Quantum Pufferfish Privacy: A Flexible Privacy Framework for Quantum Systems

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

***Abstract*— We propose a versatile privacy framework for quantum systems, termed *quantum pufferfish privacy* (QPP). Inspired by classical pufferfish privacy, our formulation generalizes and addresses limitations of quantum differential privacy by offering flexibility in specifying private information, feasible measurements, and domain knowledge. We show that QPP can be equivalently formulated in terms of the Datta–Leditzky information spectrum divergence, thus providing the first operational interpretation thereof. We reformulate this divergence as a semi-definite program and derive several properties of it, which are then used to prove convexity, composability, and post-processing of QPP mechanisms. Parameters that guarantee QPP of the depolarization mechanism are also derived. We analyze the privacy-utility tradeoff of general QPP mechanisms and, again, study the depolarization mechanism as an explicit instance. The QPP framework is then applied to privacy auditing for identifying privacy violations via a hypothesis testing pipeline that leverages quantum algorithms. Connections to quantum fairness and other quantum divergences are also explored and several variants of QPP are examined.**

***Index Terms*— Auditing privacy, privacy-utility tradeoff, pufferfish privacy, quantum differential privacy, quantum generalized divergences.**

I. Introduction

With a surging interest in quantum and hybrid classical–quantum systems, ensuring privacy of both classical and quantum data has become pivotal. Privacy-preserving data analysis has been widely studied for classical systems by means of statistical privacy frameworks. Differential privacy (DP) is an important statistical privacy framework that enables answering aggregate queries about a database while keeping individual records private [1], [2]. However, DP accounts for one type of private information only (namely, records of individual users), and it does not allow encoding domain knowledge into the framework. To address these limitations, a versatile generalization of DP, termed Pufferfish Privacy (PP), has been proposed [3]. PP allows for customizing which information is regarded as private and explicitly integrates distributional assumptions into the definition [3], [4], [5]. PP has found use in several applications, including smart metering [6], [7] and trajectory monitoring with location tracking [8], [9] (see also Figure 1 of [10] for an explicit example related to salary releases, where PP is applicable). Information-theoretic formulations of classical DP and PP have been proposed in [11] and [12], respectively.

Quantum DP (QDP) is a generalization of the classical DP notion and has been proposed in [13]. See also [14] for DP of quantum measurements and [15] for an information-theoretic interpretation of QDP. Connections to quantum stability through private learning have been studied in [16]. Moreover, [17] has explored how quantum classifiers can be made private by using the intrinsic noise of existing quantum systems. See also [18], [19], [20], [21], and [22] for applications of DP in quantum machine learning. Additionally, privacy amplification of quantum and quantum inspired algorithms has been analysed using QDP and classical DP notions in [23]. However, similar to the classical case, the versatility of QDP is limited.

In this paper, we propose a flexible privacy framework for quantum systems, termed quantum PP (QPP), that addresses these limitations. We provide a comprehensive study of QPP, encompassing properties, mechanisms, privacy-utility tradeoffs, as well as the first operational meaning of the Datta–Leditzky information spectrum divergence [24] (hereafter abbreviated as the DL divergence), which arises from our framework.

A. Motivation

We seek to address key limitations of QDP by exploring more *flexible privacy frameworks* for quantum information processing. As delineated next, flexible secrets, embedding domain knowledge, and relaxing the need for worst-case measurements are considerations central to our approach.

1) Flexible Secrets: QDP guarantees that any pair of states that are classified as neighbors are approximately indistinguishable, i.e., cannot be identified under any possible measurement. However, scenarios may arise in which one wants to hide specific properties of the states, as opposed to the state itself (e.g., whether the states possess a certain

Manuscript received 28 July 2023; revised 25 April 2024; accepted 17 May 2024. Date of publication 24 May 2024; date of current version 16 July 2024. The work of Theshani Nuradha and Mark M. Wilde was supported by the National Science Foundation under Grant 1907615 and Grant 2315398. The work of Ziv Goldfeld was supported in part by NSF under Grant CCF-2046018, Grant CDMS-2210368, and Grant CCCF-2308446; and in part by the IBM Academic Award. *(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; goldfeld@cornell.edu; wilde@cornell.edu).

Communicated by S. Fehr, Associate Editor for Sequences and Cryptography.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2024.3404927.

Digital Object Identifier 10.1109/TIT.2024.3404927

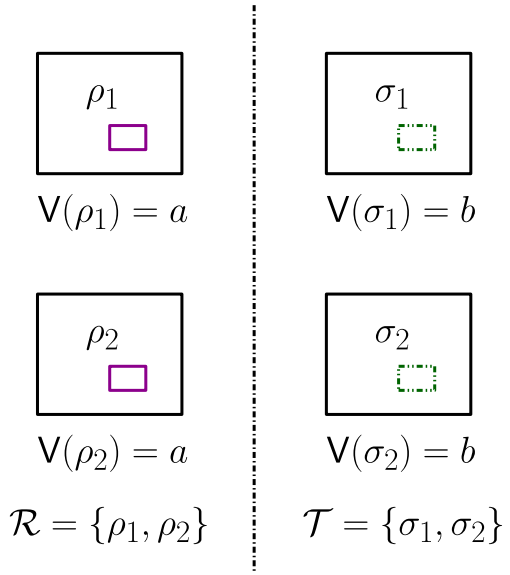


Fig. 1. Depiction of a setup where the goal is to hide whether the amount entanglement V present in the bipartite states ρ_1, ρ_2, σ_1, and σ_2 equals a or b. In this diagram, large squares represent the entire quantum state, while small rectangles correspond to a specific attribute of that state (i.e., the amount of entanglement as quantified by the function V). The specific attribute can take on one of two values, a or b, represented by solid or dotted lines, respectively. As the goal is to conceal only the entanglement level, and not necessarily the specific quantum state, we want the sets $\mathcal{R} = \{\rho_1, \rho_2\}$ and $\mathcal{T} = \{\sigma_1, \sigma_2\}$ to be indistinguishable.

symmetry or property, have any entanglement with a special subsystem, or have secret correlations with other systems). In such situations, QDP may be an overly pessimistic notion of privacy, which, in turn, hinders utility. As an example, consider hiding the amount of entanglement V present in the bipartite states in the set $\{\rho_1, \rho_2, \sigma_1, \sigma_2\}$, for which $\mathsf{V}(\rho_1) = \mathsf{V}(\rho_2) = a$ and $\mathsf{V}(\sigma_1) = \mathsf{V}(\sigma_2) = b$. As illustrated in Fig. 1, hiding whether V equals a or b amounts to making the classes $\{\rho_1, \rho_2\}$ and $\{\sigma_1, \sigma_2\}$ indistinguishable. This can be achieved by applying a QDP mechanism to the state space $\{\rho_1, \rho_2, \sigma_1, \sigma_2\}$ by choosing (ρ_i, σ_j) for all $i, j \in \{1, 2\}$ as neighbors, with the criterion that ρ and σ are neighbors if and only if $|\mathsf{V}(\rho) - \mathsf{V}(\sigma)| = |a - b|$. However, doing so provides a stricter guarantee than required. The source of the issue is the inability of QDP to account for secrets concerning collections of states (as opposed to singletons), which is the first issue we aim to address.

2) Domain Knowledge: In QDP, a worst-case privacy guarantee is provided for all neighboring states. However, one may possess knowledge about the likelihood of observing different states, e.g., via expert feedback. Referring back to the setting from Fig. 1, if we have domain knowledge such that observing the states $\rho_1, \rho_2, \sigma_1, \sigma_2$ is prescribed by the probability vector $(p/2, (1-p)/2, 1/2, 0)$, for $p \in (0, 1)$, then the requirement simplifies to the indistinguishability of $\{\rho_1, \rho_2\}$ versus $\{\sigma_1\}$. Classically, it has been demonstrated that domain knowledge can be leveraged to design privacy mechanisms with increased accuracy and utility [3], [25]. This calls for a quantum privacy framework that can also encode domain knowledge.

3) Relaxing Worst-Case Measurements: Another worst-case aspect of QDP is its account of all possible measurements. However, such a requirement might be too stringent in practice, especially in quantum systems. As an example, while a joint measurement can accurately distinguish between entangled but physically separated states, oftentimes only local operations and classical communications (LOCC) are available (e.g., as considered in quantum data-hiding protocols [26], [27], [28], [29], [30], [31], [32]). In such cases, one may achieve improved accuracy and utility by relaxing the privacy requirement to account for LOCC measurements only.

In sum, the rapid advancements in quantum technologies requires designing flexible privacy frameworks that can be adjusted to timely needs. Furnishing such a framework is the main objective of our paper.

B. Contributions

This work proposes a quantum analog of the PP framework that accounts for the three aforementioned aspects. Our formalism enables reasoning about privacy of quantum systems using information-theoretic tools. We provide a comprehensive study of QPP, encompassing properties, mechanisms, and privacy-utility tradeoffs. Our paradigm also gives rise to the first operational interpretation of the DL divergence [24]. The proposed QPP framework comprises four key ingredients:

1) the set of potential secrets,
2) the set of discriminative pairs that are required to be indistinguishable at the output of the mechanism,
3) the set of data distributions, which encodes domain knowledge on the occurrence of quantum or classical data, and
4) the set of measurements to be accounted for, which is specified based on physical, ethical, or any other constraints.

See Definition 4 for a formal definition. QPP guarantees the indistinguishability, under any allowable measurement, of sets of states formed based on the above ingredients.

After defining the operational privacy framework, we observe that when the measurement class contains all possible measurements, QPP can be equivalently posed as a DL divergence constraint. To the best of our knowledge, this provides the first operational interpretation of the DL divergence. We then derive an efficiently computable formulation of the DL divergence as a semi-definite program (SDP), which may be of independent interest. This SDP is utilized to prove properties of the DL divergence, which are then used in the analysis of QPP mechanisms. These properties include joint quasi-convexity and the data-processing inequality under positive and trace non-increasing maps (see Section IV). Our results also generalize the connection between the hockey-stick divergence and QDP, originally established in [15]. Moreover, we show that existing privacy frameworks such as classical DP [1], [2], classical PP [3], utility-optimized local DP (not subsumed by classical PP) [33], and QDP [13], [15] are special cases of our QPP framework.

We then move on to derive properties of QPP mechanisms, encompassing convexity, post-processing, and composability (both parallel and adaptive). As a specific example,

we characterize the flip parameter that guarantees QPP of the depolarization mechanism. We also describe how QPP mechanisms implementable on quantum devices can be instantiated to achieve classical PP. We consider the associated privacy-utility tradeoff for QPP mechanisms. Our utility metric captures how invertible the privacy mechanism is, which is formulated as the infimized diamond distance between a post-processing of the mechanism's output and the identity channel. We show that this utility metric can be computed as an SDP, and we analyze the privacy-utility tradeoff of the depolarization mechanism. Lastly, we study optimal privacy-utility tradeoffs of QPP mechanisms and characterize the achievable region in several settings.

Another application we consider is privacy auditing, which refers to certifying whether a black-box mechanism satisfies a target privacy guarantee. While several auditing methods are available for classical frameworks, there is currently no approach that can handle quantum data. We fill this gap by proposing the first auditing pipeline for quantum privacy mechanisms. In contrast to existing approaches for classical DP and PP, which require first relaxing the privacy notion and only then auditing, our approach audits for QDP directly. Extensions of these ideas to the QPP setting are also considered.

Finally, we explore connections between QPP, existing quantum privacy frameworks, and figures of merit. First, we examine the connection between quantum privacy and fairness [34], [35], showing that private quantum mechanisms are fair, and under certain conditions, fair algorithms are private. We also provide bounds on quantum Rényi divergences and the trace distance, which stem from QPP. This inspires relaxations of QPP that are defined via these divergences, which, in particular, provide other operational interpretations thereof as privacy metrics. Lastly, we present a variant of QPP that can incorporate entanglement into the framework with the use of reference systems.

C. Organization

The rest of our paper is organized as follows. In Section II, we introduce notation and preliminaries in quantum information theory and privacy. Section III presents the QPP framework, its equivalent formulations in terms of the DL divergence, and special cases of it. In Section IV, we focus further on the DL divergence, reformulate it as an SDP, and use this SDP to prove several properties of it. Properties of general QPP mechanisms and the depolarization mechanism are studied in Section V. We analyze the privacy-utility tradeoff of QPP mechanisms in Section VI, while the privacy auditing application is considered in Section VII. Connections to existing privacy frameworks and to other quantum divergences are explored in Section VIII. Then, we propose several relaxations and variants of QPP in Section IX. Section X summarizes our main contributions and provides concluding remarks.

II. Preliminaries and Background

A. Notation

Sets are denoted by calligraphic letters, e.g., $\mathcal{X}$. For $k, n \in \mathbb{N}$, we use $\mathcal{X}^{n\times k}$ to denote the database space of $n \times k$ matrices; columns correspond to different attributes while rows to different individuals. The (i,j)th entry of $x \in \mathcal{X}^{n\times k}$ is denoted as $x(i,j)$. The ith row and jth column of x are denoted by $x(i,\cdot)$ and $x(\cdot,j)$, respectively. We denote by $(\Omega, \mathcal{F}, \mathbb{P})$ the underlying probability space on which all random variables (RVs) are defined, with $\mathbb{E}$ designating expectation. RVs are denoted by upper case letters, e.g., X, with P_X representing the corresponding probability law. For $X \sim P_X$, we interchangeably use $\mathrm{supp}(X)$ and $\mathrm{supp}(P_X)$ for the support. The joint law of (X,Y) is denoted by P_{XY}, while $P_{Y|X}$ designates the (regular) conditional probability of Y given X. Conventions for $n\times k$-dimensional random variables are the same as for deterministic elements. The space of all Borel probability measures on $\mathcal{S} \subseteq \mathbb{R}^d$ is denoted by $\mathcal{P}(\mathcal{S})$. The Kullback–Leibler (KL) divergence between $P, Q \in \mathcal{P}(\mathcal{X})$ with $P \ll Q$ is given by $\mathsf{D}(P\|Q) \coloneqq \mathbb{E}_P\left[\ln\left(\frac{dP}{dQ}\right)\right]$, where $\frac{dP}{dQ}$ is the Radon–Nikodym derivative of P with respect to Q. For $(X,Y) \sim P_{XY}$, the mutual information between X and Y is denoted by $\mathsf{I}(X;Y) \coloneqq \mathsf{D}(P_{XY}\|P_X \otimes P_Y)$.

We now review basic concepts from quantum information theory and refer the reader to [36] and [37] for more details. A (classical or 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 it is denoted by $\mathrm{supp}(X)$. Let $\mathrm{T}(C)$ denote the transpose of C. The partial transpose of $C \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ on the subsystem A is represented as $\mathrm{T}_A(C)$. Let $\mathrm{Tr}[C]$ denote the trace of C, and let $\mathrm{Tr}_A[C]$ denote the partial trace of C over the subsystem A. The trace norm of a matrix B is defined as $\|B\|_1 \coloneqq \mathrm{Tr}[\sqrt{B^\dagger B}]$. For 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 $\rho_R \in \mathcal{L}(\mathcal{H}_R)$ on R is a PSD, unit-trace operator acting on $\mathcal{H}_R$. We denote the set of all density operators in $\mathcal{L}(\mathcal{H}_R)$ by $\mathcal{D}(\mathcal{H}_R)$. A state ρ_R of rank one is called pure, and we may choose a normalized vector $|\psi\rangle \in \mathcal{H}_R$ satisfying $\rho_R = |\psi\rangle\langle\psi|$ in this case. Otherwise, ρ_R is called a mixed state. By the spectral decomposition theorem, every mixed state can be written as a convex combination of pure, orthogonal states. A quantum channel $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ 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 adjoint of $\mathcal{N}$ by $\mathcal{N}^\dagger$. A measurement of a quantum system R is described by a positive operator-valued measure (POVM) $\{\mathrm{M}_y\}_{y\in\mathcal{Y}}$, which is defined to be a collection of PSD operators satisfying $\sum_{y\in\mathcal{Y}} \mathrm{M}_y = \mathrm{I}_{\mathcal{H}_R}$, where $\mathcal{Y}$ is a finite alphabet. The Born rule dictates that, after applying the above POVM to $\rho \in \mathcal{D}(\mathcal{H}_R)$, the probability of observing the outcome y is given by $\mathrm{Tr}[\mathrm{M}_y\rho]$.

B. Quantum Divergences

We define several quantum divergences that will be used throughout this work. We call a distinguishability measure $\mathbf{D}(\cdot\|\cdot)$ a generalized divergence [38] if it satisfies the data-processing inequality; i.e., for every channel $\mathcal{N}$, state ρ, and

PSD operator σ,

$$\mathbf{D}(\rho\|\sigma) \geq \mathbf{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))\,. \tag{1}$$

The normalized trace distance between the states ρ and σ is defined as

$$\mathsf{T}(\rho,\sigma) \coloneqq \frac{1}{2}\left\|\rho-\sigma\right\|_1, \tag{2}$$

while the fidelity between them is defined as [39]

$$\mathsf{F}(\rho,\sigma) \coloneqq \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2. \tag{3}$$

The diamond distance between the two channels $\mathcal{N},\mathcal{M} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is defined as [40]

$$\left\|\mathcal{N}-\mathcal{M}\right\|_\diamond \coloneqq \sup_{\rho_{RA}} \left\|\mathcal{N}_{A\to B}(\rho_{RA}) - \mathcal{M}_{A\to B}(\rho_{RA})\right\|_1, \tag{4}$$

where the optimization in the definition is over every reference system R and bipartite density operator ρ_{RA} (with R allowed to be arbitrarily large). It is well known, however, that it suffices to perform the optimization over pure bipartite states such that the dimension of the reference system R is equal to the dimension of the channel input system A.

The Petz–Rényi quantum relative entropy of order $\alpha \in (0,1)\cup(1,\infty)$ of a state ρ with respect to a PSD operator σ is given by [41], [42]

$$\mathsf{D}_\alpha(\rho\|\sigma) \coloneqq \frac{1}{\alpha-1}\ln \mathrm{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{5}$$

if $\alpha \in (0,1)\vee(\alpha>1\wedge \mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma))$ and ∞ otherwise. It is a generalized divergence for $\alpha\in[0,1)\cup(1,2]$ [42]. The special case of $\alpha\to 1$ is called the quantum relative entropy and amounts to

$$\mathsf{D}(\rho\|\sigma)\equiv\mathsf{D}_1(\rho\|\sigma) \coloneqq \lim_{\alpha\to1}\mathsf{D}_\alpha(\rho\|\sigma) = \mathrm{Tr}[\rho(\ln\rho-\ln\sigma)] \tag{6}$$

when $\mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma)$ and it is equal to $+\infty$ otherwise. The quantum entropy of a state ρ is defined as

$$\mathsf{S}(\rho) \coloneqq -\mathrm{Tr}[\rho\ln\rho]\,. \tag{7}$$

Equivalently, $\mathsf{S}(\rho) = -\mathsf{D}_1(\rho\|\mathrm{I})$, where I is the identity operator.

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 [43], [44]

$$\widetilde{\mathsf{D}}_\alpha(\rho\|\sigma) \coloneqq \frac{1}{\alpha-1}\ln \mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right] \tag{8}$$

if $\alpha\in(0,1)\vee(\alpha\in(1,\infty)\wedge\ \mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma))$ and ∞ otherwise. It is a generalized divergence for $\alpha\in[1/2,1)\cup(1,\infty)$ [45] (see also [46]).

Fix $\delta\in[0,1]$, a state ρ, and a PSD operator σ. The Datta–Leditzky information spectrum divergences are defined as follows [24]:

$$\underline{\mathsf{D}}^\delta(\rho\|\sigma) \coloneqq \sup\left\{\gamma\in\mathbb{R} : \mathrm{Tr}[(\rho-e^\gamma\sigma)_+]\geq 1-\delta\right\}, \tag{9a}$$

$$\overline{\mathsf{D}}^\delta(\rho\|\sigma) \coloneqq \inf\left\{\gamma\in\mathbb{R} : \mathrm{Tr}[(\rho-e^\gamma\sigma)_+]\leq \delta\right\}, \tag{9b}$$

where

$$(A)_+ \coloneqq \sum_{i:a_i\geq 0} a_i|i\rangle\langle i| \tag{10}$$

for a Hermitian operator $A = \sum_i a_i|i\rangle\langle i|$. Hereafter we abbreviate these divergences as DL divergences. Proposition 4.3 of [24] shows that

$$\underline{\mathsf{D}}^\delta(\rho\|\sigma) = \overline{\mathsf{D}}^{1-\delta}(\rho\|\sigma), \tag{11}$$

and so we can speak of a single DL divergence, which we set hereafter to be $\overline{\mathsf{D}}^\delta$ from (9b). Slightly rewriting (9), we have the equivalent representations:

$$\underline{\mathsf{D}}^\delta(\rho\|\sigma) = \ln\sup\left\{\lambda\geq 0 : \mathrm{Tr}[(\rho-\lambda\sigma)_+]\geq 1-\delta\right\} \tag{12a}$$

$$\overline{\mathsf{D}}^\delta(\rho\|\sigma) = \ln\inf\left\{\lambda\geq 0 : \mathrm{Tr}[(\rho-\lambda\sigma)_+]\leq \delta\right\}. \tag{12b}$$

The max-relative entropy of a state ρ and a PSD operator σ is defined as [47]

$$\mathsf{D}_{\max}(\rho\|\sigma) \coloneqq \ln\inf\{\lambda : \rho\leq\lambda\sigma\} \tag{13}$$

$$= \ln\sup_{0\leq \mathrm{M}\leq \mathrm{I}}\frac{\mathrm{Tr}[\mathrm{M}\rho]}{\mathrm{Tr}[\mathrm{M}\sigma]}, \tag{14}$$

and the smooth max-relative entropy is defined for $\delta\in[0,1]$ as

$$\mathsf{D}^\delta_{\max}(\rho\|\sigma) \coloneqq \inf_{\widetilde{\rho}:\frac{1}{2}\|\widetilde{\rho}-\rho\|_1\leq\delta}\mathsf{D}_{\max}(\widetilde{\rho}\|\sigma), \tag{15}$$

with the optimization taken over every state $\widetilde{\rho}$. These quantities have been given an operational meaning in [48].

The Thompson metric [49] is defined in terms of the max-relative entropy as

$$\mathsf{D}_T(\rho\|\sigma) \coloneqq \max\{\mathsf{D}_{\max}(\rho\|\sigma),\mathsf{D}_{\max}(\sigma\|\rho)\}, \tag{16}$$

and it has been given an operational meaning in [50] and [51].

C. Classical and Quantum Privacy

In this section, we provide background on the existing definitions of privacy for both classical and quantum systems, starting from classical DP and proceeding to quantum DP thereafter.

1) Classical Differential and Pufferfish Privacy: DP allows for answering queries about aggregate quantities while protecting the individual entries in a database [1]. To this end, the output of a differential privacy mechanism should be indistinguishable for neighboring databases, defined as those that differ only in a single record (row). Formally, we say that $x,x'\in\mathcal{X}^{n\times k}$ are neighbors, denoted $x\sim x'$, if $x(i,\cdot)\neq x'(i,\cdot)$ for some $i\in\{1,\ldots,n\}$, and they agree on all other rows. We also note that a randomized privacy mechanism A, as mentioned below, is described by a (regular) conditional probability distribution $P_{A|X}$ for its output given the data.

Definition 1 (Classical Differential Privacy): Fix $\varepsilon\geq 0$ and $\delta\in[0,1]$. A randomized mechanism $A:\mathcal{X}^{n\times k}\to\mathcal{Y}$ is (ε,δ)-differentially private if

$$\mathbb{P}\big(A(x)\in\mathcal{B}\big)\leq e^\varepsilon\,\mathbb{P}\big(A(x')\in\mathcal{B}\big)+\delta, \tag{17}$$

for all $x\sim x'$ with $x,x'\in\mathcal{X}^{n\times k}$ and $\mathcal{B}\subseteq\mathcal{Y}$ measurable.

As is evident from the above definition, DP aims to conceal whether any particular individual (row) is in fact part of the database or not. While being a powerful and widely applicable privacy framework, it is often appropriate to consider even broader frameworks. Pufferfish privacy [3] is a versatile generalization of DP that not only allows flexibility in the definition of secrets but also enables the integration of domain knowledge of the database space $\mathcal{X}^{n\times k}$. The PP framework consists of three components:

1) A set of secrets $\mathcal{S} \subseteq \mathcal{X}^{n\times k}$ of measurable subsets;
2) A set of secret pairs $\mathcal{Q} \subseteq \mathcal{S} \times \mathcal{S}$ that need to be indistinguishable in the (ε, δ) sense (cf., (18) below),
3) A class of data distributions $\Theta \subseteq \mathcal{P}(\mathcal{X}^{n\times k})$ that captures prior beliefs or domain knowledge.

As formulated next, PP aims to guarantee that all secret pairs in $\mathcal{Q}$ are indistinguishable with respect to the prior beliefs $P_X \in \Theta$.

Definition 2 (Classical pufferfish privacy): Fix $\varepsilon \geq 0$ and $\delta \in [0, 1]$. A randomized mechanism $A : \mathcal{X}^{n\times k} \to \mathcal{Y}$ is (ε, δ)-private in the pufferfish framework $(\mathcal{S}, \mathcal{Q}, \Theta)$ if

$$\mathbb{P}\big(A(X) \in \mathcal{B}\big|\mathcal{R}\big) \leq e^{\varepsilon}\, \mathbb{P}\big(A(X) \in \mathcal{B}\big|\mathcal{T}\big) + \delta, \tag{18}$$

for all $P_X \in \Theta$, $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$ with $P_X(\mathcal{R}), P_X(\mathcal{T}) > 0$, and $\mathcal{B} \subseteq \mathcal{Y}$ measurable.

DP from Definition 1 is a special case of PP when $\mathcal{S} = \mathcal{X}^{n\times k}$, the set $\mathcal{Q}$ contains all neighboring pairs of databases, and $\Theta = \mathcal{P}(\mathcal{X}^{n\times k})$ (i.e., there are no distributional assumptions, and privacy is guaranteed in the worst case). Other important examples that are subsumed by PP include (i) generic DP [52], which allows for arbitrary neighboring relationships, and (ii) attribute privacy [5], which privatizes global properties of a database (e.g., a column that corresponds to some sensitive information, such as salary).

2) Quantum Differential Privacy: QDP lifts the notion of DP to the space of quantum states, with the neighboring relation typically defined either in terms of closeness in trace distance [13], reachability by a single local operation [14],[1] or by quantum Wasserstein distance of order 1 [53]. We denote two states being neighbors by $\rho \sim \sigma$.

Definition 3 (Quantum differential privacy [13], [15]): Fix $\varepsilon \geq 0$ and $\delta \in [0, 1]$. Let $\mathcal{D}$ be a set of quantum states, and let $\mathcal{A}$ be a quantum algorithm (viz., a quantum channel). The algorithm $\mathcal{A}$ is (ε, δ)-differentially private if

$$\mathrm{Tr}[\mathrm{M}\mathcal{A}(\rho)] \leq e^{\varepsilon}\mathrm{Tr}[\mathrm{M}\mathcal{A}(\sigma)] + \delta. \tag{19}$$

for every measurement operator M (i.e., satisfying $0 \leq \mathrm{M} \leq \mathrm{I}$) and all $\rho, \sigma \in \mathcal{D}$ such that $\rho \sim \sigma$.

This definition reduces to classical DP for discrete-output mechanisms with an appropriate choice of the measurement set. See Remark III-C3 below for further details.

III. Quantum Pufferfish Privacy (QPP)

Inspired by the versatility of the classical PP framework, we propose a quantum variant thereof. Termed QPP, our framework allows for customizing the notion of private states, tailoring the feasible set of measurements to the application of interest, and incorporating domain knowledge of the state distribution into the model. As such, the QPP framework can generate a rich class of privacy definitions for both classical and quantum systems, and for hybrid classical–quantum systems as well.

A. Framework

The QPP framework requires a domain expert to specify four components: a set $\mathcal{S}$ of potential secrets, a set $\mathcal{Q} \subseteq \mathcal{S}\times\mathcal{S}$ of discriminative pairs, a set Θ of data distributions, and a set $\mathcal{M}$ of measurements. We expand on and explicitly define each component next.

1) Set $\mathcal{S}$ of Potential Secrets: Secrets are modeled as subsets of density operators that share a certain property (these subsets are merely singletons in the QDP case). The set $\mathcal{S}$ is a collection of such secret subsets. For example, if one aims to privatize the resource value V of a state, then the corresponding set of secrets is $\mathcal{S} = \bigcup_{i=1}^{n} \mathcal{T}_i$, where

$$\mathcal{T}_i = \big\{\rho \in \mathcal{D}(\mathcal{H}) : \mathsf{V}(\rho) = a_i\big\} \tag{20}$$

and $\{a_i\}_{i=1}^{n}$ are the possible values that V can take (recall that, in Fig. 1, we considered a setup relevant to hiding the resource value V being a or b).

2) Set $\mathcal{Q}$ of Discriminative Pairs: This is a subset of $\mathcal{S}\times\mathcal{S}$ that specifies which pairs of elements from $\mathcal{S}$ should be indistinguishable. Namely, if $(\mathcal{T}_1, \mathcal{T}_2) \in \mathcal{Q}$, then the goal of the privacy mechanism is to conceal whether the input belongs to $\mathcal{T}_1$ or $\mathcal{T}_2$. Note that $\rho \in \mathcal{T}_1 \Rightarrow \rho \notin \mathcal{T}_2$. We require that $\mathcal{Q}$ is symmetric, i.e., that $(\mathcal{T}_i, \mathcal{T}_j) \in \mathcal{Q}$ if and only if $(\mathcal{T}_j, \mathcal{T}_i) \in \mathcal{Q}$. Proceeding with the same example, we can set

$$\mathcal{Q} = \bigcup_{i\neq j} \{(\mathcal{T}_i, \mathcal{T}_j)\}. \tag{21}$$

3) Set Θ of Data Distributions: A collection of probability distributions $P_X \in \mathcal{P}(\mathcal{X})$ over a finite space $\mathcal{X}$ that indexes an ensemble of density operators $\{\rho^x\}_{x\in\mathcal{X}}$. Taking $X \sim P_X \in \Theta$, the matrix-valued random variable ρ^X models a density operator that is randomly chosen according to P_X. Proceeding with the same example, $\{\rho^x\}_{x\in\mathcal{X}} = \{\sigma : \sigma \in \mathcal{T}_i,\ \mathcal{T}_i \in \mathcal{S}\} \subset \mathcal{D}(\mathcal{H})$. The set Θ can be understood as capturing beliefs that the adversary has regarding the state of the system.

