Fidelity-Based Smooth Min-Relative Entropy: Properties and Applications

Theshani Nuradha, *Graduate Student Member, IEEE*, and Mark M. Wilde, *Fellow, IEEE*

***Abstract*— The fidelity-based smooth min-relative entropy is a distinguishability measure that has appeared in a variety of contexts in prior work on quantum information, including resource theories like thermodynamics and coherence. Here we provide a comprehensive study of this quantity. First we prove that it satisfies several basic properties, including the data-processing inequality. We also establish connections between the fidelity-based smooth min-relative entropy and other widely used information-theoretic quantities, including smooth min-relative entropy and smooth sandwiched Rényi relative entropy, of which the sandwiched Rényi relative entropy and smooth max-relative entropy are special cases. After that, we use these connections to establish the second-order asymptotics of the fidelity-based smooth min-relative entropy and all smooth sandwiched Rényi relative entropies, finding that the first-order term is the quantum relative entropy and the second-order term involves the quantum relative entropy variance. Utilizing the properties derived, we also show how the fidelity-based smooth min-relative entropy provides one-shot bounds for operational tasks in general resource theories in which the target state is mixed, with a particular example being randomness distillation. The above observations then lead to second-order expansions of the upper bounds on distillable randomness, as well as the precise second-order asymptotics of the distillable randomness of particular classical–quantum states. Finally, we establish semi-definite programs for smooth max-relative entropy and smooth conditional min-entropy, as well as a bilinear program for the fidelity-based smooth min-relative entropy, which we subsequently use to explore the tightness of a bound relating the last to the first.**

***Index Terms*— Fidelity based smoothing, quantum resource theories, randomness distillation, second-order asymptotics, smoothed Rényi divergences, smooth min-relative entropy.**

I. Introduction

A. Background

DISTINGUISHABILITY plays a fundamental role across all fields of sciences. The core toolbox in this regard involves distinguishability measures. In quantum information theory, these distinguishability measures then lead to information measures including mutual information and conditional entropy, as well as entanglement measures (see [1, Chapters 4 and 5] for a review). Furthermore, they also appear in resource theories as conversion rates [2].

The min-relative entropy is one such distinguishability measure [3], defined for a pure state $\psi := |\psi\rangle\langle\psi|$ and a positive semi-definite operator σ as

$$D_{\min}(\psi\|\sigma) := -\log_2 \operatorname{Tr}[\psi\sigma] \tag{1}$$
$$= -\log_2 F(\psi,\sigma). \tag{2}$$

In (2) above, we made use of the fidelity [4], defined generally for two positive semi-definite operators ω and τ as

$$F(\omega,\tau) := \left\|\sqrt{\omega}\sqrt{\tau}\right\|_1^2 . \tag{3}$$

In the case that σ is a state, we can interpret the expression $\operatorname{Tr}[\psi\sigma]$ in (1) as the probability that the first measurement outcome occurs when performing the measurement $\{\psi, I-\psi\}$ on the state σ. Alternatively, we can interpret the expression $F(\psi,\sigma)$ in (2) as the fidelity between the states ψ and σ. Thus, in the first case, we are interpreting ψ as a measurement operator, and in the second case, we are interpreting ψ as a state.

Given the above, there are at least two ways of generalizing the min-relative entropy when ρ is a general state. The first approach, originally introduced in [3] as the min-relative entropy, defines it as

$$D_{\min}(\rho\|\sigma) := -\log_2 \operatorname{Tr}[\Pi_\rho\sigma], \tag{4}$$

where Π_ρ denotes the projection onto the support of ρ. Clearly, this definition generalizes the expression in (1), interpreting Π_ρ as a measurement operator. As discussed in [3], this interpretation is directly linked with the operational meaning of the min-relative entropy in asymmetric hypothesis testing, as the minimum Type II error exponent if the Type I error probability is constrained to be equal to zero. This quantity has been further interpreted in a resource-theoretic manner as the maximum number of exact bits of asymmetric distinguishability that can be distilled from the pair (ρ,σ) [5]. Given the strong link between hypothesis testing and information theory [6], the min-relative entropy finds further use as the basic quantity underlying optimal rates at which zero-error distillation is possible [5], [7].

The second generalization of min-relative entropy to a general state ρ employs the formula in (2), and is defined as follows [8]:

$$D_{\min,F}(\rho\|\sigma) := -\log_2 F(\rho,\sigma). \tag{5}$$

Manuscript received 18 May 2023; revised 27 December 2023; accepted 9 March 2024. Date of publication 18 March 2024; date of current version 21 May 2024. This work was supported by the National Science Foundation under Grant 2315398. *(Corresponding author: Theshani Nuradha.)*

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850 USA (e-mail: pt388@cornell.edu; wilde@cornell.edu).

Communicated by G. Smith, Associate Editor for Quantum.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2024.3378590.

Digital Object Identifier 10.1109/TIT.2024.3378590

In the above definition and throughout, we use the extra subscript F to indicate that this generalization is based on fidelity, and we refer to it as the F-min-relative entropy. This quantity is also equal to the sandwiched Rényi relative entropy of order 1/2 [9], [10]. It is not known to have an operational meaning in hypothesis testing; however, it has appeared in a variety of previous works in quantum information theory [8], [11], [12], [13], [14].

In realistic experimental scenarios, it is pertinent to allow for approximations in terms of a smoothing parameter [15], [16], which characterizes the error that can occur in an experiment. We can then consider smoothed versions of the quantities in (4) and (5).

Let us first consider smoothing the quantity in (4). The approach employed finds its roots naturally in asymmetric hypothesis testing, given the operational scenario discussed above. With this in mind, if we relax the aforementioned Type I error probability constraint, such that it is allowed to be larger than zero, then we arrive at the smooth min-relative entropy with smoothing parameter $\varepsilon \in [0, 1]$ [17], [18], [19]:

$$D^{\varepsilon}_{\min}(\rho\|\sigma) := -\log_2 \inf_{0\leq\Lambda\leq I} \{\mathrm{Tr}[\Lambda\sigma] : \mathrm{Tr}[\Lambda\rho] \geq 1 - \varepsilon\}, \quad (6)$$

with $\{\Lambda, I - \Lambda\}$ being the measurement that distinguishes between ρ and σ. This quantity is referred to as the smooth min-relative entropy in [5], [7], [20], and [21] and as the hypothesis testing relative entropy in [19] and many other papers, including [1], [8], [22], [23], [24], [25], and [26]. Considering that [5], [8]

$$\lim_{\varepsilon\to 0} D^{\varepsilon}_{\min}(\rho\|\sigma) = D_{\min}(\rho\|\sigma), \quad (7)$$

it is clear that $D^{\varepsilon}_{\min}(\rho\|\sigma)$ is a smoothed version of $D_{\min}(\rho\|\sigma)$.

Let us now consider smoothing the quantity in (5). Since ρ is a state, the idea when smoothing is to search for a nearby subnormalized state $\widetilde{\rho}$, such that it satisfies $F(\widetilde{\rho}, \rho) \geq 1 - \varepsilon$ for a fixed smoothing parameter $\varepsilon \in [0, 1]$, and then replace ρ with $\widetilde{\rho}$ when comparing with σ. This reasoning naturally leads to the fidelity-based smooth min-relative entropy:

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) := -\log_2 \inf_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \{F(\widetilde{\rho}, \sigma) : F(\widetilde{\rho}, \rho) \geq 1 - \varepsilon\}, \quad (8)$$

where $\mathcal{D}_{\leq}$ denotes the set of sub-normalized states (see Definition 1, as well as Remark 2 for the choice of sub-normalized states). In what follows, we refer to it more simply as the smooth F-min-relative entropy. This quantity has been considered in several prior works [8], [11], [12], [13], [14]. It is interesting to compare the expressions involved in (6) and (8), where we observe that the main difference is that $D^{\varepsilon}_{\min}$ compares ρ and σ to a measurement operator Λ via a trace overlap, whereas $D^{\varepsilon}_{\min,F}$ compares ρ and σ to a subnormalized state $\widetilde{\rho}$ via the fidelity.

By building on the recent observations of [27], one contribution of the present paper is that the smooth F-min-relative entropy finds use in operational tasks such as randomness distillation. In this task, the goal is to distill a state that is close in fidelity to a maximally classically correlated state, which is a mixed state. As such, the approach to smoothing taken in (8) is more relevant in this scenario than that in (6) and can be used to obtain upper bounds on the one-shot distillable randomness of a bipartite state. More generally, and as discussed in [27], we suspect that the ideas put forward here will find use in quantum resource transformations in which the target state is a mixed state, and we provide some evidence in Section VI-F that this is the case.

More broadly, the main goal of the present paper is to provide a comprehensive study of the fidelity-based smooth min-relative entropy in (8), which, as indicated above, could be useful for understanding the fundamental limits of resource transformations in general resource theories.

B. Contributions

In this paper, we first derive several properties of the fidelity-based smooth min-relative entropy. In particular, we prove that it satisfies data processing (Theorem 1), as well as scaling, super-additivity, monotonicity, etc. (Theorem 2). We note here that the data-processing inequality was already established in [14, Theorem 3], but here we provide an independent proof. Then we proceed to establish its connections with other quantum information-theoretic quantities including sandwiched Rényi relative entropy and its smooth variants, smooth max-relative entropy, and smooth min-relative entropy.

Next, with the assistance of the derived connections, we provide a second-order asymptotic analysis for the smooth F-min-relative entropy (Theorem 4). There we find that the first-order term is the quantum relative entropy and the second-order term involves the quantum relative entropy variance. In addition, we derive the second-order behaviour of the smooth sandwiched Rényi relative entropy (Corollary 4). This corollary indicates that, in the asymptotic i.i.d. setting and up to the second order, there is no difference between all of the smooth sandwiched Rényi relative entropies for all $\alpha > 1$: they are all equivalent to the smooth max-relative entropy in this setting. Similarly, in the asymptotic i.i.d. setting and up to the second order, there is no difference between all of them for $\alpha \in [1/2, 1)$: they are all equivalent to the smooth min-relative entropy in this setting.

Furthermore, we show how the smooth F-min-relative entropy provides one-shot bounds for operational tasks in general resource-theoretic settings, with a particular analysis focusing on randomness distillation from bipartite states. We derive second-order expansions of the upper bounds on the LOCC-assisted distillable randomness (Theorem 6), as well as the precise second-order asymptotics of the distillable randomness of particular classical–quantum states (Theorem 7).

Moreover, we provide a method to compute the smooth F-min-relative entropy by means of a bilinear program (Proposition 8). We also provide semi-definite programs (SDPs) for smooth max-relative entropy and smooth conditional min-entropy (Propositions 9 and 10), which may be of independent interest.

C. Organization

The rest of our paper is organized as follows. In Section II, we introduce notation and preliminaries. The focus of Section III is on deriving some basic properties of the smooth F-min-relative entropy, including data processing. Connections to other quantum information-theoretic quantities are established in Section IV. In Section V, we study the second-order asymptotics of the smooth F-min-relative entropy. We explore how smooth F-min-relative entropy provides bounds in operational tasks related to general resource theories in Section VI. In Section VII we provide methods to compute the smooth F-min-relative entropy and related quantities, including the smooth max-relative entropy. Finally, Section VIII provides concluding remarks and future directions.

II. Preliminaries

A. Basic Concepts and Notation

We begin by reviewing basic concepts from quantum information theory and refer the reader to [1] for more details. A quantum system R is identified with a finite-dimensional Hilbert space $\mathcal{H}_R$. We denote the set of linear operators acting on $\mathcal{H}_R$ by $\mathcal{L}(\mathcal{H}_R)$. The support of a linear operator $X \in \mathcal{L}(\mathcal{H}_R)$ is defined to be the orthogonal complement of its kernel, and we denote it by $\operatorname{supp}(X)$. Let $\mathrm{T}(X)$ denote the transpose of X. The partial transpose of $C_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ on the system A is represented as $\mathrm{T}_A(C_{AB})$. Let $\mathrm{Tr}[C_{AB}]$ denote the trace of C_{AB}, and let $\mathrm{Tr}_A[C_{AB}]$ denote the partial trace of C over the system A. We use the standard notation $C_A \equiv \mathrm{Tr}_B[C_{AB}]$ and $C_B \equiv \mathrm{Tr}_A[C_{AB}]$ to denote the marginals of C_{AB}. The trace norm of an operator B is defined as $\|B\|_1 := \mathrm{Tr}[\sqrt{B^\dagger B}]$. For Hermitian operators A and B, the notation $A \geq B$ indicates that $A-B$ is a positive semi-definite (PSD) operator, while $A > B$ indicates that $A-B$ is a positive definite operator.

A quantum state $\rho_R \in \mathcal{L}(\mathcal{H}_R)$ of system R is a PSD, unit-trace operator acting on $\mathcal{H}_R$. We denote the set of all density operators acting on $\mathcal{H}_R$ as $\mathcal{D}(\mathcal{H}_R)$ (we also refer to the set of density operators by $\mathcal{D}$ when there is no ambiguity regarding the underlying Hilbert space). A rank-one state ρ_R is called pure, and in this case there exists a state vector $|\psi\rangle \in \mathcal{H}_R$ such that $\rho_R = |\psi\rangle\!\langle\psi|$. Otherwise, ρ_R is called a mixed state. By the spectral decomposition theorem, every state can be written as a convex combination of pure, orthogonal states. A quantum channel $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ is a linear, completely positive and trace-preserving (CPTP) map from $\mathcal{L}(\mathcal{H}_A)$ to $\mathcal{L}(\mathcal{H}_B)$. We denote the Hilbert–Schmidt adjoint of $\mathcal{N}$ by $\mathcal{N}^\dagger$. A measurement of a quantum system R is described by a positive operator-valued measure (POVM) $\{M_y\}_{y\in\mathcal{Y}}$, which is defined to be a collection of PSD operators satisfying $\sum_{y\in\mathcal{Y}} M_y = I_R$, where I_R is the identity operator and $\mathcal{Y}$ is a finite alphabet. The Born rule dictates that, when applying the above POVM to a state ρ, the probability of observing the outcome y is given by $\mathrm{Tr}[M_y\rho]$.

B. Divergences

First, let us recall the definition of the fidelity-based smooth min-relative entropy, which is the main distinguishability measure of interest in our paper.

Definition 1 (Fidelity-Based Smooth Min-Relative Entropy): Fix $\varepsilon \in [0,1]$. The fidelity-based smooth min-relative entropy of a state ρ and a PSD operator σ is defined as

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) := -\log_2 \inf_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \left\{ F(\widetilde{\rho},\sigma) : F(\widetilde{\rho},\rho) \geq 1-\varepsilon \right\}, \tag{9}$$

where the fidelity of PSD operators ω and τ is defined in (3) and $\mathcal{D}_{\leq}$ denotes the set of subnormalized states; i.e.,

$$\mathcal{D}_{\leq} := \{\omega : \omega \geq 0, \mathrm{Tr}[\omega] \leq 1\}. \tag{10}$$

Hereafter, we simply abbreviate this quantity as the smooth F-min-relative entropy. Recalling the definition in [8, Eq. (8)], observe that when ρ is a state and σ is a PSD operator, there is no difference between Definition 1 and the definition given in [8, Eq. (8)].

We call a distinguishability measure $\boldsymbol{D}(\cdot\|\cdot)$ a generalized divergence [28] if it satisfies the data-processing inequality; i.e., for every channel $\mathcal{N}$, state ρ, and PSD operator σ,

$$\boldsymbol{D}(\rho\|\sigma) \geq \boldsymbol{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{11}$$

Fix $\alpha \in (0,1) \cup (1,\infty)$. The sandwiched Rényi relative entropy of a state ρ and a PSD operator σ is defined as [9], [10]

$$\widetilde{D}_\alpha(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log_2 \widetilde{Q}_\alpha(\rho\|\sigma) & \text{if } \alpha \in (0,1), \text{or} \\ & \alpha \in (1,\infty),\ \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma), \\ +\infty & \text{otherwise,} \end{cases} \tag{12}$$

where

$$\widetilde{Q}_\alpha(\rho\|\sigma) := \mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right]. \tag{13}$$

It is a generalized divergence for $\alpha \in [1/2,1) \cup (1,\infty)$ [29] (see also [30], [31]), and satisfies the following α-monotonicity property [9]:

$$0 < \alpha \leq \beta \quad \Rightarrow \quad \widetilde{D}_\alpha(\rho\|\sigma) \leq \widetilde{D}_\beta(\rho\|\sigma). \tag{14}$$

For $\alpha = 1/2$, observe that

$$\widetilde{D}_{1/2}(\rho\|\sigma) = -\log_2 F(\rho,\sigma) = D_{\min,F}(\rho\|\sigma). \tag{15}$$

The special case of $\alpha \to 1$ reduces to the quantum relative entropy [9], [10]:

$$\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma), \tag{16}$$

the latter defined as [32]

$$D(\rho\|\sigma) := \mathrm{Tr}[\rho(\log_2\rho - \log_2\sigma)] \tag{17}$$

if $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ and as $+\infty$ otherwise. The relative entropy variance $V(\rho\|\sigma)$ is defined as [22], [23]

$$V(\rho\|\sigma) \coloneqq \operatorname{Tr}\left[\rho\left(\log_2\rho - \log_2\sigma\right)^2\right] - \left(D(\rho\|\sigma)\right)^2. \quad (18)$$

The Petz–Rényi relative entropy is defined for $\alpha \in (0,1) \cup (1,\infty)$, a state ρ, and a PSD operator σ as [33], [34]

$$D_\alpha(\rho\|\sigma) \coloneqq \begin{cases} \frac{1}{\alpha-1}\log_2 Q_\alpha(\rho\|\sigma) & \text{if } \alpha \in (0,1), \text{or} \\ & \alpha \in (1,\infty),\ \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma), \\ +\infty & \text{otherwise,} \end{cases} \quad (19)$$

where

$$Q_\alpha(\rho\|\sigma) \coloneqq \operatorname{Tr}[\rho^\alpha \sigma^{1-\alpha}]. \quad (20)$$

It is a generalized divergence for $\alpha \in (0,1) \cup (1,2]$ [33], [34].

The max-relative entropy of a state ρ and a PSD operator σ is defined as [3]

$$D_{\max}(\rho\|\sigma) \coloneqq \log_2 \inf_{\lambda \geq 0} \{\lambda : \rho \leq \lambda\sigma\}. \quad (21)$$

It is known from [9] that

$$D_{\max}(\rho\|\sigma) = \lim_{\alpha\to\infty} \widetilde{D}_\alpha(\rho\|\sigma). \quad (22)$$

The smooth max-relative entropy of a state ρ and a PSD operator σ is defined for $\varepsilon \in [0,1]$ as [3] (see also [35])

$$D^\varepsilon_{\max}(\rho\|\sigma) \coloneqq \inf_{\widetilde{\rho} \in \mathcal{D}_\leq} \{D_{\max}(\widetilde{\rho}\|\sigma) : F(\widetilde{\rho},\rho) \geq 1-\varepsilon\}. \quad (23)$$

We also define the following variant

$$\widehat{D}^\varepsilon_{\max}(\rho\|\sigma) \coloneqq \inf_{\widetilde{\rho} \in \mathcal{D}} \{D_{\max}(\widetilde{\rho}\|\sigma) : F(\widetilde{\rho},\rho) \geq 1-\varepsilon\}, \quad (24)$$

and note that

$$D^\varepsilon_{\max}(\rho\|\sigma) \leq \widehat{D}^\varepsilon_{\max}(\rho\|\sigma), \quad (25)$$

which follows since $\mathcal{D} \subseteq \mathcal{D}_\leq$.

III. Properties of Smooth F-Min-Relative Entropy

In this section, we derive some basic properties satisfied by the smooth F-min-relative entropy, including data processing, scaling, super-additivity, monotonicity, etc. Then in subsequent sections, we utilize these properties to obtain bounds on operational quantities arising in general resource-theoretic settings, in particular on the net rate of the distillable randomness of a bipartite state.

Before deriving these properties, let us first observe that we can always restrict the constraint in the definition of $D^\varepsilon_{\min,F}$ to be an equality constraint, by following the same line of reasoning from [36, Appendix B].

Remark 1 (Inequality Constraint in the Definition of Smooth F-Min-Relative Entropy): For $\varepsilon \in [0,1)$, the smooth F-min-relative entropy in (9) can be rewritten as

$$D^\varepsilon_{\min,F}(\rho\|\sigma) = -\log_2 \inf_{\widetilde{\rho} \in \mathcal{D}_\leq} \{F(\widetilde{\rho},\sigma) : F(\widetilde{\rho},\rho) = 1-\varepsilon\}. \quad (26)$$

Indeed, if $\widetilde{\rho}$ is such that $F(\widetilde{\rho},\rho) > 1-\varepsilon$, then we can choose a positive constant $c = (1-\varepsilon)/F(\widetilde{\rho},\rho) \in (0,1)$ such that $F(\rho',\rho) = 1-\varepsilon$, where $\rho' = c\widetilde{\rho}$ and $\rho' \in \mathcal{D}_\leq$. Furthermore, we also have that $F(\widetilde{\rho},\sigma) > F(\rho',\sigma)$, so that the objective function only decreases under this change.

A. Data Processing

In this section, we prove the data-processing inequality for the smooth F-min relative entropy. As noted previously, this finding was already established in [14, Theorem 3], but here we provide an independent proof. Before establishing the data-processing inequality, we prove the unitary invariance of the smooth F-min-relative entropy, which assists in proving the data-processing inequality.

Lemma 1 (Unitary Invariance): The smooth F-min-relative entropy $D^\varepsilon_{\min,F}$ is invariant under the action of a unitary channel $\mathcal{U}$, i.e.,

$$D^\varepsilon_{\min,F}(\rho\|\sigma) = D^\varepsilon_{\min,F}(\mathcal{U}(\rho)\|\mathcal{U}(\sigma)). \quad (27)$$

for all $\varepsilon \in [0,1)$, every state ρ, and PSD operator σ.

Proof: Let $\widetilde{\rho} \in \mathcal{D}_\leq$ satisfy $F(\widetilde{\rho},\rho) \geq 1-\varepsilon$. Then, by the unitary invariance of fidelity, we conclude that

$$F(\mathcal{U}(\widetilde{\rho}),\mathcal{U}(\rho)) = F(\widetilde{\rho},\rho). \quad (28)$$

Thus, $F(\mathcal{U}(\widetilde{\rho}),\mathcal{U}(\rho)) \geq 1-\varepsilon$, and

$$\begin{aligned} -\log_2 F(\widetilde{\rho},\sigma) &= -\log_2 F(\mathcal{U}(\widetilde{\rho}),\mathcal{U}(\sigma)) && (29)\\ &\leq D^\varepsilon_{\min,F}(\mathcal{U}(\rho)\|\mathcal{U}(\sigma)). && (30)\end{aligned}$$

Since the inequality holds for every $\widetilde{\rho} \in \mathcal{D}_\leq$ satisfying $F(\widetilde{\rho},\rho) \geq 1-\varepsilon$, we conclude that

$$D^\varepsilon_{\min,F}(\rho\|\sigma) \leq D^\varepsilon_{\min,F}(\mathcal{U}(\rho)\|\mathcal{U}(\sigma)). \quad (31)$$

To prove the opposite inequality, let $\widehat{\rho} \in \mathcal{D}_\leq$ satisfy $F(\widehat{\rho},\mathcal{U}(\rho)) \geq 1-\varepsilon$. Then, by unitary invariance of fidelity,

$$\begin{aligned} F(\widehat{\rho},\mathcal{U}(\rho)) &= F(\mathcal{U}^\dagger(\widehat{\rho}),(\mathcal{U}^\dagger\circ\mathcal{U})(\rho)) && (32)\\ &= F(\mathcal{U}^\dagger(\widehat{\rho}),\rho). && (33)\end{aligned}$$

Thus, $F(\mathcal{U}^\dagger(\widehat{\rho}),\rho) \geq 1-\varepsilon$, and

$$\begin{aligned} -\log_2 F(\widehat{\rho},\mathcal{U}(\sigma)) &= -\log_2 F(\mathcal{U}^\dagger(\widehat{\rho}),(\mathcal{U}^\dagger\circ\mathcal{U})(\sigma)) && (34)\\ &= -\log_2 F(\mathcal{U}^\dagger(\widehat{\rho}),\sigma) && (35)\\ &\leq D^\varepsilon_{\min,F}(\rho\|\sigma). && (36)\end{aligned}$$

Since the inequality holds for every $\widehat{\rho} \in \mathcal{D}_\leq$ satisfying $F(\widehat{\rho},\mathcal{U}(\rho)) \geq 1-\varepsilon$, we conclude that

$$D^\varepsilon_{\min,F}(\mathcal{U}(\rho)\|\mathcal{U}(\sigma)) \leq D^\varepsilon_{\min,F}(\rho\|\sigma). \quad (37)$$

Then (27) follows from (31) and (37). ■

Now, we are ready to present and prove the data-processing inequality for the smooth F-min-relative entropy.

