. Renner, and C. Schaffner, “The operational meaning of min- and max-entropy,” IEEE Trans. Inf. Theory, vol. 55, no. 9, pp. 4337–4347, Sep. 2009.

[46] N. Ciganovic, N. J. Beaudry, and R. Renner, “Smooth max-information measures,” in Information Theor, Berlin, Germany: Springer, 2015, pp. 5022–5036, Aug. 2020.

[47] M. Hayashi and S. Watanabe, “Uniform random number generation from Markov chains: Non-asymptotic and asymptotic analyses,” IEEE Trans. Inf. Theory, vol. 62, no. 4, pp. 1795–1822, Apr. 2016.