In the above example, we have considered a subset of density operators (i.e., $\{\rho^x\}_{x\in\mathcal{X}} \subset \mathcal{D}(\mathcal{H})$). There could be applications where we have to consider all density operators. To incorporate this, we choose the following: Fix $k \in \mathbb{N}$ and let $\mathfrak{F}_k \subset 2^{\mathcal{D}(\mathcal{H})}$ be the collection of all finite subsets of $\mathcal{D}(\mathcal{H})$ with k elements. For each $\mathcal{F} \in \mathfrak{F}_k$, we write $\mathcal{P}(\mathcal{F})$ for the class of all distributions supported on $\mathcal{F}$, and define

$$\mathcal{P}_k\big(\mathcal{D}(\mathcal{H})\big) := \bigcup_{\mathcal{F}\in\mathfrak{F}_k} \mathcal{P}(\mathcal{F}). \tag{22}$$

Every distribution $P \in \mathcal{P}_k\big(\mathcal{D}(\mathcal{H})\big)$ is supported on exactly k density operators. Note that all density operators outside of the underlying finite set comprise of the null set. We associate a random variable $X \sim P = P_X$ with each such distribution and

[1]Given two quantum states ρ and σ of n registers each, call them neighbors if it is possible to reach either σ from ρ or ρ from σ by performing a general quantum channel on a single register only.

write $\mathcal{X} = \mathrm{supp}(P_X)$ for its support. Note the slight abuse in notation, as the support of P_X changes with the distribution, which is not reflected in the generic indexing set $\mathcal{X}$. The set of data distributions in the QPP framework is now taken as $\Theta \subseteq \mathcal{P}_k\big(\mathcal{D}(\mathcal{H})\big)$ for some $k \in \mathbb{N}$.

4) Set $\mathcal{M}$ of Measurements: This set is a subset of all possible measurements, i.e., $\mathcal{M} \subseteq \{\mathrm{M} : 0 \leq \mathrm{M} \leq \mathrm{I}\}$. The choice of $\mathcal{M}$ gives the flexibility to consider only measurements that are possible under physical, legal, or ethical constraints.

Remark 1 (Designing QPP Frameworks): QPP allows system designers to explicitly encode their assumptions into the privacy framework. Setting the framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ to accurately reflect the application of interest is crucial for obtaining meaningful privacy guarantees and to optimize utility. Explicit assumptions can also help account for ethical or fairness concerns associated with quantum systems; cf. Remark 11 for a concrete example concerning quantum fairness and how it is incorporated within QPP.

Now, we are ready to present a formal definition of the quantum analog of PP, which we call QPP.

Definition 4 (Quantum Pufferfish Privacy): Fix $\varepsilon \geq 0$ and $\delta \in [0, 1]$. A quantum algorithm $\mathcal{A}$ is (ε, δ)-private in the quantum pufferfish framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ if for all $P_X \in \Theta$, $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$ with $P_X(\mathcal{R}), P_X(\mathcal{T}) > 0$, and all $\mathrm{M} \in \mathcal{M}$, the following inequality holds:

$$\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{R}})\big] \leq e^{\varepsilon}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] + \delta, \tag{23}$$

where

$$\rho^{\mathcal{R}} \coloneqq \sum_{\{x:\rho^x \in \mathcal{R}\}} q_{\mathcal{R}}(x)\rho^x, \tag{24}$$

$$q_{\mathcal{R}}(x) \coloneqq \frac{P_X(x)}{P_X(\mathcal{R})}, \tag{25}$$

$$P_X(\mathcal{R}) \coloneqq \sum_{\{x:\rho^x \in \mathcal{R}\}} P_X(x), \tag{26}$$

and $\rho^{\mathcal{T}}$ is defined similarly but with $\mathcal{T}$ instead of $\mathcal{R}$. We say that an algorithm $\mathcal{A}$ satisfies ε-QPP if it satisfies $(\varepsilon, 0)$-QPP.

Evidently, discriminative secret pairs in $\mathcal{Q}$ are indistinguishable at the output of a QPP mechanism $\mathcal{A}$ in the (ε, δ)-sense, under every measurement from the class $\mathcal{M}$.

Remark 2 (Semantics of the QPP Framework): Informally, the QPP framework provides the following privacy guarantee for fixed $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$ and $P_X \in \Theta$: For a state ρ^X chosen according to $X \sim P_X$ and input to the quantum channel $\mathcal{A}$, an adversary applying a measurement $\mathrm{M} \in \mathcal{M}$ on the channel output $\mathcal{A}(\rho^X)$ draws the same conclusions regardless of whether ρ^X belongs to $\mathcal{R}$ or $\mathcal{T}$.

Remark 3 (Incorporating Entanglement): We can incorporate entanglement in the QPP framework by introducing a reference system. Specifically, we can modify the QPP framework from $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ to $(\mathcal{S}, \mathcal{G}, \Theta, \mathcal{M}')$, where

$$\mathcal{G} \coloneqq \left\{ \begin{array}{c} (\omega_{RA}^{\mathcal{R}}, \omega_{RA}^{\mathcal{T}}) : \omega_{RA}^{\mathcal{R}}, \omega_{RA}^{\mathcal{T}} \in \mathcal{D}(\mathcal{H}_R \otimes \mathcal{H}_A), \\ \mathrm{Tr}_R\big[\omega_{RA}^{\mathcal{R}}\big] = \rho^{\mathcal{R}}, \mathrm{Tr}_R\big[\omega_{RA}^{\mathcal{T}}\big] = \rho^{\mathcal{T}}, \\ (\mathcal{R}, \mathcal{T}) \in \mathcal{Q} \end{array} \right\} \tag{27}$$

is a set of pairs of bipartite states with $\rho^{\mathcal{R}}$ and $\rho^{\mathcal{T}}$ defined similar to Definition 4. We then say that $\mathcal{A}$ is (ε, δ)-QPP in that framework if for all $P_X \in \Theta$, $\mathrm{M}' \in \mathcal{M}'$, and $(\omega_{RA}^{\mathcal{R}}, \omega_{RA}^{\mathcal{T}}) \in \mathcal{G}$, we have

$$\mathrm{Tr}\big[\mathrm{M}'(\mathcal{I} \otimes \mathcal{A})(\omega_{RA}^{\mathcal{R}})\big] \leq e^{\varepsilon}\mathrm{Tr}\big[\mathrm{M}'(\mathcal{I} \otimes \mathcal{A})(\omega_{RA}^{\mathcal{T}}))\big] + \delta. \tag{28}$$

However, it is unclear whether such a stronger privacy notion would be useful in practical applications. For example, consider $\sigma_1 \coloneqq |0\rangle\langle 0| \otimes \rho^{\mathcal{R}}$ and $\sigma_2 \coloneqq |1\rangle\langle 1| \otimes \rho^{\mathcal{T}}$ with $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$. If a measurement on the reference system can be applied, then a computational-basis measurement distinguishes σ_1 and σ_2 perfectly. Thus, it is important to choose $\mathcal{A}$ appropriately with a practically applicable $\mathcal{M}'$, such that the required indistinguishability is achieved.

We shall revisit a variant of this framework with quantum divergences in Section IX-B. The strength of the privacy framework is determined by the underlying quantum divergence. However, note that the problems discussed previously are not completely solved by the variant proposed therein.

B. Equivalent Formulation of QPP With DL Divergence

We present an equivalent formulation for (ε, δ)-QPP by means of the DL divergence from (9b). To the best of our knowledge, this provides the first operational interpretation of the DL divergence.

Proposition 1 (Equivalent Formulation of (ε, δ)-QPP): Fix the framework $(\mathcal{S}, \mathcal{Q}, \Theta, \bar{\mathcal{M}})$, with $\bar{\mathcal{M}}$ corresponding to the set of all possible measurements. Then algorithm $\mathcal{A}$ satisfies (ε, δ)-QPP with respect to the framework $(\mathcal{S}, \mathcal{Q}, \mathcal{M}, \Theta)$ if and only if for all $P_X \in \Theta$ and $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$, we have

$$\overline{\mathsf{D}}^{\delta}\big(\mathcal{A}(\rho^{\mathcal{R}})\big\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \varepsilon. \tag{29}$$

Proof: We first show that (ε, δ)-QPP implies (29). For fixed $P_X \in \Theta$ and $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$, observe that (ε, δ)-QPP corresponds to

$$\sup_{\mathrm{M} \in \bar{\mathcal{M}}} \mathrm{Tr}\big[\mathrm{M}\left(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\varepsilon}\mathcal{A}(\rho^{\mathcal{T}})\right)\big] \leq \delta. \tag{30}$$

Since

$$\sup_{\mathrm{M} \in \bar{\mathcal{M}}} \mathrm{Tr}\big[\mathrm{M}\left(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\varepsilon}\mathcal{A}(\rho^{\mathcal{T}})\right)\big] = \mathrm{Tr}\Big[\left(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\varepsilon}\mathcal{A}(\rho^{\mathcal{T}})\right)_+\Big], \tag{31}$$

as a consequence of, e.g., [15, Lemma II.1], the inequality in (30) is equivalent to

$$\mathrm{Tr}\Big[\left(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\varepsilon}\mathcal{A}(\rho^{\mathcal{T}})\right)_+\Big] \leq \delta. \tag{32}$$

By the definition in (9b), this leads to ε being a possible candidate for the optimization in $\overline{\mathsf{D}}^{\delta}\big(\mathcal{A}(\rho^{\mathcal{R}})\big\|\mathcal{A}(\rho^{\mathcal{T}})\big)$, and thus implies

$$\overline{\mathsf{D}}^{\delta}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \varepsilon. \tag{33}$$

As this holds for every $P_X \in \Theta$ and $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$, we obtain the desired implication (ε, δ)-QPP $\Rightarrow$ (29).

Next, we show that (29) implies (ε,δ)-QPP. Suppose that for all $P_X \in \Theta$ and $(\mathcal{R},\mathcal{T}) \in \mathcal{Q}$, we have $\overline{\mathsf{D}}^{\delta}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \varepsilon$. Then, for fixed $P_X \in \Theta$ and $(\mathcal{R},\mathcal{T}) \in \mathcal{Q}$, let

$$\overline{\mathsf{D}}^{\delta}\big(\mathcal{A}(\rho^{\mathcal{R}})\big\|\mathcal{A}(\rho^{\mathcal{T}})\big) = \nu, \tag{34}$$

which implies that $\operatorname{Tr}\Big[\big(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\nu}\mathcal{A}(\rho^{\mathcal{T}})\big)_+\Big] \leq \delta$. Recalling that $\nu \leq \varepsilon$ and noting that $\lambda \mapsto \operatorname{Tr}\Big[\big(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\lambda}\mathcal{A}(\rho^{\mathcal{T}})\big)_+\Big]$ is a monotonically decreasing function (cf. [24, Lemma 4.2]), we have

$$\begin{aligned}
&\sup_{\mathrm{M}\in\bar{\mathcal{M}}} \operatorname{Tr}\big[\mathrm{M}\left(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\varepsilon}\mathcal{A}(\rho^{\mathcal{T}})\right)\big] \\
&\qquad = \operatorname{Tr}\Big[\big(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\varepsilon}\mathcal{A}(\rho^{\mathcal{T}})\big)_+\Big] &(35)\\
&\qquad \leq \operatorname{Tr}\Big[\big(\mathcal{A}(\rho^{\mathcal{R}}) - e^{\nu}\mathcal{A}(\rho^{\mathcal{T}})\big)_+\Big] &(36)\\
&\qquad \leq \delta. &(37)
\end{aligned}$$

As $P_X \in \Theta$ and $(\mathcal{R},\mathcal{T}) \in \mathcal{Q}$ are arbitrary, (ε,δ)-QPP follows. ∎

In the following remark, we highlight how the DL divergence also provides a novel characterization for classical PP.

Remark 4 (Classical PP Through DL Divergence): For discrete probability distributions $p, q \in \mathcal{P}(\mathcal{Y})$, the DL divergence in Eq. (12b) reduces to

$$\overline{\mathsf{D}}_c^{\delta}(p\|q) := \ln\inf\left\{\lambda \geq 0 : \sum_{y\in\mathcal{Y}} \max\{p(y) - \lambda\, q(y), 0\} \leq \delta\right\}. \tag{38}$$

A randomized mechanism $A : \mathcal{X}^{n\times k} \to \mathcal{Y}$ is (ε,δ)-classical PP in the framework $(\mathcal{S}_c,\mathcal{Q}_c,\Theta_c)$ if for all $P_X \in \Theta_c$, $(\mathcal{R},\mathcal{T}) \in \mathcal{Q}_c$ with $P_X(\mathcal{R}), P_X(\mathcal{T}) > 0$,

$$\overline{\mathsf{D}}_c^{\delta}\big(P_{A(X)|\mathcal{R}}\|P_{A(X)|\mathcal{T}}\big) \leq \varepsilon, \tag{39}$$

where $P_{A(X)|\mathcal{R}}, P_{A(X)|\mathcal{T}}$ are the output distributions conditioned on the secret events $\mathcal{R}$ and $\mathcal{T}$, respectively. See also Remark 8 for further connections to information-theoretic quantities characterizing classical privacy frameworks.

We further note that Lemma 1 below provides a semi-definite programming characterization of the DL divergence, which reduces to a linear program in the classical case.

Remark 5 (Operational Interpretation of DL Divergence): For fixed $P_X \in \Theta$ and $(\mathcal{R},\mathcal{T}) \in \mathcal{Q}$, the DL divergence $\overline{\mathsf{D}}^{\delta}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big)$ is equal to the minimal ε that can be achieved for fixed δ via the indistinguishability condition of the QPP framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$ stated in (23).

Remark 6 (Equivalent Formulation With Hockey-Stick Divergence): Another equivalent formulation of QPP arises as a generalization of the information-theoretic equivalence for QDP [15]. Specifically, $\mathcal{A}$ is (ε,δ)-QPP with respect to the framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$, where $\bar{\mathcal{M}} = \{\mathrm{M} : 0 \leq \mathrm{M} \leq \mathrm{I}\}$, if

$$\mathsf{E}_{e^{\varepsilon}}\big(\mathcal{A}(\rho^{\mathcal{R}})\big\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \delta, \tag{40}$$

for all $P_X \in \Theta$ and $(\mathcal{R},\mathcal{T}) \in \mathcal{Q}$, where $\mathsf{E}_{\nu}(\rho\|\sigma) := \operatorname{Tr}[(\rho - \nu\sigma)_+]$ is the hockey-stick divergence for $\nu \geq 1$ [38]. Fixing P_X and $(\mathcal{R},\mathcal{T})$, the quantity $\mathsf{E}_{e^{\varepsilon}}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big)$ is the minimal δ that can be achieved for fixed ε under the indistinguishability condition from (23).

C. Reduction to Existing Privacy Frameworks

The proposed QPP framework subsumes other important privacy frameworks as special cases. These reductions are presented next.

1) Quantum DP: In QDP (Definition 3), secrets are singletons, discriminative pairs comprise states satisfying a neighboring relation, while the measurement class $\mathcal{M}$ includes all possible measurements. QPP recovers the QDP setting by making the following choices while recalling (22):[2]

$$\begin{aligned}
\mathcal{S} &= \mathcal{D},\\
\mathcal{Q} &= \{(\rho,\sigma) : \rho,\sigma \in \mathcal{D}, \rho \sim \sigma\},\\
\Theta &= \mathcal{P}_2\big(\mathcal{D}(\mathcal{H})\big),\\
\mathcal{M} &= \{\mathrm{M} : 0 \leq \mathrm{M} \leq \mathrm{I}\}.
\end{aligned} \tag{41}$$

More generally, one may add flexibility to the QDP formulation by considering other subsets Θ (i.e., $\Theta \subset \mathcal{P}_2\big(\mathcal{D}(\mathcal{H})\big)$). This can be used, for instance, to treat situations in which only certain neighboring pairs are of interest, namely, by choosing the distributions that assign positive probabilities only to those selected density operators. This can be interpreted as adding domain knowledge to the original QDP framework.

2) Quantum Local DP: In quantum local DP (QLDP) [15],[3] we choose secret pairs to be pairs of arbitrary distinct states, while the measurement class includes all possible measurements. Thus, QLDP realizes the same $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M})$ framework as QDP, except that $\mathcal{Q} = \{(\rho,\sigma) : \rho,\sigma \in \mathcal{D}\}$ for QLDP.

3) Classical PP: Consider a classical PP framework $(\mathcal{S}_c,\mathcal{Q}_c,\Theta_c)$, as specified in Definition 2. Assume that $p_X \in \Theta_c$ are discrete probability distributions over the probability space $\mathcal{P}(\mathcal{X}^{n\times k})$. Let the encoding of the database $x \in \mathcal{X}^{n\times k}$ be $\rho^x := |x\rangle\langle x|$, and denote a projective measurement operator corresponding to outcome y as $|y\rangle\langle y|$. Here note that $\{|x\rangle\}_{x\in\mathcal{X}}$ and $\{|y\rangle\}_{y\in\mathcal{Y}}$ are respective orthonormal bases formed related to the input and output alphabets of the classical PP mechanism $\mathcal{A}_c$. Then classical PP is obtained from QPP by setting

$$\begin{aligned}
\mathcal{S} &= \big\{\{\rho^x : x \in \mathcal{R}_c\} : \mathcal{R}_c \in \mathcal{S}_c\big\},\\
\mathcal{Q} &= \big\{\big(\{\rho^x : x \in \mathcal{R}_c\}, \{\rho^x : x \in \mathcal{T}_c\}\big) : (\mathcal{R}_c,\mathcal{T}_c) \in \mathcal{Q}_c\big\},\\
\Theta &= \Theta_c,\\
\mathcal{M} &= \left\{\sum_{y\in\mathcal{B}} |y\rangle\langle y| : \mathcal{B} \subseteq \mathcal{Y}\right\}.
\end{aligned} \tag{42}$$

In this scenario, assuming the output of the algorithm is discrete, we have that

$$\mathcal{A}(\rho^x) = \sum_{y\in\mathcal{Y},x'\in\mathcal{X}} p(y|x)|y\rangle\langle x'|\rho^x|x'\rangle\langle y| \tag{43}$$

where $p(y|x) = \mathbb{P}\big(\mathcal{A}_c(x) = y\big)$.

[2]For each pair of states (ρ,σ), there exists at least one probability distribution that assigns positive probability for these two states, which recovers the definitions of QDP.

[3]QLDP is also known as Local differential privacy (under the 'extreme setting', as compared to standard QDP) in Section V-A of [15].

Remark 7 (Utility-Optimized Privacy Models): As is evident from above, the measurement set corresponding to classical PP entails every subset $\mathcal{B} \subseteq \mathcal{Y}$. However, when some of the outcomes are not sensitive, we may want to relax this requirement to gain utility (cf., e.g., [33]). While classical PP does not allow for that, QPP gives extra flexibility in choosing $\mathcal{M}$ and adapting it to the application of interest. Indeed, if we only need to privatize outcomes within the set $\mathcal{Y}' \subsetneq \mathcal{Y}$, the smaller measurement set $\mathcal{M} = \left\{\sum_{y\in\mathcal{Y}'} |y\rangle\langle y| : \mathcal{B} \subseteq \mathcal{Y}'\right\}$ is sufficient.

IV. Datta–Leditzky Information Spectrum Divergence

We now focus on the DL divergence [24], whose operational interpretation in terms of QPP was provided in the previous section (see Remark 5), and we study structural properties thereof, which will be useful when analyzing the QPP framework. We first formulate a primal and dual SDP to compute the DL divergence and then use that to prove joint-quasi convexity, the data-processing inequality under positive, trace-preserving maps, and connections to the smooth max-relative entropy.

A. SDP Formulations

We now present several SDPs for computing the DL divergence in (9b), which may be of independent interest. (Recall that the other DL divergence in (9a) is easily obtained by applying the equality in (11).)

Lemma 1 (SDP Formulation of the DL Divergence): For $\delta \in (0,1)$, a state ρ, and a PSD operator σ, the following equalities hold

$$\overline{\mathsf{D}}^{\delta}(\rho\|\sigma) = \ln \inf_{\lambda, Z\geq 0} \{\lambda : \mathrm{Tr}[Z] \leq \delta,\ Z \geq \rho - \lambda\sigma\} \tag{44a}$$

$$= \ln \sup_{\mu, W\geq 0} \left\{ \begin{array}{c} \mathrm{Tr}[W\rho] - \mu\delta : \\ \mathrm{Tr}[W\sigma] \leq 1,\ W \leq \mu I \end{array} \right\}. \tag{44b}$$

Proof: Considering (12b), fix $\lambda > 0$ and first observe that

$$\mathrm{Tr}[(\rho - \lambda\sigma)_+] = \sup_{\Lambda:\, 0\leq\Lambda\leq I} \mathrm{Tr}[\Lambda\,(\rho - \lambda\sigma)]. \tag{45}$$

Indeed, this follows because, for every $0 \leq \Lambda \leq I$, we have that

$$\begin{aligned} \mathrm{Tr}[\Lambda\,(\rho - \lambda\sigma)] &= \mathrm{Tr}\left[\Lambda\left((\rho - \lambda\sigma)_+ - (\rho - \lambda\sigma)_-\right)\right] \\ &\leq \mathrm{Tr}\left[\Lambda\,(\rho - \lambda\sigma)_+\right] \\ &\leq \mathrm{Tr}\left[(\rho - \lambda\sigma)_+\right], \end{aligned} \tag{46}$$

and the inequalities above are all attained by setting Λ to be the projection onto the support of $(\rho - \lambda\sigma)_+$. The SDP dual of this quantity is given by

$$\mathrm{Tr}\left[(\rho - \lambda\sigma)_+\right] = \inf_{Z\geq 0} \{\mathrm{Tr}[Z] : Z \geq \rho - \lambda\sigma\}. \tag{47}$$

Intuitively, $Z = (\rho - \lambda\sigma)_+$ is the smallest choice of a PSD operator that satisfies the constraint $Z \geq \rho - \lambda\sigma$.

We then find from (9b), (12b), and (47) that

$$\begin{aligned} \overline{\mathsf{D}}^{\delta}_s(\rho\|\sigma) &= \ln\inf\left\{\lambda \geq 0 : \mathrm{Tr}[(\rho - \lambda\sigma)_+] \leq \delta\right\} \\ &= \ln \inf_{\lambda, Z\geq 0} \{\lambda : \mathrm{Tr}[Z] \leq \delta,\ Z \geq \rho - \lambda\sigma\}, \end{aligned} \tag{48}$$

which completes the proof of (44a).

The dual forms of these optimization problems are derived from the canonical primal and dual formulations of SDPs, which are respectively given by (cf. [37, Definition 2.20])

$$\begin{aligned} &\inf_{Y\geq 0} \left\{\mathrm{Tr}[BY] : \Phi^\dagger(Y) \geq A\right\}, \\ &\sup_{X\geq 0} \left\{\mathrm{Tr}[AX] : \Phi(X) \leq B\right\}, \end{aligned} \tag{49}$$

where A and B are Hermitian matrices and Φ is a Hermiticity-preserving superoperator. Comparing the former to (48), we make the following choices so that the general optimization problem recovers (48) (inside the logarithm):

$$Y = \begin{bmatrix} \lambda & 0 \\ 0 & Z \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{50}$$

$$\Phi^\dagger(Y) = \begin{bmatrix} -\mathrm{Tr}[Z] & 0 \\ 0 & Z + \lambda\sigma \end{bmatrix}, \tag{51}$$

$$A = \begin{bmatrix} -\delta & 0 \\ 0 & \rho \end{bmatrix}. \tag{52}$$

Then, setting

$$X = \begin{bmatrix} \mu & 0 \\ 0 & W \end{bmatrix}, \tag{53}$$

we solve for the map $\Phi(X)$ to find that

$$\begin{aligned} &\mathrm{Tr}[X\Phi^\dagger(Y)] \\ &= \mathrm{Tr}\left[\begin{bmatrix} \mu & 0 \\ 0 & W \end{bmatrix}\begin{bmatrix} -\mathrm{Tr}[Z] & 0 \\ 0 & Z + \lambda\sigma \end{bmatrix}\right] && (54) \\ &= -\mu\,\mathrm{Tr}[Z] + \mathrm{Tr}[W\,(Z + \lambda\sigma)] && (55) \\ &= \mathrm{Tr}[(W - \mu I)\,Z] + \lambda\,\mathrm{Tr}[W\sigma] && (56) \\ &= \mathrm{Tr}\left[\begin{bmatrix} \lambda & 0 \\ 0 & Z \end{bmatrix}\begin{bmatrix} \mathrm{Tr}[W\sigma] & 0 \\ 0 & W - \mu I \end{bmatrix}\right] && (57) \\ &= \mathrm{Tr}[Y\Phi(X)], && (58) \end{aligned}$$

so that

$$\Phi(X) = \begin{bmatrix} \mathrm{Tr}[W\sigma] & 0 \\ 0 & W - \mu I \end{bmatrix}. \tag{59}$$

Plugging into the dual form, we obtain

$$\begin{aligned} &\sup_{X\geq 0} \{\mathrm{Tr}[AX] : \Phi(X) \leq B\} \\ &\quad = \sup_{\mu, W\geq 0} \{\mathrm{Tr}[W\rho] - \mu\delta : \mathrm{Tr}[W\sigma] \leq 1, W \leq \mu I\}. \end{aligned} \tag{60}$$

Choose $\mu = \mu_1 \in (0,1)$ and $W = \mu_2 I$ such that $\mu_1\delta < \mu_2 < \mu_1$, as a strictly feasible solution to the above. For the other SDP formulation from (48), set λ such that $\mathrm{Tr}[(\rho - \lambda\sigma)_+] \leq \delta$, and $Z = (\rho - \lambda\sigma)_+ \geq 0$ as a feasible solution. By Slater's condition, we conclude that strong duality holds, and the primal and dual optimal values coincide. ∎

Corollary 1 (Another Formulation of the DL Divergence): DL divergence has the following equivalent formulation:

$$\overline{\mathsf{D}}^{\delta}(\rho\|\sigma) = \ln \sup_{0\leq W\leq I, \mathrm{Tr}[W\rho]\geq\delta} \frac{\mathrm{Tr}[W\rho] - \delta}{\mathrm{Tr}[W\sigma]}. \tag{61}$$

Proof: Consider the SDP formulation in (44b) and set $W' = \frac{W}{\mu}$ therein to arrive at

$$\overline{\mathsf{D}}^{\delta}(\rho\|\sigma) = \ln \sup_{\mu, W' \geq 0} \left\{ \begin{array}{c} \mu \operatorname{Tr}[W'\rho] - \mu\delta : \\ \mu \operatorname{Tr}[W'\sigma] \leq 1,\ W' \leq I \end{array} \right\} \tag{62}$$

$$= \ln \sup_{0 \leq W' \leq I, \operatorname{Tr}[W'\rho] \geq \delta} \frac{\operatorname{Tr}[W'\rho] - \delta}{\operatorname{Tr}[W'\sigma]}, \tag{63}$$

where the last equality follows from identifying that $\mu = 1/\operatorname{Tr}[W'\sigma]$ is the μ that maximizes the former, given that $\operatorname{Tr}[W'\rho] \geq \delta$. When $\operatorname{Tr}[W'\rho] < \delta$, the optimum $\mu = 0$ and the objective within the supremum becomes zero. Replacing W' by W, concludes the proof. ∎

Remark 8 (Approximate Max-Divergence): In [2], the δ-approximate max-divergence is defined as

$$\mathsf{D}_{\infty}^{\delta}(p_Y\|p_Z) \coloneqq \ln \max_{S \in \operatorname{Supp}(Y), \Pr[Y \in S] \geq \delta} \frac{\Pr[Y \in S] - \delta}{\Pr[Z \in S]}, \tag{64}$$

where Y and Z are random variables distributed according to $Y \sim p_Y$ and $Z \sim p_Z$. By substituting classical distributions into Corollary 1, we observe that the DL divergence reduces to the approximate max-divergence. Note that the approximate max-divergence has been used to characterize (ε, δ)-(classical) DP in [2, Remark 3.1]. Thus, this showcases that the equivalence we established for QPP with the DL divergence herein reduces to the existing equivalence for (classical) DP.

B. Properties

We derive several properties of the DL divergence from (12b), which are subsequently used in the analysis of the QPP framework. Basic properties of the DL divergence, including the data-processing inequality, have been proven in [24, Proposition 4.3]. Here, we generalize the data-processing inequality to hold for arbitrary positive, trace non-increasing maps (beyond the set of quantum channels) and also establish joint-quasi convexity of the DL divergence, along with its connection to the smooth max-relative entropy (recall the definition in (15)). The proofs of these properties rely on the SDP formulation from Lemma 1.