Theorem 1: The smooth F-min-relative entropy obeys the data-processing inequality:

$$D^\varepsilon_{\min,F}(\rho\|\sigma) \geq D^\varepsilon_{\min,F}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \quad (38)$$

for all $\varepsilon \in [0,1)$, every state ρ, PSD operator σ, and channel $\mathcal{N}$.

Proof: First, let us show that data processing holds under the partial trace channel. That is, for every bipartite state ρ_{AB} and bipartite positive semi-definite operator σ_{AB}:

$$D^{\varepsilon}_{\min,F}(\rho_{AB}\|\sigma_{AB}) \geq D^{\varepsilon}_{\min,F}(\rho_A\|\sigma_A). \tag{39}$$

To this end, let $\widetilde{\rho}_A \in \mathcal{D}_{\leq}$ satisfy $F(\widetilde{\rho}_A, \rho_A) \geq 1-\varepsilon$. Then, by Uhlmann's theorem [4] there exists $\widetilde{\rho}_{AB} \in \mathcal{D}_{\leq}$ such that $F(\widetilde{\rho}_{AB}, \rho_{AB}) = F(\widetilde{\rho}_A, \rho_A)$. Thus, $F(\widetilde{\rho}_{AB}, \rho_{AB}) \geq 1-\varepsilon$, and

$$\begin{aligned} -\log_2 F(\widetilde{\rho}_A, \sigma_A) &\leq -\log_2 F(\widetilde{\rho}_{AB}, \sigma_{AB}) && (40)\\ &\leq D^{\varepsilon}_{\min,F}(\rho_{AB}\|\sigma_{AB}). && (41)\end{aligned}$$

The first inequality follows from the data-processing inequality for fidelity, and the second follows from the definition of $D^{\varepsilon}_{\min,F}$. Since this inequality holds for every $\widetilde{\rho}_A \in \mathcal{D}_{\leq}$ satisfying $F(\widetilde{\rho}_A, \rho_A) \geq 1-\varepsilon$, we conclude (39).

Note that the same argument used for $D^{\varepsilon}_{\min,F}(\rho\|\sigma) \leq D^{\varepsilon}_{\min,F}(\mathcal{U}(\rho)\|\mathcal{U}(\sigma))$ (Lemma 1) holds whenever $\mathcal{U}$ is an isometric channel, due to the isometric invariance of fidelity. Since the embedding $\omega \to \omega \otimes |0\rangle\!\langle 0|$ is an isometric channel, we conclude that

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \leq D^{\varepsilon}_{\min,F}(\rho \otimes |0\rangle\!\langle 0| \| \sigma \otimes |0\rangle\!\langle 0|). \tag{42}$$

Now let $\widetilde{\rho}_{SE} \in \mathcal{D}_{\leq}$ satisfy $F(\widetilde{\rho}_{SE}, \rho_S \otimes |0\rangle\!\langle 0|_E) \geq 1-\varepsilon$. By applying Lemma 2, we conclude that

$$F(\widetilde{\rho}_{SE}, \rho_S \otimes |0\rangle\!\langle 0|_E) = F(\widetilde{\rho}^0_S, \rho_S), \tag{43}$$

where $\widetilde{\rho}^0_S := \langle 0|_E \widetilde{\rho}_{SE} |0\rangle_E$. Thus, $\widetilde{\rho}^0_S$ is a subnormalized state that satisfies

$$F\left(\widetilde{\rho}^0_S, \rho_S\right) = F(\widetilde{\rho}_{SE}, \rho_S \otimes |0\rangle\!\langle 0|) \geq 1-\varepsilon. \tag{44}$$

Furthermore, again applying Lemma 2, we conclude that

$$F(\widetilde{\rho}_{SE}, \sigma_S \otimes |0\rangle\!\langle 0|_E) = F(\widetilde{\rho}^0_S, \sigma_S). \tag{45}$$

Thus, it follows that

$$\begin{aligned} -\log_2 F(\widetilde{\rho}_{SE}, \sigma_S \otimes |0\rangle\!\langle 0|_E) &= -\log_2 F\left(\widetilde{\rho}^0, \sigma\right) && (46)\\ &\leq D^{\varepsilon}_{\min,F}(\rho\|\sigma). && (47)\end{aligned}$$

Since the inequality holds for every $\widetilde{\rho}_{SE} \in \mathcal{D}_{\leq}$ satisfying $F(\widetilde{\rho}_{SE}, \rho \otimes |0\rangle\!\langle 0|) \geq 1-\varepsilon$, we conclude that

$$D^{\varepsilon}_{\min,F}(\rho \otimes |0\rangle\!\langle 0| \| \sigma \otimes |0\rangle\!\langle 0|) \leq D^{\varepsilon}_{\min,F}(\rho\|\sigma). \tag{48}$$

Putting together (42) and (48), we find that

$$D^{\varepsilon}_{\min,F}(\rho \otimes |0\rangle\!\langle 0| \| \sigma \otimes |0\rangle\!\langle 0|) = D^{\varepsilon}_{\min,F}(\rho\|\sigma). \tag{49}$$

Finally, since, by the Stinespring dilation theorem [37] (see also [1]), every channel can be realized in terms of

1) the map $\omega \to \omega \otimes |0\rangle\!\langle 0|$,
2) a unitary channel, and
3) a partial trace,

we conclude the desired inequality in (38), after putting together (27), (39), and (49). ∎

Remark 2 (On the Choice of Subnormalized States): It is only this last step (i.e., proving (48)) in which we required the assumption of smoothing over subnormalized states in the definition of the smooth F-min-relative entropy, rather than smoothing over normalized states.

We used the following lemma in the proof of Theorem 1.

Lemma 2: Let ω_{SE} be a bipartite PSD operator, and let σ_S be a PSD operator. Then

$$F(\omega_{SE}, \sigma_S \otimes |0\rangle\!\langle 0|_E) = F(\omega^0_S, \sigma_S), \tag{50}$$

where $\omega^0_S := \langle 0|_E \omega_{SE} |0\rangle_E$.

Proof: Consider that

$$\begin{aligned} &\sqrt{F}(\omega_{SE}, \sigma_S \otimes |0\rangle\!\langle 0|_E) \\ &= \operatorname{Tr}\left[\sqrt{\sqrt{\sigma_S \otimes |0\rangle\!\langle 0|_E}\,\omega_{SE}\sqrt{\sigma_S \otimes |0\rangle\!\langle 0|_E}}\right] && (51)\\ &= \operatorname{Tr}\left[\sqrt{\sqrt{\sigma_S} \otimes |0\rangle\!\langle 0|_E\,\omega_{SE}\sqrt{\sigma_S} \otimes |0\rangle\!\langle 0|_E}\right] && (52)\\ &= \operatorname{Tr}\left[\sqrt{\sqrt{\sigma_S}\langle 0|_E\omega_{SE}|0\rangle_E\sqrt{\sigma_S} \otimes |0\rangle\!\langle 0|_E}\right] && (53)\\ &= \operatorname{Tr}\left[\sqrt{\sqrt{\sigma_S}\langle 0|_E\omega_{SE}|0\rangle_E\sqrt{\sigma_S}} \otimes |0\rangle\!\langle 0|_E\right] && (54)\\ &= \operatorname{Tr}\left[\sqrt{\sqrt{\sigma_S}\langle 0|_E\omega_{SE}|0\rangle_E\sqrt{\sigma_S}}\right] && (55)\\ &= \sqrt{F}(\langle 0|_E\omega_{SE}|0\rangle_E, \sigma_S) && (56)\\ &= \sqrt{F}\left(\omega^0_S, \sigma_S\right), && (57)\end{aligned}$$

concluding the proof. ∎

With Theorem 1 in hand, it thus follows that the smooth F-min-relative entropy is a particular kind of generalized divergence (here recall (11)). Hence it possesses some basic properties satisfied by generalized divergences, as listed next.

Corollary 1: Let ρ be a state and σ a PSD operator. The smooth F-min-relative entropy satisfies the following properties:

1) Isometric invariance: For every isometry V,

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) = D^{\varepsilon}_{\min,F}(V\rho V^{\dagger}\|V\sigma V^{\dagger}). \tag{58}$$

2) Stability: For every state τ,

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) = D^{\varepsilon}_{\min,F}(\rho \otimes \tau\|\sigma \otimes \tau). \tag{59}$$

The proof directly follows from [1, Proposition 7.14], along with Theorem 1.

B. Other Properties

In this section, we derive some other properties of the smooth F-min-relative entropy.

Theorem 2: For all $\varepsilon \in [0, 1)$, every state ρ, and PSD operator σ, the smooth F-min-relative entropy $D^{\varepsilon}_{\min,F}(\rho\|\sigma)$ satisfies the following properties:

1) Scaling: For $c > 0$, we have

$$D^{\varepsilon}_{\min,F}(\rho\|c\sigma) = D^{\varepsilon}_{\min,F}(\rho\|\sigma) + \log_2\left(\frac{1}{c}\right). \tag{60}$$

2) Monotonicity: For $\varepsilon \leq \varepsilon' \in [0, 1)$,

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \leq D^{\varepsilon'}_{\min,F}(\rho\|\sigma). \tag{61}$$

3) Superadditivity: For $\varepsilon_1, \varepsilon_2 \in [0,1)$, states ρ_1 and ρ_2, and PSD operators σ_1 and σ_2, we have

$$D^{\varepsilon_1}_{\min,F}(\rho_1\|\sigma_1) + D^{\varepsilon_2}_{\min,F}(\rho_2\|\sigma_2) \leq D^{\varepsilon'}_{\min,F}(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2), \quad (62)$$

where $\varepsilon' := \varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2$. By monotonicity, we can also choose $\varepsilon' = \varepsilon_1 + \varepsilon_2$.

4) Convexity in the second argument: Let $\{\sigma_i\}_i$ be a set of PSD operators, and let $\{p_i\}_i$ be a probability distribution. Then

$$D^{\varepsilon}_{\min,F}(\rho\|\overline{\sigma}) \leq \sum_i p_i D^{\varepsilon}_{\min,F}(\rho\|\sigma_i), \quad (63)$$

where

$$\overline{\sigma} := \sum_i p_i \sigma_i. \quad (64)$$

5) Non-negativity: For $\varepsilon \in [0,1)$, we have

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \geq \log_2\left(\frac{1}{1-\varepsilon}\right) \geq 0, \quad (65)$$

with the first inequality saturated if $\rho = \sigma$.

6) If $0 \leq \sigma \leq \sigma'$, then

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \geq D^{\varepsilon}_{\min,F}(\rho\|\sigma'). \quad (66)$$

7) Zero-error bounds: For $\varepsilon \in [0,1)$,

$$D_{\min,F}(\rho\|\sigma) \leq D^{\varepsilon}_{\min,F}(\rho\|\sigma) \quad (67)$$

and for $\varepsilon \in [0, F(\rho,\widehat{\sigma})]$, with $\widehat{\sigma} := \frac{\sigma}{\mathrm{Tr}[\sigma]}$,

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \leq -\log_2[1 - g(\varepsilon,\rho,\sigma)] - \log_2 \mathrm{Tr}[\sigma], \quad (68)$$

where

$$g(\varepsilon,\rho,\sigma) := \left(\sqrt{\varepsilon}\sqrt{F(\rho,\widehat{\sigma})} + \sqrt{1-F(\rho,\widehat{\sigma})}\sqrt{1-\varepsilon}\right)^2. \quad (69)$$

As such,

$$\lim_{\varepsilon\to 0} D^{\varepsilon}_{\min,F}(\rho\|\sigma) = D_{\min,F}(\rho\|\sigma). \quad (70)$$

Proof:

Property 1: Consider that

$$\begin{aligned} &D^{\varepsilon}_{\min,F}(\rho\|c\sigma) \\ &= \sup_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \{-\log_2 F(\widetilde{\rho}, c\sigma) : F(\widetilde{\rho},\rho) \geq 1-\varepsilon\} && (71)\\ &= \sup_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \{-\log_2 (cF(\widetilde{\rho},\sigma)) : F(\widetilde{\rho},\rho) \geq 1-\varepsilon\} && (72)\\ &= \sup_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \{-\log_2 F(\widetilde{\rho},\sigma) - \log_2 c : F(\widetilde{\rho},\rho) \geq 1-\varepsilon\} && (73)\\ &= \sup_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \{-\log_2 F(\widetilde{\rho},\sigma) : F(\widetilde{\rho},\rho) \geq 1-\varepsilon\} - \log_2 c && (74)\\ &= D^{\varepsilon}_{\min,F}(\rho\|\sigma) - \log_2 c. && (75) \end{aligned}$$

This concludes the proof.

Property 2: Let $\rho' \in \mathcal{D}_{\leq}$ satisfy $F(\rho,\rho') \geq 1-\varepsilon$. Then $F(\rho,\rho') \geq 1-\varepsilon'$ since $\varepsilon' \geq \varepsilon$. This leads to

$$-\log_2 F(\rho',\sigma) \leq D^{\varepsilon'}_{\min}(\rho\|\sigma). \quad (76)$$

The above inequality holds for every $\rho' \in \mathcal{D}_{\leq}$ such that $F(\rho,\rho') \geq 1-\varepsilon$. Thus, we have

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \leq D^{\varepsilon'}_{\min,F}(\rho\|\sigma). \quad (77)$$

Property 3: In this proof, we use the multiplicativity of fidelity with respect to tensor products:

$$F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = F(\rho_1,\sigma_1)F(\rho_2,\sigma_2). \quad (78)$$

Let $\rho_1', \rho_2' \in \mathcal{D}_{\leq}$ satisfy $F(\rho_1,\rho_1') \geq 1-\varepsilon_1$ and $F(\rho_2,\rho_2') \geq 1-\varepsilon_2$. Consider that

$$\begin{aligned} &-\log_2 F(\rho_1',\sigma_1) - \log_2 F(\rho_2',\sigma_2) && (79)\\ &= -\log_2 F(\rho_1'\otimes\rho_2', \sigma_1\otimes\sigma_2) && (80)\\ &\leq D^{\varepsilon'}_{\min}(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2), && (81) \end{aligned}$$

where the last inequality follows because

$$\begin{aligned} F(\rho_1'\otimes\rho_2', \rho_1\otimes\rho_2) &= F(\rho_1',\rho_1)\cdot F(\rho_2',\rho_2) && (82)\\ &\geq (1-\varepsilon_1)(1-\varepsilon_2) && (83)\\ &= 1-\varepsilon', && (84) \end{aligned}$$

with ε' as stated just after (62). Then, supremizing (79) over ρ_1' and ρ_2' satisfying $F(\rho_1,\rho_1') \geq 1-\varepsilon_1$ and $F(\rho_2,\rho_2') \geq 1-\varepsilon_2$, we arrive at the desired inequality in (62).

Property 4: Let $\rho' \in \mathcal{D}_{\leq}$ satisfy $F(\rho,\rho') \geq 1-\varepsilon$. By concavity of fidelity [1, Theorem 3.60] (while rescaling for a subnormalized state ρ'), we have that

$$F(\rho',\overline{\sigma}) \geq \sum_i p_i F(\rho',\sigma_i). \quad (85)$$

Then consider that

$$\begin{aligned} -\log_2 F(\rho',\overline{\sigma}) &\leq -\log_2\left[\sum_i p_i F(\rho',\sigma_i)\right] && (86)\\ &\leq \sum_i p_i\left[-\log_2 F(\rho',\sigma_i)\right] && (87)\\ &\leq \sum_i p_i D^{\varepsilon}_{\min,F}(\rho\|\sigma_i), && (88) \end{aligned}$$

where the first inequality follows from (85) and monotonicity of $-\log_2$, the second inequality from the convexity of $-\log_2$, and the last due to $F(\rho,\rho') \geq 1-\varepsilon$. Lastly, by optimizing over all ρ' satisfying the required condition, we conclude the proof.

Property 5: By the data-processing inequality derived in Theorem 1 and choosing the quantum channel $\mathcal{N}(X) = \mathrm{Tr}[X]\omega$, where X is a linear operator, and ω is a quantum state, we have

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \geq D^{\varepsilon}_{\min,F}(\omega\|\omega). \quad (89)$$

Then with the constraint $F(\widetilde{\rho},\omega) = 1-\varepsilon$ (recall Remark 1), it follows that $D^{\varepsilon}_{\min,F}(\omega\|\omega) = -\log_2(1-\varepsilon)$ for every state ω. Combining that with (89) concludes the proof.

Property 6: Let $\widetilde{\rho} \in \mathcal{D}_{\leq}$ be such that $F(\widetilde{\rho},\rho) \geq 1-\varepsilon$. By [1, Proposition 7.33], $\widetilde{D}_{1/2}(\rho\|\sigma) \geq \widetilde{D}_{1/2}(\rho\|\sigma')$ for $\sigma \leq \sigma'$. Then, by (15), we have that

$$-\log_2 F(\widetilde{\rho},\sigma) \geq -\log_2 F(\widetilde{\rho},\sigma').$$

Supremizing over $\widetilde{\rho} \in \mathcal{D}_{\leq}$ such that $F(\widetilde{\rho}, \rho) \geq 1-\varepsilon$, we arrive at the desired inequality.

Property 7: We first prove (67). For all $\varepsilon \in [0,1)$, we choose $\widetilde{\rho} = \rho$ and thus have that $F(\widetilde{\rho}, \rho) = 1 \geq 1-\varepsilon$. This leads to

$$-\log_2 F(\rho, \sigma) \leq D^{\varepsilon}_{\min,F}(\rho\|\sigma). \tag{90}$$

Now we prove (68) and first do so in the case that σ is a state. Let $\rho_1 \in \mathcal{D}_{\leq}$ satisfy $F(\rho_1, \rho) = 1-\varepsilon$. Now construct the following normalized states:

$$\rho_1' := \rho_1 \oplus (1 - \mathrm{Tr}[\rho_1]), \tag{91}$$

$$\rho' := \rho \oplus 0, \tag{92}$$

$$\sigma' := \sigma \oplus 0, \tag{93}$$

so that

$$F(\rho_1', \sigma') = F(\rho_1, \sigma), \tag{94}$$

$$F(\rho_1', \rho') = F(\rho_1, \rho), \tag{95}$$

$$F(\rho', \sigma') = F(\rho, \sigma). \tag{96}$$

By assumption, ε satisfies $0 \leq \varepsilon \leq F(\rho, \sigma)$, which implies that $\varepsilon + (1-F(\rho,\sigma)) \leq 1$. We can thus apply the refined triangular inequality for the sine distance $\sqrt{1-F}$ [35, Proposition 3.16] to find that

$$\begin{aligned}&\sqrt{1-F(\rho_1,\sigma)} \\ &\leq \sqrt{1-F(\rho_1,\rho)}\sqrt{F(\rho,\sigma)} \\ &\quad + \sqrt{1-F(\rho,\sigma)}\sqrt{F(\rho_1,\rho)} &(97)\\ &= \sqrt{\varepsilon}\sqrt{F(\rho,\sigma)} + \sqrt{1-F(\rho,\sigma)}\sqrt{1-\varepsilon}. &(98)\end{aligned}$$

After an algebraic manipulation of the inequality above, we arrive at

$$-\log_2 F(\rho_1,\sigma) \leq \\ -\log_2\left(1 - \left(\sqrt{\varepsilon}\sqrt{F(\rho,\sigma)} + \sqrt{1-F(\rho,\sigma)}\sqrt{1-\varepsilon}\right)^2\right). \tag{99}$$

This inequality holds for all ρ_1 satisfying $F(\rho_1, \rho) = 1-\varepsilon$. Then supremizing over all such candidates belonging to $\mathcal{D}_{\leq}$, we conclude that

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \leq \\ -\log_2\left(1 - \left(\sqrt{\varepsilon}\sqrt{F(\rho,\sigma)} + \sqrt{1-F(\rho,\sigma)}\sqrt{1-\varepsilon}\right)^2\right). \tag{100}$$

This concludes the proof of (68) when σ is a state.

When σ is a general PSD operator, we can write $\sigma = c\left(\frac{1}{c}\sigma\right)$ with $c := \mathrm{Tr}[\sigma]$, so that $\frac{1}{c}\sigma$ is a state. We then apply (100) to $\frac{1}{c}\sigma$ and use the scaling property (Property 1 of Theorem 2) to conclude (68) in the general case.

To conclude (70), we take the following limits of (67) and (68), respectively:

$$D_{\min,F}(\rho\|\sigma) \leq \liminf_{\varepsilon\to 0}\ D^{\varepsilon}_{\min,F}(\rho\|\sigma), \tag{101}$$

$$\begin{aligned}\limsup_{\varepsilon\to 0} D^{\varepsilon}_{\min,F}(\rho\|\sigma) &\leq -\log_2 F\left(\rho, \frac{\sigma}{c}\right) - \log_2 c &(102)\\ &= -\log_2 F(\rho,\sigma) &(103)\\ &= D_{\min,F}(\rho\|\sigma). &(104)\end{aligned}$$

This concludes the proof. ∎

IV. Relating Smooth F-Min-Relative Entropy to Other Distinguishability Measures

In this section, we establish connections between the smooth F-min-relative entropy and other information-theoretic quantities such as the sandwiched Rényi relative entropy, its smooth versions, as well as the smooth max- and min-relative entropies. Then, we utilize these connections in subsequent sections to establish the second-order asymptotics of the smooth F-min-relative entropy.

A. Relation to Sandwiched Rényi Relative Entropy and Its Smoothed Variants

The smoothed sandwiched Rényi relative entropies were defined recently in [14], and we recall their definitions here.

Definition 2: Let ρ be a state, and let σ be a PSD operator. Fix $\varepsilon \in [0,1]$ and $\alpha \in (0,1)\cup(1,\infty)$. The smooth sandwiched Rényi relative entropy is defined for $\alpha > 1$ as

$$\widetilde{D}^{\varepsilon}_{\alpha}(\rho\|\sigma) := \inf_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \left\{\widetilde{D}_{\alpha}(\widetilde{\rho}\|\sigma) : F(\widetilde{\rho},\rho) \geq 1-\varepsilon\right\}. \tag{105}$$

and for $\alpha \in (0,1)$ as

$$\widetilde{D}^{\varepsilon}_{\alpha}(\rho\|\sigma) := \sup_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \left\{\widetilde{D}_{\alpha}(\widetilde{\rho}\|\sigma) : F(\widetilde{\rho},\rho) \geq 1-\varepsilon\right\}. \tag{106}$$

In the above definitions, we use precisely the mathematical expression in (12) for evaluating $\widetilde{D}_{\alpha}(\widetilde{\rho}\|\sigma)$, even though we only defined it in (12) for (normalized) states.

Note that the smooth sandwiched Rényi relative entropy for $\alpha = 1/2$ is equivalent to the smooth F-min relative entropy (recall Definition 1).

Remark 3 (Inequality Constraint in Smooth Sandwiched Rényi Relative Entropy Definition): Note that the smooth sandwiched Rényi relative entropy can be rewritten for $\alpha > 1$ as

$$\widetilde{D}^{\varepsilon}_{\alpha}(\rho\|\sigma) := \inf_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \left\{\widetilde{D}_{\alpha}(\widetilde{\rho}\|\sigma) : F(\widetilde{\rho},\rho) = 1-\varepsilon\right\}, \tag{107}$$

and for $\alpha \in (0,1)$ as

$$\widetilde{D}^{\varepsilon}_{\alpha}(\rho\|\sigma) := \sup_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \left\{\widetilde{D}_{\alpha}(\widetilde{\rho}\|\sigma) : F(\widetilde{\rho},\rho) = 1-\varepsilon\right\}, \tag{108}$$

Indeed, let us first consider when $\alpha > 1$. If $\widetilde{\rho}$ is such that $F(\widetilde{\rho},\rho) > 1-\varepsilon$, then we can set $c = (1-\varepsilon)/F(\widetilde{\rho},\rho) \in (0,1)$ such that $F(\rho',\rho) = 1-\varepsilon$, where $\rho' = c\widetilde{\rho}$ and $\rho' \in \mathcal{D}_{\leq}$. Furthermore, we also have that $\widetilde{D}_{\alpha}(\widetilde{\rho}\|\sigma) > \widetilde{D}_{\alpha}(\rho'\|\sigma)$, so that the objective function only decreases under this change. The statement for $\alpha \in (0,1)$ follows from a similar argument.

Theorem 3: Let ρ be a state and σ a positive semi-definite operator. Let $\varepsilon_1, \varepsilon_2 \in [0,1]$ be such that $\varepsilon_1 + \varepsilon_2 \leq 1$, and let

$$\varepsilon' := \left[\sqrt{\varepsilon_1}\sqrt{1-\varepsilon_2} + \sqrt{1-\varepsilon_1}\sqrt{\varepsilon_2}\right]^2, \tag{109}$$

so that $\varepsilon' \in [0,1]$. Then for $\alpha \in (1/2, 1)$ and $\beta = \frac{\alpha}{2\alpha-1} > 1$, we have that

$$\widetilde{D}^{\varepsilon_1}_{\beta}(\rho\|\sigma) + \frac{\beta}{\beta-1}\log_2\left(\frac{1}{1-\varepsilon'}\right) \geq \widetilde{D}^{\varepsilon_2}_{\alpha}(\rho\|\sigma). \tag{110}$$

Proof: Let ρ_1 be optimal for $\widetilde{D}^{\varepsilon_1}_{\beta}(\rho\|\sigma)$, and let ρ_2 be optimal for $\widetilde{D}^{\varepsilon_2}_{\alpha}(\rho\|\sigma)$. Then it follows from Remark 3 that

$$F(\rho_i,\rho) = 1-\varepsilon_i, \tag{111}$$

for $i \in \{1,2\}$. Let

$$\rho_1' := \rho_1 \oplus (1-\operatorname{Tr}[\rho_1]) \oplus 0, \tag{112}$$
$$\rho_2' := \rho_2 \oplus 0 \oplus (1-\operatorname{Tr}[\rho_2]), \tag{113}$$
$$\rho' := \rho \oplus 0 \oplus 0. \tag{114}$$

Note that ρ_1' and ρ_2' are normalized states satisfying the equalities

$$F(\rho_1',\rho_2') = F(\rho_1,\rho_2), \tag{115}$$
$$F(\rho_i',\rho') = F(\rho_i,\rho), \tag{116}$$

for $i \in \{1,2\}$. Then, by applying the refined triangular inequality for the sine distance $\sqrt{1-F}$ of normalized states [35, Proposition 3.16], along with the assumption that $\varepsilon_1+\varepsilon_2 \le 1$), we arrive at

$$\begin{aligned}&\sqrt{1-F(\rho_1,\rho_2)}\\ &\le \sqrt{1-F(\rho_1,\rho)}\sqrt{F(\rho,\rho_2)}\\ &\quad + \sqrt{1-F(\rho,\rho_2)}\sqrt{F(\rho_1,\rho)} &(117)\\ &= \sqrt{\varepsilon_1}\sqrt{1-\varepsilon_2}+\sqrt{\varepsilon_2}\sqrt{1-\varepsilon_1}. &(118)\end{aligned}$$

Then it follows that

$$\begin{aligned}F(\rho_1,\rho_2) &\ge 1-\left[\sqrt{\varepsilon_1}\sqrt{1-\varepsilon_2}+\sqrt{\varepsilon_2}\sqrt{1-\varepsilon_1}\right]^2 &(119)\\ &= 1-\varepsilon'. &(120)\end{aligned}$$

To arrive at the desired inequality in (110), let us recall the following inequality from [5, Lemma 1]. Let ρ_1 and ρ_2 be subnormalized states, and let σ be a positive semi-definite operator such that $\operatorname{supp}(\rho_1) \subseteq \operatorname{supp}(\sigma)$. For $\alpha \in (1/2,1)$ and $\beta = \alpha/(2\alpha-1) > 1$,

$$\begin{aligned}\widetilde{D}_{\beta}(\rho_1\|\sigma) - \widetilde{D}_{\alpha}(\rho_2\|\sigma) &\ge \frac{\alpha}{1-\alpha}\log_2 F(\rho_1,\rho_2) &(121)\\ &= \frac{\beta}{\beta-1}\log_2 F(\rho_1,\rho_2). &(122)\end{aligned}$$

We note here that the proof of [5, Lemma 1] was only given therein for states, but it is clear by inspection that the same proof holds for subnormalized states. This implies that

$$\widetilde{D}_{\beta}(\rho_1\|\sigma) + \frac{\beta}{\beta-1}\log_2\left(\frac{1}{F(\rho_1,\rho_2)}\right) \ge \widetilde{D}_{\alpha}(\rho_2\|\sigma). \tag{123}$$

Then, by the inequality in (120), we have

$$\widetilde{D}_{\beta}(\rho_1\|\sigma) + \frac{\beta}{\beta-1}\log_2\left(\frac{1}{1-\varepsilon'}\right) \ge \widetilde{D}_{\alpha}(\rho_2\|\sigma). \tag{124}$$

Lastly, we conclude the proof by noting the assumption that ρ_1 and ρ_2 are optimal for $\widetilde{D}^{\varepsilon_1}_{\beta}(\rho\|\sigma)$ and $\widetilde{D}^{\varepsilon_2}_{\alpha}(\rho\|\sigma)$, respectively. ∎

Applying Theorem 3 and the limits $\varepsilon_1 \to 0$ and $\alpha \to 1/2$ (or alternatively directly from (124) with $\varepsilon_1 = 0$ and taking the limit $\alpha \to 1/2$ while employing the α-monotonicity of $\widetilde{D}_{\alpha}$), we arrive at the following corollary:

Corollary 2: For all $\varepsilon \in [0,1)$, every state ρ, PSD operator σ, and $\beta > 1$, the following inequality holds

$$\widetilde{D}_{\beta}(\rho\|\sigma) + \frac{\beta}{\beta-1}\log_2\left(\frac{1}{1-\varepsilon}\right) \ge D^{\varepsilon}_{\min,F}(\rho\|\sigma). \tag{125}$$

By applying (22) and Proposition 2, we arrive at the following inequality:

$$D_{\max}(\rho\|\sigma) + \log_2\left(\frac{1}{1-\varepsilon}\right) \ge D^{\varepsilon}_{\min,F}(\rho\|\sigma). \tag{126}$$

We can also arrive at an inequality relating the smooth F-min relative entropy to the smooth max-relative entropy. By taking the limits $\beta \to \infty$ and $\alpha \to 1/2$ in Theorem 3, while employing the α-monotonicity of $\widetilde{D}_{\beta}$, we obtain the following inequality:

Corollary 3: Let ρ be a state and σ a PSD operator. Let $\varepsilon_1,\varepsilon_2 \in [0,1]$ be such that $\varepsilon_1+\varepsilon_2 \le 1$, and let

$$\varepsilon' := \left[\sqrt{\varepsilon_1}\sqrt{1-\varepsilon_2}+\sqrt{1-\varepsilon_1}\sqrt{\varepsilon_2}\right]^2, \tag{127}$$

so that $\varepsilon' \in [0,1]$. Then

$$D^{\varepsilon_1}_{\max}(\rho\|\sigma) + \log_2\left(\frac{1}{1-\varepsilon'}\right) \ge D^{\varepsilon_2}_{\min,F}(\rho\|\sigma). \tag{128}$$

We note here that the inequality in (128) appeared in [13, Lemma III.8] via a different proof strategy. This kind of inequality was interpreted in a resource-theoretic manner in [5], [7], [38], and [39]—it would thus be interesting if it were possible to do so for (110) and (128).

B. Relation to Smooth Min-Relative Entropy

Definition 3: Recall that the smooth min-relative entropy is defined for $\varepsilon \in [0,1]$, a state ρ, and a PSD operator σ as

$$D^{\varepsilon}_{\min}(\rho\|\sigma) := -\log_2 \inf_{\Lambda \ge 0}\left\{\operatorname{Tr}[\Lambda\sigma] : \operatorname{Tr}[\Lambda\rho] \ge 1-\varepsilon, \Lambda \le I\right\}. \tag{129}$$

As mentioned before, it is also known as the hypothesis testing relative entropy.

Proposition 1: For every $\varepsilon \in (0,1)$, state ρ, and PSD σ, the following inequality holds

$$D^{\varepsilon}_{\min}(\rho\|\sigma) \le D^{\varepsilon}_{\min,F}(\rho\|\sigma) + \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{130}$$

Proof: Let Λ be an arbitrary measurement operator satisfying $\operatorname{Tr}[\Lambda\rho] \ge 1-\varepsilon$. By the gentle measurement lemma from [40, Eq. (9.202)], we know that $\operatorname{Tr}[\Lambda\rho] \ge 1-\varepsilon$ implies that

$$F(\widetilde{\rho},\rho) \ge 1-\varepsilon, \tag{131}$$

where $\widetilde{\rho} = \frac{1}{\operatorname{Tr}[\Lambda\rho]}\sqrt{\Lambda}\rho\sqrt{\Lambda}$. An alternative way to see the inequality in (131) is by the proof in Lemma 5 in Appendix A.

Now we should relate $\operatorname{Tr}[\Lambda\sigma]$ to $F(\widetilde{\rho},\sigma)$. Consider from Lemma 6 in Appendix A that

$$\begin{aligned}F(\widetilde{\rho},\sigma) &= \frac{1}{\operatorname{Tr}[\Lambda\rho]}F(\rho,\sqrt{\Lambda}\sigma\sqrt{\Lambda}) &(132)\\ &\le \frac{1}{1-\varepsilon}F(\rho,\sqrt{\Lambda}\sigma\sqrt{\Lambda}) &(133)\end{aligned}$$

$$\leq \frac{1}{1-\varepsilon} \operatorname{Tr}[\Lambda\sigma]. \tag{134}$$

The last inequality follows from data processing for fidelity under the trace channel. Then

$$-\log_2 \operatorname{Tr}[\Lambda\sigma] \leq -\log_2(F(\widetilde{\rho},\sigma)\,(1-\varepsilon)) \tag{135}$$
$$= -\log_2 F(\widetilde{\rho},\sigma) + \log_2\left(\frac{1}{1-\varepsilon}\right) \tag{136}$$
$$\leq D^{\varepsilon}_{\min,F}(\rho\|\sigma) + \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{137}$$

Since this holds for an arbitrary measurement operator satisfying $\operatorname{Tr}[\Lambda\rho] \geq 1-\varepsilon$, we conclude that

$$D^{\varepsilon}_{\min}(\rho\|\sigma) \leq D^{\varepsilon}_{\min,F}(\rho\|\sigma) + \log_2\left(\frac{1}{1-\varepsilon}\right), \tag{138}$$

which is the desired statement. ■

The following was established in [12], but we give an alternative proof for it.

Proposition 2: For every $\varepsilon \in (0,1)$, state ρ, and PSD operator σ, the following inequality holds

$$D^{\varepsilon}_{\min}(\rho\|\sigma) \leq D^{\varepsilon(2-\varepsilon)}_{\min,F}(\rho\|\sigma). \tag{139}$$

Proof: The proof is similar to the proof of Proposition 1. See Appendix B for details. ■

We note here that Proposition 1 derived in this work is useful in obtaining the second-order asymptotics presented in Section V.

Remark 4 (Lower Bound for Smooth F-Min-Relative Entropy via Petz–Rényi Relative Entropy): For $\alpha \in (0,1)$, by [41, Proposition 3], we have

$$D^{\varepsilon}_{\min}(\rho\|\sigma) \geq \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\varepsilon}\right) + D_\alpha(\rho\|\sigma). \tag{140}$$

Then by Proposition 1, we find that

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma)+\log_2\left(\frac{1}{1-\varepsilon}\right) \geq D_\alpha(\rho\|\sigma)+\frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\varepsilon}\right). \tag{141}$$

It is an open question to establish a tighter lower bound on the smooth F-min-relative entropy in terms of the Petz–Rényi relative entropy of order $\alpha \in (0,1)$.

V. Second-Order Asymptotics

In this section, we establish the second-order asymptotics of the smooth F-min-relative entropy (Theorem 4), as well as for the smooth sandwiched Rényi relative entropy (Corollary 4). Before doing so, let us first define a set of quantities that are needed in what follows. The cumulative distribution function of a standard normal random variable and its inverse are respectively given by

$$\Phi(a) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} dx\, \exp\left(\frac{-x^2}{2}\right), \tag{142}$$

$$\Phi^{-1}(\varepsilon) := \sup\{a \in \mathbb{R} \mid \Phi(a) \leq \varepsilon\}. \tag{143}$$

For $\varepsilon \in (0,1)$, recall that

$$\Phi^{-1}(1-\varepsilon) = -\Phi^{-1}(\varepsilon), \tag{144}$$

and note that $\Phi^{-1}(\varepsilon) < 0$ for $\varepsilon < 1/2$ and $\Phi^{-1}(\varepsilon) > 0$ for $\varepsilon > 1/2$. For a state ρ and PSD operator σ such that $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ and $V(\rho\|\sigma) > 0$, the following second-order expansion is known:

$$\frac{1}{n}D^{\varepsilon}_{\min}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{145}$$

This expansion gives a refined understanding of Stein's lemma for asymmetric hypothesis testing [22], [23] and has been useful in various developments toward establishing second-order asymptotic characterizations of information-theoretic tasks (see, e.g., [24], [25], [42]). For the finite-dimensional scenario, (145) was proven in [22] and [23]. For a state ρ and PSD trace-class σ acting on a separable Hilbert space, the inequality $\leq$ was established in [26] and [36], while the inequality $\geq$ was shown in [23], [26], and [43].

Theorem 4: For a state ρ, a PSD operator σ, and $\varepsilon \in (0,1)$, such that $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ and $V(\rho\|\sigma) > 0$, the following second-order expansion holds:

$$\frac{1}{n}D^{\varepsilon}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{146}$$

Proof: For the lower bound, we apply Proposition 1 to find that

$$\frac{1}{n}D^{\varepsilon}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n}) \tag{147}$$
$$\geq \frac{1}{n}D^{\varepsilon}_{\min}(\rho^{\otimes n}\|\sigma^{\otimes n}) - \frac{1}{n}\log_2\left(\frac{1}{1-\varepsilon}\right) \tag{148}$$
$$= D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{149}$$

where the last inequality follows from (145).

For the upper bound, we use Corollary 3 for $\varepsilon, \delta \in (0,1)$ such that $\varepsilon+\delta \in (0,1)$ to find that

$$\frac{1}{n}D^{\varepsilon}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n}) \leq \frac{1}{n}D^{1-\varepsilon-\delta}_{\max}(\rho^{\otimes n}\|\sigma^{\otimes n}) + \frac{1}{n}\log_2\left(\frac{1}{1-f(\varepsilon,\delta)}\right), \tag{150}$$

where

$$f(\varepsilon,\delta) := \left[\sqrt{\varepsilon}\sqrt{\varepsilon+\delta} + \sqrt{1-\varepsilon-\delta}\sqrt{1-\varepsilon}\right]^2, \tag{151}$$

so that $f(\varepsilon,\delta) \in (0,1)$. Indeed, $f(\varepsilon,\delta)$ can be understood as a classical fidelity of two binary random variables with parameters ε and $\varepsilon+\delta$, which we know is ≤ 1. We also note that

$$\frac{1}{1-f(\varepsilon,\delta)} = \frac{4\varepsilon\,(1-\varepsilon)}{\delta^2} + \frac{2\,(1-2\varepsilon)}{\delta} - \frac{1}{4\varepsilon\,(1-\varepsilon)} + O(\delta). \tag{152}$$

Thus, when we choose $\delta = 1/\sqrt{n}$ (and n sufficiently large so that $\varepsilon + \delta < 1$) for the second-order expansion, we find that

$$\log_2\left(\frac{1}{1-f(\varepsilon,\delta)}\right) = O(\log n). \tag{153}$$

Recall from [44, Theorem 4] that

$$D^{\varepsilon}_{\max}(\rho\|\sigma) \leq D^{1-\varepsilon}_{\min}(\rho\|\sigma) + \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{154}$$

Note that the choice of ε instead of $\sqrt{\varepsilon}$ as in [44, Theorem 4] is due to the fact that the smooth max-relative entropy is defined in [44] in terms of the sine distance $\sqrt{1-F(\widetilde{\rho},\rho)}$ [45], [46], [47], [48]. Combining the above inequalities together with (150), we arrive at

$$\begin{aligned}
&\frac{1}{n}D^{\varepsilon}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n}) && (155)\\
&\overset{(a)}{\leq} \frac{1}{n}D^{1-\varepsilon-\delta}_{\max}(\rho^{\otimes n}\|\sigma^{\otimes n}) + \frac{1}{n}\log_2\left(\frac{1}{1-f(\varepsilon,\delta)}\right) && (156)\\
&\overset{(b)}{\leq} \frac{1}{n}D^{\varepsilon+\delta}_{\min}(\rho^{\otimes n}\|\sigma^{\otimes n}) + \frac{1}{n}\log_2\left(\frac{1}{\varepsilon+\delta}\right)\\
&\qquad + \frac{1}{n}\log_2\left(\frac{1}{1-f(\varepsilon,\delta)}\right) && (157)\\
&\overset{(c)}{\leq} D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon+\delta) + O\left(\frac{\log n}{n}\right) && (158)\\
&\overset{(d)}{\leq} D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). && (159)
\end{aligned}$$

In the above, (a) follows from (25), (b) from (150) and (154), (c) from (145), and lastly (d) from the following argument given below. As stated in [24, Eq. (4)] in the proof of [24, Lemma 3.7]: for f a continuously differentiable function, by Taylor's theorem,

$$f\left(x \pm \frac{1}{\sqrt{n}}\right) = f(x) \pm \frac{1}{\sqrt{n}}f'(\eta), \tag{160}$$

where $\eta \in \left(x - \frac{1}{\sqrt{n}}, x\right)$ in the case $-$ and $\eta \in \left(x, x+\frac{1}{\sqrt{n}}\right)$ in the case $+$. By choosing $\delta = 1/\sqrt{n}$, the function $\Phi^{-1}(\varepsilon+\delta)$ also satisfies the said property.

With the upper bound in (159), together with the lower bound in (149), we conclude the proof. ■

Corollary 4 (Second-Order Asymptotics of Smooth Sandwiched Rényi Relative Entropy): Let ρ be a state and σ a PSD operator. Fix $\varepsilon \in (0,1)$. The smooth sandwiched Rényi relative entropy of order $\alpha > 1$ has the following second-order expansion

$$\frac{1}{n}\widetilde{D}^{\varepsilon}_{\alpha}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) - \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{161}$$

For $\alpha \in [1/2,1)$, the smooth sandwiched Rényi relative entropy has the following second-order expansion:

$$\frac{1}{n}\widetilde{D}^{\varepsilon}_{\alpha}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{162}$$

Proof: This follows from techniques similar to those in the proof of Theorem 4, as well as from the α-monotonicity of the sandwiched Rényi relative entropy and Theorem 3. See Appendix F for a proof. ■

Remark 5 (Equivalence of Smooth Relative Entropies Up to the Second-Order): Corollary 4 indicates that, in the asymptotic i.i.d. setting and up to the second order, there is no difference between all of the smooth sandwiched Rényi relative entropies for all $\alpha > 1$. That is, they are all equivalent to the smooth max-relative entropy. Similarly, in the asymptotic i.i.d. setting and up to the second order, there is no difference between all of them for $\alpha \in [1/2,1)$: they are all equivalent to the smooth min-relative entropy. This is a unifying result that should find use in future works on quantum resource theories.

A sequence $\{a_n\}_n$ is called a moderate sequence if $a_n \to 0$ and $\sqrt{n}a_n \to \infty$ when $n \to \infty$.

Proposition 3 (Moderate deviations): For a moderate sequence $\{a_n\}_n$ and $\varepsilon_n = e^{-na_n^2}$, the smooth F-min-relative entropy scales as follows:

$$\frac{1}{n}D^{\varepsilon_n}_{\min,F}\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) = D(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)}\ a_n + o(a_n). \tag{163}$$

Proof: This follows by utilizing the connections established in the last section along with the moderate deviation analysis for the smooth min-relative entropy in [49]. For completeness, we provide a proof in Appendix G. ■

Remark 6 (Moderate Deviations of Smooth Sandwiched Rényi Relative Entropy): Similar to Proposition 3, we arrive at the following scaling for smooth sandwiched Rényi relative entropy: for a moderate sequence $\{a_n\}_n$ and $\varepsilon_n = e^{-na_n^2}$,

1) $\alpha > 1$:

$$\frac{1}{n}\widetilde{D}^{\varepsilon_n}_{\alpha}\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) = D(\rho\|\sigma) + \sqrt{2V(\rho\|\sigma)}\ a_n + o(a_n). \tag{164}$$

2) $\alpha \in [1/2,1)$:

$$\frac{1}{n}\widetilde{D}^{\varepsilon_n}_{\alpha}\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) = D(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)}\ a_n + o(a_n). \tag{165}$$

The proof follows by employing the techniques in the derivation of second-order asymptotics (in Appendix F) and proceeding as in Appendix G.

VI. Application to Randomness Distillation

A. 1W-LOCC-Assisted Randomness Distillation

The distillable randomness of a bipartite state is a measure of classical correlations contained in that state. This quantity was first proposed and characterized in [50], and later studied in various guises in [27], [51], [52], [53], [54], [55], [56], and [57]. It has applications to determining the fundamental limitations on experiments in which the goal is to distill randomness from bipartite states [58], [59]. For completeness, we review the definition of a randomness distillation protocol assisted by one-way local operations and classical communications (1W-LOCC) (shown in Fig. 1).

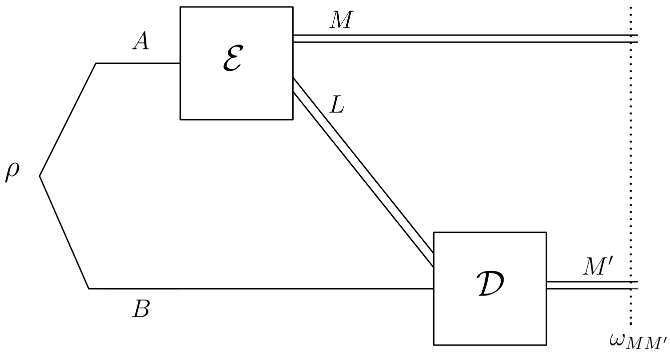

Fig. 1. One way local operations and classical communication protocol from A to B. First, the channel $\mathcal{E}_{A\to ML}$ with classical outputs L and M is applied by Alice. Then, system L is communicated to Bob via a noiseless classical channel. Bob applies the decoding channel $\mathcal{D}_{LB\to M'}$ to get the classical output M'. At the end of the protocol, the output state shared by Alice and Bob is $\omega_{MM'}$, and it should be close to a maximally classically correlated state. In the above figure, classical systems are denoted by double lines.

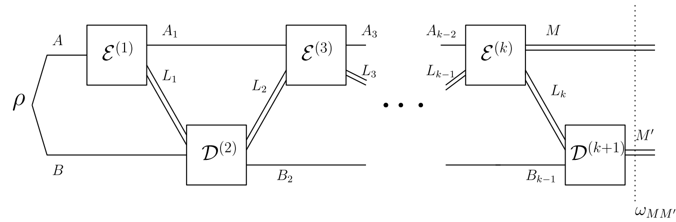

Fig. 2. General local operations and classical communication protocol from A to B: First the channel $\mathcal{E}^{(1)}_{A\to A_1L_1}$ with classical output L_1 is applied by Alice. Then, system L_1 is communicated to Bob via a noiseless classical channel. Next Bob applies the channel $\mathcal{D}^{(2)}_{L_1B\to L_2B_2}$, and the classical output L_2 is communicated to Alice. This procedure is continued for $k-1$ rounds. During the final round Alice performs $\mathcal{E}^{(k)}_{A_{k-2}L_{k-1}\to ML_k}$ where both the output systems are classical and communicates L_k to Bob. Bob completes the protocol by applying the channel $\mathcal{D}^{(k+1)}_{L_kB_{k-1}\to M'}$. At the end of the protocol, the output state shared by Alice and Bob is $\omega_{MM'}$, and it should be close to a maximally classically correlated state.

We follow the recent presentation of [27]. In Section VI-B, we recall the definition of randomness distillation assisted by general LOCC, as also considered in [27].

Let ρ_{AB} be a bipartite state. The protocol begins with Alice applying a quantum channel $\mathcal{E}_{A\to ML}$, with the output systems L and M classical. Then, system L is communicated to Bob over a noiseless classical channel. Having system L, Bob acts with the decoding channel $\mathcal{D}_{LB\to M'}$ on his systems. At the end of the protocol, the final state is

$$\omega_{MM'} := (\mathcal{D}_{LB\to M'} \circ \mathcal{E}_{A\to ML})(\rho_{AB}). \tag{166}$$

A (d,ε) randomness distillation protocol satisfies

$$F\left(\overline{\Phi}^d_{MM'}, \omega_{MM'}\right) \geq 1-\varepsilon, \tag{167}$$

where $\overline{\Phi}^d_{MM'}$ is the maximally classically correlated state of rank d:

$$\overline{\Phi}^d_{MM'} := \frac{1}{d}\sum_{i=0}^{d-1} |i\rangle\!\langle i|_M \otimes |i\rangle\!\langle i|_{M'}. \tag{168}$$

The one-shot distillable randomness of ρ_{AB} is defined as

$$R^{\varepsilon}(\rho_{AB}) := \sup_{\substack{\mathcal{E}_{A\to ML}\\ \mathcal{D}_{LB\to M'}}} \left\{ \log_2 d - \log_2 d_L : F\left(\overline{\Phi}^d_{MM'}, \omega_{MM'}\right) \geq 1-\varepsilon \right\}. \tag{169}$$

Intuitively, $R^{\varepsilon}(\rho_{AB})$ is the largest net number of maximally classically correlated random bits that can be generated from the state ρ_{AB}. Here, we need to subtract the number of bits of classical communication used in the protocol, in order to rule out the possibility of distilling an infinite number of shared random bits.