Proposition 2 (Properties of the DL Divergence): Fix $\delta \in (0,1)$, and let $\rho, \rho_1, \ldots, \rho_k$ and $\sigma, \sigma_1, \ldots, \sigma_k$ be two collections of states and PSD operators, respectively. The DL divergence in (12b) satisfies the following properties:

1) Data-processing inequality: For every positive, trace non-increasing map $\mathcal{N}$, we have

$$\overline{\mathsf{D}}^{\delta}(\rho\|\sigma) \geq \overline{\mathsf{D}}^{\delta}\big(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)\big). \tag{65}$$

2) Joint-quasi convexity: Let $p_i \in [0,1]$, for $i \in \{1, \ldots, k\}$, with $\sum_{i=1}^{k} p_i = 1$. Then

$$\overline{\mathsf{D}}^{\delta}\left(\sum_{i=1}^{k} p_i\rho_i \middle\| \sum_{i=1}^{k} p_i\sigma_i\right) \leq \max_i \overline{\mathsf{D}}^{\delta}(\rho_i\|\sigma_i), \tag{66}$$

and, more generally,

$$\overline{\mathsf{D}}^{\delta'}\left(\sum_{i=1}^{k} p_i\rho_i \middle\| \sum_{i=1}^{k} p_i\sigma_i\right) \leq \max_i \overline{\mathsf{D}}^{\delta_i}(\rho_i\|\sigma_i), \tag{67}$$

where $\delta' \coloneqq \sum_{i=1}^{k} p_i\delta_i$ with $\delta_1, \ldots, \delta_k \in (0,1)$.

3) Relation to smooth max-relative entropy:

$$\overline{\mathsf{D}}^{\delta}(\rho\|\sigma) \leq \mathsf{D}_{\max}^{\delta}(\rho\|\sigma) \leq \overline{\mathsf{D}}^{\delta'}(\rho\|\sigma) - \ln(1-\delta'), \tag{68}$$

where $\delta' \coloneqq 1 - \sqrt{1-\delta^2} \in (0,1)$, and the second inequality above can be equivalently written as

$$\mathsf{D}_{\max}^{\sqrt{\delta(2-\delta)}}(\rho\|\sigma) \leq \overline{\mathsf{D}}^{\delta}(\rho\|\sigma) - \ln(1-\delta). \tag{69}$$

4) Quasi subadditivity: Let $\delta_1, \delta_2 \in (0,1)$ satisfy $\delta_1' + \delta_2' < 1$, with $\delta_i' \coloneqq \sqrt{\delta_i(2-\delta_i)} \in (0,1)$ for $i \in \{1,2\}$. Then

$$\overline{\mathsf{D}}^{\delta_1'+\delta_2'}(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) \leq \overline{\mathsf{D}}^{\delta_1}(\rho_1\|\sigma_1) + \overline{\mathsf{D}}^{\delta_2}(\rho_2\|\sigma_2) - \ln\big((1-\delta_1)(1-\delta_2)\big). \tag{70}$$

Furthermore,

a) if $\delta_1 = \delta_2 = 0$, then

$$\overline{\mathsf{D}}^{0}(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) \leq \overline{\mathsf{D}}^{0}(\rho_1\|\sigma_1) + \overline{\mathsf{D}}^{0}(\rho_2\|\sigma_2). \tag{71}$$

b) if σ_1, σ_2 are states, then

$$\overline{\mathsf{D}}^{\delta}(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) \leq \overline{\mathsf{D}}^{\delta_1}(\rho_1\|\sigma_1) + \overline{\mathsf{D}}^{\delta_2}(\rho_2\|\sigma_2), \tag{72}$$

where

$$\delta \coloneqq \min\left\{\delta_1 + e^{\overline{\mathsf{D}}^{\delta_1}(\rho_1\|\sigma_1)}\delta_2,\ \delta_2 + e^{\overline{\mathsf{D}}^{\delta_2}(\rho_2\|\sigma_2)}\delta_1\right\}. \tag{73}$$

Proof: Property 1: The statement was proven in [24, Proposition 4.3] for a quantum channel $\mathcal{N}$ by using the inequality

$$\operatorname{Tr}[(\rho - e^{\gamma}\sigma)_+] \geq \operatorname{Tr}\big[(\mathcal{N}(\rho) - e^{\gamma}\mathcal{N}(\sigma))_+\big], \tag{74}$$

which holds for all $\gamma \in \mathbb{R}$. Here, we prove the data-processing inequality, but we generalize it to hold for a positive, trace non-increasing map $\mathcal{N}$. Our derivation relies on the SDP formulation of the DL divergence from (44a).

Let $\lambda^\star$ and $Z^\star$ be optimal choices[4] in the optimization for $\overline{\mathsf{D}}^{\delta}(\rho\|\sigma)$, so that $\overline{\mathsf{D}}^{\delta}(\rho\|\sigma) = \ln\lambda^\star$, $Z^\star \geq \rho - \lambda^\star\sigma$ with $\operatorname{Tr}[Z^\star] \leq \delta$, and $Z^\star \geq 0$ (indeed, note that the infimum is achieved with $\operatorname{Tr}[(\rho - \lambda^\star\sigma)_+] = \delta$). Since $Z^\star - (\rho - \lambda^\star\sigma) \geq 0$, it follows that $\mathcal{N}\big(Z^\star - (\rho - \lambda^\star\sigma)\big) \geq 0$ from the assumption that $\mathcal{N}$ is a positive map. Consequently, we obtain

$$Z' \coloneqq \mathcal{N}(Z^\star) \geq \mathcal{N}(\rho) - \lambda^\star\mathcal{N}(\sigma). \tag{75}$$

[4]When the DL divergence is finite, the infimum is achieved by a standard continuity plus compactness argument. The stated relations trivially hold when the DL divergence is infinite.

Furthermore, $Z' \geq 0$ since $Z^\star \geq 0$ and $\mathcal{N}$ is a positive map. Additionally, since $\mathcal{N}$ is trace non-increasing, it follows that

$$\mathrm{Tr}[Z'] \leq \mathrm{Tr}[Z^\star] \leq \delta. \tag{76}$$

Thus, $\lambda^\star$ is a feasible point for $\overline{\mathsf{D}}^\delta\big(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)\big)$. We conclude the proof by noting that the quantity $\overline{\mathsf{D}}^\delta\big(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)\big)$ involves a minimization over all such feasible points, implying the desired inequality:

$$\overline{\mathsf{D}}^\delta\big(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)\big) \leq \ln(\lambda^\star) = \overline{\mathsf{D}}^\delta(\rho\|\sigma)\,. \tag{77}$$

Note that this property can also be derived using the proof of [38, Lemma 4].

Property 2: We again consider the SDP from (44a). Let $\lambda_i^\star$ and $Z_i^\star$ be optimal for $\overline{\mathsf{D}}^\delta(\rho_i\|\sigma_i)$, so that $\overline{\mathsf{D}}^\delta(\rho_i\|\sigma_i) = \ln(\lambda_i^\star)$, $Z_i^\star \geq \rho_i - \lambda_i^\star\sigma_i$ with $\mathrm{Tr}[Z_i^\star] \leq \delta$, and $Z_i^\star \geq 0$. Define

$$Z \coloneqq \sum_{i=1}^k p_i Z_i^\star \geq \sum_{i=1}^k p_i\rho_i - \sum_{i=1}^k \lambda_i^\star p_i\sigma_i. \tag{78}$$

and observe that $\mathrm{Tr}[Z] \leq \delta$ and $Z \geq 0$. This implies that $Z \geq \sum_{i=1}^k p_i\rho_i - \max_i \lambda_i^\star \sum_{j=1}^k p_j\sigma_j$, which suggests that $\max_i \lambda_i^\star$ and Z are (candidate) infimizers in the SDP formulation of $\overline{\mathsf{D}}^\delta\Big(\sum_{i=1}^k p_i\rho_i\Big\|\sum_{i=1}^k p_i\sigma_i\Big)$. Consequently, we obtain

$$\overline{\mathsf{D}}^\delta\left(\sum_{i=1}^k p_i\rho_i\middle\|\sum_{i=1}^k p_i\sigma_i\right) \leq \ln\Big(\max_i \lambda_i^\star\Big) = \max_i \overline{\mathsf{D}}^\delta(\rho_i\|\sigma_i)\,. \tag{79}$$

The proof of the general case follows along the same lines by observing that $\mathrm{Tr}[Z] \leq \sum_{i=1}^k p_i\delta_i$.

We also present an alternative proof for this using the joint-convexity of hockey-stick divergence in Appendix I.

Property 3: From [48, Appendix B], we have that

$$\mathsf{D}_{\max}^\delta(\rho\|\sigma) = \ln \inf_{\lambda,\widetilde{\rho},Y\geq 0}\left\{\begin{array}{ll} \lambda : \widetilde{\rho} \leq \lambda\sigma, & \mathrm{Tr}[Y] \leq \delta, \\ \mathrm{Tr}[\widetilde{\rho}] = 1, & Y \geq \rho - \widetilde{\rho} \end{array}\right\}. \tag{80}$$

Let λ, Y, and $\widetilde{\rho}$ be arbitrary operators satisfying the constraints for $\mathsf{D}_{\max}^\delta(\rho\|\sigma)$. Then by combining the inequalities $\widetilde{\rho} \leq \lambda\sigma$ and $Y \geq \rho - \widetilde{\rho}$, we get

$$Y \geq \rho - \lambda\sigma. \tag{81}$$

We see that λ and Y satisfy the constraints needed for λ and Z, respectively, in the SDP for $\overline{\mathsf{D}}^\delta(\rho\|\sigma)$, whereby

$$\overline{\mathsf{D}}^\delta(\rho\|\sigma) \leq \lambda. \tag{82}$$

Since the argument holds for all λ, Y, and $\widetilde{\rho}$ satisfying the constraints in the definition of $\mathsf{D}_{\max}^\delta(\rho\|\sigma)$, we further obtain

$$\overline{\mathsf{D}}^\delta(\rho\|\sigma) \leq \mathsf{D}_{\max}^\delta(\rho\|\sigma). \tag{83}$$

The proof is concluded by invoking the following lemma (proven in Appendix II).

Lemma 2: Fix $\lambda > 0$, let ρ be a state and σ a positive semi-definite operator, and define $\delta \coloneqq \mathrm{Tr}[(\rho - \lambda\sigma)_+]$. Then

$$\mathsf{D}_{\max}^{\sqrt{\delta(2-\delta)}}(\rho\|\sigma) \leq \ln\lambda - \ln(1-\delta). \tag{84}$$

For fixed $\delta \in (0,1)$, by definition, we have $\overline{\mathsf{D}}^\delta(\rho\|\sigma) = \ln(\lambda^\star)$ with $\delta = \mathrm{Tr}[(\rho - \lambda^\star\sigma)_+]$. With that, Lemma 2 with the reparametrization $\delta \to 1 - \sqrt{1-\delta^2}$, yields

$$\mathsf{D}_{\max}^\delta(\rho\|\sigma) \leq \overline{\mathsf{D}}^{1-\sqrt{1-\delta^2}}(\rho\|\sigma) + \ln\left(\frac{1}{\sqrt{1-\delta^2}}\right). \tag{85}$$

This completes the proof.

Property 4: This follows by invoking Property 3 and using the fact that the smooth max-relative entropy satisfies subadditivity (Appendix III) with

$$\mathsf{D}_{\max}^{\delta_1+\delta_2}(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2) \leq \mathsf{D}_{\max}^{\delta_1}(\rho_1\|\sigma_1) + \mathsf{D}_{\max}^{\delta_2}(\rho_2\|\sigma_2). \tag{86}$$

Part (a) now follows by taking the limits $\delta_1 \to 0$ and $\delta_2 \to 0$ in (86), and applying Property 3.

To prove Part (b), we use the SDP formulation in (44a). Let $\overline{\mathsf{D}}^{\delta_i}(\rho_i\|\sigma_i) = \ln(\lambda_i^\star)$ for $i \in \{1,2\}$. It follows that $Z_i \geq \rho_i - \lambda_i^\star\sigma_i$ with $\mathrm{Tr}[Z_i] \leq \delta$ and $Z_i \geq 0$. Consider that

$$\begin{aligned}
&(\rho_1\otimes\rho_2) - \lambda_1^\star\lambda_2^\star(\sigma_1\otimes\sigma_2) \\
&= (\rho_1\otimes\rho_2) - \lambda_1^\star\sigma_1\otimes\rho_2 + \lambda_1^\star\sigma_1\otimes\rho_2 - \lambda_1^\star\lambda_2^\star(\sigma_1\otimes\sigma_2) \\
&= (\rho_1 - \lambda_1^\star\sigma_1)\otimes\rho_2 + \lambda_1^\star\sigma_1\otimes(\rho_2 - \lambda_2^\star\sigma_2) \\
&\leq Z_1\otimes\rho_2 + \lambda_1^\star\sigma_1\otimes Z_2 =: Z.
\end{aligned} \tag{87}$$

Observe that $Z \geq 0$ and $\mathrm{Tr}[Z] = \mathrm{Tr}[Z_1] + \lambda_1^\star\mathrm{Tr}[Z_2]$, since $\mathrm{Tr}[\rho_1] = \mathrm{Tr}[\sigma_1] = 1$. Consequently, we have $\mathrm{Tr}[Z] \leq \delta_1 + \lambda_1^\star\delta_2$, and $\lambda_1^\star\lambda_2^\star$ is a candidate infimizer. For $\delta' = \delta_1 + \lambda_1^\star\delta_2$, we now arrive at

$$\begin{aligned}
\overline{\mathsf{D}}^{\delta'}(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2) &\leq \ln(\lambda_1^\star\lambda_2^\star) && (88)\\
&= \ln(\lambda_1^\star) + \ln(\lambda_2^\star) && (89)\\
&= \overline{\mathsf{D}}^{\delta_1}(\rho_1\|\sigma_1) + \overline{\mathsf{D}}^{\delta_2}(\rho_2\|\sigma_2)\,. && (90)
\end{aligned}$$

The above holds for $\delta' = \delta_2 + \lambda_2^\star\delta_1$ as well, by adding and subtracting $\lambda_2^\star\rho_1\otimes\sigma_1$ instead of $\lambda_1^\star\sigma_1\otimes\rho_2$, and then following the same argument. ∎

V. Properties and Mechanisms for QPP

A. Properties of QPP Mechanisms

Modern guidelines for privacy frameworks [54] render properties such as convexity and post-processing (also known as transformation invariance) as basic requirements for privacy frameworks. Composability is another important property, which implies that a combination of privacy mechanisms is itself private. These properties are known to hold for the classical mutual information PP framework, and all of them, except for composability, hold for the classical PP framework; cf. [12, Theorem 2] and [3, Theorem 5.1], respectively.

Before proving these properties for the QPP framework, we discuss their operational interpretation. Convexity means that applying a QPP mechanism that is randomly chosen from a given set of such mechanisms still satisfies QPP. Post-processing ensures that passing the output of a QPP mechanism $\mathcal{A}$ through a channel $\mathcal{N}$ preserves QPP; see

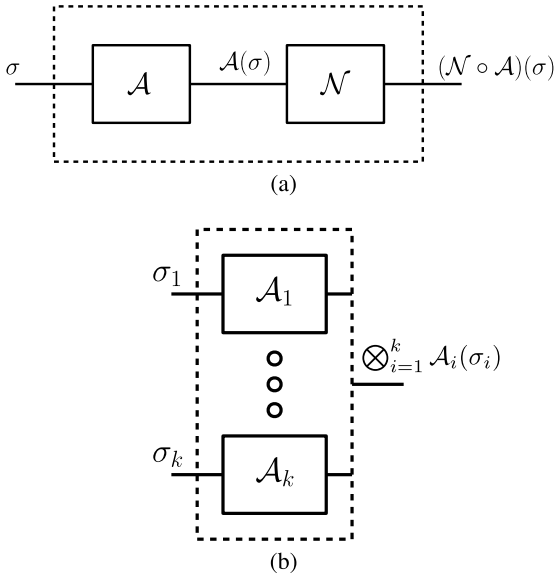


Fig. 2. Properties of QPP mechanisms: (a) refers to post-processing of QPP algorithm $\mathcal{A}$; If $\mathcal{A}$ satisfies QPP, then $\mathcal{N}\circ\mathcal{A}$ also satisfies QPP. (b) refers to parallel composition of k QPP mechanisms; composition of k mechanisms independently in a parallel fashion satisfies QPP if each $\mathcal{A}_i$ satisfies QPP.

Fig. 2a. Parallel composability is illustrated in Fig. 2b and guarantees that QPP holds after applying

$$\mathcal{A}^{(k)} \coloneqq \bigotimes_{i=1}^{k} \mathcal{A}_i = \mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \cdots \otimes \mathcal{A}_k \tag{91}$$

to the input $\rho^{X_1}\otimes\rho^{X_2}\otimes\cdots\otimes\rho^{X_k}$, with $X_i \sim P_X \in \Theta$, where each X_i is independently chosen. Informally, the semantic meaning of this property is that after applying $\mathcal{A}^{(k)}$, the same conclusions can be drawn about the input $\rho^{X_1}\otimes\rho^{X_2}\otimes\cdots\otimes\rho^{X_k}$ regardless of whether each ρ^{X_i} belongs to $\mathcal{R}_i$ or $\mathcal{T}_i$, where $(\mathcal{R}_i, \mathcal{T}_i) \in \mathcal{Q}$ for all $i \in \{1,\ldots,k\}$. In this setting, the set of discriminative pairs is taken as

$$\mathcal{Q}^{(k)} \coloneqq \left\{ (\mathcal{R}^{(k)}, \mathcal{T}^{(k)}) : \begin{array}{c} \mathcal{R}^{(k)} \coloneqq (\mathcal{R}_1,\ldots,\mathcal{R}_k), \\ \mathcal{T}^{(k)} \coloneqq (\mathcal{T}_1,\ldots,\mathcal{T}_k) \\ \forall i \in \{1,\ldots,k\}\ (\mathcal{R}_i,\mathcal{T}_i) \in \mathcal{Q} \end{array} \right\}. \tag{92}$$

Furthermore, the class of product measurements is $\bigotimes_{i=1}^{k}\mathcal{M}_i$ (i.e., the output of algorithm $\mathcal{A}_i$ is followed by a measurement from $\mathcal{M}_i$, for all $i \in \{1,\ldots,k\}$), while the set of all possible measurements on the k systems, including joint measurements, is denoted by $\bar{\mathcal{M}}^k$. We note here that one could consider other classes of limited measurements, such as local operations and classical communication (LOCC) measurements and positive-partial-transpose (PPT) measurements [55].

The formal statement of these properties is as follows.

Theorem 1 (Properties of QPP Mechanisms): The following properties hold:

1) Convexity: Let $\mathcal{A}_1,\ldots,\mathcal{A}_k$ be (ε,δ)-QPP mechanisms in the framework $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M})$. Take I to be a k-ary categorical random variable with probability distribution $(p_1,\ldots,p_k)$. Then the mechanism $\mathcal{A} \coloneqq \mathcal{A}_I$ (i.e., $\mathcal{A} = \mathcal{A}_i$ with probability p_i, for $i \in \{1,\ldots,k\}$) also satisfies (ε,δ)-QPP in the same framework $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M})$.

2) Post-processing: If a mechanism $\mathcal{A}$ satisfies (ε,δ)-QPP in the framework $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M})$, then, for a quantum channel $\mathcal{N}$, the processed mechanism $\mathcal{N}\circ\mathcal{A}$ also satisfies (ε,δ)-QPP in the framework $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M}')$, where $\mathcal{M}' \subseteq \left\{\mathrm{M}' : \mathcal{N}^\dagger(\mathrm{M}') \in \mathcal{M}\right\}$.

3) Parallel composability (non-adaptive): Let $\mathcal{A}_1,\ldots,\mathcal{A}_k$ be mechanisms such that $\mathcal{A}_i$ is (ε_i,δ_i)-QPP in the framework $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M}_i)$, for each $i \in \{1,\ldots,k\}$. Then the composed mechanism

$$\mathcal{A}^{(k)} : \bigotimes_{i=1}^{k} \sigma_i \mapsto \mathcal{A}_1(\sigma_1)\otimes\cdots\otimes\mathcal{A}_k(\sigma_k) \tag{93}$$

satisfies $\left(\sum_{i=1}^{k}\varepsilon_i, \sum_{i=1}^{k}\delta_i\right)$-QPP in the framework $\left(\mathcal{S},\mathcal{Q}^{(k)},\Theta,\bigotimes_{i=1}^{k}\mathcal{M}_i\right)$.

Proof: See Appendix IV. ∎

More broadly, parallel composition (i.e., Property 3 of Theorem 1) holds under particular separable measurements that are defined as follows:

$$\left\{ \sum_j \mathrm{M}_1^{(j)}\otimes\cdots\otimes\mathrm{M}_k^{(j)} : \forall i \sum_j \mathrm{M}_i^{(j)} \in \mathcal{M}_i,\ \forall i,j\ \mathrm{M}_i^{(j)} \geq 0 \right\}. \tag{94}$$

where product measurements considered in Theorem 1 are a special case.

The latter two properties of Theorem 1 change if one considers a measurement class that comprises all possible measurements $\bar{\mathcal{M}}^k$, as opposed to only product measurements. This is one of the main distinctions between the semi-classical and quantum cases, where, for the latter, joint measurements can infer more information and thus privacy degrades. The following theorem accounts for this latter scenario.

Theorem 2 (Properties of QPP with $\mathcal{M} = \bar{\mathcal{M}}$): The following properties hold for the case in which the measurement class is $\bar{\mathcal{M}}$:

1) Convexity: Let $\mathcal{A}_1,\ldots,\mathcal{A}_k$ be (ε,δ)-QPP mechanisms in the framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$. Take I to be a k-ary categorical random variable with parameters $(p_1,\ldots,p_k)$. Then the mechanism $\mathcal{A} \coloneqq \mathcal{A}_I$ (i.e., $\mathcal{A} = \mathcal{A}_i$ with probability p_i, for $i \in \{1,\ldots,k\}$) also satisfies (ε,δ)-QPP in the framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$.

2) Post-processing: If a mechanism $\mathcal{A}$ satisfies (ε,δ)-QPP in the framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$, then, for a quantum channel $\mathcal{N}$, the mechanism $\mathcal{N}\circ\mathcal{A}$ also satisfies (ε,δ)-QPP in the framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$.

3) Parallel composability: If $\mathcal{A}_i$ satisfies (ε_i,δ_i)-QPP in $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$ for $i \in \{1,2\}$, then the composed mechanism $\mathcal{A}_1\otimes\mathcal{A}_2$ satisfies (ε',δ')-QPP in the framework $\left(\mathcal{S},\mathcal{Q}^{(2)},\Theta,\bar{\mathcal{M}}^2\right)$ where

$$\varepsilon' \coloneqq \varepsilon_1 + \varepsilon_2 + \ln\left(\frac{1}{(1-\delta_1)(1-\delta_2)}\right), \tag{95}$$

$$\delta' \coloneqq \sqrt{\delta_1(2-\delta_1)} + \sqrt{\delta_2(2-\delta_2)}. \tag{96}$$

and $\mathcal{A}_1 \otimes \mathcal{A}_2$ also satisfies $(\varepsilon_1 + \varepsilon_2, \delta)$-QPP where

$$\delta := \min\{\delta_1 + e^{\varepsilon_1}\delta_2, \delta_2 + e^{\varepsilon_2}\delta_1\}. \tag{97}$$

Observe that, if $\delta_i = 0$ for $i \in \{1, 2\}$, then $\mathcal{A}_1 \otimes \mathcal{A}_2$ satisfies $(\varepsilon_1 + \varepsilon_2)$-QPP for the parallel composed framework.

Proof: The proof of Corollary 2 relies on properties of the DL divergence established in Proposition 2. Items 1), 2), and 3) follow from joint quasi-convexity (Property 2), data processing (Property 1), and quasi subadditivity (Property 4), respectively. ∎

Remark 9 (Comparison to Existing Results): In [15, Corollary III.3], the parallel composition of two mechanisms that satisfy $(\varepsilon_i, \delta_i)$-QDP for $i \in \{1, 2\}$ is shown to be $(\varepsilon_1 + \varepsilon_2, \delta)$-QDP, where δ is given in (97). The proof technique is, however, different from ours. Property 3 of Theorem 1 also reveals that if one considers a restricted class of measurements (e.g., product measurements), then it is possible to achieve tighter privacy guarantees (namely, $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-QDP) than those obtained when allowing all joint measurements on the two systems. Also note that δ' in (96) is independent of ε_1 and ε_2, whereas δ in (97) depends on them. Depending on the particular values that the parameters ε_i and δ_i take, for $i \in \{1, 2\}$, one of these aforementioned results provides sharper privacy guarantees.

Example 1: Here we provide an example to illustrate the distinction between the case of joint measurements and product measurements, and more generally PPT measurements (which contain the set of LOCC measurements, as well as product measurements). Let α_d be the maximally mixed state on the antisymmetric subspace of two d-dimensional systems, and let σ_d be the maximally mixed state on the symmetric subspace, i.e.,

$$\alpha_d := \frac{I - F}{d\,(d-1)}, \tag{98}$$

$$\sigma_d := \frac{I + F}{d\,(d+1)}, \tag{99}$$

where $F := \sum_{i,j} |i\rangle\langle j| \otimes |j\rangle\langle i|$ is the unitary swap operator. These states are orthogonal and thus perfectly distinguishable by a joint measurement. Indeed, this measurement is given by $\{\Pi^{\alpha_d}, \Pi^{\sigma_d}\}$, where $\Pi^{\alpha_d} := (I - F)/2$ and $\Pi^{\sigma_d} := (I + F)/2$. By setting $\mathrm{M} = \Pi^{\alpha_d}$, we find that $\mathrm{Tr}[\mathrm{M}\alpha_d] = 1$ and $\mathrm{Tr}[\mathrm{M}\sigma_d] = 0$.

We can consider a QPP framework with $\mathcal{Q} = \{(\alpha_d, \sigma_d)\}$ and the set of measurements to be $\bar{\mathcal{M}}$. In this case, we only have QPP (i.e., the inequality $\mathrm{Tr}[\mathrm{M}\alpha_d] \leq e^{\varepsilon}\,\mathrm{Tr}[\mathrm{M}\sigma_d] + \delta$ is satisfied) by setting $\varepsilon \geq 0$ arbitrary and $\delta \geq 1$, which is a weak privacy guarantee (or really no privacy at all).

However, we can alternatively restrict the measurement operators to PPT measurement operators, i.e., those M which satisfy $0 \leq \mathrm{M} \leq I$ and $0 \leq \mathrm{M}^{\Gamma} \leq I$, where the Γ superscript denotes the partial transpose. In this case, we find for every such PPT measurement operator M that

$$\mathrm{Tr}[\mathrm{M}\alpha_d] = \mathrm{Tr}[\mathrm{M}^{\Gamma}\alpha_d^{\Gamma}] \tag{100}$$

$$= \mathrm{Tr}[\mathrm{M}^{\Gamma}\left(I - d\,\Phi^d\right)/\left(d\,(d-1)\right)] \tag{101}$$

$$\leq \mathrm{Tr}[\mathrm{M}^{\Gamma}\left(I + d\,\Phi^d\right)/\left(d\,(d-1)\right)] \tag{102}$$

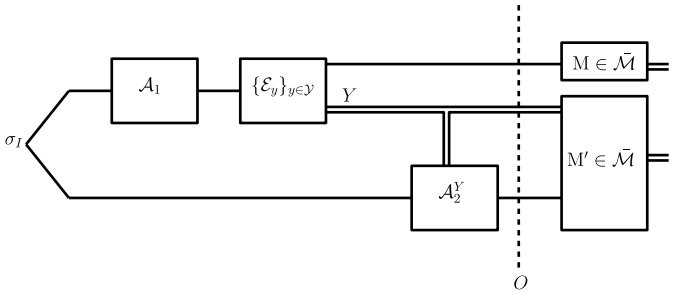


Fig. 3. Setup for adaptive composition: On the top system, the channel $\mathcal{A}_1$ is followed by the quantum instrument $\{\mathcal{E}_y\}_{y\in\mathcal{Y}}$, and then the random classical outcome Y is used to choose the channel $\mathcal{A}_2^Y$. In this setting, we analyse how well an adversary can learn properties of the input state σ_I by applying measurements on the output state.

$$= \mathrm{Tr}[\mathrm{M}\,(I + F)\,/\,(d\,(d-1))] \tag{103}$$

$$= \frac{d+1}{d-1}\,\mathrm{Tr}[\mathrm{M}\sigma_d]. \tag{104}$$

The first equality follows because the partial transpose is its own adjoint, and the second equality follows by introducing the maximally entangled state $\Phi^d := \frac{1}{d}\sum_{i,j} |i\rangle\langle j| \otimes |i\rangle\langle j|$. The inequality follows because $0 \leq \mathrm{M}^{\Gamma}$. By applying the above inequality, we conclude that (ε, δ)-QPP holds with $\varepsilon = \ln\left(\frac{d+1}{d-1}\right)$ and $\delta = 0$, so that privacy improves as dimension increases.

1) Adaptive Composability: Adaptive composition refers to the case when each subsequently composed mechanism is chosen based on the outputs of the preceding ones. The goal is to quantify the overall privacy leakage at the output of the adaptively composed mechanism. This idea has been studied in detail for classical privacy settings [2], and here we explore it for QPP.

We first focus on the setting depicted in Fig. 3. Fix $X_i \sim P_X \in \Theta$ for $i \in \{1, 2\}$, which are independently chosen, and let the input state be

$$\sigma_I := \rho^{X_1} \otimes \rho^{X_2}. \tag{105}$$

On the top subsystem in Fig. 3, the channel $\mathcal{A}_1$ is followed by the quantum instrument $\{\mathcal{E}_y\}_{y\in\mathcal{Y}}$, which is a collection of completely positive maps such that the sum map

$$\overline{\mathcal{E}} := \sum_{y\in\mathcal{Y}} \mathcal{E}_y \tag{106}$$

is trace preserving [56], [57], [58]. Depending on the measurement outcome y, the channel $\mathcal{A}_2^y$ is chosen and applied to the bottom subsystem. The combined output state at stage O, as marked in the figure, is

$$\sigma_O := \sum_{y\in\mathcal{Y}} \mathcal{E}^y\left(\mathcal{A}_1(\rho^{X_1})\right) \otimes |y\rangle\langle y| \otimes \mathcal{A}_2^y(\rho^{X_2}). \tag{107}$$

We focus on adaptive composition of two quantum mechanisms in the above described setting. Suppose that $\mathcal{A}_1$ is an $(\varepsilon_1, \delta_1)$-QPP mechanism in the framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M}_1)$. Suppose furthermore that, for each outcome $y \in \mathcal{Y}$, the mechanism $\mathcal{A}_2^y$ satisfies $(\varepsilon_2, \delta_2)$-QPP in the framework

$(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M}_2)$ in the following sense: for all $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$ and $\mathrm{M} \in \mathcal{M}_2$,

$$\mathrm{Tr}\big[\mathrm{M}\mathcal{A}_2^y(\rho^{\mathcal{R}})\big] \leq e^{\varepsilon_2}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}_2^y(\rho^{\mathcal{T}})\big] + \delta_2. \tag{108}$$

Under adaptive composition, we want to guarantee indistinguishability of pairs of states

$$\sigma_I^{\mathcal{R}} \coloneqq \rho^{\mathcal{R}_1} \otimes \rho^{\mathcal{R}_2} \quad \text{and} \quad \sigma_I^{\mathcal{T}} \coloneqq \rho^{\mathcal{T}_1} \otimes \rho^{\mathcal{T}_2}, \tag{109}$$

where $(\mathcal{R}_i, \mathcal{T}_i) \in \mathcal{Q}$ for $i \in \{1, 2\}$. This means, informally, that the adversary would draw the same conclusions regardless of whether ρ^{X_i} belongs to $\mathcal{R}_i$ or $\mathcal{T}_i$, for $i \in \{1, 2\}$, when the initial input σ_I to the system in Fig. 3 is given by (105). The following proposition provides parameters under which QPP of the adaptively composed mechanism is guaranteed.

Proposition 3 (Adaptive Composition of QPP): Fix the framework $(\mathcal{S}, \mathcal{Q}, \Theta, \bar{\mathcal{M}})$. Suppose that $\mathcal{A}_1$ satisfies $(\varepsilon_1, \delta_1)$-QPP and $\mathcal{A}_2^y$ satisfies $(\varepsilon_2, \delta_2)$-QPP for every measurement outcome y, as in (108). Then the mechanism in Fig. 3 satisfies $(\varepsilon_1 + \varepsilon_2, \delta_2 + \delta_1|\mathcal{Y}|)$-QPP in the framework $(\mathcal{S}, \mathcal{Q} \times \mathcal{Q}, \Theta, \bar{\mathcal{M}} \otimes \bar{\mathcal{M}})$ where $|\mathcal{Y}|$ denotes the cardinality of the set $\mathcal{Y}$.

Proof: See Appendix V. ∎

Note that, when $\delta_i = 0$ for $i \in \{1, 2\}$, the privacy parameters are additive. However, when $\delta_i \neq 0$, the privacy parameter $\delta_2 + \delta_1|\mathcal{Y}|$ degrades linearly with an increasing number of measurement outcomes.

Remark 10 (Composability With Correlated States): In Property 3 of Theorem 1 and Proposition 3, we considered the case in which two mechanisms composed in parallel, receive independent inputs (i.e., the input being $\rho^{X_1} \otimes \rho^{X_2}$ where $X_i \sim P_X \in \Theta$ for $i = \{1, 2\}$, which are chosen independently). In Appendix VI we study the setting in which the inputs are correlated. There, we observe that QPP shares similar properties related to composability of classical PP frameworks, where the class of Θ plays a key role in composability to hold in general.

In Proposition 3, we assume a local structure of measurements conducted in the process, as shown in Fig. 3. This assumption is mainly motivated by technical considerations, as we can treat the resulting setting using our existing set of tools. Exploration of advanced adaptive composition techniques, which holds for more general classes of measurements is an interesting avenue for future work. In Section IX-B, we present a variant of QPP where adaptive composition holds for general measurements and strategies (refer to Fig. 8 and Remark 25).

B. Mechanisms for QPP

We propose mechanisms to achieve ε-QPP and (ε, δ)-QPP using the depolarization channel. In addition, we provide a general procedure to generate (ε, δ)-(classical) PP mechanisms using a quantum mechanism satisfying (ε, δ)-QPP.

1) Depolarization Mechanism: Let

$$\mathcal{A}_{\mathrm{Dep}}^p(\rho) \coloneqq (1-p)\rho + \frac{p}{d}I, \tag{110}$$

where $p \in [0, 1]$ and d is the dimension of the Hilbert space on which ρ acts.

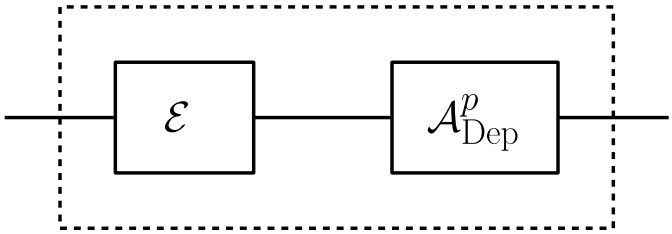


Fig. 4. Depolarization mechanism to achieve QPP: This corresponds to a channel $\mathcal{E}$ followed by a depolarizing channel. Note that we can choose $\mathcal{E} = \mathcal{I}$ to be the identity channel as well.

Theorem 3 (ε-QPP depolarization mechanism): Fix $p \in [0, 1]$ and a privacy framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$. Let $\mathcal{E}$ be a quantum channel. Then $\mathcal{A}_{\mathrm{Dep}}^p(\mathcal{E}(\cdot))$ (in Fig. 4) is ε-QPP if

$$p \geq \frac{dK}{dK + e^{\varepsilon} - 1}, \tag{111}$$

where

$$K \coloneqq \sup_{\mathrm{M} \in \mathcal{M}} \frac{\|\mathrm{M}\|_\infty}{\mathrm{Tr}[\mathrm{M}]} \times \sup_{\Theta, (\mathcal{R}, \mathcal{T}) \in \mathcal{Q}} \frac{\big\|\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big\|_1}{2}. \tag{112}$$

This further implies that the depolarization channel with parameter p achieves ε-QPP whenever

$$\varepsilon \geq \ln\left(1 + \frac{(1-p)dK}{p}\right). \tag{113}$$

Proof: Fix $P_X \in \Theta$, $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$, and $\mathrm{M} \in \mathcal{M}$, and consider that

$$\begin{aligned}
&\frac{\mathrm{Tr}\Big[\mathrm{M}\mathcal{A}_{\mathrm{Dep}}^p\big(\mathcal{E}(\rho^{\mathcal{R}})\big)\Big]}{\mathrm{Tr}\Big[\mathrm{M}\mathcal{A}_{\mathrm{Dep}}^p\big(\mathcal{E}(\rho^{\mathcal{T}})\big)\Big]} - 1 \\
&= \frac{(1-p)\mathrm{Tr}\big[\mathrm{M}\mathcal{E}(\rho^{\mathcal{R}})\big] + \frac{p}{d}\mathrm{Tr}[\mathrm{M}]}{(1-p)\mathrm{Tr}[\mathrm{M}\mathcal{E}(\rho^{\mathcal{T}})] + \frac{p}{d}\mathrm{Tr}[\mathrm{M}]} - 1 && (114)\\
&= \frac{(1-p)\mathrm{Tr}\big[\mathrm{M}\big(\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big)\big]}{(1-p)\mathrm{Tr}[\mathrm{M}\mathcal{E}(\rho^{\mathcal{T}})] + \frac{p}{d}\mathrm{Tr}[\mathrm{M}]} && (115)\\
&\leq \frac{(1-p)\big|\mathrm{Tr}\big[\mathrm{M}\big(\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big)\big]\big|}{\frac{p}{d}\mathrm{Tr}[\mathrm{M}]} && (116)
\end{aligned}$$

Given the above, if

$$\varepsilon \geq \ln\left(1 + \frac{d(1-p)\big|\mathrm{Tr}\big[\mathrm{M}\big(\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big)\big]\big|}{p\mathrm{Tr}[\mathrm{M}]}\right), \tag{117}$$

then

$$\frac{\mathrm{Tr}\Big[\mathrm{M}\mathcal{A}_{\mathrm{Dep}}^p\big(\mathcal{E}(\rho^{\mathcal{R}})\big)\Big]}{\mathrm{Tr}\Big[\mathrm{M}\mathcal{A}_{\mathrm{Dep}}^p\big(\mathcal{E}(\rho^{\mathcal{T}})\big)\Big]} \leq e^{\varepsilon}. \tag{118}$$

Recalling that $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$ if and only if $(\mathcal{T}, \mathcal{R}) \in \mathcal{Q}$, the roles of $\rho^{\mathcal{R}}$ and $\rho^{\mathcal{T}}$ can be interchanged, and we conclude that

$$e^{-\varepsilon} \leq \frac{\mathrm{Tr}\Big[\mathrm{M}\mathcal{A}_{\mathrm{Dep}}^p\big(\mathcal{E}(\rho^{\mathcal{R}})\big)\Big]}{\mathrm{Tr}\Big[\mathrm{M}\mathcal{A}_{\mathrm{Dep}}^p\big(\mathcal{E}(\rho^{\mathcal{T}})\big)\Big]}. \tag{119}$$

Consider that

$$\mathrm{Tr}\big[\mathrm{M}\big(\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big)\big] \leq \|\mathrm{M}\|_\infty \frac{\big\|\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big\|_1}{2}. \tag{120}$$

Indeed, consider the following Jordan–Hahn decomposition $\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}}) = P - Q$, where P and Q are the positive and negative parts of $\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})$, respectively, satisfying $P, Q \geq 0$ and $PQ = 0$. Then

$$\begin{aligned} &\mathrm{Tr}\big[\mathrm{M}\big(\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big)\big] \\ &= \mathrm{Tr}[\mathrm{M}(P-Q)] && (121)\\ &\leq \mathrm{Tr}[\mathrm{M}P] && (122)\\ &\leq \|\mathrm{M}\|_\infty \frac{\big\|\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big\|_1}{2}, && (123)\end{aligned}$$

where the last inequality follows from Hölder's inequality and because $\big\|\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big\|_1 = \mathrm{Tr}[P] + \mathrm{Tr}[Q] = 2\mathrm{Tr}[P]$ since $\mathrm{Tr}\big[\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big] = 0 = \mathrm{Tr}[P-Q]$.

Collecting all terms and supremizing over $\mathcal{M}$ and Θ and all secret pairs of $\mathcal{Q}$ yields the desired result. ∎

Note that the parameter K derived from Theorem 3 represents the domain knowledge accessible and incorporated into the privacy model of the $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ QPP framework.

Corollary 2 (ε-QDP With Domain Knowledge): Fix $p \in [0,1]$, and a privacy framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ for QDP that encodes domain knowledge. Let $\mathcal{E}$ be a quantum channel. Then $\mathcal{A}^p_{\mathrm{Dep}}(\mathcal{E}(\cdot))$ is ε-QPP with

$$\varepsilon \geq \ln\left(1 + \frac{(1-p)d}{2p} k' \sup_{\mathrm{M}\in\mathcal{M}} \frac{\|\mathrm{M}\|_\infty}{\mathrm{Tr}[\mathrm{M}]}\right), \tag{124}$$

where

$$k' := \sup_{(\rho^{x_1}, \rho^{x_2}) \in \mathcal{W}_\Theta} \big\|\mathcal{E}(\rho^{x_1}) - \mathcal{E}(\rho^{x_2})\big\|_1, \tag{125}$$

$$\mathcal{W}_\Theta := \{(\rho^{x_1}, \rho^{x_2}) \in \mathcal{Q} \mid \exists P_X \in \Theta \;\; P_X(x_1), P_X(x_2) > 0\}. \tag{126}$$

Note that the domain knowledge encoded into the QDP framework may guide towards an improved accuracy/utility, as opposed to considering all neighboring states as secret pairs and all possible measurements. For the QDP framework without domain knowledge, [13, Theorem 3] shows that

$$\varepsilon \geq \ln\left(1 + \frac{(1-p)d}{2p} \sup_{\rho\sim\sigma} \|\rho - \sigma\|_1\right) \tag{127}$$

is a sufficient condition to ensure $(\varepsilon, 0)$-QDP. Compared with that due to the condition

$$\begin{aligned} \sup_{\mathrm{M}\in\mathcal{M}} \frac{\|\mathrm{M}\|_\infty}{\mathrm{Tr}[\mathrm{M}]} \times \sup_{(\rho^{x_1}, \rho^{x_2}) \in \mathcal{W}_\Theta} &\big\|\mathcal{E}(\rho^{x_1}) - \mathcal{E}(\rho^{x_2})\big\|_1 \\ &\leq \sup_{\rho\sim\sigma} \|\rho - \sigma\|_1, \end{aligned} \tag{128}$$

a QDP framework that has the capability to incorporate domain knowledge may cause less perturbation to the useful channel output of $\mathcal{E}$ in some cases. The rightmost inequality holds because $\mathcal{W}_\Theta$ includes only the neighboring pairs of states such that their occurrence has a positive probability, while $\rho \sim \sigma$ denotes all possible neighboring pairs. Furthermore, we always have that $\frac{\|\mathrm{M}\|_\infty}{\mathrm{Tr}[\mathrm{M}]} \leq 1$ for every measurement operator M.

Remark 11 (Local DP): For the setup in Remark III-C2, Theorem 3 reduces to

$$p \geq \frac{d}{d + e^\varepsilon - 1}, \tag{129}$$

with the choice of the identity channel instead of $\mathcal{E}$ in Fig. 4. This occurs because $\|\rho - \sigma\|_1 \leq 2$, with equality for pairs of orthogonal states, and $\|\mathrm{M}\|_\infty \leq \mathrm{Tr}[\mathrm{M}]$, with equality holding whenever M is a rank-one measurement operator. This is analogous to a version of the randomized response technique used to achieve classical local DP [59], [60], [61]. For a finite alphabet $\mathcal{X}$ with cardinality $|\mathcal{X}|$, the randomized response mechanism outputs the true value with probability $1-q$, and it outputs a randomly chosen realization with probability $q/|\mathcal{X}|$. Then, if

$$q \geq \frac{|\mathcal{X}|}{|\mathcal{X}| + e^\varepsilon - 1}, \tag{130}$$

ε-local differential privacy is achieved. This analogy further suggests that the depolarization mechanism can be considered as a quantum version of the randomized response mechanism that achieves classical privacy guarantees.

Considering the scenario in which we want to provide a privacy guarantee for all possible measurements (i.e., $\mathcal{M} = \bar{\mathcal{M}}$), next we derive the parameter p to achieve (ε, δ)-QPP.

Proposition 4 ((ε, δ)-QPP Depolarization Mechanism): Fix $p \in [0,1]$ and the privacy framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ with $\mathcal{M} = \{\mathrm{M} : 0 \leq \mathrm{M} \leq \mathrm{I}\}$. Let $\mathcal{E}$ be a quantum channel. Then $\mathcal{A}^p_{\mathrm{Dep}}(\mathcal{E}(\cdot))$ is ε-QPP if

$$p \geq \max\left\{0, \frac{d(K'-\delta)}{dK' + e^\varepsilon - 1}\right\}, \tag{131}$$

where

$$K' := \sup_{\Theta, (\mathcal{R},\mathcal{T})\in\mathcal{Q}} \frac{\big\|\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big\|_1}{2}. \tag{132}$$

Proof: The proof follows from the use of the equivalent formulation through the hockey-stick divergence and the properties of this divergence. By [15, Lemma IV.I], we have

$$\begin{aligned} &\mathsf{E}_{e^\varepsilon}\big(\mathcal{A}_{\mathrm{Dep}}(\mathcal{E}(\rho^{\mathcal{R}})) \,\|\, \mathcal{A}_{\mathrm{Dep}}(\mathcal{E}(\rho^{\mathcal{T}}))\big) \\ &\qquad \leq (1 - e^\varepsilon)\frac{p}{d} + (1-p)\mathsf{E}_{e^\varepsilon}\big(\mathcal{E}(\rho^{\mathcal{R}}) \| \mathcal{E}(\rho^{\mathcal{T}})\big). \end{aligned} \tag{133}$$

We also have the property [15, Lemma II.4]

$$\mathsf{E}_{e^\varepsilon}\big(\mathcal{E}(\rho^{\mathcal{R}}) \| \mathcal{E}(\rho^{\mathcal{T}})\big) \leq \frac{\big\|\mathcal{E}(\rho^{\mathcal{R}}) - \mathcal{E}(\rho^{\mathcal{T}})\big\|_1}{2}. \tag{134}$$

Combining these relations, and supremizing over Θ, and secret pairs $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$, we can choose

$$\delta \geq (1 - e^\varepsilon)\frac{p}{d} + (1-p)K'. \tag{135}$$

Then, rearranging the terms we arrive at

$$p \geq \frac{d(K'-\delta)}{dK' + e^\varepsilon - 1}. \tag{136}$$

Since $p \geq 0$, when $K' - \delta \leq 0$, we set $p = 0$. ∎

2) Classical PP Mechanisms From QPP Mechanisms: The QPP formalism provides a direct methodology to design classical PP mechanisms with the assistance of QPP mechanisms.[5]

[5]In the context of classical PP mechanisms, the main attempt has been the introduction of Wasserstein mechanisms, based on the infinity order Wasserstein distance and its modifications. In particular, [4] and [62] introduced these mechanisms to achieve $(\varepsilon, 0)$-PP and (ε, δ)-PP, respectively. However, it is important to note that these approaches may encounter computational intractability challenges.

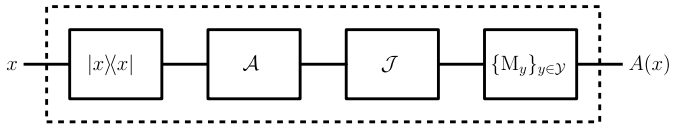


Fig. 5. Generation of classical PP mechanisms from QPP mechanism $\mathcal{A}$: First the classical data is encoded using quantum encoding techniques, then the QPP mechanism $\mathcal{A}$, and if needed any other channel $\mathcal{J}$, and finally the measurement channel.

In this case, we use quantum encoding to convert classical data to quantum data. We denote the quantum encoding of classical data $x \in \mathcal{X}^{n\times k}$ as $\rho^x := |x\rangle\langle x|$ (recall that $p_X \in \Theta_c$ are discrete probability distributions over the probability space $\mathcal{P}(\mathcal{X}^{n\times k})$ and this leads to a finite collection of $\{\rho^x\}_x$ quantum encodings). Then, we ensure the privacy for the quantum data (quantum encoding) such that the privacy is ensured for the underlying classical data.

Proposition 5 ((ε,δ)- (classical) PP mechanism): Given an (ε,δ)-QPP mechanism $\mathcal{A}$ within the framework $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M})$ when $\rho^x := |x\rangle\langle x|$ with

$$\begin{aligned}\mathcal{S} &= \{\{\rho^x : x \in \mathcal{R}_c\} : \mathcal{R}_c \in \mathcal{S}_c\},\\ \mathcal{Q} &= \{(\{\rho^x : x \in \mathcal{R}_c\}, \{\rho^x : x \in \mathcal{T}_c\}) : (\mathcal{R}_c, \mathcal{T}_c) \in \mathcal{Q}_c\}\},\\ \Theta &= \{\{p_X(x), \rho^x\}_x : p_X \in \Theta_c\},\\ \mathcal{M} &= \{\mathrm{M} : 0 \le \mathrm{M} \le \mathrm{I}\},\end{aligned} \tag{137}$$

any post-processing of $\mathcal{A}$ by a quantum channel $\mathcal{J}$ followed by applying a POVM $\{\mathrm{M}_y\}_{y\in\mathcal{Y}}$ denoted as $A: \mathcal{X}^{n\times k} \to \mathcal{Y}$ as shown in Fig. 5 is (ε,δ)-PP in the framework $(\mathcal{S}_c,\mathcal{Q}_c,\Theta_c)$.

Furthermore, for a selected post-processing $\mathcal{J}$ and POVM $\{\mathrm{M}_y\}_{y\in\mathcal{Y}}$, it is sufficient for $\mathcal{M}_c^{\mathcal{J}} \subseteq \mathcal{M}$ for A to be (ε,δ)-PP, where

$$\mathcal{M}_c^{\mathcal{J}} := \left\{ \mathcal{J}^\dagger\left(\sum_{y\in\mathcal{B}} \mathrm{M}_y\right) : \mathcal{B} \in \mathcal{Y} \right\}. \tag{138}$$

Proof: Fix $\mathcal{B} \subseteq \mathcal{Y}$. Consider that

$$\begin{aligned}&\mathbb{P}(A(X) \in \mathcal{B}|\mathcal{R}_c)\\ &= \frac{\mathbb{P}(\{A(X)\in\mathcal{B}\}\cap\mathcal{R}_c)}{\mathbb{P}(\mathcal{R}_c)} &(139)\\ &\overset{(a)}{=} \frac{\sum_{x\in\mathcal{R}_c} p(x)\mathbb{P}(A(x)\in\mathcal{B})}{P_X(\mathcal{R})} &(140)\\ &\overset{(b)}{=} \frac{\sum_{x\in\mathcal{R}_c} p(x)\sum_{y\in\mathcal{B}}\mathbb{P}(A(x)=y)}{P_X(\mathcal{R})} &(141)\\ &\overset{(c)}{=} \frac{\sum_{\rho^x\in\mathcal{R}} p(x)\sum_{y\in\mathcal{B}}\mathrm{Tr}[\mathrm{M}_y\mathcal{J}\circ\mathcal{A}(\rho^x)]}{P_X(\mathcal{R})} &(142)\\ &\overset{(d)}{=} \mathrm{Tr}\left[\sum_{y\in\mathcal{B}}\mathrm{M}_y\mathcal{J}\circ\mathcal{A}\left(\sum_{\rho^x\in\mathcal{R}}\frac{p(x)}{P_X(\mathcal{R})}\rho^x\right)\right] &(143)\\ &\overset{(e)}{=} \mathrm{Tr}\left[\mathcal{J}^\dagger\left(\sum_{y\in\mathcal{B}}\mathrm{M}_y\right)\mathcal{A}\left(\rho^{\mathcal{R}}\right)\right] &(144)\\ &\overset{(f)}{=} \mathrm{Tr}\left[\mathrm{M}\mathcal{A}(\rho^{\mathcal{R}})\right], &(145)\end{aligned}$$

where: (a) from $\mathcal{R} := \{\rho^x : x \in \mathcal{R}_c\}$; (b) from $\mathcal{B}$ being a collection of $y \in \mathcal{Y}$; (c) from M_y being the measurement applied to obtain the outcome y; (d) from the linearity of trace operator and quantum channels $\mathcal{A},\mathcal{J}$; (e) from the definition of $\rho^{\mathcal{R}}$, and $\mathcal{J}^\dagger$ being the adjoint of $\mathcal{J}$; and finally (f) from $\mathrm{M} := \mathcal{J}^\dagger\left(\sum_{y\in\mathcal{B}}\mathrm{M}_y\right)$.

Similarly, $\mathbb{P}(A(X)\in\mathcal{B}|\mathcal{T}_c) = \mathrm{Tr}\left[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\right]$. Then with the assumption that $\mathcal{A}$ is (ε,δ)-QPP for $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M})$ mentioned in the proposition statement, we have

$$\mathbb{P}(A(X)\in\mathcal{B}|\mathcal{R}) \le e^{\varepsilon}\,\mathbb{P}(A(X)\in\mathcal{B}|\mathcal{T}) + \delta. \tag{146}$$

concluding the proof. ∎

Depolarization is a common kind of noise considered in quantum information processing. Thus, this offers a method for designing classical PP mechanisms by combining the results presented in Proposition 5 and Theorem 3. However, it is essential to recognize that quantum encoding of classical data would require additional computational resources, shifting the complexity of the mechanism design phase to the encoding phase.

VI. Quantifying Privacy-Utility Tradeoff

In this section, we aim to assess the utility achievable through the implementation of a privatization mechanism while adhering to privacy constraints and characterize the inherent tradeoffs involved in this process. To achieve this, we define a utility metric grounded in an operational approach and demonstrate its representation via an SDP. Subsequently, we leverage this metric to conduct an in-depth analysis of privacy-utility tradeoffs, with a specific emphasis on the depolarization mechanism.

A. Utility Metric

Let $\mathcal{A}$ denote a privacy mechanism. We focus on assessing the potential of reversing the effects of $\mathcal{A}$ by applying a post-processing mechanism $\mathcal{B}$ to recover the initial input state to $\mathcal{A}$ up to an error $1-\gamma$. We define γ-utility in terms of how distinguishable $\mathcal{B}\circ\mathcal{A}$ is from the identity channel, employing the normalized diamond distance as the distinguishability measure.

Definition 5 (γ-utility): Let $\mathcal{A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_C)$ be a privacy mechanism, and fix $\gamma \in [0,1]$. We say that $\mathcal{A}$ satisfies γ-utility if

$$\mathsf{U}(\mathcal{A}) := 1 - \inf_{\mathcal{B}} \frac{1}{2}\left\|\mathcal{I}-\mathcal{B}\circ\mathcal{A}\right\|_\diamond \ge \gamma, \tag{147}$$

where the infimum is taken over every quantum channel $\mathcal{B} : \mathcal{L}(\mathcal{H}_C) \to \mathcal{L}(\mathcal{H}_D)$.

The defined utility metric can be reformulated as an SDP by using the dual SDP form of the diamond distance [63, Section 4] and rewriting the quantum channel $\mathcal{B}$ in terms of its Choi matrix $\Gamma_{CD}^{\mathcal{B}}$, as well as translating the conditions for $\mathcal{B}$ to be a channel to conditions on its Choi matrix, namely $\Gamma_{CD}^{\mathcal{B}} \ge 0$ and $\mathrm{Tr}_D\left[\Gamma_{CD}^{\mathcal{B}}\right] = I_c$.

Proposition 6 (SDP Formulation of γ-Utility): The γ-utility of a privacy mechanism $\mathcal{A}$ can be formulated as the

following SDP:

$$\mathsf{U}(\mathcal{A}) = 1 - \inf_{\substack{\mu \geq 0 \\ Z_{AD} \geq 0 \\ \Gamma^{\mathcal{B}}_{CD} \geq 0}} \left\{ \begin{array}{c} \mu : \\ Z_{AD} \geq \Gamma_{AD} - \Gamma^{\mathcal{B}\circ\mathcal{A}}_{AD}, \\ \mu I_A \geq \mathrm{Tr}_D[Z_{AD}], \\ \mathrm{Tr}_D\left[\Gamma^{\mathcal{B}}_{CD}\right] = I_C \end{array} \right\} \geq \gamma, \tag{148}$$

where

$$\Gamma^{\mathcal{B}\circ\mathcal{A}}_{AD} := \mathrm{Tr}_C\left[\left(I_A \otimes \Gamma^{\mathcal{B}}_{CD}\right)\left(\mathrm{T}_C(\Gamma^{\mathcal{A}}_{AC}) \otimes I_D\right)\right], \tag{149}$$

and Γ represents the Choi matrix with the subscripts showing the input and the output system of the channel while the superscript indicating the channel considered, with no superscript for the identity channel.[6]

Note that a similar SDP formulation for approximate degradability where the identity channel in Definition 5 is replaced by the complementary channel is presented in [64, Proposition 9].

Remark 12 (Characterizing Optimal Privacy-Utility Tradeoffs): The optimal utility attained by an (ε, δ)-QPP mechanism can be characterized as an SDP. To achieve this, we combine the equivalent formulation of QPP via the DL divergence from Proposition 1 with the SDP formulation of DL divergence in Lemma 1. Combining this with the SDP formulated from Proposition 6 enables computing the privacy requirements and quantifying utility together. Additionally, we determine the optimal privacy parameters for fixed utility requirements. For a comprehensive discussion of this point, please refer to Appendix VII. The utilization of the SDP derived for the DL divergence in this operational task highlights an advantage of the equivalent formulation for QPP using the DL divergence.

B. Analysis of Depolarization Mechanism

We now instantiate $\mathcal{A}$ as the depolarizing channel with parameter p, denoted as $\mathcal{A}^p_{\mathrm{Dep}}$ (as defined in (110)), and proceed to analyze $\mathcal{U}(\mathcal{A}^p_{\mathrm{Dep}})$.

Proposition 7 (Utility from depolarization mechanism): Fix $p \in [0,1]$. The depolarization mechanism satisfies γ-utility if and only if

$$\mathsf{U}(\mathcal{A}^p_{\mathrm{Dep}}) = 1 - \frac{p(d^2-1)}{d^2} \geq \gamma. \tag{150}$$

Proof: The proof below relies on observing that the optimization term (in the utility metric) is minimized by setting $\mathcal{B} = \mathcal{I}$, and then evaluating $\left\|\mathcal{I} - \mathcal{A}^p_{\mathrm{Dep}}\right\|_\diamond$ using the Choi states of the channels $\mathcal{A}^p_{\mathrm{Dep}}$ and $\mathcal{I}$, due to the joint covariance of the two channels under unitaries (i.e., $\mathcal{A}^p_{\mathrm{Dep}} \circ \mathcal{U} = \mathcal{U} \circ \mathcal{A}^p_{\mathrm{Dep}}$ and $\mathcal{I} \circ \mathcal{U} = \mathcal{U} \circ \mathcal{I}$ for every unitary channel $\mathcal{U}$).