B. LOCC-Assisted Randomness Distillation

Here we review general LOCC protocols for randomness distillation [27] (shown in Fig. 2). A general LOCC-assisted randomness distillation protocol starts with Alice performing the channel $\mathcal{E}^{(1)}_{A\to A_1L_1}$, with system L_1 being classical and communicated to Bob. Then, Bob performs the channel $\mathcal{D}^{(2)}_{L_1B\to L_2B_2}$, with system L_2 being classical and communicated to Alice. The above procedure continues for k rounds. We denote the rest of the channels for Alice and Bob as $\{\mathcal{E}^{(i)}_{A_{i-2}L_{i-1}\to A_iL_i}\}_i$ for $i \in \{3,5,\ldots\}$ and $\{\mathcal{D}^{(i)}_{L_{i-1}B_{i-2}\to L_iB_i}\}_i$ for $i \in \{4,\ldots\}$, respectively. Without loss of generality, we consider the last two channels of the protocol to be $\mathcal{E}^{(k)}_{A_{k-2}L_{k-1}\to ML_k}$ and $\mathcal{D}^{(k+1)}_{L_kB_{k-1}\to M'}$. The state shared by Alice and Bob at the end of this protocol is

$$\omega_{MM'} := \left(\mathcal{D}^{(k+1)} \circ \mathcal{E}^{(k)} \circ \ldots \circ \mathcal{D}^{(2)} \circ \mathcal{E}^{(1)}\right)(\rho_{AB}). \tag{170}$$

The above-mentioned protocol has ε error if it satisfies

$$p_{\text{err}}(\mathcal{P}^k) := 1 - F\left(\omega_{MM'}, \overline{\Phi}^d_{MM'}\right) \leq \varepsilon, \tag{171}$$

where $\mathcal{P}^k$ corresponds to the protocol described above. With that, the one-shot distillable randomness from ρ_{AB} assisted by LOCC is defined as

$$R^{\varepsilon}_{\leftrightarrow}(\rho_{AB}) := \sup_{k\in\mathbb{N},\mathcal{P}^k} \left\{ \log_2 d - \sum_{i=1}^{k} \log_2 d_{L_i} : p_{\text{err}}(\mathcal{P}^k) \leq \varepsilon \right\}. \tag{172}$$

Since general LOCC assistance contains 1W-LOCC assistance as a special case, we obtain the following bound for every state ρ_{AB} and $\varepsilon \in [0,1]$:

$$R^{\varepsilon}(\rho_{AB}) \leq R^{\varepsilon}_{\leftrightarrow}(\rho_{AB}). \tag{173}$$

C. Γ-Upper Bound on One-Shot Distillable Randomness

Let us recall the definition of the classical correlation measure γ, defined recently in [27] for a PSD bipartite operator σ_{AB} as

$$\gamma(\sigma_{AB}) := \inf_{K_A,L_B,V_{AB}\in\text{Herm}} \left\{ \begin{array}{c} \operatorname{Tr}[K_A\otimes L_B] : \\ \operatorname{T}_B(V_{AB}\pm\sigma_{AB}) \geq 0, \\ K_A\otimes L_B \pm V_{AB} \geq 0 \end{array} \right\}. \tag{174}$$

Some intuition for this quantity was not discussed in [27], and so we provide some briefly here. For a classical correlation measure, it is desirable for it to indeed measure correlations, meaning that it should be equal to zero for a product state, be greater than zero for a state that is not product, and should not increase under the action of local channels. For this purpose, mutual information and its variants are helpful, but they measure quantum correlations in addition to classical correlations [60]. Mutual information can be written as the minimum quantum relative entropy between the state of interest and the set of product states [40, Exercise 11.8.2]:

$$I(A;B)_\rho = \inf_{\substack{\sigma_A \in \mathcal{D}(\mathcal{H}_A),\\ \sigma_B \in \mathcal{D}(\mathcal{H}_B)}} D(\rho_{AB} \| \sigma_A \otimes \sigma_B). \tag{175}$$

Using the max-relative entropy in place of the quantum relative entropy, one can define the max-mutual information as [61]

$$\begin{aligned} I_{\max}(A;B)_\rho &:= \inf_{\substack{\sigma_A \in \mathcal{D}(\mathcal{H}_A),\\ \sigma_B \in \mathcal{D}(\mathcal{H}_B)}} D_{\max}(\rho_{AB} \| \sigma_A \otimes \sigma_B) \qquad (176)\\ &= \log_2 \inf_{K_A, L_B \geq 0} \{\operatorname{Tr}[K_A \otimes L_B] : \rho_{AB} \leq K_A \otimes L_B\}. \qquad (177)\end{aligned}$$

As mentioned above, this quantity measures quantum correlations in addition to classical correlations; for example, it is equal to $2\log_2 d$ for a maximally entangled state of Schmidt rank d, as we prove in Appendix C (we also prove the equality above there). Given that the classical correlations contained in or the distillable randomness extractable from such a state is equal to $\log_2 d$, the mutual information is twice as large as it should be for this case. To address this problem, recall that the logarithmic negativity of a bipartite state is defined as follows [62], [63]:

$$\log_2 \|T_B(\rho_{AB})\|_1 = \log_2 \inf_{\substack{M_{AB},\\ N_{AB} \geq 0}} \left\{ \begin{array}{c} \operatorname{Tr}[M_{AB} + N_{AB}] : \\ M_{AB} \geq T_B(\rho_{AB}), \\ N_{AB} \geq -T_B(\rho_{AB}) \end{array} \right\}, \tag{178}$$

where the second expression follows from [1, Proposition 3.53]. It is equal to $\log_2 d$ for a maximally entangled state and equal to zero for a maximally classically correlated state of rank d. This latter property is also undesirable for a classical correlation measure, which should be equal to $\log_2 d$ for such a state. The basic idea behind the measure in (174) is to combine features of the max-mutual information and logarithmic negativity into a single measure (the features being the optimization over product positive semi-definite operators and the use of the partial transpose). The resulting measure is equal to $\log_2 d$ for both a maximally entangled state of Schmidt rank d and a maximally classically correlated state of rank d, and it satisfies a number of desirable properties expected of a classical correlation measure, as shown in [27].

We can then use the definition in (174) and the general construction in [27, Eq. (4)] to define the following classical correlation measure for a bipartite state ρ_{AB}, relevant for us here:

$$\Gamma^\varepsilon_{\min,F}(A;B)_\rho := \inf_{\substack{\sigma_{AB} \geq 0:\\ \gamma(\sigma_{AB}) \leq 1}} D^\varepsilon_{\min,F}(\rho_{AB} \| \sigma_{AB}). \tag{179}$$

Theorem 5: Fix $\varepsilon \in (0,1)$. The following bound holds for the one-shot LOCC-assisted distillable randomness of a bipartite state ρ_{AB}:

$$R^\varepsilon_{\leftrightarrow}(\rho_{AB}) \leq \Gamma^\varepsilon_{\min,F}(A;B)_\rho. \tag{180}$$

Proof: Let us begin by proving the bound

$$R^\varepsilon(\rho_{AB}) \leq \Gamma^\varepsilon_{\min,F}(A;B)_\rho, \tag{181}$$

and then we discuss afterward how to generalize it to get (180). The proof of (181) follows similarly to the proof of [27, Eq. (41)], by some properties satisfied by $D^\varepsilon_{\min,F}(\cdot\|\cdot)$, which include data processing (see Theorem 1) and scaling (see Property 1 of Theorem 2). These properties of smooth F-min-relative entropy result in $\Gamma^\varepsilon_{\min,F}(A;B)_\rho$ satisfying the properties presented in [27]; out of those we use Proposition 1 on symmetry, Proposition 2 on data processing under local channels, and Proposition 7 on scaling.

By Lemma 9 of [27], we have

$$\sup_{\substack{\tau_{MM'} \geq 0:\\ \gamma(\tau_{MM'}) \leq 1}} F\left(\overline{\Phi}^d_{MM'}, \tau_{MM'}\right) \leq \frac{1}{d}. \tag{182}$$

Then, considering the constraint $F\left(\overline{\Phi}^d_{MM'}, \omega_{MM'}\right) \geq 1-\varepsilon$ from (171), we arrive at

$$\begin{aligned} \log_2 d &\leq -\log_2 F\left(\overline{\Phi}^d_{MM'}, \tau_{MM'}\right) \qquad (183)\\ &\leq D^\varepsilon_{\min,F}(\omega_{MM'} \| \tau_{MM'}). \qquad (184)\end{aligned}$$

The above inequality holds for all $\tau_{MM'} \geq 0$ satisfying $\gamma(\tau_{MM'}) \leq 1$. Thus we have

$$\begin{aligned} \log_2 d &\leq \inf_{\substack{\tau_{MM'} \geq 0:\\ \gamma(\tau_{MM'}) \leq 1}} D^\varepsilon_{\min,F}(\omega_{MM'} \| \tau_{MM'}) \qquad (185)\\ &= \Gamma^\varepsilon_{\min,F}(M;M')_\omega. \qquad (186)\end{aligned}$$

By the data-processing inequality for the smooth-min relative entropy (Theorem 1) applied to the channel $\mathcal{D}_{BL\to M'}$, as well as Proposition 2 of [27], we get

$$\Gamma^\varepsilon_{\min,F}(M;M')_\omega \leq \Gamma^\varepsilon_{\min,F}(M;BL)_{\mathcal{E}(\rho)}. \tag{187}$$

Then applying Proposition 7 of [27] with the assistance of the scaling property of $D^\varepsilon_{\min,F}(\cdot\|\cdot)$ (Theorem 2), we find that

$$\Gamma^\varepsilon_{\min,F}(M;BL)_{\mathcal{E}(\rho)} \leq \log_2 d_L + \Gamma^\varepsilon_{\min,F}(LM;B)_{\mathcal{E}(\rho)}. \tag{188}$$

Next, again applying data processing under the local channel $\mathcal{E}_{A\to LM}$, we conclude that

$$\Gamma^\varepsilon_{\min,F}(LM;B)_{\mathcal{E}(\rho)} \leq \Gamma^\varepsilon_{\min,F}(A;B)_\rho. \tag{189}$$

Putting everything together, we arrive at

$$\log_2 d - \log_2 d_L \leq \Gamma^\varepsilon_{\min,F}(A;B)_\rho. \tag{190}$$

Since the above inequality holds for an arbitrary (d,ε) randomness distillation protocol, we conclude the desired bound in (181).

The proof of (180) then follows by iterating the same reasoning as in the proof above, while going backward through the protocol $\mathcal{P}^k$ defined in Section VI-B. See the end of the proof of [27, Theorem 11] for similar reasoning. ■

Remark 7 (Comparison with other existing bounds): Note that the bound provided in the previous work (Theorem 11 of [27]) is a consequence of Theorem 5. In particular, this previous work established the following bound: For $\alpha > 1$

$$R^{\varepsilon}_{\leftrightarrow}(\rho_{AB}) \leq \inf_{\substack{\sigma_{AB}\geq 0,\\ \gamma(\sigma_{AB})\leq 1}} \widetilde{D}_{\alpha}(\rho_{AB}\|\sigma_{AB}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right). \tag{191}$$

Together with the relationship between the sandwiched Rényi relative entropy and the smooth F-min relative entropy derived in Proposition 2, it can be seen that the bound derived in Theorem 5 implies the previous bound.

D. Smooth F-Min- and Min-Mutual-Information Bounds on One-Shot Distillable Randomness

In this section, we define the smooth F-min-mutual information of a state and establish several of its properties, thus justifying it as a correlation measure for bipartite states. We also discuss how it leads to an alternative upper bound on the one-shot distillable randomness.

Let us define the following mutual-information-like correlation measure, which we call the smooth F-min-mutual information:

$$I^{\varepsilon}_{\min,F}(A;B)_{\rho} := \inf_{\substack{\sigma_A,\sigma_B\geq 0:\\ \mathrm{Tr}[\sigma_A]\leq 1,\mathrm{Tr}[\sigma_B]\leq 1}} D^{\varepsilon}_{\min,F}(\rho_{AB}\|\sigma_A\otimes\sigma_B). \tag{192}$$

Since every subnormalized product state $\sigma_A \otimes \sigma_B$ satisfies $\gamma(\sigma_A \otimes \sigma_B) \leq 1$ (as a consequence of Propositions 4 and 6 of [27]), we conclude that the following bound holds for every state ρ_{AB}:

$$\Gamma^{\varepsilon}_{\min,F}(A;B)_{\rho} \leq I^{\varepsilon}_{\min,F}(A;B)_{\rho}. \tag{193}$$

Next, we prove various properties of $I^{\varepsilon}_{\min,F}(A;B)_{\rho}$.

Lemma 3: Let ρ_{AB} be a state. Then $I^{\varepsilon}_{\min,F}(A;B)_{\rho}$ satisfies the following properties.

1) Symmetry:

$$I^{\varepsilon}_{\min,F}(A;B)_{\rho} = I^{\varepsilon}_{\min,F}(B;A)_{\rho} \tag{194}$$

2) Data processing under local channels: Let $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$ be quantum channels, then

$$I^{\varepsilon}_{\min,F}(A;B)_{\rho} \geq I^{\varepsilon}_{\min,F}(A';B')_{\omega}, \tag{195}$$

where $\omega_{AA'} := (\mathcal{N}_{A\to A'}\otimes\mathcal{M}_{B\to B'})(\rho_{AB})$.

3) Classical communication bound: Let ρ_{XAB} be a tripartite state:

$$\rho_{XAB} := \sum_x p(x)|x\rangle\langle x|_X \otimes \rho^x_{AB}, \tag{196}$$

where $\{p(x)\}_x$ is a probability distribution and $\{\rho^x_{AB}\}_x$ is a set of states. Then, we have

$$I^{\varepsilon}_{\min,F}(AX;B)_{\rho} \leq \log_2 d_X + I^{\varepsilon}_{\min,F}(A;BX)_{\rho}. \tag{197}$$

Proof: Symmetry: This follows because $\sigma_B\otimes\sigma_A \geq 0$ if and only if $\sigma_A\otimes\sigma_B \geq 0$, and by the unitary invariance of $D^{\varepsilon}_{\min,F}(\cdot\|\cdot)$. In particular, by applying the unitary SWAP operation, we have

$$D^{\varepsilon}_{\min,F}(\rho_{AB}\|\sigma_A\otimes\sigma_B) = D^{\varepsilon}_{\min,F}(\rho_{BA}\|\sigma_B\otimes\sigma_A). \tag{198}$$

Then, by definition of $I^{\varepsilon}_{\min,F}(A;B)_{\rho}$, it satisfies symmetry.

Data processing under local channels: By the data-processing inequality for the smooth F-min-relative entropy (Theorem 1), we have

$$D^{\varepsilon}_{\min,F}(\rho_{AB}\|\sigma_A\otimes\sigma_B) \geq D^{\varepsilon}_{\min,F}(\omega_{A'B'}\|\mathcal{N}_{A\to A'}(\sigma_A)\otimes\mathcal{M}_{B\to B'}(\sigma_B)). \tag{199}$$

Since $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$ are quantum channels, we find that $\mathcal{N}_{A\to A'}(\sigma_A)\otimes\mathcal{M}_{B\to B'}(\sigma_B) \geq 0$, $\mathrm{Tr}[\mathcal{N}_{A\to A'}(\sigma_A)] \leq 1$ and $\mathrm{Tr}[\mathcal{M}_{B\to B'}(\sigma_B)] \leq 1$. With that, we conclude that

$$D^{\varepsilon}_{\min,F}(\rho_{AB}\|\sigma_A\otimes\sigma_B) \geq \inf_{\substack{\tau_{A'},\tau_{B'}\geq 0:\\ \mathrm{Tr}[\tau_{A'}]\leq 1,\mathrm{Tr}[\tau_{B'}]\leq 1}} D^{\varepsilon}_{\min,F}(\omega_{A'B'}\|\tau_{A'}\otimes\tau_{B'}). \tag{200}$$

Then, optimizing over $\sigma_A,\sigma_B \geq 0, \mathrm{Tr}[\sigma_A]\leq 1$, and $\mathrm{Tr}[\sigma_B]\leq 1$, we arrive at the desired conclusion.

Classical communication bound: Fix σ_A and σ_{BX} such that $\sigma_A,\sigma_{BX} \geq 0, \mathrm{Tr}[\sigma_A]\leq 1, \mathrm{Tr}[\sigma_{BX}]\leq 1$. By scaling of smooth F-min-relative entropy in Theorem 2, we have

$$\begin{aligned}\log_2 d_X + D^{\varepsilon}_{\min,F}(\rho_{ABX}\|\sigma_A\otimes\sigma_{BX}) \\ = D^{\varepsilon}_{\min,F}\left(\rho_{ABX}\middle\|\sigma_A\otimes\frac{\sigma_{BX}}{d_X}\right).\end{aligned} \tag{201}$$

Denote the completely dephasing channel as

$$\overline{\Delta}_X(\cdot) := \sum_x |x\rangle\langle x|(\cdot)|x\rangle\langle x|, \tag{202}$$

and set $\sum_x \tilde{\sigma}^x_B \otimes |x\rangle\langle x| := \overline{\Delta}_X(\sigma_{BX})$. Consider that

$$\begin{aligned}&D^{\varepsilon}_{\min,F}\left(\rho_{ABX}\middle\|\sigma_A\otimes\frac{\sigma_{BX}}{d_X}\right)\\ &\geq D^{\varepsilon}_{\min,F}\left(\overline{\Delta}_X(\rho_{ABX})\middle\|\sigma_A\otimes\frac{\overline{\Delta}_X(\sigma_{BX})}{d_X}\right) &(203)\\ &= D^{\varepsilon}_{\min,F}\left(\rho_{AXB}\middle\|\sigma_A\otimes\frac{1}{d_X}\sum_x|x\rangle\langle x|\otimes\tilde{\sigma}^x_B\right) &(204)\\ &\geq D^{\varepsilon}_{\min,F}\left(\rho_{AXB}\middle\|\sigma_A\otimes\frac{I}{d_X}\otimes\sum_x\tilde{\sigma}^x_B\right) &(205)\\ &\geq \inf_{\substack{\tau_{AX}\otimes\tau_B\geq 0:\\ \mathrm{Tr}[\tau_{AX}]\leq 1,\mathrm{Tr}[\tau_B]\leq 1}} D^{\varepsilon}_{\min,F}(\rho_{AXB}\|\tau_{AX}\otimes\tau_B), &(206)\end{aligned}$$

where the first inequality follows from data processing under the dephasing channel, the equality from the unitary invariance of smooth F-min-relative entropy with the SWAP operator to interchange the B and X systems, the next inequality from Property 6 of Theorem 2 (since $|x\rangle\langle x| \leq I$ for every x), and the final inequality because $\mathrm{Tr}[\sum_x\tilde{\sigma}^x] \leq 1$ given that $\mathrm{Tr}[\sigma_{BX}] \leq 1$ and $\mathrm{Tr}[\sigma_A\otimes I/d_X] \leq 1$, given that $\mathrm{Tr}[\sigma_A] \leq 1$. ■

Motivated by the expression for the distillable randomness of classical–quantum states from [50], we present the following upper bound on the one-shot distillable randomness.

Indeed, it can be understood as a one-shot generalization of the quantity from [50, Eq. (24)].

Proposition 4: Fix $\varepsilon \in (0,1)$. The following upper bound holds for the one-shot distillable randomness of a classical-quantum (cq) state ρ_{XB}:

$$R^{\varepsilon}(\rho_{XB}) \leq I^{\varepsilon}_{\min,F}(X;B)_{\rho}. \tag{207}$$

Furthermore, for a bipartite quantum state ρ_{AB}, we have

$$R^{\varepsilon}(\rho_{AB}) \leq \sup_{\mathcal{M}:A\to X} I^{\varepsilon}_{\min,F}(X;B)_{\omega}, \tag{208}$$

where $\omega_{XB} := \mathcal{M}_{A\to X}(\rho_{AB})$ and the supremum is over every measurement channel $\mathcal{M}_{A\to X}$ that takes a quantum input A and outputs a classical system X, i.e., of the form $\mathcal{M}_{A\to X}(\cdot) = \sum_x \operatorname{Tr}[M^x_A(\cdot)]|x\rangle\!\langle x|_X$, where $\{M^x_A\}_x$ is a POVM.

Proof: The first inequality in (207) follows from similar reasoning as in the proof of Theorem 5. For that to follow, Lemmas 3 and 4 establish the required properties for $I^{\varepsilon}_{\min,F}(A;B)_{\rho}$. Note that these lemmas are proved for a general bipartite state, and these also hold for the special case of cq states.

The second inequality in (208) follows because the inequality in (207) holds for an arbitrary cq state formed after a measurement on the A system, and every such one-way LOCC protocol for randomness distillation consists of a measurement on Alice's system as the first step. ■

Remark 8 (Classical Output State): Notice that for a fixed measurement channel $\mathcal{M}$ acting on the system A, applying the completely dephasing channel $\overline{\Delta}_X$ in (202) we have

$$D^{\varepsilon}_{\min,F}(\mathcal{M}_{A\to X}(\rho_{AB})\|\sigma_X\otimes\sigma_B) \geq D^{\varepsilon}_{\min,F}\left(\mathcal{M}_{A\to X}(\rho_{AB})\|\overline{\Delta}_X(\sigma_X)\otimes\sigma_B\right), \tag{209}$$

which follows due to the data-processing inequality for the smooth F-min-relative entropy and $\overline{\Delta}_X\circ\mathcal{M}_{A\to X} = \mathcal{M}_{A\to X}$. This shows that the infimum in the definition of $I^{\varepsilon}_{\min,F}$ is achieved by a state that is classical on X.

The following lemma was used in the proof of Proposition 4:

Lemma 4: The following bound holds:

$$\sup_{\substack{\sigma_A\otimes\sigma_B\geq 0:\\ \operatorname{Tr}[\sigma_A]\leq 1,\operatorname{Tr}[\sigma_B]\leq 1}} F\left(\overline{\Phi}^d_{AB},\sigma_A\otimes\sigma_B\right) \leq \frac{1}{d}, \tag{210}$$

where $\overline{\Phi}^d_{AB}$ is the maximally classically correlated state.

Proof: We give two proofs for this statement. A first proof follows from the observation that every product state $\sigma_A\otimes\sigma_B$ satisfies $\gamma(\sigma_A\otimes\sigma_B) = 1$, by applying Proposition 4 of [27]. Then we obtain the inequality $\gamma(\sigma_A\otimes\sigma_B) \leq 1$ for subnormalized states by applying Proposition 6 of [27]. This then proves that the set of subnormalized product states is contained in the set $\{\sigma_{AB} \geq 0 : \gamma(\sigma_{AB}) \leq 1\}$, so that the desired statement follows from Lemma 9 of [27].

As an alternative proof, let σ_A,σ_B satisfy the constraints $\sigma_A\otimes\sigma_B \geq 0, \operatorname{Tr}[\sigma_A] \leq 1$, and $\operatorname{Tr}[\sigma_B] \leq 1$. By the data-processing inequality for fidelity, consider that

$$F\left(\overline{\Phi}^d_{AB},\sigma_A\otimes\sigma_B\right) \leq F\left(\overline{\Phi}^d_{AB},\overline{\Delta}_A(\sigma_A)\otimes\overline{\Delta}_B(\sigma_B)\right), \tag{211}$$

where $\overline{\Delta}$ is a dephasing channel defined as

$$\overline{\Delta}(\cdot) := \sum_{m=0}^{d-1} |m\rangle\!\langle m|(\cdot)|m\rangle\!\langle m|. \tag{212}$$

With that, we have

$$\begin{aligned}
&\overline{\Delta}_A(\sigma_A)\otimes\overline{\Delta}_B(\sigma_B)\\
&= \sum_{m=0}^{d-1} |m\rangle\!\langle m|\sigma_A|m\rangle\!\langle m| \otimes \sum_{\ell=0}^{d-1} |\ell\rangle\!\langle \ell|\sigma_B|\ell\rangle\!\langle \ell| &(213)\\
&= \sum_{m,\ell=0}^{d-1} \langle m|\sigma_A|m\rangle\langle \ell|\sigma_B|\ell\rangle |m\rangle\!\langle m|\otimes|\ell\rangle\!\langle \ell|. &(214)
\end{aligned}$$

Then, we obtain

$$\begin{aligned}
&F\left(\overline{\Phi}^d_{AB},\overline{\Delta}_A(\sigma_A)\otimes\overline{\Delta}_B(\sigma_B)\right)\\
&= \left(\sum_{m=0}^{d-1}\sqrt{\frac{1}{d}\langle m|\sigma_A|m\rangle\langle m|\sigma_B|m\rangle}\right)^2 &(215)\\
&\leq \frac{1}{d}\sum_{m=0}^{d-1}\langle m|\sigma_A|m\rangle\sum_{m=0}^{d-1}\langle m|\sigma_B|m\rangle &(216)\\
&= \frac{1}{d}\operatorname{Tr}[\sigma_A]\operatorname{Tr}[\sigma_B] &(217)\\
&\leq \frac{1}{d}, &(218)
\end{aligned}$$

where the first equality follows from the fidelity reducing to a classical fidelity, the first inequality by Cauchy–Schwarz, and last inequality from the assumptions that $\operatorname{Tr}[\sigma_A] \leq 1$ and $\operatorname{Tr}[\sigma_B] \leq 1$. Finally, we complete the proof of Lemma 4 by supremizing over σ_A and σ_B satisfying the required constraints. ■

For a cq state with a uniform classical probability distribution, in what follows we derive a lower bound for the one-shot distillable randomness. We obtain this by devising an achievable protocol that makes use of position-based coding [64] and the square-root measurement at the decoder. Let us recall that the smooth min-mutual information of a bipartite state ρ_{AB} is defined for $\varepsilon\in[0,1]$ as [19]

$$I^{\varepsilon}_{\min}(A;B)_{\rho} := D^{\varepsilon}_{\min}(\rho_{AB}\|\rho_A\otimes\rho_B). \tag{219}$$

Proposition 5 (Lower bound): Fix $\varepsilon \in (0,1)$ and $\eta \in (0,\varepsilon)$. For a cq state ρ_{XB} of the form

$$\rho_{XB} := \frac{1}{L}\sum_{x=1}^{L}|x\rangle\!\langle x|_X\otimes\rho^x_B, \tag{220}$$

the one-shot distillable randomness of ρ_{XB} is bounded from below as follows:

$$R^{\varepsilon}(\rho_{XB}) \geq \left\lfloor I^{\varepsilon-\eta}_{\min}(X;B)_{\rho} - \log_2\left(\frac{4\varepsilon}{\eta^2}\right)\right\rfloor. \tag{221}$$

Proof: We prove this lower bound by devising an achievable one-way protocol from Alice to Bob. The proof below follows by employing the idea behind the protocol presented in the proof of [65, Theorem 6]. Prior to the initiation of the protocol, Alice and Bob share the state ρ_{XB}, Alice has access to the X system, and Bob has access to the B system.