Consider that

$$\begin{aligned} &\left\|\mathcal{I} - \mathcal{B} \circ \mathcal{A}^p_{\mathrm{Dep}}\right\|_\diamond \\ &\overset{(a)}{=} \left\|\mathcal{U} \circ (\mathcal{I} - \mathcal{B} \circ \mathcal{A}^p_{\mathrm{Dep}}) \circ \mathcal{U}^\dagger\right\|_\diamond &(151)\\ &\overset{(b)}{=} \left\|\mathcal{I} - \mathcal{U} \circ \mathcal{B} \circ \mathcal{U}^\dagger \circ \mathcal{A}^p_{\mathrm{Dep}})\right\|_\diamond &(152)\\ &\overset{(c)}{=} \int \mathrm{d}\mathcal{U} \left\|\mathcal{I} - \mathcal{U} \circ \mathcal{B} \circ \mathcal{U}^\dagger \circ \mathcal{A}^p_{\mathrm{Dep}})\right\|_\diamond &(153)\\ &\overset{(d)}{\geq} \left\|\mathcal{I} - \left(\int \mathrm{d}\mathcal{U}\, \mathcal{U} \circ \mathcal{B} \circ \mathcal{U}^\dagger\right) \circ \mathcal{A}^p_{\mathrm{Dep}})\right\|_\diamond, &(154)\end{aligned}$$

where: (a) follows from the unitary invariance of the diamond norm with $\mathcal{U}$ representing a unitary channel; (b) from the commutative property of $\mathcal{A}^p_{\mathrm{Dep}}$ with every unitary channel; (c) with $\mathrm{d}\mathcal{U}$ denoting the Haar measure over the unitary group and from the left-hand side being independent of $\mathcal{U}$; and (d) from the convexity of the diamond norm.

Next, observe that $\mathcal{B}^\star := \int \mathrm{d}\mathcal{U}\, \mathcal{U} \circ \mathcal{B} \circ \mathcal{U}^\dagger$ is a quantum channel, and it is in fact equal to a depolarization channel [65]. Then, $\mathcal{B}^\star = \mathcal{A}^q_{\mathrm{Dep}}$ for some $q \in [0,1]$.

The composition of two depolarization channels

$$\mathcal{B}^\star = \mathcal{A}^q_{\mathrm{Dep}} \circ \mathcal{A}^p_{\mathrm{Dep}} \tag{155}$$

is also a depolarization channel with parameter $p^\star := 1-(1-p)(1-q)$. The minimum value is attained by the choice $q = 0$, where $\mathcal{A}^q_{\mathrm{Dep}}$ in that case corresponds to the identity channel.

With that, we arrive at

$$\inf_{\mathcal{B}} \left\|\mathcal{I} - \mathcal{B} \circ \mathcal{A}^p_{\mathrm{Dep}}\right\|_\diamond = \left\|\mathcal{I} - \mathcal{A}^p_{\mathrm{Dep}}\right\|_\diamond. \tag{156}$$

With the property of joint covariance of $\mathcal{I}$ and $\mathcal{A}^p_{\mathrm{Dep}}$ under unitaries [37, Proposition 7.82], we simplify this to

$$\left\|\mathcal{I} - \mathcal{A}^p_{\mathrm{Dep}}\right\|_\diamond = \frac{1}{d}\left\|\Gamma_{AD} - \Gamma^{\mathcal{A}^p_{\mathrm{Dep}}}_{AD}\right\|_1. \tag{157}$$

Then consider that

$$\begin{aligned} &\frac{1}{d}\left\|\Gamma_{AD} - \Gamma^{\mathcal{A}^p_{\mathrm{Dep}}}_{AD}\right\|_1 \\ &\overset{(a)}{=} \frac{1}{d}\left\|\Gamma_{AD} - \left((1-p)\Gamma_{AD} + \frac{p}{d} I_{d^2}\right)\right\|_1 &(158)\\ &= \frac{p}{d}\left\|\Gamma_{AD} - \frac{1}{d} I_{d^2}\right\|_1 &(159)\\ &= p\left\|\frac{\Gamma_{AD}}{d}\left(1 - \frac{1}{d^2}\right) - \frac{1}{d^2}\left(I_{d^2} - \frac{\Gamma_{AD}}{d}\right)\right\|_1 &(160)\\ &\overset{(b)}{=} p\left(1 - \frac{1}{d^2}\right)\left(\left\|\frac{\Gamma_{AD}}{d}\right\|_1 + \left\|\frac{I_{d^2} - \frac{\Gamma_{AD}}{d}}{d^2 - 1}\right\|_1\right) &(161)\\ &\overset{(c)}{=} 2p\left(1 - \frac{1}{d^2}\right), &(162)\end{aligned}$$

where: (a) from $\Gamma^{\mathcal{A}^p_{\mathrm{Dep}}}_{AD} = (1-p)\Gamma_{AD} + \frac{p}{d} I_{d^2}$; (b) from $\frac{\Gamma_{AD}}{d}$, and $I_{d^2} - \frac{\Gamma_{AD}}{d}$ being orthogonal; and (c) from trace norm of quantum states being equal to one.

Combining the above chain of arguments together completes the proof. ∎

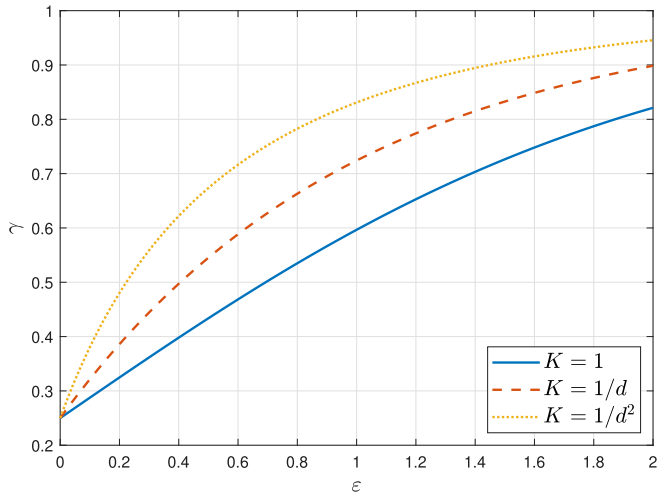
Next, we focus on understanding the privacy-utility tradeoff with respect to the parameter p governing a depolarization mechanism. From Proposition 7, to achieve γ-utility, we require that

$$p \leq \frac{(1-\gamma)d^2}{(d^2-1)}. \tag{163}$$

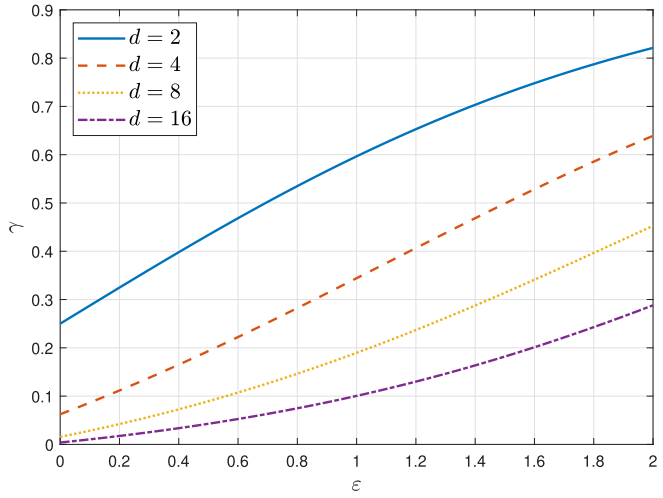
Conversely, to achieve ε-QPP in the chosen privacy framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$, from Theorem 3, it suffices for p to satisfy

$$p \geq \frac{dK}{dK + e^\varepsilon - 1}. \tag{164}$$

[6] The Choi matrix of the composed channel $\mathcal{B} \circ \mathcal{A}$ is denoted by $\Gamma^{\mathcal{B}\circ\mathcal{A}}_{AD}$ and (149) follows from [37, Eq. (3.2.22)].



(a)



(b)

Fig. 6. (a) For fixed $d = 2$, the figure depicts the optimum utility γ for ε achievable with the depolarization mechanism in Theorem 3. The value of K encodes the domain knowledge available, where $K = 1$ corresponds to no such additional information being available. (b) For fixed $K = 1$, the figure depicts the optimum utility γ for ε achievable with the depolarization mechanism in Theorem 3 for $d \in \{2, 4, 8, 16\}$.

These two inequalities provide insights into the privacy-utility tradeoff associated with the depolarization mechanism. Consequently, it is essential to carefully tune the parameter p based on the desired utility, characterized by γ, as well as the privacy parameter ε.

1) Effect of Domain Knowledge: Fig. 6a illustrates the optimal utility achievable using the ε-QPP depolarization mechanism presented in Theorem 3. Notably, as the value of K reduces, the attainable utility region expands. The parameter K derived from Theorem 3 represents the domain knowledge accessible and incorporated into the privacy model of the $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ QPP framework. This observation underscores the significance of incorporating such domain knowledge to enhance utility gains while simultaneously ensuring the necessary privacy assurances.

2) Effect of Dimension: In Fig. 6, we observe a prominent privacy-utility tradeoff as the dimension increases for the depolarization mechanism presented in Theorem 3. Regarding the utility of the depolarization mechanism (given by $1 - \frac{p(d^2-1)}{d^2}$), we can always establish the following lower bound for every d:

$$1 - \frac{p(d^2-1)}{d^2} \geq 1 - p, \tag{165}$$

where this lower bound is attained as $d \to \infty$. However, the achievable privacy level ε in (113) degrades at most by an order of $\ln(d)$. Hence, it is crucial to identify the optimal privacy parameters achieved by private mechanisms, particularly in high-dimensional scenarios.

Remark 13 (Application Specific Privacy-Utility Tradeoffs): In the previous analysis concerning the depolarization mechanism in Fig. 4, we chose $\mathcal{E} = \mathcal{I}$, the identity channel. However, it would be an interesting future work to explore the utility for user-specific $\mathcal{E}$ channels. Specifically, we can choose $1 - \frac{1}{2}\inf_{\mathcal{B}\in\mathrm{CPTP}} \|\mathcal{E} - \mathcal{B} \circ \mathcal{A}^p_{\mathrm{Dep}} \circ \mathcal{E}\|_\diamond$ as the utility metric. If $\mathcal{E}$ possesses certain symmetries, one can potentially utilize arguments akin to those presented in the proof of Proposition 7. This investigation could shed light on tailoring privacy mechanisms to specific application needs, leading to more effective privacy-utility tradeoffs.

VII. Auditing Privacy Frameworks

Auditing for privacy aims to detect violations in privacy guarantees and reject incorrect algorithms (see [10], [66], [67], [68] for classical approaches). In this section, our focus is on utilizing quantum information theory tools and quantum algorithms to audit the privacy of quantum systems. Specifically, we concentrate on auditing algorithms for QDP guarantees, and it should be noted that these ideas can be extended to audit algorithms for privacy guarantees demanded by QPP (see Remark 16).

The main idea behind auditing classical privacy frameworks (DP and PP) involves translating the privacy requirement into a weaker privacy notion that can be efficiently computed. For example, in [10], sliced mutual information based DP is used to audit for DP. By doing so, algorithms failing to meet the privacy conditions imposed by the relaxed privacy notion are concluded to violate the original privacy requirement. However, a pitfall of this approach is the inability to quantify the gap between the constraints stemming from the original DP or PP notion and the relaxed privacy notions. In other words, even if we verify that the relaxed privacy notion is satisfied, we cannot determine whether the original privacy requirement is also satisfied. In contrast, in this work, we focus on QDP without translating it into a relaxed privacy notion.

A. Techniques for Auditing QDP

1) Using Semi-Definite Programs: By leveraging the equivalent formulation from Proposition 1 and adopting the specific choices of $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ as provided in Remark III-C1, for (ε, δ)-QDP, we have that

$$\sup_{\rho\sim\sigma} \overline{\mathsf{D}}^\delta(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \varepsilon. \tag{166}$$

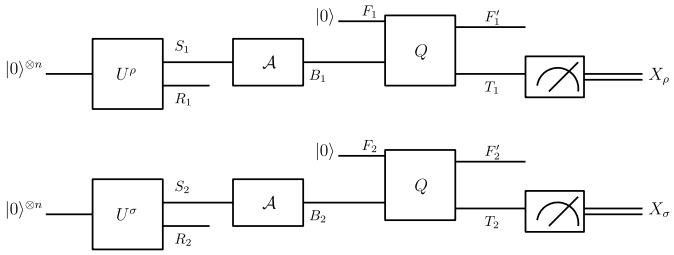


Fig. 7. Quantum circuit assisted in estimating QDP: U^ρ, U^σ are the unitaries used to prepare ρ and σ by tracing out R_1, R_2 systems, respectively. Then $\mathcal{A}$ is applied on the systems S_i for $i \in \{1,2\}$. The unitary Q takes inputs F_i, B_i and outputs F_i', T_i, where F_i and T_i are qubit systems. Finally, each of the T_i systems is measured and the (classical) output random variable is denoted as X_ρ for $i=1$ and X_σ for $i=2$. Here $X_\rho, X_\sigma \in \{0,1\}$. This procedure is repeated a sufficient number of times, and the outcomes of the trials are used to estimate $\mathbb{P}(X_\rho = 0)$ and $\mathbb{P}(X_\sigma = 1)$.

Then, we can compute the left-hand side above by using the SDP formulation of $\overline{\mathsf{D}}^\delta(\cdot\|\cdot)$ presented in Lemma 1. This approach is particularly beneficial for low-dimensional setups as the time complexity of SDP computation is polynomial in the dimension of the quantum states. However, it is essential to note that the complexity of this approach grows exponentially with the number of qubits, making it less feasible for higher-dimensional systems.

Additionally, using the equivalent formulation in Remark 6, consider that (ε,δ)-QDP is equivalent to

$$\sup_{\rho\sim\sigma} \mathsf{E}_{e^\varepsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \delta, \tag{167}$$

where

$$\mathsf{E}_\gamma(\rho\|\sigma) \coloneqq \mathrm{Tr}[(\rho-\gamma\sigma)_+] \tag{168}$$

$$= \frac{1}{2}\|\rho-\gamma\sigma\|_1 + \frac{(1-\gamma)}{2}, \tag{169}$$

with $\gamma \geq 1$ for quantum states ρ and σ [15, Eq. (II.2)]. As shown in (45) and (47), the quantity on the right-hand side of (168) can be evaluated by means of an SDP. Then, auditing QDP reduces to computing $\mathsf{E}_\gamma\big(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)\big)$ for $\rho\sim\sigma$. However, similar to the previous approach, the time complexity of this SDP also grows exponentially with the number of qubits. Thus, computing these SDPs remains challenging for higher-dimensional quantum systems.

2) Using Quantum Circuits: Another approach is to borrow the results of [69] and [70] and use the connection of $\mathsf{E}_\gamma(\rho\|\sigma)$ to the trace distance given in (169). Despite this connection, evaluating $\mathsf{E}_\gamma(\rho\|\sigma)$ remains computationally challenging, even for quantum computers [69], [70]. Nevertheless, there are proposals for evaluating the trace distance using variational quantum algorithms [71], [72] (which however do not give particular runtimes), and for cases in which the quantum states have low rank [73].

In the subsequent analysis, we explore an approach from [71] and [72] (via variational algorithms with parameterized quantum circuits) to estimate the quantity $\|\mathcal{A}(\rho)-e^\varepsilon\mathcal{A}(\sigma)\|_1$. Using such an estimate, for a fixed value of the privacy parameter ε, we can validate on which values of δ the needed guarantees are satisfied.

Firstly, let us focus on how to estimate $\mathsf{E}_{e^\varepsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma))$ for a fixed $\rho\sim\sigma$ and ε. With the ideas developed in [71, Algorithm 14], we discuss how a process similar to a quantum interactive proof can be used for estimating the privacy level. For that, refer to the quantum circuit in Fig. 7: the unitaries U^ρ and U^σ are used to prepare the states ρ and σ, by tracing out systems R_1 and R_2, respectively. Then the algorithm $\mathcal{A}$ is applied on the systems S_i for $i\in\{1,2\}$. The unitary Q takes inputs F_i, B_i and outputs F_i', T_i, where F_i and T_i are qubit systems. Finally, both of the T_i systems are measured in the standard basis $\{|0\rangle,|1\rangle\}$, and the (classical) output random variable is denoted as X_ρ for $i=1$ and X_σ for $i=2$. Here X_ρ and X_σ take values in $\{0,1\}$. This procedure is repeated a sufficient number of times, and we use the results to estimate $\mathbb{P}(X_\rho=0)$ and $\mathbb{P}(X_\sigma=1)$.

Next, consider a scenario in which one could maximize the following utility function over all possible choices of Q:

$$g(Q,\rho,\sigma,\mathcal{A},\varepsilon) \coloneqq \frac{1}{e^\varepsilon+1}\mathbb{P}(X_\rho=0) + \frac{e^\varepsilon}{e^\varepsilon+1}\mathbb{P}(X_\sigma=1). \tag{170}$$

In quantum complexity terminology, this action could be conducted by a quantum prover who has unbounded computational power (we discuss how to relax this assumption in Remark 14). From [71, Eq. (128)] and the discussion therein, we conclude that

$$f(\rho,\sigma,\mathcal{A},\varepsilon) \coloneqq \sup_Q g(Q,\rho,\sigma,\mathcal{A},\varepsilon) \tag{171}$$

$$= \frac{1}{2}\left(1+\frac{1}{e^\varepsilon+1}\|\mathcal{A}(\rho)-e^\varepsilon\mathcal{A}(\sigma)\|_1\right), \tag{172}$$

where the optimization is over every unitary Q.

If

$$f(\rho,\sigma,\mathcal{A},\varepsilon) \leq \frac{1}{2}\left(1+\frac{2\delta+e^\varepsilon-1}{e^\varepsilon+1}\right), \tag{173}$$

then $\mathsf{E}_{e^\varepsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \delta$, due to (168). Then, by changing the role of ρ and σ in (170), we obtain $f(\sigma,\rho,\mathcal{A},\varepsilon)$. Next, we select the maximum out of these quantities

$$\widehat{f}(\rho,\sigma,\mathcal{A},\varepsilon) \coloneqq \max\{f(\rho,\sigma,\mathcal{A},\varepsilon), f(\sigma,\rho,\mathcal{A},\varepsilon)\}. \tag{174}$$

To this end, if

$$\sup_{\rho\sim\sigma}\widehat{f}(\rho,\sigma,\mathcal{A},\varepsilon) \leq \frac{1}{2}\left(1+\frac{2\delta+e^\varepsilon-1}{e^\varepsilon+1}\right), \tag{175}$$

it can be verified that $\mathcal{A}$ is (ε,δ)-QDP.

Remark 14 (Relaxing the Computationally Unbounded Assumption): As the number of qubits increases, classical methods (e.g., using SDPs) often become intractable due to the exponential growth in computational complexity. So in light of this, the above approach is desired with the increasing dimension of the quantum system.

However, finding the Q that achieves the optimum utility in (170) is practically infeasible. To relax this assumption, we replace the action of the prover (who finds this Q) with a parameterized circuit Q_θ. Then we can use (170) as a utility function for training a variational quantum algorithm [74], [75] to estimate a lower bound for (171). With that we obtain a lower bound on $\mathsf{E}_{e^\varepsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma))$, which we denote as $\mathsf{E}^{\mathrm{LB}}_{e^\varepsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma))$. A lower bound gives a sufficient

condition for ruling out algorithms that do not satisfy (ε, δ)-QDP. This claim follows because the estimated lower bound $\mathsf{E}^{\mathrm{LB}}_{e^\varepsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) > \delta$ implies that $\mathsf{E}_{e^\varepsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) > \delta$.

Remark 15 (Neighboring Pairs): To verify (ε, δ)-QDP, it is required to compute whether $\widehat{f}(\rho, \sigma, \mathcal{A}, \varepsilon) \leq \frac{1}{2}\left(1 + \frac{2\delta + e^\varepsilon - 1}{e^\varepsilon + 1}\right)$ for all neighboring pairs $(\rho, \sigma) \in \mathcal{Q}$. However, checking this requirement has increasing computational complexity as the cardinality of the set $\mathcal{Q}$ increases. If the privacy requirements can be relaxed so as to encode domain knowledge as in QPP framework, the effective set $\mathcal{Q}$ may be a small set in some applications of interest. For an example, consider an application where hypothesis testing is carried out between the states ρ and σ under privacy constraints where $\mathcal{Q} = \{(\rho, \sigma), (\sigma, \rho)\}$.

Remark 16 (Auditing QPP): [7] To audit for (ε, δ)-QPP in the framework $(\mathcal{S}, \mathcal{Q}, \Theta, \bar{\mathcal{M}})$, one can use the ideas described above by choosing $\rho^{\mathcal{R}}$ and $\rho^{\mathcal{T}}$ for all $P_X \in \Theta, (\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$ instead of ρ and σ. In that case, the complexity of the approach relies on the set of $\mathcal{Q}$ as well as distributions contained in Θ.

B. Formal Guarantees for Auditing QDP

The question of quantifying the success of a privacy auditing approach, specifically in correctly accepting and rejecting an algorithm with given privacy requirements, is a crucial consideration in privacy auditing research. The authors of [10] and [68] have worked towards answering this question for auditing classical DP, but using a relaxed privacy definition.

To tackle this for the quantum setting, we propose a hypothesis testing-based auditing pipeline tailored specifically for QDP (also for QPP). In this pipeline, we use the trace-norm estimation quantum algorithm proposed in [73]. This quantum algorithm provides an estimation with at most α additive error from the exact value, with high probability, which allows us to achieve the desired significance in the hypothesis test.

Let us define

$$\mathsf{T}^\varepsilon(\rho, \sigma, \mathcal{A}) \coloneqq \frac{1}{e^\varepsilon + 1} \left\|\mathcal{A}(\rho) - e^\varepsilon \mathcal{A}(\sigma)\right\|_1. \tag{176}$$

Fix $(\rho, \sigma) \in \mathcal{Q}$. If algorithm $\mathcal{A}$ satisfies (ε, δ)-QDP, then $\mathsf{E}_{e^\varepsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \delta$. By applying (168), we arrive at

$$\mathsf{T}^\varepsilon(\rho, \sigma, \mathcal{A}) \leq \frac{2\delta + e^\varepsilon - 1}{e^\varepsilon + 1} =: \mathsf{g}(\varepsilon, \delta). \tag{177}$$

Next, by estimating the quantity on the left-hand side, and using $\mathsf{g}(\varepsilon, \delta)$ as a threshold, we design an auditing pipeline for QDP by means of the following null and alternative hypotheses:

$$H_0 : \max\{\mathsf{T}^\varepsilon(\rho, \sigma, \mathcal{A}), \mathsf{T}^\varepsilon(\sigma, \rho, \mathcal{A})\} \leq \mathsf{g}(\varepsilon, \delta), \tag{178}$$

$$H_1 : \max\{\mathsf{T}^\varepsilon(\rho, \sigma, \mathcal{A}), \mathsf{T}^\varepsilon(\sigma, \rho, \mathcal{A})\} > \mathsf{g}(\varepsilon, \delta). \tag{179}$$

Let the estimates of $\mathsf{T}^\varepsilon(\rho, \sigma, \mathcal{A})$ and $\mathsf{T}^\varepsilon(\sigma, \rho, \mathcal{A})$ from a randomized algorithm (in our analysis we use the algorithm corresponding to [73, Corollary 3.4]) be $\hat{\mathsf{T}}^\varepsilon(\rho, \sigma, \mathcal{A})$ and $\hat{\mathsf{T}}^\varepsilon(\sigma, \rho, \mathcal{A})$, respectively. We choose

$$\hat{\mathsf{T}}^\varepsilon_{\max}(\rho, \sigma, \mathcal{A}) \coloneqq \max\{\hat{\mathsf{T}}^\varepsilon(\rho, \sigma, \mathcal{A}), \hat{\mathsf{T}}^\varepsilon(\sigma, \rho, \mathcal{A})\} \tag{180}$$

as our test statistic.

[7] Note that the following ideas can also be extended to auditing for variants of QPP in Section IX as well (See also Remark 21).

1) Estimation of the Test Statistic:

Lemma 3 (Estimating Trace Distance Using Samples of ρ, σ — Restatement of [73, Corollary 3.4]): Given access to identical copies of d-dimensional quantum states ρ and σ, there is a quantum algorithm that estimates the normalized trace distance $\mathsf{T}(\rho, \sigma)$ (recall (2)) within additive error α and with probability not less than $1 - \beta$, by using

$$O\left(\log\left(\frac{1}{\beta}\right) \frac{r^2}{\alpha^5} \log^2\left(\frac{r}{\alpha}\right) \log^2\left(\frac{1}{\alpha}\right)\right) \tag{181}$$

samples of ρ and σ, where r is an upper bound on the rank of ρ and σ.

Lemma 3 is obtained by using the existing result in Corollary 3.4 of [73], and combining its argument in Theorem 2.6 therein on estimating $\mathrm{Tr}[A\rho]$ within additive error α with probability $1 - \beta$, by using $O\left(\frac{1}{\alpha^2} \log\left(\frac{1}{\beta}\right)\right)$ identical samples of ρ. The algorithm proposed in [73] is designed based on the following idea. Let $V \coloneqq (\rho - \sigma)/2$. Consider its singular value decomposition as $V = W\Sigma U^\dagger$. Then the trace distance can be expressed by the following identity:

$$\mathsf{T}(\rho, \sigma) = \frac{1}{2}\left(\mathrm{Tr}[\rho\, \mathrm{sgn}(V)] - \mathrm{Tr}[\sigma\, \mathrm{sgn}(V)]\right), \tag{182}$$

where $\mathrm{sgn}(V) \coloneqq W\mathrm{sgn}(\Sigma)U^\dagger$, and $\mathrm{sgn}(\cdot)$ is the sign function. Then, $\mathrm{Tr}[\rho\, \mathrm{sgn}(V)]$ and $\mathrm{Tr}[\sigma\, \mathrm{sgn}(V)]$ can be estimated separately, by combining the techniques of quantum singular value transformation [76] and the Hadamard test [77]. To this end, to implement unitary block-encodings of ρ and σ approximately, the technique of density matrix exponentiation [78] is used.

The same techniques can be employed to compute $\mathsf{T}^\varepsilon(\rho, \sigma, \mathcal{A})$ since

$$\mathsf{T}^\varepsilon(\rho, \sigma, \mathcal{A}) = \frac{1}{e^\varepsilon + 1}\left(\mathrm{Tr}[\mathcal{A}(\rho)\ \mathrm{sgn}(\mathrm{V}^\varepsilon)] - e^\varepsilon \mathrm{Tr}[\mathcal{A}(\sigma)\ \mathrm{sgn}(\mathrm{V}^\varepsilon)]\right), \tag{183}$$

where $\mathrm{V}^\varepsilon \coloneqq (\mathcal{A}(\rho) - e^\varepsilon \mathcal{A}(\sigma))/(e^\varepsilon + 1)$.

2) Type-1 Error Analysis: We arrive at the following bound on the Type-1 error of the proposed hypothesis testing pipeline.

Proposition 8 (Type-I Error): Fix arbitrary $\alpha, \delta > 0$ and consider the above hypothesis testing pipeline. Then

$$\sup_{\rho \sim \sigma} \mathbb{P}\left(\hat{\mathsf{T}}^\varepsilon_{\max}(\rho, \sigma, \mathcal{A}) > \mathsf{g}(\varepsilon, \delta) + \alpha \,\middle|\, H_0\right) \leq \beta, \tag{184}$$

if the algorithm from Lemma 3 has access to

$$O\left(\log\left(\frac{1}{\beta}\right) \frac{r^2}{\alpha^5} \log^2\left(\frac{r}{\alpha}\right) \log^2\left(\frac{1}{\alpha}\right)\right) \tag{185}$$

identical copies of the states ρ and σ, such that $\rho \sim \sigma$ and where

$$r \coloneqq \sup_{\rho \sim \sigma} \max\{\mathrm{rank}(\mathcal{A}(\rho)), \mathrm{rank}(\mathcal{A}(\sigma))\}. \tag{186}$$

Proof: Fix ρ and σ such that $\rho \sim \sigma$. Under the null hypothesis and the assumption that $\hat{\mathsf{T}}^\varepsilon_{\max}(\rho, \sigma, \mathcal{A}) = \hat{\mathsf{T}}^\varepsilon(\rho, \sigma, \mathcal{A})$, we have that

$$\mathbb{P}\left(\hat{\mathsf{T}}^\varepsilon_{\max}(\rho, \sigma, \mathcal{A}) > \mathsf{g}(\varepsilon, \delta) + \alpha\right)$$

$$= \mathbb{P}\Big(\hat{\mathsf{T}}^{\varepsilon}(\rho,\sigma,\mathcal{A}) > \mathsf{g}(\varepsilon,\delta)+\alpha\Big) \tag{187}$$

$$\overset{(a)}{\leq} \mathbb{P}\Big(\hat{\mathsf{T}}^{\varepsilon}(\rho,\sigma,\mathcal{A}) - \mathsf{T}^{\varepsilon}(\rho,\sigma,\mathcal{A}) > \alpha\Big) \tag{188}$$

$$\leq \mathbb{P}\Big(|\hat{\mathsf{T}}^{\varepsilon}(\rho,\sigma,\mathcal{A}) - \mathsf{T}^{\varepsilon}(\rho,\sigma,\mathcal{A})| > \alpha\Big) \tag{189}$$

$$\overset{(b)}{\leq} \beta, \tag{190}$$

where: (a) follows since $\mathsf{T}^{\varepsilon}(\rho,\sigma,\mathcal{A}) \leq \mathsf{g}(\varepsilon,\delta)$ under the null hypothesis; (b) from the high probability statement in Lemma 3. Similarly, the above inequality holds when $\hat{\mathsf{T}}^{\varepsilon}_{\max}(\rho,\sigma,\mathcal{A}) = \hat{\mathsf{T}}^{\varepsilon}(\sigma,\rho,\mathcal{A})$ concluding the proof. ∎

Proposition 8 provides a bound on the number of samples of the states required to achieve type-I error (significance) of β. In that case, we would use a threshold of $g(\varepsilon,\delta)+\alpha$ for accepting the null hypothesis, such that the null hypothesis is accepted when the test statistic is less than or equal to $g(\varepsilon,\delta)+\alpha$.