The protocol begins with Alice picking an index $m \in \mathcal{M}$ uniformly at random and placing it in a classical register M. She then labels her X system of ρ_{XB} as X_m. She prepares $|\mathcal{M}|$ independent instances of the classical state

$$\rho_X = \frac{1}{L}\sum_{x=1}^{L} |x\rangle\!\langle x|_X \tag{222}$$

and labels them as $X_1, \ldots, X_{m-1}, X_{m+1}, \ldots, X_{|\mathcal{M}|}$. She sends the registers $X_1, \ldots, X_{|\mathcal{M}|}$, in this order, over a classical channel to Bob, while keeping a copy of each of them in her laboratory (denote the copies by $X'_1, \ldots, X'_{|\mathcal{M}|}$). For a fixed value of m, the reduced state of Bob has the following form:

$$\rho_{X_1} \otimes \cdots \otimes \rho_{X_{m-1}} \otimes \rho_{X_m B} \otimes \rho_{X_{m+1}} \otimes \cdots \otimes \rho_{|\mathcal{M}|}, \tag{223}$$

and his goal is to employ a decoding measurement to figure out which X system is correlated with his B system. This reduced state has exactly the form considered in position-based coding [64] (see also [41], [42] and in particular [41, Eq. (3.5)]). As such, at this point, we can invoke those results to conclude that as long as

$$\log_2 |\mathcal{M}| = \left\lfloor I_{\min}^{\varepsilon-\eta}(X;B)_\rho - \log_2\left(\frac{4\varepsilon}{\eta^2}\right)\right\rfloor, \tag{224}$$

it is possible for Bob to decode the index m with an error probability $\leq \varepsilon$. Furthermore, Bob can make use of the square-root measurement construction to perform the decoding, and due to the permutation symmetry of the protocol and measurement, it follows that the error probability in decoding each index m is equal to the same fixed value p_{err}, for some $p_{\text{err}} \in [0,1]$ such that $p_{\text{err}} \leq \varepsilon$ (see [41, Eq. (3.10)]). After performing the decoding, he places his measurement outcome in a classical register M'. Thus, the final state of registers M and M' at the end of the protocol is given by

$$\omega_{MM'} := (1-p_{\text{err}})\overline{\Phi}_{MM'} + p_{\text{err}}\frac{I_{MM'} - \overline{\Phi}_{MM'}}{|\mathcal{M}|-1}, \tag{225}$$

for which we have that

$$\frac{1}{2}\left\|\overline{\Phi}_{MM'} - \omega_{MM'}\right\|_1 = 1 - F(\overline{\Phi}_{MM'}, \omega_{MM'}) = p_{\text{err}} \leq \varepsilon. \tag{226}$$

Since the state ρ_X is uniform, the cost for Alice to communicate each of the registers $X_1, \ldots, X_{|\mathcal{M}|}$ over a classical channel is precisely equal to the amount of distillable randomness contained in the register pairs (X_1, X'_1), …, $(X_{|\mathcal{M}|}, X'_{|\mathcal{M}|})$. Thus, the net number of random shared bits generated by this protocol is equal to $\log_2|\mathcal{M}|$. ■

E. Second-Order Expansions for Randomness Distillation

Now we show how to utilize the second-order asymptotics of the smooth F-min-relative entropy, as presented in Theorem 4, to obtain upper and lower bounds on the rate at which randomness distillation is possible. Before doing so, let us define the following quantities:

$$\Gamma(A;B)_\rho := \inf_{\substack{\sigma_{AB}\geq 0:\\ \gamma(\sigma_{AB})\leq 1}} D(\rho_{AB}\|\sigma_{AB}), \tag{227}$$

where $D(\rho_{AB}\|\sigma_{AB})$ is the quantum relative entropy from (17). Then, denote $\Pi_\gamma \subseteq \{\sigma_{AB} \geq 0 : \gamma(\sigma_{AB}) \leq 1\}$ as the set of PSD operators achieving the infimum in $\Gamma(A;B)_\rho$. From that, we define the following variance quantity:

$$V_\Gamma^\varepsilon(A;B)_\rho := \begin{cases} \inf_{\sigma_{AB}\in\Pi_\gamma} V(\rho_{AB}\|\sigma_{AB}) & \text{if } \varepsilon \geq \frac{1}{2}, \\ \sup_{\sigma_{AB}\in\Pi_\gamma} V(\rho_{AB}\|\sigma_{AB}) & \text{if } \varepsilon < \frac{1}{2}. \end{cases} \tag{228}$$

Theorem 6 (Upper bound): Fix $\varepsilon \in (0,1)$. The LOCC-assisted distillable randomness of a bipartite state ρ_{AB} is bounded from above as follows:

$$\frac{1}{n}R_{\leftrightarrow}^{\varepsilon}(\rho_{AB}^{\otimes n}) \leq \Gamma(A;B)_\rho + \sqrt{\frac{1}{n}V_\Gamma^\varepsilon(A;B)_\rho}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{229}$$

Proof: Applying Theorem 5 and choosing σ_{AB} to be an optimum in (228), we find that

$$\begin{aligned} &\frac{1}{n}R_{\leftrightarrow}^{\varepsilon}(\rho_{AB}^{\otimes n}) \\ &\leq \frac{1}{n}\Gamma_{\min,F}^{\varepsilon}(A^n;B^n)_{\rho^{\otimes n}} && (230)\\ &\leq \frac{1}{n}D_{\min,F}^{\varepsilon}(\rho_{AB}^{\otimes n}\|\sigma_{AB}^{\otimes n}) && (231)\\ &= \Gamma(A;B)_\rho + \sqrt{\frac{1}{n}V_\Gamma^\varepsilon(A;B)_\rho}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). && (232) \end{aligned}$$

The second inequality follows from the definition of $\Gamma_{\min,F}^{\varepsilon}$, and the last equality follows from the second-order expansion from Theorem 4. ■

In the above proof, once we restrict the second argument in $\Gamma_{\min,F}^{\varepsilon}(A^n;B^n)_{\rho^{\otimes n}}$ to be a tensor-power state, it follows from [66, Lemma 63] that the choices we have made are optimal.

Remark 9 (Computational efficiency): We note here that we can further relax the upper bound in Theorem 6 to find an efficiently computable upper bound, by following an approach similar to that discussed in [27, Section VI].

Now let us recall the definition of the mutual information variance of a bipartite state ω_{AB} [25, Section 2.2] as

$$V(A;B)_\omega := V(\omega_{AB}\|\omega_A \otimes \omega_B), \tag{233}$$

where the relative entropy variance V is defined in (18).

Proposition 6 (Upper bound for cq states): Fix $\varepsilon \in (0,1)$. The LOCC-assisted distillable randomness of a cq state ρ_{XB}

is bounded from above as follows:

$$\frac{1}{n}R^{\varepsilon}_{\leftrightarrow}(\rho_{XB}^{\otimes n}) \leq I(X;B)_\rho + \sqrt{\frac{1}{n}V(X;B)_\rho}\ \Phi^{-1}(\varepsilon) + O\!\left(\frac{\log n}{n}\right). \tag{234}$$

Proof: Applying Theorem 5, consider that

$$\begin{aligned}\frac{1}{n}R^{\varepsilon}_{\leftrightarrow}(\rho_{XB}^{\otimes n}) &\leq \frac{1}{n}\Gamma^{\varepsilon}_{\min,F}(X^n;B^n)_{\rho_{XB}^{\otimes n}} &(235)\\ &\leq \frac{1}{n}D^{\varepsilon}_{\min,F}(\rho_{XB}^{\otimes n}\|(\rho_X\otimes\rho_B)^{\otimes n}) &(236)\\ &= D(\rho_{XB}\|\rho_X\otimes\rho_B) \\ &\quad + \sqrt{\frac{1}{n}V(\rho_{XB}\|\rho_X\otimes\rho_B)}\ \Phi^{-1}(\varepsilon) + O\!\left(\frac{\log n}{n}\right) &(237)\\ &= I(X;B)_\rho + \sqrt{\frac{1}{n}V(X;B)_\rho}\ \Phi^{-1}(\varepsilon) + O\!\left(\frac{\log n}{n}\right), &(238)\end{aligned}$$

where the second inequality follows because $\gamma(\rho_X\otimes\rho_B) = 1$ for product states [27, Proposition 4], the first equality by applying Theorem 4, and the last equality from the definitions of mutual information and mutual information variance. ■

Proposition 7 (Lower bound for uniform cq states): Fix $\varepsilon\in(0,1)$. For a cq state of the form

$$\rho_{XB} := \frac{1}{M}\sum_{x=1}^{M}|x\rangle\!\langle x|\otimes\rho_B^x, \tag{239}$$

the 1W-LOCC-assisted distillable randomness of ρ_{XB} satisfies the following:

$$\frac{1}{n}R^{\varepsilon}(\rho_{XB}^{\otimes n}) \geq I(X;B)_\rho + \sqrt{\frac{1}{n}V(X;B)_\rho}\ \Phi^{-1}(\varepsilon) + O\!\left(\frac{\log n}{n}\right). \tag{240}$$

Proof: First by applying Proposition 5 for the state $\rho_{XB}^{\otimes n}$ and choosing $\eta = 1/\sqrt{n}$, consider that

$$\begin{aligned}\frac{1}{n}R^{\varepsilon}(\rho_{XB}^{\otimes n}) &\geq \frac{1}{n}I^{\varepsilon-\eta}_{\min}(X^n;B^n)_{\rho^{\otimes n}} - \frac{1}{n}\log\!\left(\frac{4\varepsilon}{\eta^2}\right) - \frac{1}{n} &(241)\\ &= D(\rho_{XB}\|\rho_X\otimes\rho_B) + \sqrt{\frac{1}{n}V(\rho_{XB}\|\rho_X\otimes\rho_B)}\ \Phi^{-1}(\varepsilon-\eta) \\ &\quad + O\!\left(\frac{\log n}{n}\right) - \frac{1}{n}\log\!\left(\frac{4\varepsilon}{\eta^2}\right) &(242)\\ &= I(X;B)_\rho + \sqrt{\frac{1}{n}V(X;B)_\rho}\ \Phi^{-1}(\varepsilon) + O\!\left(\frac{\log n}{n}\right), &(243)\end{aligned}$$

where the first equality follows from (221) and the second equality by (145). For the latter, we choose n sufficiently large so that $\eta = 1/\sqrt{n}\in(0,\varepsilon)$ and invoke a standard step in [22, Footnote 6] applied to $\Phi^{-1}(\varepsilon-\eta)$, which follows from Taylor's theorem: for f continuously differentiable, c is a positive constant, and $n\geq n_0$, the following equality holds

$$\sqrt{n}f(x-c/\sqrt{n}) = \sqrt{n}f(x) - cf'(a) \tag{244}$$

for some $a\in[x-c/\sqrt{n},x]$ (note that the reasoning we used to arrive at (159), in the proof of Theorem 4 is a special case of this argument). The last equality follows from the definitions of mutual information and mutual information variance. ■

Theorem 7: Fix $\varepsilon\in(0,1)$. For a cq state of the form

$$\rho_{XB} := \frac{1}{M}\sum_{x=1}^{M}|x\rangle\!\langle x|\otimes\rho_B^x, \tag{245}$$

the 1W-LOCC-assisted and LOCC-assisted distillable randomness of ρ_{XB} satisfy the following:

$$\begin{aligned}\frac{1}{n}R^{\varepsilon}(\rho_{XB}^{\otimes n}) &= \frac{1}{n}R^{\varepsilon}_{\leftrightarrow}(\rho_{XB}^{\otimes n})\\ &= I(X;B)_\rho + \sqrt{\frac{1}{n}V(X;B)_\rho}\ \Phi^{-1}(\varepsilon) + O\!\left(\frac{\log n}{n}\right). &(246)\end{aligned}$$

Proof: First by applying (173), we get

$$\frac{1}{n}R^{\varepsilon}(\rho_{XB}^{\otimes n}) \leq \frac{1}{n}R^{\varepsilon}_{\leftrightarrow}(\rho_{XB}^{\otimes n}). \tag{247}$$

Then, by applying Proposition 6, we arrive at the desired upper bound. By obtaining a matching lower bound from Proposition 7, we conclude the proof. ■

We suspect that Theorem 7 holds more generally for all cq states. To establish this finding, it seems that one would need to devise a protocol that has comparable performance to that given in the proof of Proposition 5.

Remark 10: (Impact of Feedback on Distillable Randomness): Theorem 7 indicates that feedback does not improve the distillable randomness of a cq state of the form in (245), even up to the second order for this class of states. This finding is in distinction to the findings of [67] for channel coding, in which it was shown that feedback can improve the classical communication rate up to the second order for channels with compound dispersion. The example in [67, Remark 1] provides an interesting case study. If we choose the uniform distribution over the six channel input symbols, this leads to a fixed bipartite classical state shared between Alice and Bob, for which its distillable randomness cannot be improved by feedback, even up to the second order. However, in the channel coding setting, the sender can adjust the input distribution based on feedback from the receiver, and this is the mechanism underlying the improved performance found in [67].

We now evaluate the upper bound from Theorem 6 and the lower bound from Proposition 7 for a particular example. Suppose that Alice and Bob share an isotropic state of the form $\rho_{AB}^{(d,p)} := (1-p)\Phi^d_{AB} + p\frac{I_{AB}}{d^2}$, where

$$\Phi^d_{AB} := \frac{1}{d}\sum_{i,j}|i\rangle\!\langle j|_A\otimes|i\rangle\!\langle j|_B. \tag{248}$$

One scheme for obtaining a lower bound on the distillable randomness of this state, as considered in [27], is for Alice to

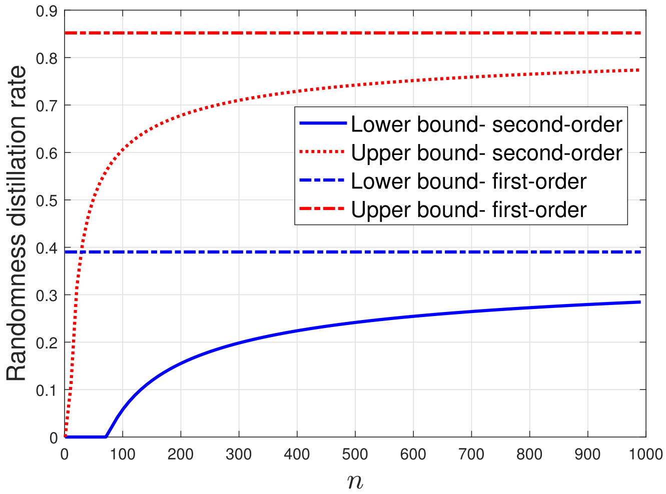

Fig. 3. For fixed $\varepsilon = 0.0001$, the plots depicts upper and lower bounds on the randomness distillation rate of an isotropic state $\rho_{AB} = (1-p)\Phi^d_{AB} + p\frac{I_{AB}}{d^2}$ with $p = 0.3$ and $d = 2$. The horizontal lines depict asymptotic values of the upper and lower bounds, while the curves depict the lower bound from Proposition 7 and the upper bound from Theorem 6.

measure her system A in the computational basis. Note that this measurement procedure has the same effect as applying a completely dephasing channel on system A. The resulting state is of the form in (245), and so then Proposition 7 applies for obtaining a lower bound on its distillable randomness. In Fig. 3, we show how the upper bound (from Theorem 6) and the lower bound (from Proposition 7) on the LOCC-assisted distillable randomness vary within the finite n regime for fixed $\varepsilon = 0.0001$, when choosing $\rho^{(d,p)}_{AB}$ with $p = 0.3$ and $d = 2$.

F. Applications to General Resource Theories

In this section, we discuss how our approach can be generalized beyond randomness distillation, to the distillation of mixed states in a general resource theory (see [2] for a review). Let us consider the following general scenario: let $\mathbb{O}$ denote the set of free channels, and let $\mathbb{F}$ denote the set of free states. Suppose that the goal of a protocol is to start from an arbitrary state ρ and apply a free channel $\mathcal{M} \in \mathbb{O}$, in order to approximately distill a state from the set $\{\tau^d\}_{d\in\mathbb{Z}^+}$, where d denotes the amount of the resource being distilled (a concrete example of such a set is the set of maximally classical correlated states, defined from (168)). The one-shot distillable resource of ρ is then defined for $\varepsilon \in [0, 1]$ as follows:

$$G^{\varepsilon}(\rho) := \sup_{\mathcal{M}\in\mathbb{O}} \{\log_2 d : F(\mathcal{M}(\rho), \tau^d) \geq 1 - \varepsilon\}. \tag{249}$$

Now suppose that the following generalization of (182) holds:

$$\sup_{\sigma\in\mathbb{F}} F(\tau^d, \sigma) \leq \frac{1}{d}. \tag{250}$$

Then the ideas presented in Section VI can be extended to this more general scenario. In particular, we arrive at the following upper bound on the one-shot distillable resource:

$$G^{\varepsilon}(\rho) \leq \inf_{\sigma\in\mathbb{F}} D^{\varepsilon}_{\min,F}(\rho\|\sigma), \tag{251}$$

which follows because

$$\begin{aligned}\log_2 d &\leq \inf_{\sigma'\in\mathbb{F}} -\log_2 F(\tau^d, \sigma') && (252)\\ &\leq \inf_{\sigma'\in\mathbb{F}} D^{\varepsilon}_{\min,F}(\mathcal{M}(\rho)\|\sigma') && (253)\\ &\leq \inf_{\sigma\in\mathbb{F}} D^{\varepsilon}_{\min,F}(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) && (254)\\ &\leq \inf_{\sigma\in\mathbb{F}} D^{\varepsilon}_{\min,F}(\rho\|\sigma), && (255)\end{aligned}$$

where the first inequality follows from (250), the second from the definition of the smooth F-min-relative entropy, the third inequality from the assumption that $\mathcal{M}$ is a free channel (thus preserving the set of free states), and the last inequality from the data-processing inequality for the smooth F-min-relative entropy (Theorem 1).

We can also obtain a second-order upper bound on the asymptotic distillable resource, by following the approach from Section VI-E. Indeed, let us define the following quantities:

$$D_{\mathbb{F}}(\rho) := \inf_{\sigma\in\mathbb{F}} D(\rho\|\sigma), \tag{256}$$

where $D(\rho\|\sigma)$ is the quantum relative entropy from (17). Then, denote $\Pi_{\mathbb{F}} \subseteq \mathbb{F}$ as the set of states achieving the infimum in $D_{\mathbb{F}}(\rho)$. From that, we define the following variance quantity:

$$V^{\varepsilon}_{\mathbb{F}}(\rho) := \begin{cases} \inf_{\sigma\in\Pi_{\mathbb{F}}} V(\rho\|\sigma) & \text{if } \varepsilon \geq \frac{1}{2}, \\ \sup_{\sigma\in\Pi_{\mathbb{F}}} V(\rho\|\sigma) & \text{if } \varepsilon < \frac{1}{2}. \end{cases} \tag{257}$$

Fix $\varepsilon \in (0, 1)$. Then the distillable resource of a state ρ is bounded from above as follows:

$$\frac{1}{n} G^{\varepsilon}(\rho^{\otimes n}) \leq D_{\mathbb{F}}(\rho) + \sqrt{\frac{1}{n} V^{\varepsilon}_{\mathbb{F}}(\rho)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{258}$$

VII. Computational Analysis

In this section, we provide techniques based on semi-definite programs (SDPs) to quantify the smooth F-min-relative entropy and other related quantities, including the smooth max-relative entropy and smooth conditional min-entropy.

First, we present a bilinear program to evaluate the smooth F-min-relative entropy, which follows from the SDP formulation of fidelity in [68].

Proposition 8: Given a state ρ, a PSD operator σ, and $\varepsilon \in (0, 1)$, the smooth F-min-relative entropy can be written in terms of the following optimization:

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) = -2\log_2 a^{\star}, \tag{259}$$

where

$$a^{\star} = \frac{1}{2} \inf_{\substack{\widetilde{\rho}\geq 0,\\ Y,Z\geq 0,\\ X\in\mathcal{L}(\mathcal{H})}} \left\{ \begin{array}{c} \operatorname{Tr}[Y\widetilde{\rho}] + \operatorname{Tr}[Z\sigma] : \\ \begin{bmatrix} Y & I \\ I & Z \end{bmatrix} \geq 0, \\ \operatorname{Re}[\operatorname{Tr}[X]] \geq \sqrt{1-\varepsilon}, \\ \begin{bmatrix} \widetilde{\rho} & X \\ X^{\dagger} & \rho \end{bmatrix} \geq 0, \\ \operatorname{Tr}[\widetilde{\rho}] \leq 1 \end{array} \right\}. \tag{260}$$

Proof: From the definition of the smooth F-min relative entropy, consider that

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) = -\log_2(a^\star)^2, \tag{261}$$

where

$$a^\star := \inf_{\widetilde{\rho}\in\mathcal{D}_{\leq}} \left\{\sqrt{F}(\widetilde{\rho},\sigma) : \sqrt{F}(\widetilde{\rho},\rho) \geq \sqrt{1-\varepsilon}\right\}. \tag{262}$$

Recall that the root fidelity $\sqrt{F}(\rho,\sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1$ has the following primal and dual SDP characterizations [68] (see also [1, Proposition 6.6]):

$$\sqrt{F}(\rho,\sigma) = \sup_{X\in\mathcal{L}(\mathcal{H})} \left\{\mathrm{Re}[\mathrm{Tr}[X]] : \begin{bmatrix}\rho & X\\ X^\dagger & \sigma\end{bmatrix} \geq 0\right\} \tag{263}$$

$$= \frac{1}{2}\inf_{Y,Z\geq 0} \left\{\mathrm{Tr}[Y\rho] + \mathrm{Tr}[Z\sigma] : \begin{bmatrix}Y & I\\ I & Z\end{bmatrix} \geq 0\right\}. \tag{264}$$

We then find the following expression for the negative root fidelity:

$$-\sqrt{F}(\rho,\widetilde{\rho}) = -\sup_{X\in\mathcal{L}(\mathcal{H})} \left\{\mathrm{Re}[\mathrm{Tr}[X]] : \begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix} \geq 0\right\} \tag{265}$$

$$= \inf_{X\in\mathcal{L}(\mathcal{H})} \left\{-\mathrm{Re}[\mathrm{Tr}[X]] : \begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix} \geq 0\right\}. \tag{266}$$

With these expressions in hand, consider from (262) and (264) that

$$a^\star = \frac{1}{2}\inf_{\widetilde{\rho}\geq 0, Y,Z\geq 0} \left\{\begin{array}{c}\mathrm{Tr}[Y\widetilde{\rho}] + \mathrm{Tr}[Z\rho] :\\ \sqrt{F}(\widetilde{\rho},\rho) \geq \sqrt{1-\varepsilon},\\ \begin{bmatrix}Y & I\\ I & Z\end{bmatrix} \geq 0,\\ \mathrm{Tr}[\widetilde{\rho}] \leq 1\end{array}\right\} \tag{267}$$

$$= \frac{1}{2}\inf_{\widetilde{\rho}\geq 0, Y,Z\geq 0} \left\{\begin{array}{c}\mathrm{Tr}[Y\widetilde{\rho}] + \mathrm{Tr}[Z\rho] :\\ -\sqrt{F}(\widetilde{\rho},\rho) \leq -\sqrt{1-\varepsilon},\\ \begin{bmatrix}Y & I\\ I & Z\end{bmatrix} \geq 0,\\ \mathrm{Tr}[\widetilde{\rho}] \leq 1\end{array}\right\} \tag{268}$$

$$= \frac{1}{2}\inf_{\substack{\widetilde{\rho}\geq 0,\\ Y,Z\geq 0,\\ X\in\mathcal{L}(\mathcal{H})}} \left\{\begin{array}{c}\mathrm{Tr}[Y\widetilde{\rho}] + \mathrm{Tr}[Z\sigma] :\\ \begin{bmatrix}Y & I\\ I & Z\end{bmatrix} \geq 0,\\ -\mathrm{Re}[\mathrm{Tr}[X]] \leq -\sqrt{1-\varepsilon},\\ \begin{bmatrix}\widetilde{\rho} & X\\ X^\dagger & \rho\end{bmatrix} \geq 0,\\ \mathrm{Tr}[\widetilde{\rho}] \leq 1\end{array}\right\}. \tag{269}$$

In the last line, we replaced the inequality $-\sqrt{F}(\widetilde{\rho},\rho) \leq -\sqrt{1-\varepsilon}$ with the optimization in (266), to conclude the proof. ■

Note that the optimization problem in (260) is not an SDP, but it is rather a bilinear program, due to the bilinear term $\mathrm{Tr}[Y\widetilde{\rho}]$ in the objective function in (260). Thus, to estimate the smooth F-min relative entropy, we can employ the seesaw or mountain-climbing algorithm [69], which results in a lower bound on the smooth F-min-relative entropy.