Remark 17 (Computational Complexity With Rank r): From Proposition 8, it is evident that the copy complexity of the algorithm grows as $O(r^2\log^2(r))$. Let $\mathcal{A}:\mathcal{L}(\mathcal{H}_A)\to\mathcal{L}(\mathcal{H}_B)$ be a quantum channel. Then, $r\leq d_B$, where d_B is the dimension of the Hilbert space $\mathcal{H}_B$. To handle computational complexity, one possibility is to compose $\mathcal{A}$ with another quantum channel $\mathcal{N}$ that translates the space to a low-dimensional setting. However, due to the data-processing inequality for the trace distance, it will only provide a lower bound. With that, it is possible to reject algorithms if $\mathsf{T}^{\varepsilon}(\rho,\sigma,\mathcal{N}\circ\mathcal{A}) > g(\varepsilon,\delta)$, which implies that $\mathsf{T}^{\varepsilon}(\rho,\sigma,\mathcal{A}) > g(\varepsilon,\delta)$. Consequently, it may lead to limitations similar to classical auditing approaches that use relaxed privacy notions, since the contraction gap between $\mathsf{T}^{\varepsilon}(\rho,\sigma,\mathcal{N}\circ\mathcal{A})$, and $\mathsf{T}^{\varepsilon}(\rho,\sigma,\mathcal{A})$ is hard to quantify. In the quantification of the gap, finding the contraction coefficient $\eta_{\mathcal{N}}$ of the channel $\mathcal{N}$ would be useful if $\eta_{\mathcal{N}}<1$ (recall that the contraction coefficient of a channel $\mathcal{N}$ under a generalized divergence $\mathbf{D}$, as given in Eq. (1), is defined as $\eta_{\mathcal{N}} := \sup_{\rho,\sigma\in\mathcal{D},\mathbf{D}(\rho\|\sigma)\neq 0}\frac{\mathbf{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))}{\mathbf{D}(\rho\|\sigma)}$).

In summary, we proposed a hypothesis testing pipeline for auditing the privacy of quantum systems, offering formal guarantees on auditing QDP using quantum algorithms designed for estimating trace distance. However, an essential task for further investigation is analyzing the Type-II error of this approach. This analysis would allow us to quantify the power of the test and assess its ability to correctly accept algorithms with the desired privacy requirements, which is left for future work.

VIII. Information-Theoretic Bounds From QPP and Connections to Quantum Fairness

In this section, we begin by investigating several information-theoretic bounds that stem from an algorithm satisfying QPP constraints. We then establish connections between QPP and quantum fairness [34], [35] using these bounds. Later on we utilize the derived bounds to assess the relative strength of the QPP variants introduced in Section IX.

A. Information-Theoretic Bounds From QPP

In [15], it was highlighted that finding bounds on quantum relative entropy and mutual information resulting from QDP is an interesting open problem. We address this question in a general setting, encompassing QDP as a special case. We offer bounds for relative entropy and Holevo information, along with bounds for Rényi relative entropies and trace distance. For the rest of the discussion, we adopt the fixed privacy framework to be $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$.

Proposition 9 (Bounds on Quantum Rényi Relative Entropy and Quantum Relative Entropy Due to QPP): Fix $\alpha>1$. If $\mathcal{A}$ is ε-QPP (i.e., $(\varepsilon,0)$-QPP) in the framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$ for all $P_X\in\Theta$ and $(\mathcal{R},\mathcal{T})\in\mathcal{Q}$, then

$$\mathsf{D}_{\alpha}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \min\left\{\frac{\varepsilon^2\alpha}{2},\varepsilon\right\}, \tag{191}$$

where $\mathsf{D}_{\alpha}(\cdot\|\cdot)$ is an arbitrary quantum Rényi relative entropy satisfying data processing.[8] Furthermore,

$$\mathsf{D}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \min\left\{\frac{\varepsilon^2}{2},\varepsilon\right\}, \tag{192}$$

where D is the quantum relative entropy in (6).

Proof: ε-QPP of the framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$ implies that for all $P_X\in\Theta$ and $(\mathcal{R},\mathcal{T})\in\mathcal{Q}$

$$\mathsf{D}_T\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \varepsilon, \tag{193}$$

where D_T is the Thompson metric from (16). This follows by the definition of the max-relative entropy defined in (14) and Thompson metric in (16), and recalling the definition of QPP framework for $\delta=0$ and $\mathcal{M}=\bar{\mathcal{M}}$ (see Definition 4). By this implication and the fact that every quantum Rényi-divergence D_{α} of order α satisfying data processing is bounded from above by $\mathsf{D}_{\max}$ [79, Eq. (4.36)], ε-QPP implies that for all $P_X\in\Theta$ and $(\mathcal{R},\mathcal{T})\in\mathcal{Q}$,

$$\mathsf{D}_{\alpha}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \varepsilon. \tag{194}$$

By a different argument via the maximal extension presented in Lemma 4, we can further arrive at the following, for all $P_X\in\Theta$, $(\mathcal{R},\mathcal{T})\in\mathcal{Q}$, and $\varepsilon\alpha\leq 2$,

$$\mathsf{D}_{\alpha}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big) \leq \frac{\varepsilon^2\alpha}{2}. \tag{195}$$

This completes the proof of the first inequality.

Next, noting that the Petz–Rényi relative entropy in (5) satisfies data processing for $\alpha\in(0,1)\cup(1,2]$, and then taking the limit $\alpha\to 1^{+}$, we arrive at the bound on quantum relative entropy by using the equality in (6). ∎

Lemma 4: Fix $\alpha>1$, and ρ,σ PSD operators. For $\alpha\mathsf{D}_T(\rho\|\sigma)\leq 2$, the following inequality holds:

$$\mathsf{D}_{\alpha}(\rho\|\sigma) \leq \frac{\alpha}{2}(\mathsf{D}_T(\rho\|\sigma))^2, \tag{196}$$

where $\mathsf{D}_{\alpha}(\rho\|\sigma)$ is an arbitrary quantum Rényi relative entropy satisfying data processing.

Proof: See Appendix VIII. ∎

Remark 18 (Operational Interpretation of Thompson Metric): An operational interpretation of the Thompson metric has appeared in symmetric postselected hypothesis testing (a setting allowing for an inconclusive outcome along with

[8] Note that Petz–Rényi in (5) satisfies data processing for $\alpha\in(0,1)\cup(1,2]$ and sandwiched Rényi in (8) satisfies data processing for $\alpha\in[1/2,1)\cup(1,\infty)$.

two general conclusive outcomes and postselecting on the conclusive outcomes) as the asymptotic error exponent of discriminating two quantum states ρ and σ [51], as well as in the resource theory of symmetric distinguishability [50]. Here by referring to (193), the QPP framework also provides another operational interpretation of the Thompson metric. In the framework $(\mathcal{S}, \mathcal{Q}, \Theta, \bar{\mathcal{M}})$, for fixed $P_X \in \Theta$ and $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$,[9] the Thompson metric given by $\mathsf{D}_T\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big)$ is equal to the minimal ε needed to achieve ε-QPP.

Proposition 10 (Bounds on Trace Norm): If $\mathcal{A}$ is ε-QPP, then

$$\sup_{\Theta,(\mathcal{R},\mathcal{T})\in\mathcal{Q}} \left\|\mathcal{A}(\rho^{\mathcal{R}}) - \mathcal{A}(\rho^{\mathcal{T}})\right\|_1 \leq \min\left\{\varepsilon, \sqrt{2\varepsilon}\right\}, \tag{197}$$

and if $\mathcal{A}$ is (ε, δ)-QPP, we have

$$\sup_{\Theta,(\mathcal{R},\mathcal{T})\in\mathcal{Q}} \left\|\mathcal{A}(\rho^{\mathcal{R}}) - \mathcal{A}(\rho^{\mathcal{T}})\right\|_1 \leq 2 - \frac{4(1-\delta)}{e^{\varepsilon}+1}. \tag{198}$$

Proof: The first inequality holds by applying the quantum Pinsker inequality ($\frac{1}{2}\|\rho-\sigma\|_1^2 \leq \mathsf{D}(\rho\|\sigma)$) [80, Theorem 1.15] and Proposition 9.

For the second inequality: $(0, \delta')$-QPP is equivalent to

$$\sup_{\Theta,(\mathcal{R},\mathcal{T})\in\mathcal{Q}} \frac{\left\|\mathcal{A}(\rho^{\mathcal{R}}) - \mathcal{A}(\rho^{\mathcal{T}})\right\|_1}{2} \leq \delta'. \tag{199}$$

Then, adapting Lemma 5 and fixing ε' therein to zero leads to the desired result. ∎

Lemma 5: Fix $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ privacy framework. Then, we have

$$(\varepsilon, \delta)\text{-QPP} \implies (\varepsilon', \delta')\text{-QPP}, \tag{200}$$

where $\varepsilon' < \varepsilon$ with

$$\delta' := 1 - \frac{(e^{\varepsilon'}+1)(1-\delta)}{(e^{\varepsilon}+1)}. \tag{201}$$

Proof: The proof follows similarly to the proof of [11, Property 3] for classical DP and is presented in Appendix IX. ∎

Remark 19 (Bounds on Holevo Information): In Appendix X, we provide bounds on the Holevo information in the settings of QDP and QLDP (recall this privacy notion from Remark III-C2).

B. Quantum Fairness and QPP

We now demonstrate that quantum fairness can be viewed as a special case of QPP, which should encourage the design of customized fairness models via the QPP framework.

Quantum fairness seeks to treat all input states equally, meaning that all pairs of input states that are close in some metric (e.g., close in trace distance) should yield similar outcomes when processed by an algorithm [34]. For a quantum decision model $\mathcal{A} = \{\mathcal{E}, \{\mathrm{M}_i\}_{i\in\mathcal{O}}\}$, where a quantum channel $\mathcal{E}$ is followed by a POVM $\{\mathrm{M}_i\}_{i\in\mathcal{O}}$ (i.e., a quantum algorithm described by a quantum to classical channel), quantum fairness is defined in [34] as follows.

[9]By definition also for $(\mathcal{T}, \mathcal{R}) \in \mathcal{Q}$.

Definition 6 ((α, β)-Fairness [34]): Suppose we are given a quantum decision model $\mathcal{A} = \{\mathcal{E}, \{\mathrm{M}_i\}_{i\in\mathcal{O}}\}$, two distance metrics $D(\cdot\|\cdot)$ and $d(\cdot\|\cdot)$ on $\mathcal{D}(\mathcal{H})$ and $\mathcal{D}(\mathcal{O})$ respectively. Fix $0 < \alpha, \beta \leq 1$. Then the decision model $\mathcal{A}$ is (α, β) fair if for all $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ such that $D(\rho\|\sigma) \leq \alpha$, then

$$d(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \beta. \tag{202}$$

Proposition 11 (Fairness Guarantee From QPP): Let $D(\rho\|\sigma) = \|\rho-\sigma\|_1/2$ and $d(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) = \frac{1}{2}\sum_i |\mathrm{Tr}[\mathrm{M}_i\mathcal{E}(\rho-\sigma)]|$. Fix

$$\begin{aligned}
\mathcal{S} &= \{\rho : \rho \in \mathcal{D}(\mathcal{H})\},\\
\mathcal{Q} &= \{(\rho,\sigma) : \rho,\sigma \in \mathcal{D}(\mathcal{H}),\ D(\rho\|\sigma) \leq \alpha\},\\
\Theta &= \mathcal{P}_2\big(\mathcal{D}(\mathcal{H})\big),\\
\mathcal{M} &= \{\mathrm{M} : 0 \leq \mathrm{M} \leq \mathrm{I}\}.
\end{aligned} \tag{203}$$

If $\mathcal{E}$ satisfies ε-QPP with $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ above, then $\mathcal{A} = \{\mathcal{E}, \{\mathrm{M}_i\}_{i\in\mathcal{O}}\}$ is $(\alpha, \sqrt{\varepsilon'/2})$-fair, where $\varepsilon' = \min\{\varepsilon, \varepsilon^2/2\}$.

Proof: From Proposition 10, $\mathcal{E}$ being ε-QPP implies

$$\|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1 \leq \min\{\varepsilon, \sqrt{2\varepsilon}\} = \sqrt{2\varepsilon'}, \tag{204}$$

for ρ and σ such that $D(\rho\|\sigma) \leq \alpha$.

Then, consider the measurement channel that performs the following transformation:

$$\mathcal{E}(\rho) \to \sum_{i\in\mathcal{O}} \mathrm{Tr}[\mathrm{M}_i\mathcal{E}(\rho)]\,|i\rangle\langle i|. \tag{205}$$

It follows from the data-processing inequality for the trace distance that

$$\left\|\sum_{i\in\mathcal{O}} \left(\mathrm{Tr}[\mathrm{M}_i\mathcal{E}(\rho)] - \mathrm{Tr}[\mathrm{M}_i\mathcal{E}(\sigma)]\right)|i\rangle\langle i|\right\|_1 \leq \sqrt{2\varepsilon'}. \tag{206}$$

This leads to

$$d(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) = \frac{1}{2}\sum_i |\mathrm{Tr}[\mathrm{M}_i\mathcal{E}(\rho-\sigma)]| \leq \sqrt{\frac{\varepsilon'}{2}}, \tag{207}$$

concluding the proof. ∎

This shows how ε-QPP also provides fairness to the decision models of interest. However, fairness guarantees may not be sufficient to guarantee privacy in general. Next, we show that when $\mathcal{M}$ satisfies a relationship related to the POVM of the quantum decision model where fairness needs to be ensured, fair algorithms also act as privacy mechanisms.

Proposition 12 (Privacy Guarantees Obtained From Fairness): Consider the same privacy framework as in (203) but with the modification $\mathcal{M} = \{\cup_{i\in\mathcal{B}}\mathrm{M}_i | \mathcal{B} \subseteq \mathcal{O}\}$. If $\mathcal{A} = \{\mathcal{E}, \{\mathrm{M}_i\}_{i\in\mathcal{O}}\}$ is (α, β)-fair, then $\mathcal{E}$ is $(\varepsilon, 2\beta)$-QPP for every $\varepsilon \geq 0$.

Proof: Since $\mathcal{A}$ is (α, β)-fair, it follows that

$$\frac{1}{2}\sum_{i\in\mathcal{O}} |\mathrm{Tr}[\mathrm{M}_i\mathcal{E}(\rho-\sigma)]| \leq \beta. \tag{208}$$

Then, for every $\mathcal{B} \subseteq \mathcal{O}$

$$\left|\mathrm{Tr}\left[\sum_{i\in\mathcal{B}} \mathrm{M}_i\mathcal{E}(\rho-\sigma)\right]\right| \leq \sum_{i\in\mathcal{B}} |\mathrm{Tr}[\mathrm{M}_i\mathcal{E}(\rho-\sigma)]| \tag{209}$$

$$\leq \sum_{i\in\mathcal{O}} |\mathrm{Tr}[\mathrm{M}_i\mathcal{E}(\rho-\sigma)]| \tag{210}$$

$$\leq 2\beta, \tag{211}$$

where the first inequality follows from the triangular inequality, the second from $\mathcal{B}\subseteq\mathcal{O}$, and the last from (208). ∎

IX. Variants of Quantum Pufferfish Privacy Framework

We now propose variants of QPP via generalized divergences (as defined in (1)).[10] We provide an operational interpretation of generalized divergences as privacy metrics and characterize the relative strength between them.

Here, we focus on QPP in the framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$ and formulate QPP based on generalized divergences, which we denote by $(\mathbf{D},\varepsilon)$-QPP, where $\mathbf{D}$ is a placeholder for the generalized divergence being used.

Definition 7 ($(\mathbf{D},\varepsilon)$-QPP): Fix a privacy framework $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$ and $\varepsilon>0$. An algorithm $\mathcal{A}$ is $(\mathbf{D},\varepsilon)$-QPP if

$$\sup_{\Theta,(\mathcal{R},\mathcal{T})\in\mathcal{Q}} \mathbf{D}\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big)\leq\varepsilon, \tag{212}$$

where $\mathbf{D}$ is an arbitrary generalized divergence and $\rho^{\mathcal{R}}$ and $\rho^{\mathcal{T}}$ are defined as in Definition 4. Note that these two density matrices depend on elements of Θ and $\mathcal{Q}$.

Indeed, Definition 7 encompasses variants of classical DP [81], [82] and QDP [15], which rely on Rényi divergences as the generalized divergence along with the appropriate choice of QPP framework $(\mathcal{S},\mathcal{Q},\Theta,\mathcal{M})$ stated in Remark III-C3 and Remark III-C1, respectively.

Remark 20 (Properties of $(\mathbf{D},\varepsilon)$-QPP): Post-processing holds, by the definition of generalized divergences. Convexity follows if $\mathbf{D}$ satisfies the direct-sum property, as defined in [37, Eq. (4.3.7)], from which it follows that $\mathbf{D}$ is jointly convex [37, Proposition 4.15].

Parallel composability: If $\mathcal{A}_i$ satisfies (ε_i,δ_i)-QPP in $(\mathcal{S},\mathcal{Q},\Theta,\bar{\mathcal{M}})$ for $i\in\{1,\ldots,k\}$, then the composed mechanism, as defined in Theorem 1, satisfies $\left(\mathbf{D},\sum_{i=1}^k\varepsilon_i\right)$-QPP in the framework $\left(\mathcal{S},\mathcal{Q}^{(k)},\Theta,\bar{\mathcal{M}}^k\right)$, if $\mathbf{D}$ satisfies subadditivity (i.e., if $\mathbf{D}(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2)\leq\mathbf{D}(\rho_1\|\sigma_1)+\mathbf{D}(\rho_2\|\sigma_2)$ for all states ρ_1, ρ_2, σ_1, and σ_2). It is worth noting that by employing this privacy notion based on generalized divergences, we achieve improved composability results, even in scenarios involving joint measurements (recall property 3 of Corollary 2).

Remark 21 (Auditing Variants of QPP): The methodologies proposed in Section VII can be used to audit the variants of QPP (based on generalized divergences) as well. In this regard, quantum algorithms and procedures for estimating respective generalized divergences (e.g., Rényi relative entropies) would be useful. This motivates the development of novel techniques for estimating them efficiently and accurately, beyond those already established in [83].

[10] Due to the definition of generalized divergences, the privacy notions defined based on them inherently satisfy post-processing.

A. Variants Based on Rényi Divergences

First, let us recall definitions of the following quantities. The measured Rényi divergence of order $\alpha\in(0,1)\cup(1,\infty)$ is defined as [84, Eqs. (3.116)–(3.117)]

$$\check{\mathsf{D}}_\alpha(\rho\|\sigma) := \sup_{\mathcal{M}} \mathsf{D}^c_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)), \tag{213}$$

where

$$\mathsf{D}^c_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) := \frac{1}{\alpha-1}\ln\left(\sum_{x\in\mathcal{X}}(p(x))^\alpha(q(x))^{1-\alpha}\right), \tag{214}$$

with $p(x):=\mathrm{Tr}[\mathrm{M}_x\rho]$ and $q(x):=\mathrm{Tr}[\mathrm{M}_x\sigma]$ for $\mathcal{M}$ corresponding to a POVM $\{\mathrm{M}_x\}_{x\in\mathcal{X}}$. The Rényi preparation divergence of order $\alpha\in(0,1)\cup(1,\infty)$ is defined as [85]

$$\hat{\mathsf{D}}_\alpha(\rho\|\sigma) := \inf_{P,Q,\mathcal{P}} \mathsf{D}^c_\alpha(P\|Q), \tag{215}$$

where $\mathcal{P}$ is a classical–quantum channel, $\mathcal{P}(P)=\rho$, $\mathcal{P}(Q)=\sigma$, and the classical Rényi divergence is defined as

$$\mathsf{D}^c_\alpha(P\|Q) := \frac{1}{\alpha-1}\ln\left(\sum_{x\in\mathcal{X}}(P(x))^\alpha(Q(x))^{1-\alpha}\right). \tag{216}$$

The quantities in (213) and (215) satisfy the data-processing inequality for all $\alpha\in(0,1)\cup(1,\infty)$ by construction, following from this property holding for the underlying classical divergence. Moreover, the following bounds hold

$$\check{\mathsf{D}}_\alpha(\rho\|\sigma)\leq\mathsf{D}_\alpha(\rho\|\sigma)\leq\hat{\mathsf{D}}_\alpha(\rho\|\sigma). \tag{217}$$

where D_α is an arbitrary quantum Rényi divergence that satisfies data processing [86, Eq. (3.7)].

Remark 22 (Relative Strength of Privacy Metrics): Choosing the preparation divergence $\hat{\mathsf{D}}_\alpha=\mathbf{D}$ in $(\mathbf{D},\varepsilon)$-QPP gives a stronger privacy metric that satisfies the post-processing property for the family of quantum Rényi divergences of order α, while $\check{\mathsf{D}}_\alpha=\mathbf{D}$ gives a weaker privacy metric from that same family of divergences. That is, we have that

$$\sup_{\Theta,(\mathcal{R},\mathcal{T})\in\mathcal{Q}} \check{\mathsf{D}}_\alpha\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big)$$

$$\leq \sup_{\Theta,(\mathcal{R},\mathcal{T})\in\mathcal{Q}} \mathsf{D}_\alpha\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big) \tag{218}$$

$$\leq \sup_{\Theta,(\mathcal{R},\mathcal{T})\in\mathcal{Q}} \hat{\mathsf{D}}_\alpha\big(\mathcal{A}(\rho^{\mathcal{R}})\|\mathcal{A}(\rho^{\mathcal{T}})\big), \tag{219}$$

so that

$$(\hat{\mathsf{D}}_\alpha,\varepsilon)\text{-QPP} \implies (\mathsf{D}_\alpha,\varepsilon)\text{-QPP} \implies (\check{\mathsf{D}}_\alpha,\varepsilon)\text{-QPP}. \tag{220}$$

The above relations follow by the direct application of (217) and Definition 7.

Remark 23 (Operational Interpretation as Privacy Metrics): Choose

$$\mathcal{S}=\{\rho,\sigma\}, \tag{221}$$

$$\mathcal{Q}=\{(\rho,\sigma)\}, \tag{222}$$

$$\Theta=\{\{P_X(x),\rho^x\}_{x\in\mathcal{X}}: P_X\in\mathcal{P}(\mathcal{X}),\ \rho^x\in\{\rho,\sigma\}\}, \tag{223}$$

$$\mathcal{M} = \bar{\mathcal{M}}. \tag{224}$$

The strongest privacy metric that can be generated from the family of quantum Rényi divergences of order α, which is also a generalized divergence, is devised by setting $\mathbf{D} = \hat{\mathsf{D}}_\alpha$. In the chosen QPP framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$, the value of ε that divides the region where $(\mathsf{D}, \varepsilon)$-QPP is achieved and the region where it is violated, for an identity channel, is $\hat{\mathsf{D}}_\alpha(\rho\|\sigma)$.

The sandwiched Rényi relative entropy $\widetilde{\mathsf{D}}_\alpha$ in (8) satisfies data processing for $\alpha \in [1/2, 1) \cup (1, \infty)$ [45], [46], and the quantum relative entropy D in (6) also satisfies data processing [87]. Thus, both of these are candidates for a generalized divergence.

Proposition 13: Fix $(\mathcal{S}, \mathcal{Q}, \Theta, \bar{\mathcal{M}})$, $\alpha \in [1/2, 1) \cup (1, \infty)$, and $\delta \in (0, 1)$. Then, for an algorithm $\mathcal{A}$, we have that

$$\varepsilon\text{-QPP} \implies (\widetilde{\mathsf{D}}_\alpha, \varepsilon')\text{-QPP} \implies (\varepsilon^\star, \delta)\text{-QPP}, \tag{225}$$

where

$$\varepsilon' \coloneqq \min\left\{\varepsilon, \frac{\varepsilon^2 \alpha}{2}\right\}, \tag{226}$$

$$\varepsilon^\star \coloneqq \varepsilon' + \frac{1}{\alpha - 1} \ln\left(\frac{1}{\delta^2}\right) + \ln\left(\frac{1}{1 - \delta^2}\right). \tag{227}$$

We also have

$$\varepsilon\text{-QPP} \implies (\mathsf{D}, \varepsilon'')\text{-QPP} \implies (\hat{\varepsilon}, \delta)\text{-QPP}, \tag{228}$$

where

$$\varepsilon'' \coloneqq \min\left\{\varepsilon, \frac{\varepsilon^2}{2}\right\}, \tag{229}$$

$$K \coloneqq \sup_{\Theta, (\mathcal{R}, \mathcal{T}) \in \mathcal{Q}} \mathsf{T}\big(\mathcal{A}(\rho^{\mathcal{R}}), \mathcal{A}(\rho^{\mathcal{T}})\big) \tag{230}$$

$$\hat{\varepsilon} \coloneqq \frac{1}{\delta^2}(\varepsilon' + K) + \ln\left(\frac{1}{1 - \delta^2}\right). \tag{231}$$

Proof: The first implication in both (225) and (228) follows from Proposition 9. Then for the second implication, recall from item 3 of Proposition 2 that $\overline{\mathsf{D}}^\delta(\rho\|\sigma) \leq \mathsf{D}^\delta_{\max}(\rho\|\sigma)$. Then applying [48, Propositions 5 and 6], we arrive at:

$$\overline{\mathsf{D}}^\delta(\rho\|\sigma) \leq \widetilde{\mathsf{D}}_\alpha(\rho\|\sigma) + \frac{1}{\alpha - 1} \ln\left(\frac{1}{\delta^2}\right) + \ln\left(\frac{1}{1 - \delta^2}\right), \tag{232}$$

$$\overline{\mathsf{D}}^\delta(\rho\|\sigma) \leq \frac{1}{\delta^2}(\mathsf{D}(\rho\|\sigma) + \mathsf{T}(\rho, \sigma)) + \ln\left(\frac{1}{1 - \delta^2}\right). \tag{233}$$

Finally, to conclude the proof, invoke Proposition 1 to establish the required relationship to the QPP framework from there. ■

Note that the chain of implications in (225) holds for every Rényi divergence D_α satisfying data processing, beyond just $\widetilde{\mathsf{D}}_\alpha$, because data processing is the key property to adapt Proposition 9 and [48, Proposition 6] (see Eqs. (K51) and (K52) therein).

Remark 24 (Comparison to Existing Results for QDP): In the special case of QDP, the dependence on the (ε, α) parameters in (225) provides a strict improvement over previous results. Specifically, Lemmas V.4 and V.5 of [15] show that

$$\varepsilon\text{-QDP} \implies (\mathsf{D}_\alpha, \varepsilon)\text{-QDP} \implies (\bar{\varepsilon}, \delta)\text{-QDP}, \tag{234}$$

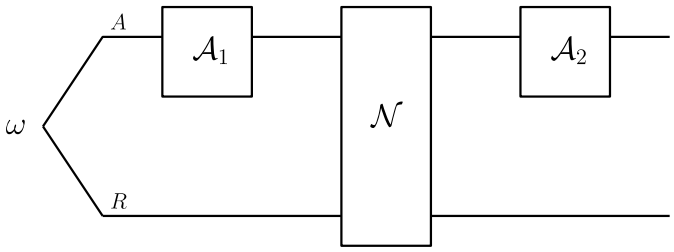


Fig. 8. Adaptive composition with reference systems: First $\mathcal{A}_1$ is applied on the upper system, then the quantum channel $\mathcal{N}$ on both the systems, and lastly $\mathcal{A}_2$ on the upper system. Then the adaptive composition is QPP if $\mathcal{A}_1$ and $\mathcal{A}_2$ are.

with

$$\bar{\varepsilon} \coloneqq \varepsilon + \frac{\ln\left(1/(1 - \sqrt{1 - \delta^2})\right)}{\alpha - 1} \approx \varepsilon + \frac{\ln(2/\delta^2)}{\alpha - 1}, \tag{235}$$

where the approximation holds for small δ. Our first implication in (225) is tighter compared to this since $\varepsilon' \leq \varepsilon$ and the second implication is also tighter (i.e, $\varepsilon^\star \leq \bar{\varepsilon}$) if we choose $\delta^2 \leq 1 - 2^{\left(-\frac{1}{\alpha - 1}\right)}$ in the small δ regime.

B. Variant Incorporating Entanglement

In this subsection, we introduce a novel variant of QPP that incorporates reference systems. This extension potentially allows us to explore the impact of entanglement on the privacy of the system of interest.

Definition 8 ($(\mathbf{D}^R, \varepsilon)$-QPP With Reference Systems): With the same setup $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ in Definition 4, a quantum algorithm $\mathcal{A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is $(\varepsilon, \mathbf{D})$-QPP with reference systems if

$$\sup_{\Theta, (\omega^{\mathcal{R}}_{RA}, \omega^{\mathcal{T}}_{RA}) \in \mathcal{G}} \mathbf{D}\big((\mathcal{I}_R \otimes \mathcal{A})(\omega^{\mathcal{R}}_{RA}) \| (\mathcal{I}_R \otimes \mathcal{A})(\omega^{\mathcal{T}}_{RA})\big) \leq \varepsilon, \tag{236}$$

where the set $\mathcal{G}$ is defined in (27).

Since there is no restriction on the reference systems, the supremum in Definition 8 is also taken over the dimension of the reference system R, which is an unbounded set. However, following from isometric invariance of generalized divergences, together with purification and the Schmidt decomposition, it suffices to take the supremum over pure bipartite states with the reference system R isomorphic to the channel input system A.

Remark 25 (Properties): Similar to variants of QPP (recall Remark 20), this variant incorporating reference systems also satisfies post-processing, convexity, and parallel composability.

Adaptive composition: Let $\mathcal{A}_i$ be a $(\mathbf{D}^R, \varepsilon_i)$-QPP algorithm for $i \in \{1, 2\}$. Consider the adaptive composition of $\mathcal{A}_1$ and $\mathcal{A}_2$, as shown in Fig. 8: First $\mathcal{A}_1$ is applied on the upper system, then the quantum channel $\mathcal{N}$ acts on both systems, and lastly $\mathcal{A}_2$ acts on the upper system. Adaptive composition of $\mathcal{A}_1$ and $\mathcal{A}_2$ also satisfies $(\mathbf{D}^R, \varepsilon_1)$-QPP. This shows that the adaptive composition illustrated in Fig. 8 does not degrade privacy, showcasing the strength of the privacy framework in Definition 8.

Next, we observe that Definition 8 is a stronger privacy guarantee than Definition 7, which does not take into account reference systems.

Corollary 3 (Strength Compared to $(\mathbf{D},\varepsilon)$-QPP): $(\mathbf{D}^R,\varepsilon)$-QPP implies $(\mathbf{D},\varepsilon)$-QPP.
This follows from the data-processing inequality for the underlying generalized divergence $\mathbf{D}$, choosing the channel as the partial trace over the reference system R. It highlights that in certain scenarios where the system of interest is entangled with other reference systems, the general QPP guarantees defined in earlier sections may not be sufficient.

The choice of $\mathbf{D}$ could heavily affect the design of useful privacy frameworks with entanglement (recall the example in Remark 3 with the choice $\mathbf{D} = \mathsf{D}_{\max}$, along with the equivalence of $\mathsf{D}_{\max}$ to ε-QPP, as given in (193)). Therefore, careful consideration of the appropriate generalized divergence is essential for developing effective and meaningful privacy frameworks that account for entanglement effects.

X. Concluding Remarks and Future Directions

This work proposed QPP as a flexible privacy framework for quantum systems. We showed that QPP is captured exactly by the DL divergence, endowing the latter with an operational interpretation of the DL divergence. The DL divergence representation was used to study properties of QPP mechanisms and characterize privacy-utility tradeoffs. As a concrete case study, we explored the depolarization QPP mechanisms and characterized the parameter values to achieve privacy. A methodology for auditing quantum privacy was also developed.

Future research directions are abundant and include privacy-utility analysis of specific quantum estimation tasks, designing efficient quantum algorithms that achieve QPP, analysing the type-II error of our quantum privacy auditing pipeline, providing tight characterizations of parallel composability of QPP mechanisms, and devising efficient methods for computing the DL divergence using quantum algorithms to enable quantum privacy auditing. In the longer term, the proposed framework could lay the foundations for privacy-preserving learning in quantum systems.

Appendix I
Alternative Proof for Joint-Quasi Convexity of DL Divergence

By adapting the strong convexity of Hockey-Stick divergence (Proposition II.5 of [15]), with the substitutions $p = q$ and $\gamma_2 = 1$ and $\gamma_1 = \gamma$ therein, we obtain

$$\operatorname{Tr}\left[\left(\sum_{i=1}^k p_i\rho_i - \gamma\sum_{i=1}^k p_i\sigma_i\right)_+\right] \le \sum_{i=1}^k p_i \operatorname{Tr}\left[(\rho_i-\gamma\sigma_i)_+\right] \le \max_i \operatorname{Tr}\left[(\rho_i-\gamma\sigma_i)_+\right]. \tag{237}$$

Then by assuming $\max_i \operatorname{Tr}\left[(\rho_i-\gamma\sigma_i)_+\right] \le \delta$, γ is a candidate for the optimization of $\overline{\mathsf{D}}^\delta\left(\sum_{i=1}^k p_i\rho_i \middle\| \sum_{i=1}^k p_i\sigma_i\right)$. This leads to

$$\overline{\mathsf{D}}^\delta\left(\sum_{i=1}^k p_i\rho_i \middle\| \sum_{i=1}^k p_i\sigma_i\right) \le \ln(\gamma). \tag{238}$$

The above holds for all γ such that $\max_i \operatorname{Tr}\left[(\rho_i-\gamma\sigma_i)_+\right] \le \delta$. Then, optimizing over such γ, we arrive at joint quasi-convexity:

$$\overline{\mathsf{D}}^\delta\left(\sum_{i=1}^k p_i\rho_i \middle\| \sum_{i=1}^k p_i\sigma_i\right) \le \max_i \overline{\mathsf{D}}^\delta(\rho_i\|\sigma_i). \tag{239}$$

Appendix II
Proof of Lemma 2

The proof given below is closely related to the proof of [79, Lemma 6.21], but there are some subtle differences and so we provide it here for completeness.

Let

$$\Sigma := (\rho-\lambda\sigma)_+, \tag{240}$$
$$G := (\lambda\sigma)^{1/2}(\lambda\sigma+\Sigma)^{-1/2}. \tag{241}$$

Note that

$$\begin{aligned} 0 &\le G^\dagger G && (242)\\ &= (\lambda\sigma+\Sigma)^{-1/2}(\lambda\sigma)(\lambda\sigma+\Sigma)^{-1/2} && (243)\\ &\le (\lambda\sigma+\Sigma)^{-1/2}(\lambda\sigma+\Sigma)(\lambda\sigma+\Sigma)^{-1/2} && (244)\\ &\le I. && (245)\end{aligned}$$

From the fact that $\rho-\lambda\sigma \le \Sigma$, it follows that

$$\rho \le \lambda\sigma + \Sigma. \tag{246}$$

Define the following state:

$$\widetilde{\rho} := \frac{G\rho G^\dagger}{\operatorname{Tr}[G^\dagger G\rho]}. \tag{247}$$

Consider that

$$\begin{aligned} &1-\operatorname{Tr}[G\rho G^\dagger]\\ &= \operatorname{Tr}[(I-G^\dagger G)\rho] && (248)\\ &\le \operatorname{Tr}[(I-G^\dagger G)(\lambda\sigma+\Sigma)] && (249)\\ &= \operatorname{Tr}[(I-G^\dagger G)(\lambda\sigma+\Sigma)] && (250)\\ &= \operatorname{Tr}[\lambda\sigma+\Sigma]\\ &\quad - \operatorname{Tr}\left[(\lambda\sigma+\Sigma)^{-1/2}(\lambda\sigma)(\lambda\sigma+\Sigma)^{-1/2}(\lambda\sigma+\Sigma)\right] && (251)\\ &= \operatorname{Tr}[\lambda\sigma+\Sigma]-\operatorname{Tr}[\lambda\sigma] && (252)\\ &= \operatorname{Tr}[\Sigma] && (253)\\ &= \delta && (254)\end{aligned}$$

This implies that

$$\operatorname{Tr}[G\rho G^\dagger] \ge 1-\delta. \tag{255}$$

Then it follows that

$$\begin{aligned} \widetilde{\rho} &= \frac{G\rho G^\dagger}{\operatorname{Tr}[G^\dagger G\rho]} && (256)\\ &\le \frac{G(\lambda\sigma+\Sigma)G^\dagger}{\operatorname{Tr}[G^\dagger G\rho]} && (257)\\ &= \frac{\lambda\sigma}{\operatorname{Tr}[G^\dagger G\rho]} && (258)\end{aligned}$$

$$\leq \frac{\lambda\sigma}{1-\delta}. \tag{259}$$

Let $\psi_{RA} = \sqrt{\rho_A}\Gamma_{RA}\sqrt{\rho_A}$ be the canonical purification of ρ_A, with $\Gamma_{RA} \coloneqq \sum_{i,j} |i\rangle\langle j|_R \otimes |i\rangle\langle j|_A$, and let $\widetilde{\psi}_{RA} = \frac{G_A\psi_{RA}G_A^\dagger}{\mathrm{Tr}[G^\dagger G\rho]}$ purify $\widetilde{\rho}$. Then

$$\begin{aligned}
\sqrt{F(\rho,\widetilde{\rho})} &\geq \frac{1}{\sqrt{\mathrm{Tr}[G^\dagger G\rho]}} |\langle\psi|_{RA} I_R \otimes G_A|\psi\rangle_{RA}| && (260)\\
&\geq |\langle\psi|_{RA} I_R \otimes G_A|\psi\rangle_{RA}| && (261)\\
&= |\langle\Gamma|_{RA} I_R \otimes \sqrt{\rho_A}G_A\sqrt{\rho_A}|\Gamma\rangle_{RA}| && (262)\\
&= |\mathrm{Tr}[G\rho]| && (263)\\
&\geq \mathrm{Re}[\mathrm{Tr}[G\rho]] && (264)\\
&= \mathrm{Tr}[\overline{G}\rho] && (265)\\
&= 1 - \mathrm{Tr}[\left(I-\overline{G}\right)\rho], && (266)
\end{aligned}$$

where

$$\overline{G} \coloneqq \frac{G+G^\dagger}{2}. \tag{267}$$

The first inequality follows from Uhlmann's theorem for fidelity, and the second follows because $\mathrm{Tr}[G^\dagger G\rho] \leq 1$. Observe that $\overline{G} \leq I$ because $\|G\|_\infty \leq 1$ and by applying the triangle inequality. So this means that $I - \overline{G} \geq 0$. Now consider that

$$\begin{aligned}
&\mathrm{Tr}[\left(I-\overline{G}\right)\rho]\\
&\leq \mathrm{Tr}[\left(I-\overline{G}\right)(\lambda\sigma+\Sigma)] && (268)\\
&= \mathrm{Tr}[\lambda\sigma+\Sigma] - \mathrm{Tr}[\overline{G}(\lambda\sigma+\Sigma)] && (269)\\
&= \mathrm{Tr}[\lambda\sigma+\Sigma]\\
&\quad - \frac{1}{2}\mathrm{Tr}\left[\left((\lambda\sigma)^{1/2}(\lambda\sigma+\Sigma)^{-1/2} + (\lambda\sigma+\Sigma)^{-1/2}(\lambda\sigma)^{1/2}\right)\times\right.\\
&\qquad \left.(\lambda\sigma+\Sigma)\right] && (270)\\
&= \mathrm{Tr}[\lambda\sigma+\Sigma] - \frac{1}{2}\mathrm{Tr}\left[(\lambda\sigma)^{1/2}(\lambda\sigma+\Sigma)^{-1/2}(\lambda\sigma+\Sigma)\right]\\
&\quad - \frac{1}{2}\mathrm{Tr}\left[(\lambda\sigma+\Sigma)^{-1/2}(\lambda\sigma)^{1/2}(\lambda\sigma+\Sigma)\right] && (271)\\
&= \mathrm{Tr}[\lambda\sigma+\Sigma] - \frac{1}{2}\mathrm{Tr}\left[(\lambda\sigma)^{1/2}(\lambda\sigma+\Sigma)^{1/2}\right]\\
&\quad - \frac{1}{2}\mathrm{Tr}\left[(\lambda\sigma)^{1/2}(\lambda\sigma+\Sigma)^{1/2}\right] && (272)\\
&= \mathrm{Tr}[\lambda\sigma+\Sigma] - \mathrm{Tr}\left[(\lambda\sigma)^{1/2}(\lambda\sigma+\Sigma)^{1/2}\right] && (273)\\
&\leq \mathrm{Tr}[\lambda\sigma+\Sigma] - \mathrm{Tr}\left[(\lambda\sigma)^{1/2}(\lambda\sigma)^{1/2}\right] && (274)\\
&= \mathrm{Tr}[\lambda\sigma+\Sigma] - \mathrm{Tr}[\lambda\sigma] && (275)\\
&= \mathrm{Tr}[\Sigma] && (276)\\
&= \delta. && (277)
\end{aligned}$$

So all of this implies that

$$\sqrt{\mathsf{F}(\rho,\widetilde{\rho})} \geq 1-\delta, \tag{278}$$

and in turn that

$$\mathsf{F}(\rho,\widetilde{\rho}) \geq (1-\delta)^2. \tag{279}$$

By applying the inequality

$$\frac{1}{2}\|\rho-\widetilde{\rho}\|_1 \leq \sqrt{1-\mathsf{F}(\rho,\widetilde{\rho})}, \tag{280}$$

we conclude that

$$\begin{aligned}
\frac{1}{2}\|\rho-\widetilde{\rho}\|_1 &\leq \sqrt{1-(1-\delta)^2} && (281)\\
&= \sqrt{1-(1-2\delta+\delta^2)} && (282)\\
&= \sqrt{2\delta-\delta^2} && (283)\\
&= \sqrt{\delta(2-\delta)}. && (284)
\end{aligned}$$

Putting everything together, we see that $\widetilde{\rho}$ is a quantum state satisfying

$$\begin{aligned}
\frac{1}{2}\|\rho-\widetilde{\rho}\|_1 &\leq \sqrt{\delta(2-\delta)}, && (285)\\
\widetilde{\rho} &\leq \frac{\lambda\sigma}{1-\delta}. && (286)
\end{aligned}$$

This means that $\widetilde{\rho}$ and $\frac{\lambda}{1-\delta}$ are feasible for $\mathsf{D}_{\max}^{\sqrt{\delta(2-\delta)}}(\rho\|\sigma)$, and so it follows that

$$\begin{aligned}
\mathsf{D}_{\max}^{\sqrt{\delta(2-\delta)}}(\rho\|\sigma) &\leq \ln\left(\frac{\lambda}{1-\delta}\right) && (287)\\
&= \ln\lambda + \ln\left(\frac{1}{1-\delta}\right). && (288)
\end{aligned}$$

This concludes the proof.

Appendix III
Subadditivity of Smooth Max-Relative Entropy

Lemma 6: Given $\delta_1, \delta_2 \in [0,1]$ such that $\delta_1 + \delta_2 \leq 1$, states ρ_1 and ρ_2, and PSD operators σ_1 and σ_2, the following subadditivity relation holds

$$\mathsf{D}_{\max}^{\delta_1+\delta_2}(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2) \leq \mathsf{D}_{\max}^{\delta_1}(\rho_1\|\sigma_1) + \mathsf{D}_{\max}^{\delta_2}(\rho_2\|\sigma_2). \tag{289}$$

Proof: Let $\overline{\rho}_i$ and λ_i be optimal choices for $D_{\max}^{\delta_i}(\rho_i\|\sigma_i)$, for $i \in \{1,2\}$. Then, consider that

$$\overline{\rho}_1\otimes\overline{\rho}_2 \leq \lambda_1\sigma_1\otimes\overline{\rho}_2 \leq \lambda_1\sigma_1\otimes\lambda_2\sigma_2 = \lambda_1\lambda_2\sigma_1\otimes\sigma_2. \tag{290}$$

Furthermore, consider that

$$\begin{aligned}
&\frac{1}{2}\|\overline{\rho}_1\otimes\overline{\rho}_2 - \rho_1\otimes\rho_2\|_1\\
&= \frac{1}{2}\|\overline{\rho}_1\otimes\overline{\rho}_2 - \overline{\rho}_1\otimes\rho_2 + \overline{\rho}_1\otimes\rho_2 - \rho_1\otimes\rho_2\|_1 && (291)\\
&= \frac{1}{2}\|\overline{\rho}_1\otimes(\overline{\rho}_2-\rho_2) + (\overline{\rho}_1-\rho_1)\otimes\rho_2\|_1 && (292)\\
&\leq \frac{1}{2}\|\overline{\rho}_1\otimes(\overline{\rho}_2-\rho_2)\|_1 + \frac{1}{2}\|(\overline{\rho}_1-\rho_1)\otimes\rho_2\|_1 && (293)\\
&= \frac{1}{2}\|\overline{\rho}_1\|_1\|\overline{\rho}_2-\rho_2\|_1 + \frac{1}{2}\|\overline{\rho}_1-\rho_1\|_1\|\rho_2\|_1 && (294)\\
&= \frac{1}{2}\|\overline{\rho}_2-\rho_2\|_1 + \frac{1}{2}\|\overline{\rho}_1-\rho_1\|_1 && (295)\\
&\leq \delta_1+\delta_2, && (296)
\end{aligned}$$

where (291) follows from the triangular inequality for the trace norm and the final inequality from the assumption that $\overline{\rho}_i$ are the optimizers for $D_{\max}^{\delta_i}(\rho_i\|\sigma_i)$, for $i \in \{1,2\}$.

Finally we have shown that $\overline{\rho}_1\otimes\overline{\rho}_2$ and $\lambda_1\lambda_2$ are candidates for the optimization for $D_{\max}^{\delta_1+\delta_2}(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2)$, thus concluding the proof. ∎

APPENDIX IV
PROOF OF THEOREM 1

For (1): Fix $(\rho^{\mathcal{R}}, \rho^{\mathcal{T}})$ in (23), $\mathrm{M} \in \mathcal{M}$. Consider that

$$\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{R}})\big] = \mathrm{Tr}\left[\mathrm{M}\sum_{i=1}^{k} p_i \mathcal{A}_i(\rho^{\mathcal{R}})\right] \tag{297}$$

$$= \sum_{i=1}^{k} p_i \mathrm{Tr}\big[\mathrm{M}\mathcal{A}_i(\rho^{\mathcal{R}})\big] \tag{298}$$

$$\overset{(a)}{\leq} \sum_{i=1}^{k} p_i \left(e^{\varepsilon}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}_i(\rho^{\mathcal{T}})\big] + \delta\right) \tag{299}$$

$$= \sum_{i=1}^{k} p_i e^{\varepsilon}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}_i(\rho^{\mathcal{T}})\big] + \delta \tag{300}$$

$$= e^{\varepsilon}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] + \delta, \tag{301}$$

where (a) follows due to each $\mathcal{A}_i$ being (ε, δ)-QPP. This inequality holds for every $(\rho^{\mathcal{R}}, \rho^{\mathcal{T}})$, and so it holds for all such pairs generated from $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$.

For (2): Fix $0 \leq \mathrm{M}' \leq \mathrm{I}$ such that $\mathrm{M}' \in \mathcal{M}'$ as stated in the property. With that assumption, there exists $\mathrm{M} \in \mathcal{M}$ such that $\mathrm{M} = \mathcal{N}^{\dagger}(\mathrm{M}')$. Consider that

$$\mathrm{Tr}\big[\mathrm{M}'\mathcal{N}\big(\mathcal{A}(\rho^{\mathcal{R}})\big)\big] = \mathrm{Tr}\big[\mathcal{N}^{\dagger}(\mathrm{M}')\mathcal{A}(\rho^{\mathcal{R}})\big] \tag{302}$$

$$= \mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{R}})\big], \tag{303}$$

where $\mathcal{N}^{\dagger}$ is the adjoint of $\mathcal{N}$, implying that

$$0 \leq \mathcal{N}^{\dagger}(\mathrm{M}') = \mathrm{M} \leq \mathrm{I} \tag{304}$$

because $\mathcal{N}^{\dagger}$ is positive and unital by the assumption that $\mathcal{N}$ is a quantum channel. Similarly, we have that $\mathrm{Tr}\big[\mathrm{M}'\mathcal{N}\left(\mathcal{A}(\rho^{\mathcal{T}})\right)\big] = \mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big]$, and we conclude that the processed mechanism satisfies (ε, δ)-QPP with the choice of $\mathcal{M}' \subseteq \{\mathrm{M}' : \mathcal{N}^{\dagger}(\mathrm{M}') \in \mathcal{M}\}$.

For (3): Fix $(\mathcal{R}_i, \mathcal{T}_i) \in \mathcal{Q}$ for $i \in \{1, \ldots, k\}$, and $\bigotimes_{i=1}^{k} \mathrm{M}_i \in \bigotimes_{i=1}^{k} \mathcal{M}_i$. Denote

$$\mathcal{A}^{(k)}(\rho^{\mathcal{R}^{(k)}}) \coloneqq \mathcal{A}_1(\rho^{\mathcal{R}_1}) \otimes \mathcal{A}_2(\rho^{\mathcal{R}_2}) \otimes \ldots \mathcal{A}_k(\rho^{\mathcal{R}_k}) \tag{305}$$

and $\mathcal{A}^{(k)}(\rho^{\mathcal{T}^{(k)}})$ similarly by replacing $\mathcal{R}$ with $\mathcal{T}$.

Fix $i \in \{1, \ldots, k\}$. Consider that $\mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big] \leq 1 \leq 1 + \delta_i$ because $\mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big]$ is a probability. Combining with the inequality $\mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big] \leq e^{\varepsilon_i}\mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{T}_i})\big] + \delta_i$, which holds from the assumption that QPP holds, we conclude that

$$\mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big] \leq \min\left\{e^{\varepsilon_i}\mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{T}_i})\big], 1\right\} + \delta_i. \tag{306}$$

Consider that

$$\prod_{i=1}^{k} \mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big]$$

$$\leq \left(\min\left\{1, e^{\varepsilon_1}\mathrm{Tr}\big[\mathrm{M}_1\mathcal{A}_1(\rho^{\mathcal{R}_1})\big]\right\} + \delta_1\right)\prod_{i=2}^{k} \mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big] \tag{307}$$

$$\leq \min\left\{1, e^{\varepsilon_1}\mathrm{Tr}\big[\mathrm{M}_1\mathcal{A}_1(\rho^{\mathcal{R}_1})\big]\right\}\prod_{i=2}^{k} \mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big] + \delta_1 \tag{308}$$

$$\leq \min\left\{1, e^{\varepsilon_1}\mathrm{Tr}\big[\mathrm{M}_1\mathcal{A}_1(\rho^{\mathcal{R}_1})\big]\right\} \times \left(\min\left\{1, e^{\varepsilon_2}\mathrm{Tr}\big[\mathrm{M}_2\mathcal{A}_2(\rho^{\mathcal{T}_2})\big]\right\} + \delta_2\right)\prod_{i=3}^{k} \mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big] + \delta_1 \tag{309}$$

$$\leq \prod_{j=1}^{2} \min\left\{e^{\varepsilon_j}\mathrm{Tr}\big[\mathrm{M}_j\mathcal{A}_j(\rho^{\mathcal{T}_j})\big], 1\right\}\prod_{i=3}^{k} \mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{R}_i})\big] + \delta_1 + \delta_2 \tag{310}$$

$$\leq e^{\sum_{i=1}^{k}\varepsilon_i}\prod_{i=1}^{k} \mathrm{Tr}\big[\mathrm{M}_i\mathcal{A}_i(\rho^{\mathcal{T}_i})\big] + \sum_{i=1}^{k}\delta_i, \tag{311}$$

where the last inequality follows by proceeding with similar expansions for each remaining term of the product as carried out in the first three steps.

APPENDIX V
PROOF OF PROPOSITION 3

Fix $\mathrm{M}', \mathrm{M} \in \bar{\mathcal{M}}$, $P_X \in \Theta$, and $(\mathcal{R}_1, \mathcal{T}_1), (\mathcal{R}_2, \mathcal{T}_2) \in \mathcal{Q}$. Let $\mathrm{M}_y \coloneqq (\langle y| \otimes \mathrm{I})\mathrm{M}'(|y\rangle \otimes \mathrm{I})$ and note that M_y is a measurement operator in $\bar{\mathcal{M}}$. Recall the definition of the channel

$$\overline{\mathcal{E}} \coloneqq \sum_{y \in \mathcal{Y}} \mathcal{E}^y. \tag{312}$$

Consider that

$$\mathrm{Tr}\left[(\mathrm{M} \otimes \mathrm{M}')\left(\sum_{y \in \mathcal{Y}} \mathcal{E}^y(\mathcal{A}_1(\rho^{\mathcal{R}_1})) \otimes |y\rangle\langle y| \otimes \mathcal{A}_2^y(\rho^{\mathcal{R}_2})\right)\right]$$

$$= \sum_{y \in \mathcal{Y}} \mathrm{Tr}\big[\mathrm{M}\mathcal{E}^y(\mathcal{A}_1(\rho^{\mathcal{R}_1})) \otimes \mathrm{M}'\big(|y\rangle\langle y| \otimes \mathcal{A}_2^y(\rho^{\mathcal{R}_2})\big)\big] \tag{313}$$

$$= \sum_{y \in \mathcal{Y}} \mathrm{Tr}[\mathrm{M}\mathcal{E}^y(\mathcal{A}_1(\rho^{\mathcal{R}_1}))]\,\mathrm{Tr}\big[\mathrm{M}'\big(|y\rangle\langle y| \otimes \mathcal{A}_2^y(\rho^{\mathcal{R}_2})\big)\big] \tag{314}$$

$$= \sum_{y \in \mathcal{Y}} \mathrm{Tr}[\mathrm{M}\mathcal{E}^y(\mathcal{A}_1(\rho^{\mathcal{R}_1}))]\,\mathrm{Tr}[\mathrm{M}'_y\mathcal{A}_2^y(\rho^{\mathcal{R}_2})] \tag{315}$$

$$\overset{(a)}{\leq} \sum_{y \in \mathcal{Y}} \mathrm{Tr}[\mathrm{M}\mathcal{E}^y(\mathcal{A}_1(\rho^{\mathcal{R}_1}))] \times \left(\min\left\{1, e^{\varepsilon_2}\,\mathrm{Tr}[\mathrm{M}'_y\mathcal{A}_2^y(\rho^{\mathcal{T}_2})]\right\} + \delta_2\right) \tag{316}$$

$$= \sum_{y \in \mathcal{Y}} \mathrm{Tr}[\mathcal{E}^{y\dagger}(\mathrm{M})\mathcal{A}_1(\rho^{\mathcal{R}_1})] \times \left(\min\left\{1, e^{\varepsilon_2}\,\mathrm{Tr}[\mathrm{M}'_y\mathcal{A}_2^y(\rho^{\mathcal{T}_2})]\right\} + \delta_2\right) \tag{317}$$

$$\overset{(b)}{=} \mathrm{Tr}[\mathrm{M}\overline{\mathcal{E}}(\mathcal{A}_1(\rho^{\mathcal{R}_1}))]\delta_2 + \sum_{y \in \mathcal{Y}} \left(\mathrm{Tr}[\mathcal{E}^{y\dagger}(\mathrm{M})\mathcal{A}_1(\rho^{\mathcal{R}_1})]\min\left\{1, e^{\varepsilon_2}\,\mathrm{Tr}[\mathrm{M}'_y\mathcal{A}_2^y(\rho^{\mathcal{T}_2})]\right\}\right) \tag{318}$$

$$\overset{(c)}{\leq} \delta_2 + \sum_{y \in \mathcal{Y}} \mathrm{Tr}[\mathcal{E}^{y\dagger}(\mathrm{M})\mathcal{A}_1(\rho^{\mathcal{R}_1})]\min\left\{1, e^{\varepsilon_2}\,\mathrm{Tr}[\mathrm{M}'_y\mathcal{A}_2^y(\rho^{\mathcal{T}_2})]\right\} \tag{319}$$

$$\overset{(d)}{\leq} \delta_2 + \sum_{y\in\mathcal{Y}} \left(e^{\varepsilon_1} \operatorname{Tr}[\mathcal{E}^{y\dagger}(\mathrm{M})\mathcal{A}_1(\rho^{\mathcal{T}_1})] + \delta_1\right) \times \min\left\{1, e^{\varepsilon_2} \operatorname{Tr}[\mathrm{M}'_y \mathcal{A}_2^y(\rho^{\mathcal{T}_2})]\right\} \tag{320}$$

$$\leq \sum_{y\in\mathcal{Y}} \left(e^{\varepsilon_1} \operatorname{Tr}[\mathcal{E}^{y\dagger}(\mathrm{M})\mathcal{A}_1(\rho^{\mathcal{T}_1})] e^{\varepsilon_2} \operatorname{Tr}[\mathrm{M}'_y \mathcal{A}_2^y(\rho^{\mathcal{T}_2})] + \delta_1\right) + \delta_2 \tag{321}$$

$$= \sum_{y\in\mathcal{Y}} \left(e^{\varepsilon_2} e^{\varepsilon_1} \operatorname{Tr}[\mathcal{E}^{y\dagger}(\mathrm{M})\mathcal{A}_1(\rho^{\mathcal{T}_1})] \operatorname{Tr}[\mathrm{M}'_y \mathcal{A}_2^y(\rho^{\mathcal{T}_2})] + \delta_1\right) + \delta_2 \tag{322}$$

$$= e^{\varepsilon'} \times \operatorname{Tr}\left[(\mathrm{M}\otimes\mathrm{M}')\left(\sum_{y\in\mathcal{Y}} \mathcal{E}^y(\mathcal{A}_1(\rho^{\mathcal{T}_1})) \otimes |y\rangle\langle y| \otimes \mathcal{A}_2^y(\rho^{\mathcal{T}_2})\right)\right] + \delta_2 + \delta_1 |\mathcal{Y}|, \tag{323}$$

where: (a) from $\operatorname{Tr}[\mathrm{M}'_y \mathcal{A}_2^y(\rho^{\mathcal{R}_2})] \leq 1 \leq 1 + \delta_2$ and $\mathcal{A}_2^y$ being $(\varepsilon_2, \delta_2)$-QPP; (b) and (c) from

$$\operatorname{Tr}\left[\sum_{y\in\mathcal{Y}} \mathcal{E}^{y\dagger}(\mathrm{M})\mathcal{A}_1(\rho^{\mathcal{T}_1})\right] = \operatorname{Tr}\left[\mathrm{M}\sum_{y\in\mathcal{Y}} \mathcal{E}^y \mathcal{A}_1(\rho^{\mathcal{T}_1})\right] \tag{324}$$

$$= \operatorname{Tr}[\mathrm{M}\bar{\mathcal{E}}(\mathcal{A}_1(\rho^{\mathcal{T}_1}))] \tag{325}$$

$$\leq 1; \tag{326}$$

and (d) from $\mathcal{A}_1$ being $(\varepsilon_1, \delta_1)$-QPP and the fact that $\mathcal{E}^{y\dagger}(\mathrm{M})$ is a measurement operator in $\bar{\mathcal{M}}$.