We apply the seesaw algorithm as follows. Set $k = 0$. For fixed $\widetilde{\rho} = \widetilde{\rho}_k$, the optimization in (260) is an SDP with the additional constraint $\widetilde{\rho} = \widetilde{\rho}_k$. By solving that SDP, we can find a Y that achieves the optimum for the objective function, which we denote as Y_k. Next, by fixing $Y = Y_k$, we solve the respective SDP and find the optimum $\widetilde{\rho}_{k+1}$. Then, this iterative process is continued for a fixed number of iterations. In a finite (yet not necessarily polynomial) number of iterations, the algorithm converges to the optimum value [69]. In practice, it is guaranteed that we arrive at an upper bound on (260) by this method, which will in turn provide a lower bound on the smooth F-min-relative entropy. Note that one can employ the more advanced algorithm from [70] as well for this purpose.

We use the seesaw method described above to investigate the tightness of the lower bound from Corollary 3. In particular, we obtain various upper bounds for the smooth F-min-relative entropy by varying δ in

$$D^{\varepsilon}_{\min,F}(\rho\|\sigma) \leq D^{1-\varepsilon-\delta}_{\max}(\rho\|\sigma) + \log_2\left(\frac{1}{1-f(\varepsilon,\delta)}\right), \tag{270}$$

where $f(\varepsilon,\delta)$ is given in (151).

To compute the upper bound in (270), it is required to compute the smooth max-relative entropy. For this purpose, we derive an SDP for the smooth max-relative entropy with fidelity smoothing, which may be of independent interest. We consider two variants of that: $\widehat{D}^{\varepsilon}_{\max}(\rho\|\sigma)$ with smoothing over normalized states and $D^{\varepsilon}_{\max}(\rho\|\sigma)$ with smoothing over sub-normalized states. Then, we use the latter variant to compute the right-hand side of (270).

Proposition 9: The smooth max-relative entropy with fidelity smoothing over normalized states has the following primal and dual SDP characterizations for a state ρ and PSD operator σ:

$$\widehat{D}^{\varepsilon}_{\max}(\rho\|\sigma)$$

$$= \log_2 \inf_{\widetilde{\rho}\geq 0, \lambda\geq 0, X\in\mathcal{L}(\mathcal{H})} \left\{\begin{array}{c}\lambda : \widetilde{\rho} \leq \lambda\sigma, \mathrm{Tr}[\widetilde{\rho}] = 1,\\ \mathrm{Re}[\mathrm{Tr}[X]] \geq \sqrt{1-\varepsilon},\\ \begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix} \geq 0\end{array}\right\} \tag{271}$$

$$= \log_2 \sup_{W,\nu,Z\geq 0, \mu\in\mathbb{R}} \left\{\begin{array}{c}\mu + 2\nu\sqrt{1-\varepsilon} - \mathrm{Tr}[Z\rho] :\\ \mathrm{Tr}[W\sigma] \leq 1,\\ \begin{bmatrix}Z & \nu I\\ \nu I & W - \mu I\end{bmatrix} \geq 0\end{array}\right\}. \tag{272}$$

Proof: See Appendix D. ■

Following a similar approach, we obtain an SDP for $D^{\varepsilon}_{\max}(\rho\|\sigma)$.

Proposition 10: The smooth max-relative entropy with fidelity smoothing over sub-normalized states has the following primal and dual SDP characterizations for a state ρ, PSD operator σ, and $\varepsilon \in (0,1)$:

$$D^{\varepsilon}_{\max}(\rho\|\sigma)$$

$$= \log_2 \inf_{\widetilde{\rho}\geq 0, \lambda\geq 0, X\in\mathcal{L}(\mathcal{H})} \left\{\begin{array}{c}\lambda : \widetilde{\rho} \leq \lambda\sigma, \mathrm{Tr}[\widetilde{\rho}] \leq 1,\\ \mathrm{Re}[\mathrm{Tr}[X]] \geq \sqrt{1-\varepsilon},\\ \begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix} \geq 0\end{array}\right\}, \tag{273}$$

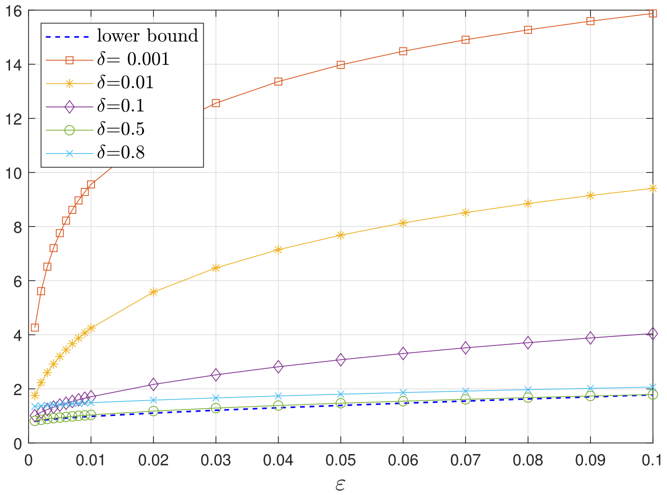

Fig. 4. This plot shows the tightness of the lower bound on $D^{\varepsilon}_{\min,F}(\rho\|\sigma)$ obtained from running the seesaw algorithm for 10 iterations for each ε, where ρ, σ are random quantum states in a Hilbert space of dimension 2. By changing δ, the upper bound $D^{1-\varepsilon-\delta}_{\max}(\rho\|\sigma) + \log_2\left(\frac{1}{1-f(\varepsilon,\delta)}\right)$ is also shown.

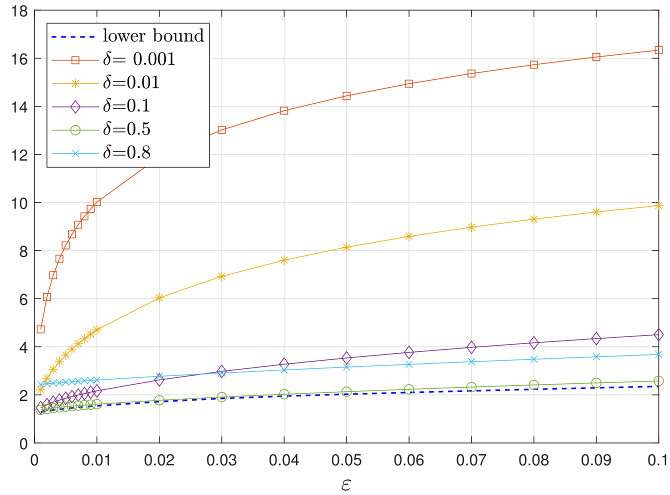

Fig. 5. This plot shows the tightness of the lower bound on $D^{\varepsilon}_{\min,F}(\rho\|\sigma)$ obtained from running the seesaw algorithm for ten iterations for each ε, where ρ, σ are random quantum states in a Hilbert space of dimension 4. By changing δ, the upper bound $D^{1-\varepsilon-\delta}_{\max}(\rho\|\sigma) + \log_2\left(\frac{1}{1-f(\varepsilon,\delta)}\right)$ is also shown.

$$= \log_2 \sup_{W,\nu,Z,\mu\geq 0} \left\{ \begin{array}{c} -\mu + 2\nu\sqrt{1-\varepsilon} - \mathrm{Tr}[Z\rho] : \\ \mathrm{Tr}[W\sigma] \leq 1, \\ \begin{bmatrix} Z & \nu I \\ \nu I & W+\mu I \end{bmatrix} \geq 0 \end{array} \right\}. \tag{274}$$

Proof: We omit the proof since it is similar to the proof of Proposition 9, except that we remove the constraint $\mathrm{Tr}[\tilde{\rho}] \geq 1$. ∎

Utilizing the methods explained above (specifically in Propositions 8 and 10), Figure 4 shows the tightness of the lower bound on the smooth F-min-relative entropy of two random quantum states in a Hilbert space of dimension two, by employing the seesaw method for ten iterations. In addition, Figure 5 presents the obtained results under the same setting, for two random quantum states in a Hilbert space of dimension four.

Due to the close link between smooth max-relative entropy and smooth conditional min-entropy, our SDP results apply to this latter quantity as well. We highlight this observation below, as it may be of independent interest for future work.

Remark 11 (Conditional smooth min-entropy): By Definitions 11 and 12 of [71], for $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$, the smooth conditional min-entropy is defined as

$$H^{\varepsilon}_{\min}(A|B)_{\rho} := \max_{\widetilde{\rho}_{AB}\in\mathcal{B}^{\varepsilon}(\rho_{AB})} \max_{\sigma_B\in\mathcal{D}(\mathcal{H}_B)} -D_{\max}(\widetilde{\rho}_{AB}\|I_A\otimes\sigma_B), \tag{275}$$

where

$$\mathcal{B}^{\varepsilon}(\rho) := \{\tau \in \mathcal{D}_{\leq} : P(\tau,\rho) \leq \varepsilon\}, \tag{276}$$

with $P(\tau,\rho) := \sqrt{1-F(\tau,\rho)}$. Note that we can equivalently define smooth conditional min-entropy by

$$H^{\varepsilon}_{\min}(A|B)_{\rho} := -\min_{\sigma_B\in\mathcal{D}(\mathcal{H}_B)} D^{\varepsilon^2}_{\max}(\rho_{AB}\|I_A\otimes\sigma_B), \tag{277}$$

with the association of smooth max-relative entropy in (23).

An SDP for $H^{\varepsilon}_{\min}(A|B)_{\rho}$ was first given in [72] if a purification of ρ_{AB} is available. In this work, Proposition 10 together with (277) lead to an SDP for smooth conditional min-entropy directly in terms of the state ρ_{AB}, as presented in the next corollary.

Corollary 5: Let ρ_{AB} be a state and fix $\varepsilon \in (0,1)$. The smooth conditional min-entropy has the following SDP characterization:

$$H^{\varepsilon}_{\min}(A|B)_{\rho}$$

$$= -\log_2 \inf_{\substack{\widetilde{\rho}_{AB}\geq 0,\\ S_B\geq 0,\\ X_{AB}\in\mathcal{L}(\mathcal{H})}} \left\{ \begin{array}{c} \mathrm{Tr}[S_B] : \widetilde{\rho}_{AB} \leq I_A\otimes S_B, \\ \mathrm{Tr}[\widetilde{\rho}_{AB}] \leq 1, \\ \mathrm{Re}[\mathrm{Tr}[X_{AB}]] \geq \sqrt{1-\varepsilon^2}, \\ \begin{bmatrix} \rho_{AB} & X_{AB} \\ X^{\dagger}_{AB} & \widetilde{\rho}_{AB} \end{bmatrix} \geq 0 \end{array} \right\} \tag{278}$$

$$= -\log_2 \sup_{\substack{W_{AB}\geq 0,\\ Z_{AB}\geq 0,\\ \nu\geq 0,\mu\geq 0}} \left\{ \begin{array}{c} -\mu + 2\nu\sqrt{1-\varepsilon^2} - \mathrm{Tr}[Z_{AB}\rho_{AB}] : \\ \mathrm{Tr}_A[W_{AB}] \leq I_B, \\ \begin{bmatrix} Z_{AB} & \nu I_{AB} \\ \nu I_{AB} & W_{AB}+\mu I_{AB} \end{bmatrix} \geq 0 \end{array} \right\}. \tag{279}$$

Proof: We can write

$$H^{\varepsilon}_{\min}(A|B)_{\rho}$$

$$= -\log_2 \inf_{\substack{\widetilde{\rho}_{AB}\geq 0,\\ \sigma_B\geq 0,\\ \lambda\geq 0,\\ X_{AB}\in\mathcal{L}(\mathcal{H})}} \left\{ \begin{array}{c} \lambda : \widetilde{\rho}_{AB} \leq \lambda I_A\otimes\sigma_B, \\ \mathrm{Tr}[\widetilde{\rho}_{AB}] \leq 1, \\ \mathrm{Re}[\mathrm{Tr}[X_{AB}]] \geq \sqrt{1-\varepsilon^2}, \\ \begin{bmatrix} \rho_{AB} & X_{AB} \\ X^{\dagger}_{AB} & \widetilde{\rho}_{AB} \end{bmatrix} \geq 0, \\ \mathrm{Tr}[\sigma_B] = 1 \end{array} \right\} \tag{280}$$

as a direct result of the SDP for smooth max-relative entropy in Proposition 10 and its connection to smooth conditional min-entropy in (277). The characterization in (278) follows from the substitution $S_B = \lambda\sigma_B$ in (280). The dual SDP for $H^{\varepsilon}_{\min}(A|B)_\rho$ in (279) is proved in Appendix E. ∎

VIII. Conclusion

In conclusion, we provided a comprehensive study of the fidelity-based smooth min-relative entropy. In particular, we proved some of its basic properties, including a proof of data processing independent from that of [14, Theorem 3], which is of interest for several operational tasks. With the use of the derived relationships between smooth F-min-relative entropy and other quantum information-theoretic quantities, we derived the second-order asymptotics of the smooth F-min-relative entropy and all smooth sandwiched Rényi relative entropies. We explored applications where this quantity arises, with a focus on randomness distillation, establishing an upper bound on the one-shot distillable randomness that improves upon the prior one from [27]. Furthermore, we obtained a second-order expansion of this upper bound, as well as the precise second-order asymptotics of the distillable randomness of particular classical–quantum states. Lastly, we designed methods to estimate the smooth F-min-relative entropy and showed that the estimates are sufficiently tight for some examples. To that end, we presented SDPs to compute the smooth max-relative entropy and smooth conditional min-entropy, which may be of independent interest.

Some future research directions are as follows. It would be interesting to analyze the third-order asymptotics and large deviations associated with the smooth F-min-relative entropy. It might also be possible to devise efficient computational methods for it and tighter connections to other information-theoretic quantities. To that end, understanding which parameter ε_1 achieves the tightest bound in Corollary 3 when ε_2 is fixed is a possible direction. We have shown here that the smooth F-min-relative entropy is the core quantity underlying an upper bound on the one-shot distillable randomness, but finding an information-theoretic task that provides an operational interpretation of the smooth F-min-relative entropy itself is an open research question.

Another interesting future direction is to follow the observation made after (8) in the introduction. Indeed, we see that the main difference between the hypothesis testing relative entropy and the smooth F-min relative entropy is the replacement of the quantities $\mathrm{Tr}[\Lambda\rho]$ and $\mathrm{Tr}[\Lambda\sigma]$ with $F(\tilde{\rho},\rho)$ and $F(\tilde{\rho},\sigma)$, respectively. Since $\mathrm{Tr}[\Lambda\rho]$ and $\mathrm{Tr}[\Lambda\sigma]$ are related to Type I and Type II error probabilities in quantum hypothesis testing, one could consider a variant of quantum hypothesis testing in which the "error probabilities" are then related to $F(\tilde{\rho},\rho)$ and $F(\tilde{\rho},\sigma)$. Specifically, one could consider $1 - F(\tilde{\rho},\rho)$ to be analogous to a Type I error probability and $F(\tilde{\rho},\sigma)$ to be analogous to a Type II error probability. Under this perspective, Theorem 4 demonstrates that we have already determined the second-order asymptotics of this variant of the traditional asymmetric quantum hypothesis testing task and that they are the same as the second-order asymptotics in the standard setting of asymmetric quantum hypothesis testing. What remains open is to determine the asymptotics of a variant of symmetric hypothesis testing. That is, for fixed $\lambda \in (0,1)$, what is the following quantity equal to?

$$\lim_{n\to\infty} -\frac{1}{n}\ln p_e^n \tag{281}$$

where

$$p_e^n := \inf_{\tilde{\rho}^{(n)}\in\mathcal{D}_\leq} \lambda(1 - F(\tilde{\rho}^{(n)},\rho^{\otimes n})) + (1-\lambda)F(\tilde{\rho}^{(n)},\sigma^{\otimes n}). \tag{282}$$

Based on known results in symmetric quantum hypothesis testing [73], [74] and the aforementioned coincidence for the asymmetric setting, one might guess that (281) would be equal to the quantum Chernoff divergence, but this remains an intriguing open question for future work.

Appendix A
Supplementary Lemmas

Lemma 5: Let ρ be a state, Λ a measurement operator, and set $\widetilde{\rho} := \frac{\sqrt{\Lambda}\rho\sqrt{\Lambda}}{\mathrm{Tr}[\Lambda\rho]}$. Then

$$F(\widetilde{\rho},\rho) \geq \mathrm{Tr}[\Lambda\rho]. \tag{283}$$

Proof: Consider that

$$\begin{aligned}
F(\widetilde{\rho},\rho) &= \left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\widetilde{\rho}\sqrt{\rho}}\right]\right)^2 &(284)\\
&= \frac{1}{\mathrm{Tr}[\Lambda\rho]}\left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sqrt{\Lambda}\rho\sqrt{\Lambda}\sqrt{\rho}}\right]\right)^2 &(285)\\
&= \frac{1}{\mathrm{Tr}[\Lambda\rho]}\left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sqrt{\Lambda}\sqrt{\rho}\sqrt{\rho}\sqrt{\Lambda}\sqrt{\rho}}\right]\right)^2 &(286)\\
&= \frac{1}{\mathrm{Tr}[\Lambda\rho]}\left(\mathrm{Tr}\left[\sqrt{\left(\sqrt{\rho}\sqrt{\Lambda}\sqrt{\rho}\right)^2}\right]\right)^2 &(287)\\
&= \frac{1}{\mathrm{Tr}[\Lambda\rho]}\left(\mathrm{Tr}\left[\sqrt{\rho}\sqrt{\Lambda}\sqrt{\rho}\right]\right)^2 &(288)\\
&= \frac{1}{\mathrm{Tr}[\Lambda\rho]}\left(\mathrm{Tr}\left[\sqrt{\Lambda}\rho\right]\right)^2 &(289)\\
&\geq \frac{1}{\mathrm{Tr}[\Lambda\rho]}\left(\mathrm{Tr}\left[\Lambda\rho\right]\right)^2 &(290)\\
&= \mathrm{Tr}[\Lambda\rho], &(291)
\end{aligned}$$

concluding the proof. ∎

Lemma 6: Let ρ be a state, Λ a measurement operator, and set $\widetilde{\rho} := \frac{\sqrt{\Lambda}\rho\sqrt{\Lambda}}{\mathrm{Tr}[\Lambda\rho]}$. Then

$$F(\widetilde{\rho},\sigma) = \frac{1}{\mathrm{Tr}[\Lambda\rho]}F(\rho,\sqrt{\Lambda}\sigma\sqrt{\Lambda}). \tag{292}$$

Proof: Let $|\psi^\sigma\rangle$ purify σ, and let $|\psi^\rho\rangle$ purify ρ. Let U denote a unitary acting on the purifying system. Consider from Uhlmann's theorem [4] that

$$\begin{aligned}
&F(\widetilde{\rho},\sigma)\\
&= \sup_U \frac{1}{\mathrm{Tr}[\Lambda\rho]}\left|\langle\psi^\rho|\left(\sqrt{\Lambda}\otimes I\right)(I\otimes U)|\psi^\sigma\rangle\right|^2 &(293)\\
&= \frac{1}{\mathrm{Tr}[\Lambda\rho]}\sup_U \left|\langle\psi^\rho|\sqrt{\Lambda}\otimes U|\psi^\sigma\rangle\right|^2 &(294)\\
&= \frac{1}{\mathrm{Tr}[\Lambda\rho]}\sup_U \left|\langle\psi^\rho|(I\otimes U)\left(\sqrt{\Lambda}\otimes I\right)|\psi^\sigma\rangle\right|^2 &(295)
\end{aligned}$$

$$= \frac{1}{\mathrm{Tr}[\Lambda\rho]} F(\rho, \sqrt{\Lambda}\sigma\sqrt{\Lambda}). \tag{296}$$

An alternative way of seeing this follows from

$$F(\widetilde{\rho}, \sigma) = \left(\mathrm{Tr}\left[\sqrt{\sqrt{\sigma}\widetilde{\rho}\sqrt{\sigma}}\right]\right)^2 \tag{297}$$
$$= \frac{1}{\mathrm{Tr}[\Lambda\rho]} \left(\mathrm{Tr}\left[\sqrt{\sqrt{\sigma}\sqrt{\Lambda}\rho\sqrt{\Lambda}\sqrt{\sigma}}\right]\right)^2 \tag{298}$$
$$= \frac{1}{\mathrm{Tr}[\Lambda\rho]} \left(\mathrm{Tr}\left[\sqrt{\sqrt{\sigma}\sqrt{\Lambda}\sqrt{\rho}\sqrt{\rho}\sqrt{\Lambda}\sqrt{\sigma}}\right]\right)^2 \tag{299}$$
$$= \frac{1}{\mathrm{Tr}[\Lambda\rho]} \left(\left\|\sqrt{\sigma}\sqrt{\Lambda}\sqrt{\rho}\right\|_1\right)^2 \tag{300}$$
$$= \frac{1}{\mathrm{Tr}[\Lambda\rho]} \left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sqrt{\Lambda}\sqrt{\sigma}\sqrt{\sigma}\sqrt{\Lambda}\sqrt{\rho}}\right]\right)^2 \tag{301}$$
$$= \frac{1}{\mathrm{Tr}[\Lambda\rho]} \left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sqrt{\Lambda}\sigma\sqrt{\Lambda}\sqrt{\rho}}\right]\right)^2 \tag{302}$$
$$= \frac{1}{\mathrm{Tr}[\Lambda\rho]} F(\rho, \sqrt{\Lambda}\sigma\sqrt{\Lambda}). \tag{303}$$

This concludes the proof. ■

Appendix B
Proof of Proposition 2

The proof is quite similar to the proof of the Proposition 1, but we give it here for completeness. Let Λ be an arbitrary measurement operator satisfying $\mathrm{Tr}[\Lambda\rho] \geq 1-\varepsilon$. By the gentle measurement lemma, we know that $\mathrm{Tr}[\Lambda\rho] \geq 1-\varepsilon$ implies that

$$F(\widetilde{\rho}, \rho) \geq 1 - \varepsilon(2-\varepsilon), \tag{304}$$

where $\widetilde{\rho} = \sqrt{\Lambda}\rho\sqrt{\Lambda}$. To see the inequality in (304), consider that

$$F(\widetilde{\rho}, \rho) = \left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\widetilde{\rho}\sqrt{\rho}}\right]\right)^2 \tag{305}$$
$$= \left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sqrt{\Lambda}\rho\sqrt{\Lambda}\sqrt{\rho}}\right]\right)^2 \tag{306}$$
$$= \left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sqrt{\Lambda}\sqrt{\rho}\sqrt{\rho}\sqrt{\Lambda}\sqrt{\rho}}\right]\right)^2 \tag{307}$$
$$= \left(\mathrm{Tr}\left[\sqrt{\left(\sqrt{\rho}\sqrt{\Lambda}\sqrt{\rho}\right)^2}\right]\right)^2 \tag{308}$$
$$= \left(\mathrm{Tr}\left[\sqrt{\rho}\sqrt{\Lambda}\sqrt{\rho}\right]\right)^2 \tag{309}$$
$$= \left(\mathrm{Tr}\left[\sqrt{\Lambda}\rho\right]\right)^2 \tag{310}$$
$$\geq (\mathrm{Tr}[\Lambda\rho])^2 \tag{311}$$
$$\geq (1-\varepsilon)^2 \tag{312}$$
$$= 1 - \varepsilon\,(2-\varepsilon)\,. \tag{313}$$

Now we should relate $\mathrm{Tr}[\Lambda\sigma]$ to $F(\widetilde{\rho}, \sigma)$. Consider from Uhlmann's theorem [4] that

$$F(\widetilde{\rho}, \sigma) = \sup_U \left|\langle\psi^\rho|\left(\sqrt{\Lambda}\otimes I\right)(I\otimes U)\,|\psi^\sigma\rangle\right|^2 \tag{314}$$
$$= \sup_U \left|\langle\psi^\rho|\sqrt{\Lambda}\otimes U|\psi^\sigma\rangle\right|^2 \tag{315}$$
$$= \sup_U \left|\langle\psi^\rho|\,(I\otimes U)\left(\sqrt{\Lambda}\otimes I\right)|\psi^\sigma\rangle\right|^2 \tag{316}$$
$$= F(\rho, \sqrt{\Lambda}\sigma\sqrt{\Lambda}) \tag{317}$$
$$\leq \mathrm{Tr}[\Lambda\sigma]. \tag{318}$$

The last inequality follows from data processing for fidelity under the trace channel. Then

$$-\log_2 \mathrm{Tr}[\Lambda\sigma] \leq -\log_2 F(\widetilde{\rho}, \sigma) \tag{319}$$
$$\leq D_{\min,F}^{\varepsilon(2-\varepsilon)}(\rho\|\sigma). \tag{320}$$

Since this holds for an arbitrary measurement operator satisfying $\mathrm{Tr}[\Lambda\rho] \geq 1-\varepsilon$, we conclude that

$$D_{\min}^{\varepsilon}(\rho\|\sigma) \leq D_{\min,F}^{\varepsilon(2-\varepsilon)}(\rho\|\sigma), \tag{321}$$

which is the desired statement.

Appendix C
Properties of Max-Mutual Information

In this appendix, we prove the equality in (177), we establish the bound $I_{\max}(A;B)_\rho \leq 2\log_2 \min\{d_A, d_B\}$, and we prove the equality $I_{\max}(A;B)_\Phi = 2\log_2 d$ for a maximally entangled state Φ_{AB} of Schmidt rank d. To begin, consider that

$$I_{\max}(A;B)_\rho$$
$$= \inf_{\substack{\sigma_A\in\mathcal{D}(\mathcal{H}_A),\\ \sigma_B\in\mathcal{D}(\mathcal{H}_B)}} D_{\max}(\rho_{AB}\|\sigma_A\otimes\sigma_B) \tag{322}$$
$$= \log_2 \inf_{\substack{\lambda\geq 0,\ \sigma_A\in\mathcal{D}(\mathcal{H}_A),\\ \sigma_B\in\mathcal{D}(\mathcal{H}_B)}} \{\lambda : \rho_{AB} \leq \lambda\sigma_A\otimes\sigma_B\} \tag{323}$$
$$= \log_2 \inf_{K_A, L_B\geq 0} \{\mathrm{Tr}[K_A\otimes L_B] : \rho_{AB} \leq K_A\otimes L_B\}, \tag{324}$$

where the last equality follows from the substitution $\lambda\sigma_A\otimes\sigma_B = K_A\otimes L_B$ and the fact that $\mathrm{Tr}[K_A\otimes L_B] = \mathrm{Tr}[\lambda\sigma_A\otimes\sigma_B] = \lambda$. Now consider that, for every state ρ_{AB}, the operator inequality $\rho_{AB} \leq d_A I_A\otimes\rho_B$ holds, because

$$\frac{1}{d_A^2}\rho_{AB} \leq \frac{1}{d_A^2}\sum_{i=1}^{d_A^2} U_A^i\rho_{AB}\left(U_A^i\right)^\dagger = \frac{I_A}{d_A}\otimes\rho_B, \tag{325}$$

where $\{U_A^i\}_{i=1}^{d_A^2}$ is a set of Heisenberg–Weyl unitaries. So this implies that the choices $K_A = d_A I_A$ and $L_B = \rho_B$ are feasible, leading to the claimed upper bound $I_{\max}(A;B)_\rho \leq 2\log_2 \min d_A$. By a symmetric argument, the following bound holds $I_{\max}(A;B)_\rho \leq 2\log_2 \min d_B$. Finally, for a maximally entangled state, we have $I_{\max}(A;B)_\Phi \leq 2\log_2 d$ by the upper bound just derived. We also have

$$I_{\max}(A;B)_\Phi = \inf_{\substack{\sigma_A\in\mathcal{D}(\mathcal{H}_A),\\ \sigma_B\in\mathcal{D}(\mathcal{H}_B)}} D_{\max}(\Phi_{AB}\|\sigma_A\otimes\sigma_B) \tag{326}$$
$$\geq \inf_{\substack{\sigma_A\in\mathcal{D}(\mathcal{H}_A),\\ \sigma_B\in\mathcal{D}(\mathcal{H}_B)}} D(\Phi_{AB}\|\sigma_A\otimes\sigma_B) \tag{327}$$
$$= I(A;B)_\Phi \tag{328}$$
$$= 2\log_2 d, \tag{329}$$

where the inequality follows from (14) and the second equality from [40, Exercise 11.8.2]. So we conclude that $I_{\max}(A;B)_\Phi = 2\log_2 d$.

Appendix D
Proof of Proposition 9

First, let us verify that strong duality holds for the primal and dual SDPs. Consider the following feasible choices for the primal SDP: $\widetilde{\rho}=\rho$, λ such that $\rho \leq \lambda\sigma$ (one possible choice would be $\lambda = 2^{\widehat{D}_{\max}(\rho\|\sigma)}$), and $X = \sqrt{1-\varepsilon}\,\rho$. This follows because

$$\begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix} = \begin{bmatrix}\rho & \sqrt{1-\varepsilon}\,\rho\\ \sqrt{1-\varepsilon}\,\rho & \rho\end{bmatrix} \tag{330}$$
$$= \begin{bmatrix}1 & \sqrt{1-\varepsilon}\\ \sqrt{1-\varepsilon} & 1\end{bmatrix}\otimes\rho \tag{331}$$
$$\geq 0. \tag{332}$$

In addition, choosing μ, ν, W, and Z to satisfy $\mu>0$, $\nu>0$, $\mu+\nu<1/\operatorname{Tr}[\sigma]$, $W=(\mu+\nu)I$, and $Z=\nu I$ leads to strictly feasible choices for the dual program. Thus, strong duality holds, due to Slater's condition.

Considering the known SDP for negative root fidelity [68], we have

$$-\sqrt{F}(\rho,\widetilde{\rho}) = -\sup_{X\in\mathcal{L}(\mathcal{H})}\left\{\operatorname{Re}[\operatorname{Tr}[X]] : \begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix}\geq 0\right\} \tag{333}$$
$$= \inf_{X\in\mathcal{L}(\mathcal{H})}\left\{-\operatorname{Re}[\operatorname{Tr}[X]] : \begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix}\geq 0\right\}. \tag{334}$$

With that we find that

$$\widehat{D}^{\varepsilon}_{\max}(\rho\|\sigma)$$
$$= \log_2 \inf_{\substack{\widetilde{\rho}\geq 0,\lambda\geq 0,\\ X\in\mathcal{L}(\mathcal{H})}}\left\{\begin{array}{c}\lambda : \widetilde{\rho}\leq\lambda\sigma, \operatorname{Tr}[\widetilde{\rho}]=1,\\ -\operatorname{Re}[\operatorname{Tr}[X]] \leq -\sqrt{1-\varepsilon},\\ \begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix}\geq 0\end{array}\right\} \tag{335}$$
$$= \log_2 \inf_{\substack{\widetilde{\rho}\geq 0,\lambda\geq 0,\\ X\in\mathcal{L}(\mathcal{H})}}\left\{\begin{array}{c}\lambda : \widetilde{\rho}\leq\lambda\sigma, \operatorname{Tr}[\widetilde{\rho}]=1,\\ \operatorname{Re}[\operatorname{Tr}[X]] \geq \sqrt{1-\varepsilon},\\ \begin{bmatrix}\rho & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix}\geq 0\end{array}\right\}. \tag{336}$$

Then, recall the standard form of dual and primal SDPs:

$$\sup_{Z\geq 0}\left\{\operatorname{Tr}[AZ] : \Phi(Z)\leq B\right\}, \tag{337}$$
$$\inf_{Y\geq 0}\left\{\operatorname{Tr}[BY] : \Phi^\dagger(Y)\geq A\right\}. \tag{338}$$

To that end, we find that

$$Y = \begin{bmatrix}\lambda & 0 & 0\\ 0 & Z & X\\ 0 & X^\dagger & \widetilde{\rho}\end{bmatrix},\quad B = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}, \tag{339}$$

$$\Phi^\dagger(Y) = \begin{bmatrix}\lambda\sigma-\widetilde{\rho} & 0 & 0 & 0 & 0 & 0\\ 0 & \operatorname{Tr}[\widetilde{\rho}] & 0 & 0 & 0 & 0\\ 0 & 0 & -\operatorname{Tr}[\widetilde{\rho}] & 0 & 0 & 0\\ 0 & 0 & 0 & \operatorname{Re}[\operatorname{Tr}[X]] & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & X\\ 0 & 0 & 0 & 0 & X^\dagger & \widetilde{\rho}\end{bmatrix}, \tag{340}$$

$$A = \begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & -1 & 0 & 0 & 0\\ 0 & 0 & 0 & \sqrt{1-\varepsilon} & 0 & 0\\ 0 & 0 & 0 & 0 & -\rho & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}. \tag{341}$$

Setting

$$Z = \begin{bmatrix}W & 0 & 0 & 0 & 0 & 0\\ 0 & \mu_1 & 0 & 0 & 0 & 0\\ 0 & 0 & \mu_2 & 0 & 0 & 0\\ 0 & 0 & 0 & \nu & 0 & 0\\ 0 & 0 & 0 & 0 & Z_1 & V\\ 0 & 0 & 0 & 0 & V^\dagger & Z_2\end{bmatrix}, \tag{342}$$

we find that

$$\begin{aligned}&\operatorname{Tr}[Z\Phi^\dagger(Y)]\\ &= \operatorname{Tr}[W(\lambda\sigma-\widetilde{\rho})] + (\mu_1-\mu_2)\operatorname{Tr}[\widetilde{\rho}] + \nu\operatorname{Re}[\operatorname{Tr}[X]]\\ &\qquad + \operatorname{Tr}\left[\begin{bmatrix}Z_1 & V\\ V^\dagger & Z_2\end{bmatrix}\begin{bmatrix}0 & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix}\right]\end{aligned} \tag{343}$$
$$\begin{aligned}&= \lambda\operatorname{Tr}[W\sigma] + \operatorname{Tr}[((\mu_1-\mu_2)I - W)\widetilde{\rho}]\\ &\qquad + \nu\operatorname{Re}[\operatorname{Tr}[X]] + \operatorname{Tr}\left[VX^\dagger + V^\dagger X + Z_2\widetilde{\rho}\right]\end{aligned} \tag{344}$$
$$\begin{aligned}&= \lambda\operatorname{Tr}[W\sigma] + \operatorname{Tr}[((\mu_1-\mu_2)I - W + Z_2)\widetilde{\rho}]\\ &\qquad + \operatorname{Re}[\operatorname{Tr}[(\nu I + 2V^\dagger)X]]\end{aligned} \tag{345}$$
$$\begin{aligned}&= \operatorname{Tr}\left[\begin{bmatrix}0 & 0\\ 0 & (\mu_1-\mu_2)I - W + Z_2\end{bmatrix}\begin{bmatrix}Z & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix}\right]\\ &\quad + \operatorname{Tr}\left[\begin{bmatrix}0 & \frac{\nu}{2}I+V\\ \frac{\nu}{2}I+V^\dagger & 0\end{bmatrix}\begin{bmatrix}Z & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix}\right] + \lambda\operatorname{Tr}[W\sigma]\end{aligned} \tag{346}$$
$$\begin{aligned}&= \operatorname{Tr}\left[\begin{bmatrix}0 & \frac{\nu}{2}I+V\\ \frac{\nu}{2}I+V^\dagger & (\mu_1-\mu_2)I - W + Z_2\end{bmatrix}\begin{bmatrix}Z & X\\ X^\dagger & \widetilde{\rho}\end{bmatrix}\right]\\ &\qquad + \lambda\operatorname{Tr}[W\sigma].\end{aligned} \tag{347}$$

This implies that

$$\Phi(Z) = \begin{bmatrix}\operatorname{Tr}[W\sigma] & 0\\ 0 & \begin{bmatrix}0 & \frac{\nu}{2}I+V\\ \frac{\nu}{2}I+V^\dagger & (\mu_1-\mu_2)I-W+Z_2\end{bmatrix}\end{bmatrix} \tag{348}$$

and we find that the dual SDP is given by

$$\sup_{Z\geq 0}\left\{\operatorname{Tr}[AZ] : \Phi(Z)\leq B\right\}$$
$$= \sup_{W,\mu_1,\mu_2,\nu\geq 0}\left\{\begin{array}{c}(\mu_1-\mu_2) + \nu\sqrt{1-\varepsilon} - \operatorname{Tr}[Z_1\rho] :\\ \operatorname{Tr}[W\sigma]\leq 1,\\ \begin{bmatrix}0 & \frac{\nu}{2}I+V\\ \frac{\nu}{2}I+V^\dagger & (\mu_1-\mu_2)I - W + Z_2\end{bmatrix}\leq 0,\\ \begin{bmatrix}Z_1 & V\\ V^\dagger & Z_2\end{bmatrix}\geq 0\end{array}\right\} \tag{349}$$
$$= \sup_{W,\nu\geq 0,\mu\in\mathbb{R}}\left\{\begin{array}{c}\mu + \nu\sqrt{1-\varepsilon} - \operatorname{Tr}[Z_1\rho] :\\ \operatorname{Tr}[W\sigma]\leq 1,\\ \begin{bmatrix}0 & \frac{\nu}{2}I+V\\ \frac{\nu}{2}I+V^\dagger & \mu I - W + Z_2\end{bmatrix}\leq 0,\\ \begin{bmatrix}Z_1 & V\\ V^\dagger & Z_2\end{bmatrix}\geq 0\end{array}\right\} \tag{350}$$

$$= \sup_{W,\nu\geq 0,\mu\in\mathbb{R}} \left\{ \begin{array}{c} \mu + 2\nu\sqrt{1-\varepsilon} - \operatorname{Tr}[Z_1\rho] : \\ \operatorname{Tr}[W\sigma] \leq 1, \\ \begin{bmatrix} 0 & \nu I + V \\ \nu I + V^\dagger & \mu I - W + Z_2 \end{bmatrix} \leq 0, \\ \begin{bmatrix} Z_1 & V \\ V^\dagger & Z_2 \end{bmatrix} \geq 0 \end{array} \right\} \tag{351}$$

Consider that

$$\begin{aligned} &\begin{bmatrix} 0 & \nu I + V \\ \nu I + V^\dagger & \mu I - W + Z_2 \end{bmatrix} \leq 0 \\ &\Leftrightarrow \quad \begin{bmatrix} 0 & V \\ V^\dagger & Z_2 \end{bmatrix} \leq \begin{bmatrix} 0 & -\nu I \\ -\nu I & W - \mu I \end{bmatrix} \qquad (352) \\ &\Leftrightarrow \quad \begin{bmatrix} Z_1 & V \\ V^\dagger & Z_2 \end{bmatrix} \leq \begin{bmatrix} Z_1 & -\nu I \\ -\nu I & W - \mu I \end{bmatrix}. \qquad (353) \end{aligned}$$

Thus, we can eliminate the matrix variable $\begin{bmatrix} Z_1 & V \\ V^\dagger & Z_2 \end{bmatrix}$, and the SDP reduces to

$$\begin{aligned} &\sup_{W,\nu,Z_1\geq 0,\mu\in\mathbb{R}} \left\{ \begin{array}{c} \mu + 2\nu\sqrt{1-\varepsilon} - \operatorname{Tr}[Z_1\rho] : \\ \operatorname{Tr}[W\sigma] \leq 1, \\ \begin{bmatrix} Z_1 & -\nu I \\ -\nu I & W - \mu I \end{bmatrix} \geq 0 \end{array} \right\} \\ &= \sup_{W,\nu,Z\geq 0,\mu\in\mathbb{R}} \left\{ \begin{array}{c} \mu + 2\nu\sqrt{1-\varepsilon} - \operatorname{Tr}[Z\rho] : \\ \operatorname{Tr}[W\sigma] \leq 1, \\ \begin{bmatrix} Z & -\nu I \\ -\nu I & W - \mu I \end{bmatrix} \geq 0 \end{array} \right\} \qquad (354) \\ &= \sup_{W,\nu,Z\geq 0,\mu\in\mathbb{R}} \left\{ \begin{array}{c} \mu + 2\nu\sqrt{1-\varepsilon} - \operatorname{Tr}[Z\rho] : \\ \operatorname{Tr}[W\sigma] \leq 1, \\ \begin{bmatrix} Z & \nu I \\ \nu I & W - \mu I \end{bmatrix} \geq 0 \end{array} \right\}. \qquad (355) \end{aligned}$$

APPENDIX E
SDP DUAL OF SMOOTH CONDITIONAL MIN-ENTROPY

In Corollary 5, we presented the primal SDP of the smooth conditional min-entropy. The dual SDP for smooth conditional min-entropy is as follows:

$$H^\varepsilon_{\min}(A|B)_\rho = - \log_2 \sup_{\substack{W_{AB}\geq 0, \\ Z_{AB}\geq 0, \\ \nu\geq 0,\mu\geq 0}} \left\{ \begin{array}{c} -\mu + 2\nu\sqrt{1-\varepsilon^2} - \operatorname{Tr}[Z_{AB}\rho_{AB}] : \\ \operatorname{Tr}_A[W_{AB}] \leq I_B, \\ \begin{bmatrix} Z_{AB} & \nu I_{AB} \\ \nu I_{AB} & W_{AB} + \mu I_{AB} \end{bmatrix} \geq 0 \end{array} \right\}, \tag{356}$$

Before proving the dual, let us verify that strong duality holds for the primal and dual SDPs. Consider the following feasible choices for the primal SDP: $\widetilde{\rho}_{AB} = \rho_{AB}$, $S_B = d_B\rho_B$, and $X_{AB} = (\sqrt{1-\varepsilon^2})\rho_{AB}$, where d_B is the dimension of the B system. This follows because the operator inequality $\rho_{AB} \leq I_A \otimes d_B\rho_B$ holds for every state ρ_{AB} (see, e.g., just after Eq. (34) in [8]). Furthermore,

$$\begin{aligned} \begin{bmatrix} \rho_{AB} & X_{AB} \\ X^\dagger_{AB} & \widetilde{\rho}_{AB} \end{bmatrix} &= \begin{bmatrix} \rho_{AB} & \sqrt{1-\varepsilon^2}\rho_{AB} \\ \sqrt{1-\varepsilon^2}\rho_{AB} & \rho_{AB} \end{bmatrix} \qquad (357) \\ &= \begin{bmatrix} 1 & \sqrt{1-\varepsilon^2} \\ \sqrt{1-\varepsilon^2} & 1 \end{bmatrix} \otimes \rho_{AB} \qquad (358) \\ &\geq 0. \qquad (359) \end{aligned}$$

In addition, choosing μ, ν, W_{AB}, and Z_{AB} to satisfy $\nu > \mu > 0$, $d_A(\nu - \mu) < 1$, $W_{AB} = (\nu - \mu)I_{AB}$, $Z_{AB} = \nu I_{AB}$, and $W_{AB} = (\nu - \mu)I_{AB}$ leads to strictly feasible choices for the dual program. Thus, strong duality holds, due to Slater's condition.

Recall the primal of smooth conditional min-entropy given in (280) and the standard form of SDPs:

$$\sup_{Z\geq 0} \{\operatorname{Tr}[AZ] : \Phi(Z) \leq B\}, \tag{360}$$

$$\inf_{Y\geq 0} \{\operatorname{Tr}[BY] : \Phi^\dagger(Y) \geq A\}. \tag{361}$$

In standard form, this SDP is given by

$$Y = \begin{bmatrix} S_B & 0 & 0 \\ 0 & W_{AB} & X_{AB} \\ 0 & X^\dagger_{AB} & \widetilde{\rho}_{AB} \end{bmatrix}, \quad B = \begin{bmatrix} I_B & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{362}$$

$$\Phi^\dagger(Y) = \begin{bmatrix} L_{AB} & 0 & 0 & 0 & 0 \\ 0 & -\operatorname{Tr}[\widetilde{\rho}_{AB}] & 0 & 0 & 0 \\ 0 & 0 & \operatorname{Re}[\operatorname{Tr}[X_{AB}]] & 0 & 0 \\ 0 & 0 & 0 & 0 & X_{AB} \\ 0 & 0 & 0 & X^\dagger_{AB} & \widetilde{\rho}_{AB} \end{bmatrix}, \tag{363}$$

where $L_{AB} = I_A \otimes S_B - \widetilde{\rho}_{AB}$, and

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{1-\varepsilon^2} & 0 & 0 \\ 0 & 0 & 0 & -\rho_{AB} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{364}$$

Setting

$$Z = \begin{bmatrix} Z^1_{AB} & 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & \nu & 0 & 0 \\ 0 & 0 & 0 & Z^2_{AB} & V_{AB} \\ 0 & 0 & 0 & V^\dagger_{AB} & Z^3_{AB} \end{bmatrix}, \tag{365}$$

we find that

$$\begin{aligned} &\operatorname{Tr}[Z\Phi^\dagger(Y)] \\ &= \operatorname{Tr}[Z^1_{AB}(I_A \otimes S_B - \widetilde{\rho}_{AB})] - \mu\operatorname{Tr}[\widetilde{\rho}_{AB}] \\ &\quad + \operatorname{Tr}\left[\begin{bmatrix} Z^2_{AB} & V_{AB} \\ V^\dagger_{AB} & Z^3_{AB} \end{bmatrix}\begin{bmatrix} 0 & X_{AB} \\ X^\dagger_{AB} & \widetilde{\rho}_{AB} \end{bmatrix}\right] \\ &\quad + \nu\operatorname{Re}[\operatorname{Tr}[X_{AB}]] \qquad (366) \\ &= \operatorname{Tr}[\operatorname{Tr}_A[Z^1_{AB}]S_B] + \nu\operatorname{Re}[\operatorname{Tr}[X_{AB}]] + 2\operatorname{Re}[\operatorname{Tr}[V^\dagger_{AB}X_{AB}]] \\ &\quad + \operatorname{Tr}[\left(-\mu I_{AB} - Z^1_{AB} + Z^3_{AB}\right)\widetilde{\rho}_{AB}] \qquad (367) \\ &= \operatorname{Tr}[\operatorname{Tr}_A[Z^1_{AB}]S_B] \\ &\quad + \operatorname{Tr}\left[\begin{bmatrix} 0 & L_{AB} \\ L^\dagger_{AB} & -\mu I_{AB} - Z^1_{AB} + Z^3_{AB} \end{bmatrix}\begin{bmatrix} W_{AB} & X_{AB} \\ X^\dagger_{AB} & \widetilde{\rho}_{AB} \end{bmatrix}\right], \qquad (368) \end{aligned}$$

with the shorthand

$$L_{AB} := \frac{\nu}{2}I_{AB} + V_{AB}. \tag{369}$$

This implies that

$$\Phi(Z) = \begin{bmatrix} \mathrm{Tr}_A[Z^1_{AB}] & 0 & 0 \\ 0 & 0 & L_{AB} \\ 0 & L^\dagger_{AB} & -\mu I_{AB} - Z^1_{AB} + Z^3_{AB} \end{bmatrix}. \tag{370}$$

Then we find that the dual SDP is given by

$$\sup_{Z\geq 0} \{\mathrm{Tr}[AZ] : \Phi(Z) \leq B\} \tag{371}$$

$$= \sup_{\substack{Z^1_{AB}\geq 0,\\ \mu\geq 0,\\ \nu\geq 0}} \left\{ \begin{array}{c} -\mu + \nu\sqrt{1-\varepsilon^2} - \mathrm{Tr}[Z^2_{AB}\rho_{AB}] : \\ \mathrm{Tr}_A[Z^1_{AB}] \leq I_B, \\ \begin{bmatrix} 0 & L_{AB} \\ L^\dagger_{AB} & -\mu I_{AB} - Z^1_{AB} + Z^3_{AB} \end{bmatrix} \leq 0 \\ \begin{bmatrix} Z^2_{AB} & V_{AB} \\ V^\dagger_{AB} & Z^3_{AB} \end{bmatrix} \geq 0 \end{array} \right\} \tag{372}$$

$$= \sup_{\substack{Z^1_{AB}\geq 0,\\ \nu\geq 0,\\ \mu\geq 0}} \left\{ \begin{array}{c} -\mu + \nu\sqrt{1-\varepsilon^2} - \mathrm{Tr}[Z^2_{AB}\rho_{AB}] : \\ \mathrm{Tr}_A[Z^1_{AB}] \leq I_B, \\ \begin{bmatrix} 0 & L_{AB} \\ L^\dagger_{AB} & -\mu I_{AB} - Z^1_{AB} + Z^3_{AB} \end{bmatrix} \leq 0 \\ \begin{bmatrix} Z^2_{AB} & V_{AB} \\ V^\dagger_{AB} & Z^3_{AB} \end{bmatrix} \geq 0 \end{array} \right\} \tag{373}$$

$$= \sup_{\substack{Z^1_{AB}\geq 0,\\ \nu\geq 0,\\ \mu\geq 0}} \left\{ \begin{array}{c} -\mu + 2\nu\sqrt{1-\varepsilon^2} - \mathrm{Tr}[Z^2_{AB}\rho_{AB}] : \\ \mathrm{Tr}_A[Z^1_{AB}] \leq I_B, \\ \begin{bmatrix} 0 & L'_{AB} \\ (L'_{AB})^\dagger & -\mu I_{AB} - Z^1_{AB} + Z^3_{AB} \end{bmatrix} \leq 0 \\ \begin{bmatrix} Z^2_{AB} & V_{AB} \\ V^\dagger_{AB} & Z^3_{AB} \end{bmatrix} \geq 0 \end{array} \right\}, \tag{374}$$

where

$$L'_{AB} \coloneqq \nu I_{AB} + V_{AB}. \tag{375}$$

Then, as in derivations related to (353), we find that

$$\begin{bmatrix} 0 & \nu I_{AB} + V_{AB} \\ \nu I_{AB} + V^\dagger_{AB} & -\mu I_{AB} - Z^1_{AB} + Z^3_{AB} \end{bmatrix} \leq 0 \tag{376}$$

$$\Leftrightarrow \begin{bmatrix} Z^2_{AB} & V_{AB} \\ V^\dagger_{AB} & Z^3_{AB} \end{bmatrix} \leq \begin{bmatrix} Z^2_{AB} & -\nu I_{AB} \\ -\nu I_{AB} & Z^1_{AB} + \mu I_{AB} \end{bmatrix}, \tag{377}$$

and we can again eliminate variables to reduce the SDP to

$$\sup_{Z^1_{AB},\nu\geq 0,\mu\geq 0} \left\{ \begin{array}{c} -\mu + 2\nu\sqrt{1-\varepsilon^2} - \mathrm{Tr}[Z^2_{AB}\rho_{AB}] : \\ \mathrm{Tr}_A[Z^1_{AB}] \leq I_B, \\ \begin{bmatrix} Z^2_{AB} & -\nu I_{AB} \\ -\nu I_{AB} & Z^1_{AB} + \mu I_{AB} \end{bmatrix} \geq 0 \end{array} \right\} \tag{378}$$

$$= \sup_{Z^1_{AB},\nu\geq 0,\mu\geq 0} \left\{ \begin{array}{c} -\mu + 2\nu\sqrt{1-\varepsilon^2} - \mathrm{Tr}[Z^2_{AB}\rho_{AB}] : \\ \mathrm{Tr}_A[Z^1_{AB}] \leq I_B, \\ \begin{bmatrix} Z^2_{AB} & \nu I_{AB} \\ \nu I_{AB} & Z^1_{AB} + \mu I_{AB} \end{bmatrix} \geq 0 \end{array} \right\} \tag{379}$$

We then do a final rewriting as

$$\sup_{\substack{W_{AB},Z_{AB}\geq 0,\\ \nu\geq 0,\mu\geq 0}} \left\{ \begin{array}{c} -\mu + 2\nu\sqrt{1-\varepsilon^2} - \mathrm{Tr}[Z_{AB}\rho_{AB}] : \\ \mathrm{Tr}_A[W_{AB}] \leq I_B, \\ \begin{bmatrix} Z_{AB} & \nu I_{AB} \\ \nu I_{AB} & W_{AB} + \mu I_{AB} \end{bmatrix} \geq 0 \end{array} \right\}. \tag{380}$$

Appendix F
Second-Order Asymptotics of Smooth Sandwiched Rényi Relative Entropy

Proposition 11: Fix $\varepsilon \in (0,1)$. For a state ρ and a PSD operator σ, the following second-order expansions hold:

1) For $\alpha > 1$:

$$\frac{1}{n}\widetilde{D}^\varepsilon_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) - \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{381}$$

2) For $\alpha \in [1/2, 1)$:

$$\frac{1}{n}\widetilde{D}^\varepsilon_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{382}$$

Proof: Part (1):

For the upper bound, due to the β-monotonicity of sandwiched Rényi relative entropy for $\beta > 1$, we have

$$\frac{1}{n}\widetilde{D}^\varepsilon_\beta(\rho^{\otimes n}\|\sigma^{\otimes n})$$
$$\leq \frac{1}{n}D^\varepsilon_{\max}(\rho^{\otimes n}\|\sigma^{\otimes n}) \tag{383}$$
$$\overset{(a)}{\leq} \frac{1}{n}D^{1-\varepsilon}_{\min}(\rho^{\otimes n}\|\sigma^{\otimes n}) + \frac{1}{n}\log_2\left(\frac{1}{1-\varepsilon}\right) \tag{384}$$
$$\overset{(b)}{=} D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(1-\varepsilon) + O\left(\frac{\log n}{n}\right) \tag{385}$$
$$\overset{(c)}{=} D(\rho\|\sigma) - \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right) \tag{386}$$

where: (a) follows from (154); (b) from (145); and (c) from $\Phi^{-1}(1-\varepsilon) = -\Phi^{-1}(\varepsilon)$.

For the lower bound, we use Theorem 3 for $\varepsilon, \delta \in (0,1)$ such that $\varepsilon + \delta \in (0,1)$, as well as the $\alpha \to 1/2$ limit of the lower bound, by α-monotonicity of $\widetilde{D}_\alpha$, and then consider

$$\frac{1}{n}\widetilde{D}^\varepsilon_\beta(\rho^{\otimes n}\|\sigma^{\otimes n})$$
$$\geq \frac{1}{n}D^{1-\varepsilon-\delta}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n}) - \frac{1}{n}\frac{\beta}{\beta-1}\log_2\left(\frac{1}{1-f(\varepsilon,\delta)}\right), \tag{387}$$

where $f(\varepsilon,\delta)$ is given in (151), and we know from (153) that $-\log(1-f(\varepsilon,\delta)) = O(\log n)$ for $\delta = 1/\sqrt{n}$ and sufficiently large n.

Now, using Theorem 4, consider that

$$\frac{1}{n}D^{1-\varepsilon-\delta}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n})$$
$$= D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(1-\varepsilon-\delta) + O\left(\frac{\log n}{n}\right) \tag{388}$$
$$= D(\rho\|\sigma) - \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon+\delta) + O\left(\frac{\log n}{n}\right) \tag{389}$$
$$= D(\rho\|\sigma) - \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{390}$$

where the last equality holds from similar reasoning used to arrive at (159), which was used in the proof of Theorem 4 by the choice $\delta = 1/\sqrt{n}$.

Then, combining the above inequality with (387), we obtain the desired lower bound. Finally together with (386), we complete the proof for the case $\beta > 1$.

Part (2): For $\alpha \in [1/2, 1)$, from the α-monotonicity of the sandwiched Rényi relative entropy, we have

$$\frac{1}{n}\widetilde{D}^{\varepsilon}_{\alpha}(\rho^{\otimes n}\|\sigma^{\otimes n})$$
$$\geq \frac{1}{n}D^{\varepsilon}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n}) \tag{391}$$
$$= D(\rho\|\sigma) + \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{392}$$

where the equality follows from Theorem 4.

For the upper bound, using Theorem 3, consider that

$$\frac{1}{n}\widetilde{D}^{\varepsilon}_{\alpha}(\rho^{\otimes n}\|\sigma^{\otimes n})$$
$$\leq \frac{1}{n}\widetilde{D}^{1-\varepsilon-\delta}_{\beta}(\rho^{\otimes n}\|\sigma^{\otimes n}) + \frac{1}{n}\frac{\beta}{\beta-1}\log\left(\frac{1}{1-f(\varepsilon,\delta)}\right) \tag{393}$$
$$= D(\rho\|\sigma) - \sqrt{\frac{1}{n}V(\rho\|\sigma)}\ \Phi^{-1}(1-\varepsilon-\delta) + O\left(\frac{\log n}{n}\right) + \frac{1}{n}\frac{\beta}{\beta-1}\log\left(\frac{1}{1-f(\varepsilon,\delta)}\right), \tag{394}$$

where the last inequality follows from the second-order expansion obtained in Part (1). Now observing that $-\log(1-f(\varepsilon,\delta)) = O(\log n)$ for $\delta = 1/\sqrt{n}$ and sufficiently large n, and following the same reasoning, we used to arrive at (159), in the proof of Theorem 4, we get the matching upper bound.

Combining the obtained upper bound and (392), we conclude the proof. ∎

Appendix G
Moderate Deviation Analysis for Smooth F-Min-Relative Entropy (Proof of Proposition 3)

In this appendix, we prove Proposition 3. Recall again that a sequence $\{a_n\}_n$ is called a moderate sequence if $a_n \to 0$ and $\sqrt{n}a_n \to \infty$ when $n \to \infty$. From [49], we have the following scaling of the smooth min-relative entropy under moderate deviations, where $\varepsilon_n := e^{-na_n^2}$,

$$\frac{1}{n}D^{\varepsilon_n}_{\min}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)}\ a_n + o(a_n). \tag{395}$$

Proof: [Proof of Proposition 3] For the lower bound, we employ Proposition 1 to find that

$$\frac{1}{n}D^{\varepsilon_n}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n})$$
$$\geq \frac{1}{n}D^{\varepsilon_n}_{\min}(\rho^{\otimes n}\|\sigma^{\otimes n}) - \frac{1}{n}\log_2\left(\frac{1}{1-\varepsilon_n}\right) \tag{396}$$
$$= D(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)}\ a_n + o(a_n), \tag{397}$$

where the last equality holds by (395). Then, we arrive at

$$\frac{1}{n}D^{\varepsilon_n}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n}) \geq D(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)}\ a_n + o(a_n), \tag{398}$$

along with $\frac{1}{n}\log_2\left(\frac{1}{1-\varepsilon_n}\right) = o\left(\frac{1}{n}\right)$ leading to $\frac{1}{n}\log_2\left(\frac{1}{1-\varepsilon_n}\right) = o(a_n)$ since $na_n^2 \to \infty$.

For the upper bound, similar to the proof of Theorem 4, specifically using the relationship derived in (157), we have

$$\frac{1}{n}D^{\varepsilon_n}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n})$$
$$\leq \frac{1}{n}D^{\varepsilon_n+\delta}_{\min}(\rho^{\otimes n}\|\sigma^{\otimes n}) + \frac{1}{n}\log_2\left(\frac{1}{\varepsilon_n+\delta}\right) + \frac{1}{n}\log_2\left(\frac{1}{1-f(\varepsilon_n,\delta)}\right). \tag{399}$$

Since $\delta \in (0,1)$, by choosing $\delta = \varepsilon_n$, $\frac{1}{n}\log_2\left(\frac{1}{1-f(\varepsilon_n,\varepsilon_n)}\right) = o(a_n)$ and $\frac{1}{n}\log_2\left(\frac{1}{2\varepsilon_n}\right) = o(a_n)$. Also observe that

$$2\ e^{-na_n^2} = \exp\left(-n\left(a_n^2 - \frac{\ln 2}{n}\right)\right) = e^{-nb_n^2}, \tag{400}$$

where $b_n := a_n + o(a_n)$. Collecting these observations together, we obtain the upper bound

$$\frac{1}{n}D^{\varepsilon_n}_{\min,F}(\rho^{\otimes n}\|\sigma^{\otimes n}) \leq D(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)}\ a_n + o(a_n). \tag{401}$$

We conclude the proof by combining (398) and (401). ∎

Acknowledgment

The authors acknowledge helpful discussions with Mario Berta, Dhrumil Patel, Soorya Rethinasamy, and Aaron Wagner.

References

[1] S. Khatri and M. M. Wilde, "Principles of quantum communication theory: A modern approach," 2020, *arXiv:2011.04672v1*.
[2] E. Chitambar and G. Gour, "Quantum resource theories," *Rev. Mod. Phys.*, vol. 91, no. 2, Apr. 2019, Art. no. 025001.
[3] N. Datta, "Min-and max-relative entropies and a new entanglement monotone," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2816–2826, Jun. 2009.
[4] A. Uhlmann, "The 'transition probability' in the state space of a*algebra," *Rep. Math. Phys.*, vol. 9, no. 2, pp. 273–279, Apr. 1976.
[5] X. Wang and M. M. Wilde, "Resource theory of asymmetric distinguishability," *Phys. Rev. Res.*, vol. 1, no. 3, Dec. 2019, Art. no. 033170.
[6] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 405–417, Jul. 1974.
[7] X. Wang and M. M. Wilde, "Resource theory of asymmetric distinguishability for quantum channels," *Phys. Rev. Res.*, vol. 1, no. 3, Dec. 2019, Art. no. 033169.

[8] F. Dupuis, L. Kraemer, P. Faist, J. M. Renes, and R. Renner, "Generalized entropies," in *Proc. Int. Congr. Math. Phys.* Singapore: World Scientific, 2014, pp. 134–153.

[9] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: A new generalization and some properties," *J. Math. Phys.*, vol. 54, no. 12, Dec. 2013, Art. no. 122203.

[10] 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, Oct. 2014.

[11] P. Faist, "Quantum coarse-graining: An information-theoretic approach to thermodynamics," Ph.D. thesis, ETH Zurich, 2016.

[12] Q. Zhao, Y. Liu, X. Yuan, E. Chitambar, and A. Winter, "One-shot coherence distillation: Towards completing the picture," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6441–6453, Oct. 2019.

[13] N. Ramakrishnan, M. Tomamichel, and M. Berta, "Moderate deviation expansion for fully quantum tasks," *IEEE Trans. Inf. Theory*, vol. 69, no. 8, pp. 5041–5059, Jan. 2023.

[14] R. Rubboli and M. Tomamichel, "Fundamental limits on correlated catalytic state transformations," *Phys. Rev. Lett.*, vol. 129, no. 12, Sep. 2022, Art. no. 120506.

[15] R. Renner and S. Wolf, "Smooth Renyi entropy and applications," in *Proc. Int. Symp. Inf. Theory*, Jul. 2004, p. 233.

[16] R. Renner, "Security of quantum key distribution," Ph.D. thesis, Dept. Physics, ETH Züürich, Zürich, 2005. [Online]. Available: https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/72791/1/eth-28331-01.pdf

[17] F. Buscemi and N. Datta, "The quantum capacity of channels with arbitrarily correlated noise," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1447–1460, Mar. 2010.

[18] F. G. S. L. Brandão and N. Datta, "One-shot rates for entanglement manipulation under non-entangling maps," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1754–1760, Mar. 2011.

[19] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Phys. Rev. Lett.*, vol. 108, no. 20, May 2012, Art. no. 200501.

[20] K. Bu, U. Singh, S.-M. Fei, A. K. Pati, and J. Wu, "Maximum relative entropy of coherence: An operational coherence measure," *Phys. Rev. Lett.*, vol. 119, no. 15, Oct. 2017, Art. no. 150405.

[21] Z.-W. Liu, "On quantum randomness and quantum resources," Ph.D. thesis, Dept. Phys., Massachusetts Inst. Technol., MA, USA, 2018.

[22] 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.

[23] K. Li, "Second-order asymptotics for quantum hypothesis testing," *Ann. Statist.*, vol. 42, no. 1, pp. 171–189, 2014.

[24] N. Datta and F. Leditzky, "Second-order asymptotics for source coding, dense coding, and pure-state entanglement conversions," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 582–608, Jan. 2015.

[25] N. Datta, M. Tomamichel, and M. M. Wilde, "On the second-order asymptotics for entanglement-assisted communication," *Quantum Inf. Process.*, vol. 15, pp. 2569–2591, Jan. 2016.

[26] N. Datta, Y. Pautrat, and C. Rouzé, "Second-order asymptotics for quantum hypothesis testing in settings beyond i.I.d.—Quantum lattice systems and more," *J. Math. Phys.*, vol. 57, no. 6, Jun. 2016, Art. no. 062207.

[27] L. Lami, B. Regula, X. Wang, and M. M. Wilde, "Upper bounds on the distillable randomness of bipartite quantum states," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Apr. 2023, pp. 203–208.

[28] N. Sharma and N. Ahmad Warsi, "On the strong converses for the quantum channel capacity theorems," 2012, *arXiv:1205.1712*.

[29] R. L. Frank and E. H. Lieb, "Monotonicity of a relative Rényi entropy," *J. Math. Phys.*, vol. 54, Jan. 2013, Art. no. 122201.

[30] M. M. Wilde, "Optimized quantum f-divergences and data processing," *J. Phys. A, Math. Theor.*, vol. 51, no. 37, Sep. 2018, Art. no. 374002.

[31] M. M. Wilde, "Optimized quantum F-divergences," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2481–2485.

[32] H. Umegaki, "Conditional expectation in an operator algebra. IV. Entropy and information," *Kodai Math. J.*, vol. 14, no. 2, pp. 59–85, Jan. 1962.

[33] D. Petz, "Quasi-entropies for states of a von Neumann algebra," *Publications Res. Inst. Math. Sci.*, vol. 21, no. 4, pp. 787–800, Aug. 1985.

[34] D. Petz, "Quasi-entropies for finite quantum systems," *Rep. Math. Phys.*, vol. 23, no. 1, pp. 57–65, 1986.

[35] M. Tomamichel, *Quantum Information Processing With Finite Resources: Mathematical Foundations*, vol. 5. Cham, Switzerland: Springer, 2015.

[36] E. Kaur and M. M. Wilde, "Upper bounds on secret-key agreement over lossy thermal bosonic channels," *Phys. Rev. A, Gen. Phys.*, vol. 96, no. 6, Dec. 2017, Art. no. 062318.

[37] W. F. Stinespring, "Positive functions on $C^\star$-Algebras," *Proc. Amer. Math. Soc.*, vol. 6, pp. 211–216, Jan. 1955.

[38] M. M. Wilde, "Second law of entanglement dynamics for the non-asymptotic regime," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2021, pp. 1–6.

[39] R. Takagi, B. Regula, and M. M. Wilde, "One-shot yield-cost relations in general quantum resource theories," *PRX Quantum*, vol. 3, no. 1, Mar. 2022, Art. no. 010348.

[40] M. M. Wilde, *Quantum Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[41] H. Qi, Q. Wang, and M. M. Wilde, "Applications of position-based coding to classical communication over quantum channels," *J. Phys. A, Math. Theor.*, vol. 51, no. 44, Nov. 2018, Art. no. 444002.

[42] M. M. Wilde, "Position-based coding and convex splitting for private communication over quantum channels," *Quantum Inf. Process.*, vol. 16, no. 10, p. 264, Oct. 2017.

[43] S. K. Oskouei, S. Mancini, and M. M. Wilde, "Union bound for quantum information processing," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 475, no. 2221, Jan. 2019, Art. no. 20180612.

[44] A. Anshu, M. Berta, R. Jain, and M. Tomamichel, "A minimax approach to one-shot entropy inequalities," *J. Math. Phys.*, vol. 60, no. 12, Dec. 2019, Art. no. 122201.

[45] A. E. Rastegin, "Relative error of state-dependent cloning," *Phys. Rev. A, Gen. Phys.*, vol. 66, no. 4, Oct. 2002, Art. no. 042304.

[46] A. E. Rastegin, "A lower bound on the relative error of mixed-state cloning and related operations," *J. Opt. B, Quantum Semiclass. Opt.*, vol. 5, no. 6, pp. S647–S650, Dec. 2003.

[47] A. Gilchrist, N. K. Langford, and M. A. Nielsen, "Distance measures to compare real and ideal quantum processes," *Phys. Rev. A, Gen. Phys.*, vol. 71, no. 6, Jun. 2005, Art. no. 062310.

[48] A. E. Rastegin, "Sine distance for quantum states," 2006, *arXiv:quant-ph/0602112*.

[49] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel, "Moderate deviation analysis for classical communication over quantum channels," *Commun. Math. Phys.*, vol. 355, no. 3, pp. 1283–1315, Nov. 2017.

[50] I. Devetak and A. Winter, "Distilling common randomness from bipartite quantum states," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3183–3196, Dec. 2004.

[51] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, "Thermodynamical approach to quantifying quantum correlations," *Phys. Rev. Lett.*, vol. 89, no. 18, Oct. 2002, Art. no. 180402.

[52] I. Devetak, "Distillation of local purity from quantum states," *Phys. Rev. A, Gen. Phys.*, vol. 71, no. 6, Jun. 2005, Art. no. 062303.

[53] I. Devetak, A. W. Harrow, and A. J. Winter, "A resource framework for quantum Shannon theory," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4587–4618, Oct. 2008.

[54] H. Krovi and I. Devetak, "Local purity distillation with bounded classical communication," *Phys. Rev. A, Gen. Phys.*, vol. 76, no. 1, Jul. 2007, Art. no. 012321.

[55] G. Manzano, F. Plastina, and R. Zambrini, "Optimal work extraction and thermodynamics of quantum measurements and correlations," *Phys. Rev. Lett.*, vol. 121, no. 12, Sep. 2018, Art. no. 120602.

[56] B. Morris, L. Lami, and G. Adesso, "Assisted work distillation," *Phys. Rev. Lett.*, vol. 122, no. 13, Apr. 2019, Art. no. 130601.

[57] S. Chakraborty, A. Nema, and F. Buscemi, "One-shot purity distillation with local noisy operations and one-way classical communication," 2022, *arXiv:2208.05628*.

[58] Y. Zhang et al., "Experimental low-latency device-independent quantum randomness," *Phys. Rev. Lett.*, vol. 124, no. 1, Jan. 2020, Art. no. 010505.

[59] L. K. Shalm et al., "Device-independent randomness expansion with entangled photons," *Nature Phys.*, vol. 17, no. 4, pp. 452–456, Apr. 2021.

[60] B. Groisman, S. Popescu, and A. Winter, "Quantum, classical, and total amount of correlations in a quantum state," *Phys. Rev. A, Gen. Phys.*, vol. 72, no. 3, Sep. 2005, Art. no. 032317.

[61] N. Ciganović, N. J. Beaudry, and R. Renner, "Smooth max-information as one-shot generalization for mutual information," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1573–1581, Mar. 2014.

[62] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, "Volume of the set of separable states," *Phys. Rev. A, Gen. Phys.*, vol. 58, no. 2, pp. 883–892, Aug. 1998.
[63] G. Vidal and R. F. Werner, "Computable measure of entanglement," *Phys. Rev. A, Gen. Phys.*, vol. 65, Feb. 2002, Art. no. 032314.
[64] A. Anshu, R. Jain, and N. A. Warsi, "Building blocks for communication over noisy quantum networks," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1287–1306, Feb. 2019.
[65] S. Khatri, E. Kaur, S. Guha, and M. M. Wilde, "Second-order coding rates for key distillation in quantum key distribution," 2019, *arXiv:1910.03883*.
[66] 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.
[67] A. B. Wagner, N. V. Shende, and Y. Altuğ, "A new method for employing feedback to improve coding performance," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6660–6681, May 2020.
[68] J. Watrous, "Simpler semidefinite programs for completely bounded norms," *Chicago J. Theor. Comput. Sci.*, vol. 10, no. 8, pp. 1–19, 2013.
[69] H. Konno, "A cutting plane algorithm for solving bilinear programs," *Math. Program.*, vol. 11, no. 1, pp. 14–27, Dec. 1976.
[70] S. Huber, R. König, and M. Tomamichel, "Jointly constrained semidefinite bilinear programming with an application to Dobrushin curves," *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 2934–2950, May 2020.
[71] M. Tomamichel, R. Colbeck, and R. Renner, "Duality between smooth min- and max-entropies," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4674–4681, Sep. 2010.
[72] C. Schaffner, B. Terhal, and S. Wehner, "Robust cryptography in the noisy-quantum-storage model," *Quantum Inf. Comput.*, vol. 9, no. 11, pp. 963–996, Nov. 2009.
[73] M. Nussbaum and A. Szkoła, "The Chernoff lower bound for symmetric quantum hypothesis testing," *Ann. Statist.*, vol. 37, no. 2, pp. 1040–1057, Apr. 2009.
[74] K. M. R. Audenaert et al., "Discriminating states: The quantum Chernoff bound," *Phys. Rev. Lett.*, vol. 98, no. 16, Apr. 2007, Art. no. 160501.

Theshani Nuradha (Graduate Student Member, IEEE) received the B.Sc. degree (Hons.) from the Department of Electronic and Telecommunication Engineering (ENTC), University of Moratuwa, Sri Lanka, in December 2018. She is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering, Cornell University, USA. During the bachelor's studies, she was a full-time Research Intern with Singapore University of Technology and Design, Singapore, in 2017. Prior to joining Cornell University, she was a Lecturer on contract with ENTC from 2019 to 2020.

Mark M. Wilde (Fellow, IEEE) received the Ph.D. degree in Electrical Engineering from the University of Southern California, Los Angeles, CA, USA. He is currently an Associate Professor of Electrical and Computer Engineering with Cornell University. His current research interests include quantum Shannon theory, quantum computation, quantum optical communication, quantum computational complexity theory, and quantum error correction. He was a recipient of the National Science Foundation Career Development Award. He was a co-recipient of the 2018 AHP-Birkhauser Prize, awarded to "the most remarkable contribution" published in the journal *Annales Henri Poincaré*. He is an Outstanding Referee of the American Physical Society.