Appendix VI
Composability With Classically Correlated States

In Property 3 of Theorem 1 and Proposition 3, we considered the case in which two mechanisms, composed either in parallel or adaptively, receive independent inputs (i.e., the input being $\rho^{X_1} \otimes \rho^{X_2}$ where $X_i \sim P_X \in \Theta$ for $i = \{1, 2\}$, which are chosen independently). We now focus on the setting in which the inputs are classically correlated. The input is chosen as a separable state of the form

$$\sigma_I \coloneqq \sum_{z\in\mathcal{Z}} q(z)\, \omega^z \otimes \tau^z, \tag{327}$$

where q represents a probability distribution with $q(z) \geq 0$ and $\sum_{z\in\mathcal{Z}} q(z) = 1$, and ω^z and τ^z are quantum states for all $z \in \mathcal{Z}$.[11] One special case of interest is as follows:

$$\sigma_I \coloneqq \sum_{x\in\mathcal{X}} P_X(x)\, \rho^x \otimes \rho^x. \tag{328}$$

In this setting, QDP ensures indistinguishability of the input states

$$\sigma_I^1 \coloneqq \sum_{z\in\mathcal{Z}} q(z)\, \omega_1^z \otimes \tau_1^z \quad \text{and} \quad \sigma_I^2 \coloneqq \sum_{z\in\mathcal{Z}} q(z)\, \omega_2^z \otimes \tau_2^z, \tag{329}$$

where $\omega_1^z \sim \omega_2^z$ and $\tau_1^z \sim \tau_2^z$ are neighbors for all $z \in \mathcal{Z}$.

[11] Note that (327) covers the case of having input states of the form $\sigma_I \coloneqq \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} q(x,y)\, \omega^x \otimes \tau^y$ where for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ ω^x and τ^y are states, by considering z to be an index for multiple variables, i.e., setting $z = (x, y)$.

We consider an instance of the QPP framework, called flexible QDP, where $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M})$ is such that Θ and $\mathcal{M}$ are chosen based on user needs, while the other parameters are as given in Remark III-C1. Flexible QDP then satisfies the following composability properties.

Corollary 4 (Composability of Flexible QDP): Let the initial input to the two mechanisms $\mathcal{A}_1$ and $\mathcal{A}_2$ be of the form $\sum_{z\in\mathcal{Z}} q(z)\, \omega^z \otimes \tau^z$. The following composability properties hold for the QDP framework.

Parallel Composability: Consider the parallel composed mechanism $\sum_{z\in\mathcal{Z}} q(z)\, \mathcal{A}_1(\omega^z) \otimes \mathcal{A}_2(\tau^z)$.

1) If $\mathcal{A}_i$ is $(\varepsilon_i, \delta_i)$-QDP in the framework $(\mathcal{S}, \mathcal{Q}, \Theta, \mathcal{M}_i)$, for $i \in \{1, 2\}$, then the composed mechanism satisfies $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-QDP in $\left(\mathcal{S}, \mathcal{Q}^{(2)}, \Theta, \bigotimes_{i=1}^2 \mathcal{M}_i\right)$
2) If $\mathcal{A}_i$ is $(\varepsilon_i, \delta_i)$-QDP in the framework $(\mathcal{S}, \mathcal{Q}, \Theta, \bar{\mathcal{M}})$, for $i \in \{1, 2\}$, then the composed mechanism satisfies (ε', δ')-QDP in $\left(\mathcal{S}, \mathcal{Q}^{(2)}, \Theta, \bar{\mathcal{M}}^2\right)$ with

$$\varepsilon' \coloneqq \varepsilon_1 + \varepsilon_2 + \ln\left(\frac{1}{(1-\delta_1)(1-\delta_2)}\right), \tag{330}$$

$$\delta' \coloneqq \sqrt{\delta_1(2-\delta_1)} + \sqrt{\delta_2(2-\delta_2)}. \tag{331}$$

and also satisfies $(\varepsilon_1 + \varepsilon_2, \delta)$ in the same framework with $\delta \coloneqq \min\{\delta_1 + e^{\varepsilon_1}\delta_2, \delta_2 + e^{\varepsilon_2}\delta_1\}$.

Adaptive Composability: Suppose that $\mathcal{A}_1$ satisfies $(\varepsilon_1, \delta_1)$-QDP and $\mathcal{A}_2$ chosen adaptively satisfies $(\varepsilon_2, \delta_2)$-QDP, as in (108). Then, the composed mechanism in Fig. 3 with σ_I in (327) satisfies $(\varepsilon_1 + \varepsilon_2, \delta_2 + \delta_1|\mathcal{Y}|)$ in the framework $(\mathcal{S}, \mathcal{Q} \times \mathcal{Q}, \Theta, \bar{\mathcal{M}} \otimes \bar{\mathcal{M}})$.

Proof: Item 1 in the parallel composability part follows by a similar argument as given in the proof of Property 3 from Theorem 1. For the proof of Item 2, first, we use quasi-convexity of the DL divergence (property 2 in Proposition 2) and then adapt Item 3 of Corollary 2. The adaptive composition result follows along the same lines as the proof of Proposition 3 for fixed z, and then averaging over all $z \in \mathcal{Z}$ gives the desired result. ∎

Remark 26 (Extensions Beyond Flexible QDP): Corollary 4 does not hold for the general QPP framework. Indeed, it fails to hold, for instance, for the classical PP framework [3, Theorem 9.1]. Nevertheless, Corollary 4 can be extended to account for input states $\sum_{x\in\mathcal{X}} P_X(x)\, \rho^x \otimes \rho^x$ subjected to additional structural assumptions on the class of admissible distributions:

$$\Theta \subseteq \left\{ P_X \in \mathcal{P}(\mathcal{X}) : \begin{array}{c} \forall\ (\mathcal{R}, \mathcal{T}) \in \mathcal{Q},\ \exists\ x, x' \in \mathcal{X} \\ \text{s.t. } q_{\mathcal{R}}(x) = q_{\mathcal{T}}(x') = 1 \end{array} \right\} \tag{332}$$

where $q_{\mathcal{R}}$ and $q_{\mathcal{T}}$ are defined as in Definition 4. The classical version of this condition for PP is known as *"universally composable scenarios"* [3, Corollary 9.4].

Appendix VII
Characterizing Optimal Privacy-Utility Tradeoff

In this Appendix, we focus on identifying the optimal utility that can be obtained by applying an (ε, δ)-QPP mechanism. Here, we first focus on the setting in which

$\mathcal{Q} = \{(\mathcal{R}_1, \mathcal{R}_2), (\mathcal{R}_2, \mathcal{R}_1)\}$, $\bar{\mathcal{M}} = \{\mathrm{M} : 0 \leq \mathrm{M} \leq \mathrm{I}\}$, and $\Theta = \{P_X\}$, but the following ideas can be extended to the case when $\mathcal{Q}$ is an arbitrary finite set and Θ includes a finite number of probability distributions. However, the computational complexity involved in identifying the optimal utility increases with the cardinality of the set $\mathcal{Q}$ and Θ, due to the addition of more constraints to the optimization problem.

To incorporate privacy requirements, we use the equivalent formulation of QPP via the DL divergence presented in Proposition 1. To this end, first, we employ the SDP formulated in Lemma 1 to compute the relevant DL divergence and then use that in the optimization of utility. We showcase the use of this SDP in characterizing optimal utility next.

Proposition 14 (Optimal Utility for Fixed Privacy Constraints): The optimal utility, as quantified by the γ-utility metric, for every privacy mechanism that is (ε, δ)-QPP in the $(\mathcal{S}, \mathcal{Q}, \Theta, \bar{\mathcal{M}})$ framework, where $\mathcal{Q} = \{(\mathcal{R}_1, \mathcal{R}_2), (\mathcal{R}_2, \mathcal{R}_1)\}$, is given by the following:

$$\mathsf{U}(\varepsilon, \delta, \mathcal{R}_1, \mathcal{R}_2) \coloneqq 1 - \inf_{\substack{\mu \geq 0 \\ Z_{AD} \geq 0 \\ \Gamma^{\mathcal{B}}_{CD} \geq 0 \\ \Gamma^{\mathcal{A}}_{AC} \geq 0 \\ \lambda_1 \geq 0, Y_1 \geq 0 \\ \lambda_2 \geq 0, Y_2 \geq 0}} \left\{ \begin{array}{c} \mu : \\ Z_{AD} \geq \Gamma_{AD} - \mathrm{Tr}_C\left[\Gamma^{\mathcal{B}}_{CD} \mathrm{T}_C(\Gamma^{\mathcal{A}}_{AC})\right], \\ \mu I_A \geq \mathrm{Tr}_D[Z_{AD}], \\ \mathrm{Tr}_D\left[\Gamma^{\mathcal{B}}_{CD}\right] = I_C, \\ \mathrm{Tr}_D\left[\Gamma^{\mathcal{A}}_{AC}\right] = I_A, \\ \ln(\lambda_1) \leq \varepsilon, \\ \mathrm{Tr}[Y_1] \leq \delta, \\ Y_1 \geq \mathrm{Tr}_A\left[\left(\mathrm{T}(\rho^{\mathcal{R}_1}) \otimes I_C\right)\Gamma^{\mathcal{A}}_{AC}\right] \\ -\lambda_1 \mathrm{Tr}_A\left[\left(\mathrm{T}(\rho^{\mathcal{R}_2}) \otimes I_C\right)\Gamma^{\mathcal{A}}_{AC}\right], \\ \ln(\lambda_2) \leq \varepsilon, \\ \mathrm{Tr}[Y_2] \leq \delta, \\ Y_2 \geq \mathrm{Tr}_A\left[\left(\mathrm{T}(\rho^{\mathcal{R}_2}) \otimes I_C\right)\Gamma^{\mathcal{A}}_{AC}\right] \\ -\lambda_2 \mathrm{Tr}_A\left[\left(\mathrm{T}(\rho^{\mathcal{R}_1}) \otimes I_C\right)\Gamma^{\mathcal{A}}_{AC}\right] \end{array} \right\}. \tag{333}$$

Proof: The proof follows from the SDP formulation of the γ-utility given in Proposition 6, and the privacy constraints (i.e., $\max\{\overline{\mathsf{D}}^{\delta}\left(\mathcal{A}(\rho^{\mathcal{R}_1}) \| \mathcal{A}(\rho^{\mathcal{R}_2})\right), \overline{\mathsf{D}}^{\delta}\left(\mathcal{A}(\rho^{\mathcal{R}_2}) \| \mathcal{A}(\rho^{\mathcal{R}_1})\right)\} \leq \varepsilon$) imposed through the SDP formulation of DL divergence presented in Lemma 1. We also used the fact that for a superoperator $\mathcal{A}$ from system A to C, the following equality holds [37, Eq. (3.2.14)]

$$\mathcal{A}(\rho^{\mathcal{R}_1}) = \mathrm{Tr}_A\left[\left(\mathrm{T}(\rho^{\mathcal{R}_1}) \otimes I_C\right)\Gamma^{\mathcal{A}}_{AC}\right]. \tag{334}$$

■

Remark 27 (Privacy Constraints via Equivalent Formulation Through Hockey-Stick Divergence): Instead of using the DL divergence, we can also encode the privacy constraints through the equivalent formulation in Remark 6. To this end, the dual formulation of the hockey-stick divergence (can be obtained by (47)), as

$$\mathsf{E}_{\lambda}(\rho \| \sigma) = \inf_{Z \geq 0} \left\{\mathrm{Tr}[Z] : Z \geq \rho - \lambda\sigma\right\}, \tag{335}$$

can also be incorporated to compute the optimum utility.

Remark 28 (Optimal Privacy Parameters for Fixed Utility): To find out the optimal (minimal) privacy parameter $\varepsilon^{\star}$ for a given mechanism $\mathcal{A}$ with the utility constraint γ, and fixed tolerance δ, first we compute the following quantity:

$$\lambda_1^{\star}(\mathcal{A}, \gamma, \delta) \coloneqq \inf_{\substack{\lambda \geq 0 \\ Z_{AD} \geq 0 \\ \Gamma^{\mathcal{B}}_{CD} \geq 0 \\ Y_1 \geq 0}} \left\{ \begin{array}{c} \lambda : \\ Z_{AD} \geq \Gamma_{AD} - \mathrm{Tr}_C\left[\Gamma^{\mathcal{B}}_{CD} \mathrm{T}_C(\Gamma^{\mathcal{A}}_{AC})\right], \\ (1-\gamma) I_A \geq \mathrm{Tr}_D[Z_{AD}], \\ \mathrm{Tr}_D\left[\Gamma^{\mathcal{B}}_{CD}\right] = I_C, \\ \mathrm{Tr}[Y_1] \leq \delta, \\ Y_1 \geq \mathrm{Tr}_A\left[\left(\mathrm{T}(\rho^{\mathcal{R}_1}) \otimes I_C\right)\Gamma^{\mathcal{A}}_{AC}\right] \\ -\lambda \mathrm{Tr}_A\left[\left(\mathrm{T}(\rho^{\mathcal{R}_2}) \otimes I_C\right)\Gamma^{\mathcal{A}}_{AC}\right] \end{array} \right\}. \tag{336}$$

Similarly $\lambda_2^{\star}$ can be obtained by exchanging $\rho^{\mathcal{R}_1}$ and $\rho^{\mathcal{R}_2}$. Then the optimal value is given by

$$\varepsilon^{\star}(\mathcal{A}, \gamma, \delta) \coloneqq \ln(\max\{\lambda_1^{\star}(\mathcal{A}, \gamma, \delta), \lambda_2^{\star}(\mathcal{A}, \gamma, \delta)\}). \tag{337}$$

The optimal (minimal) δ for a fixed ε with a utility constraint can be obtained by encoding the privacy constraint through the dual form of hockey-stick divergence, as given in Remark 27.

Appendix VIII
Proof of Lemma 4

The proof follows analogously to the classical version of this bound in [82, Proposition 3.3], along with the upper bound for an arbitrary $\mathsf{D}_{\alpha}(\cdot\|\cdot)$ satisfying data processing. Set $\alpha > 1$. Then, for such $\mathsf{D}_{\alpha}(\cdot\|\cdot)$, from [79, Equation 4.34], which is obtained by choosing a specific preparation channel, we have that

$$\mathsf{D}_{\alpha}(\rho\|\sigma) \leq \frac{1}{\alpha - 1} \log \mathrm{Tr}\left[\sigma^{1/2}\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)^{\alpha}\sigma^{1/2}\right]. \tag{338}$$

Let us use the following substitution:

$$\mathsf{D}_T(\rho\|\sigma) = \varepsilon. \tag{339}$$

Then we have $\mathsf{D}_{\max}(\rho\|\sigma) \leq \varepsilon$ and $\mathsf{D}_{\max}(\sigma\|\rho) \leq \varepsilon$. Moreover, with the definition of $\mathsf{D}_{\max}(\cdot\|\cdot)$ in (13), we have $\rho \leq e^{\varepsilon}\sigma$ and $\sigma \leq e^{\varepsilon}\rho$. Then we find that

$$e^{-\varepsilon}\mathrm{I} \leq \sigma^{-1/2}\rho\sigma^{-1/2} \leq e^{\varepsilon}\mathrm{I}. \tag{340}$$

Suppose that $\sigma^{-1/2}\rho\sigma^{-1/2}$ has the following spectral decomposition $\sum_i t_i |\phi_i\rangle\langle\phi_i|$. Then $e^{-\varepsilon} \leq t_i \leq e^{\varepsilon}$, and so for all i, $\exists \lambda_i \in [0,1]$ such that

$$t_i = \lambda_i e^{\varepsilon} + (1 - \lambda_i)e^{-\varepsilon}. \tag{341}$$

Consider that

$$\begin{aligned} & e^{(\alpha-1)\mathsf{D}_{\alpha}(\rho\|\sigma)} \\ & \leq \mathrm{Tr}\left[\sigma^{1/2}\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)^{\alpha}\sigma^{1/2}\right] && (342) \\ & = \mathrm{Tr}\left[\sigma^{1/2}\sum_i \left(\lambda_i e^{\varepsilon} + (1-\lambda_i)e^{-\varepsilon}\right)^{\alpha} |\phi_i\rangle\langle\phi_i| \, \sigma^{1/2}\right] && (343) \\ & = \sum_i \left(\lambda_i e^{\varepsilon} + (1-\lambda_i)e^{-\varepsilon}\right)^{\alpha} \mathrm{Tr}[\sigma|\phi_i\rangle\langle\phi_i|] && (344) \\ & \leq \sum_i \left(\lambda_i e^{\varepsilon\alpha} + (1-\lambda_i)e^{-\varepsilon\alpha}\right) \mathrm{Tr}[\sigma|\phi_i\rangle\langle\phi_i|] && (345) \\ & = e^{\varepsilon\alpha}c_1 + e^{-\varepsilon\alpha}c_2 && (346) \end{aligned}$$

where the first inequality follows from the inequality in (338), the second from the convexity of the function $x \mapsto x^\alpha$ for $\alpha > 1$, and the definitions

$$c_1 := \sum_i \lambda_i \mathrm{Tr}[\sigma|\phi_i\rangle\langle\phi_i|], \tag{347}$$

$$c_2 := \sum_i (1-\lambda_i)\mathrm{Tr}[\sigma|\phi_i\rangle\langle\phi_i|]. \tag{348}$$

In (346), we arrive at a function of α (i.e., $e^{\varepsilon\alpha}c_1 + e^{-\varepsilon\alpha}c_2$). Observing that $c_1+c_2 = 1$, we can find c_1 and c_2 by evaluating this function of α at $\alpha = 1$, which turns out to be equal to one because

$$\sum_i t_i \mathrm{Tr}[\sigma|\phi_i\rangle\langle\phi_i|] = \mathrm{Tr}[\rho] = 1, \tag{349}$$

where t_i is given in (341). Proceeding with this we get $c_1 = \frac{1-e^{-\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}}$. Then collecting all these relations and simplifying we obtain,

$$e^{(\alpha-1)\mathsf{D}_\alpha(\rho\|\sigma)} \leq \frac{\sinh(\alpha\varepsilon) - \sinh((\alpha-1)\varepsilon)}{\sinh(\varepsilon)}. \tag{350}$$

Together with [82, Lemma B.1], and the assumption that $\alpha\varepsilon \leq 2$, we can further bound (350) from above by $e^{\alpha(\alpha-1)\varepsilon^2/2}$. With the substitution in (339) we conclude the proof.

Appendix IX
Proof of Lemma 5

Fix $P_X \in \Theta$ and $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$. By assumption, we have that

$$\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{R}})\big] \leq e^{\varepsilon}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] + \delta. \tag{351}$$

With the choice for δ' as in the Lemma statement (i.e., (201)), we have $\delta - \delta' = (1-\delta)(e^{\varepsilon'} - e^{\varepsilon})/(e^{\varepsilon}+1)$. Plugging this in, we find that

$$\begin{aligned}
&\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{R}})\big] \\
&\leq e^{\varepsilon'}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] + \delta' + (\delta - \delta') \\
&\quad + (e^{\varepsilon} - e^{\varepsilon'})\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] \qquad (352)\\
&= e^{\varepsilon'}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] + \delta' \\
&\quad + (e^{\varepsilon} - e^{\varepsilon'})\left(\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] - \frac{1-\delta}{e^{\varepsilon}+1}\right). \qquad (353)
\end{aligned}$$

Since $\varepsilon' \leq \varepsilon$, we get the desired inequality if $\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] \leq \frac{1-\delta}{e^{\varepsilon}+1}$.

By choosing the measurement operator $\mathrm{I} - \mathrm{M}$, we also have by assumption that

$$\mathrm{Tr}\big[(\mathrm{I}-\mathrm{M})\mathcal{A}(\rho^{\mathcal{T}})\big] \leq e^{\varepsilon}\mathrm{Tr}\big[(\mathrm{I}-\mathrm{M})\mathcal{A}(\rho^{\mathcal{R}})\big] + \delta. \tag{354}$$

Rewriting (354), we arrive at

$$\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{R}})\big] \leq 1 - e^{-\varepsilon}(1-\delta) + e^{-\varepsilon}\ \mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big]. \tag{355}$$

Similar to the previous manipulations, we get

$$\begin{aligned}
&\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{R}})\big] \leq e^{\varepsilon'}\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] + \delta' \\
&\quad + (e^{\varepsilon'} - e^{-\varepsilon})\left(-\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] + \frac{1-\delta}{e^{\varepsilon}+1}\right). \qquad (356)
\end{aligned}$$

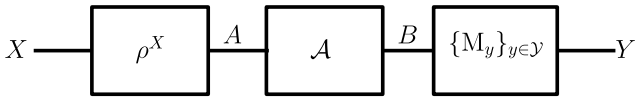


Fig. 9. Setup relevant to Proposition 15: X is a classical random variable that determines ρ^X, which is the input into the channel $\mathcal{A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$. Then Y is a random variable describing the outcome after applying the POVM $\{\mathrm{M}_y\}_{y\in\mathcal{Y}}$. Note that here the classical systems related to X, Y are also given by the same system labels X, Y.

Since $\varepsilon' \leq \varepsilon$, we arrive at the desired inequality when $\mathrm{Tr}\big[\mathrm{M}\mathcal{A}(\rho^{\mathcal{T}})\big] \geq \frac{1-\delta}{e^{\varepsilon}+1}$. By these two arguments, the desired inequality holds for either of the cases, proving its validity.

A similar inequality holds for every $(\mathcal{R}, \mathcal{T}) \in \mathcal{Q}$ and $P_X \in \Theta$. Thus, the desired implication has been proved.

Appendix X
Bounds on Holevo Information from QDP and QLDP

Let $X \sim P_X$ be a random variable, which can take values in an alphabet $\mathcal{X}$. Depending on X, the state ρ^X is chosen from the set $\{\rho^1, \ldots, \rho^{|\mathcal{X}|}\}$. Then the state ρ^X is sent through a quantum channel $\mathcal{A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ satisfying QLDP (recall this notion from Remark III-C2). Afterwards the goal is to identify X by performing a measurement described by the POVM $\{\mathrm{M}_y\}_{y\in\mathcal{Y}}$, which realizes the output Y. The flow diagram relevant to this setup is shown in Fig. 9.

Here, we focus on how much information about X can be learned from the output of the quantum privacy mechanism $\mathcal{A}(\rho^X)$ and the classical output Y, with an emphasis on the quantities $\mathsf{I}(X;B)_\sigma$ and $\mathsf{I}(X;Y)$, respectively, where

$$\sigma_{XB} = \sum_{x\in\mathcal{X}} P_X(x)\,|x\rangle\langle x| \otimes \mathcal{A}(\rho^x). \tag{357}$$

We define the Holevo information of the classical–quantum state σ_{XB} as

$$\begin{aligned}
\mathsf{I}(X;B)_\sigma &:= \mathsf{S}\left(\sum_{x\in\mathcal{X}} P_X(x)\,\mathcal{A}(\rho^x)\right) \\
&\quad - \sum_{x\in\mathcal{X}} P_X(x)\,\mathsf{S}(\mathcal{A}(\rho^x)). \qquad (358)
\end{aligned}$$

By data processing of the Holevo information, we have that $\mathsf{I}(X;Y) \leq \mathsf{I}(X;B)_\sigma$. Next, we provide bounds for $\mathsf{I}(X;B)_\sigma$ when $\mathcal{A}$ satisfies ε-quantum local DP (QLDP). Recall that, to achieve QLDP, the algorithm $\mathcal{A}$ is designed so that the output of $\mathcal{A}$ may not release much information about the input states (recall Remark III-C2).

Proposition 15 (Bounds on Holevo Information Due to QLDP): Let $\mathcal{A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a quantum channel. If $\mathcal{A}$ satisfies ε-QLDP, then

$$\mathsf{I}(X;B)_\sigma \leq \min\left\{\varepsilon, \frac{\varepsilon^2}{2}\right\}. \tag{359}$$

Furthermore, if $\mathcal{A}$ is (ε, δ)-QLDP, then

$$\mathsf{I}(X;B)_\sigma \leq \delta' \ln(d-1) + \mathsf{h}(\delta'), \tag{360}$$

if $\delta' \in [0, 1-1/d]$, where $\delta' = 1 - 2(1-\delta)/(e^{\varepsilon}+1)$, d is the dimension of the Hilbert space $\mathcal{H}_B$, and

$$\mathsf{h}(\delta') := -\delta'\ln(\delta') - (1-\delta')\ln(1-\delta') \tag{361}$$

is the binary Shannon entropy in nats.

Proof: The following proof ideas are inspired by the works on mutual information based classical DP and PP presented in [11, Theorem 1] and [10, Theorem 1], respectively. For the proof of the first part, consider that

$$\begin{aligned}
&\mathsf{I}(X;B)_\sigma \\
&= \sum_{x\in\mathcal{X}} P_X(x)\, \mathsf{D}\left(\mathcal{A}(\rho^x)\middle\|\sum_{x'\in\mathcal{X}} P_X(x')\mathcal{A}(\rho^{x'})\right) && (362)\\
&\leq \sum_{x\in\mathcal{X}}\sum_{x'\in\mathcal{X}} P_X(x)P_X(x')\ \mathsf{D}\Big(\mathcal{A}(\rho^x)\Big\|\mathcal{A}(\rho^{x'})\Big) && (363)\\
&\leq \min\left\{\varepsilon, \frac{\varepsilon^2}{2}\right\}, && (364)
\end{aligned}$$

where the first inequality follows from the joint convexity of quantum relative entropy [88]. The final inequality follows because $\mathcal{A}$ satisfies ε-QLDP, as well as from Proposition 9 and the setup for QLDP in Remark III-C2.

For the second part, consider that

$$\begin{aligned}
\mathsf{I}(X;B)_\sigma &= \mathsf{S}(\mathcal{A}(\bar\rho)) - \sum_{x\in\mathcal{X}} P_X(x)\mathsf{S}(\mathcal{A}(\rho^x)) && (365)\\
&\leq \delta' \ln(d-1) + \mathsf{h}(\delta'), && (366)
\end{aligned}$$

where

$$\bar\rho = \sum_{x\in\mathcal{X}} P_X(x)\rho^x, \tag{367}$$

and the final inequality follows from the fact that $\|\mathcal{A}(\bar\rho) - \mathcal{A}(\rho^x)\|_1 \leq 2\delta'$ due to $\mathcal{A}$ satisfying (ε,δ)-QPP along with Proposition 10. Then, to arrive at the final expression, we use the entropy continuity result known as the Fannes-Audenaert Inequality [89]:

$$|\mathsf{S}(\mathcal{A}(\bar\rho)) - \mathsf{S}(\mathcal{A}(\rho^x))| \leq \delta' \ln(d-1) + \mathsf{h}(\delta'), \tag{368}$$

concluding the proof. ∎

Next, we extend the earlier setup to the case in which we generate n i.i.d. random variables from the distribution of P_X. Then, $\{X_1,\ldots,X_n\}$ forms a database. Depending on the value of the database, we choose $\rho^{X^n} \coloneqq \rho^{X_1}\otimes\cdots\otimes\rho^{X_n}$. Then ρ^{X^n} is passed through a quantum algorithm $\mathcal{A}$ that satisfies QDP, followed by a POVM $\{\mathrm{M}_y\}_{y\in\mathcal{Y}}$ that generates the random outcome Y taking values in $\mathcal{Y}$. In this formulation, we take the convention that $\rho \sim \sigma$ (i.e., ρ and σ are neighbors) in the QDP framework if $\mathrm{Tr}_i[\rho] = \mathrm{Tr}_i[\sigma]$ for all $i\in\{1,\ldots,n\}$. Let

$$\sigma_{X^nB} = \sum_{x\in\mathcal{X}} P_{X^n}(x^n)\,|x^n\rangle\langle x^n|\otimes\mathcal{A}(\rho^{x^n}). \tag{369}$$

Proposition 16 (Bounds on Mutual Information Due to QDP): Let $\mathcal{A}:\mathcal{L}(\mathcal{H}_A)\to\mathcal{L}(\mathcal{H}_B)$ be a quantum channel. For all $i\in\{1,\ldots,n\}$, suppose that X_i is drawn i.i.d. from the distribution P_X. If $\mathcal{A}$ satisfies ε-QDP, then

$$\sup_{i\in\{1,\ldots,n\},P_X} \mathsf{I}(X_i;B|X^{n\setminus i})_\sigma \leq \min\left\{\varepsilon,\frac{\varepsilon^2}{2}\right\}. \tag{370}$$

Furthermore, if $\mathcal{A}$ satisfies (ε,δ)-QDP, then

$$\sup_{i\in\{1,\ldots,n\},P_X} \mathsf{I}(X_i;B|X^{n\setminus i})_\sigma \leq \delta'\ln(d-1)+\mathsf{h}(\delta'), \tag{371}$$

if $\delta'\in[0,1-1/d]$, where $\delta' = 1-2(1-\delta)/(e^\varepsilon+1)$.

Proof: First, let us consider $\mathsf{I}(X_i;B|X^{n\setminus i}=z^{n\setminus i})_\sigma$. Define the shorthands

$$\begin{aligned}
\rho^{n\setminus i} &\coloneqq \rho_1^{z_1}\otimes\ldots\rho_{i-1}^{z_{i-1}}\otimes\rho_{i+1}^{z_{i+1}}\otimes\rho_n^{z_n}, && (372)\\
\omega &\coloneqq \sum_{x\in\mathcal{X}} P_{X_i}(x)\rho_i^x\otimes\rho^{n\setminus i}, && (373)\\
\omega^x &\coloneqq \rho_i^x\otimes\rho^{n\setminus i}. && (374)
\end{aligned}$$

Then $\omega\sim\omega^x$ because $\mathrm{Tr}_i[\sigma]=\mathrm{Tr}_i[\sigma^x]$.

For the first part, it thus follows that

$$\begin{aligned}
&\mathsf{I}(X_i;B|X^{n\setminus i}=z^{n\setminus i})_\sigma \\
&\quad= \sum_{x\in\mathcal{X}} P_{X_i}(x)\ \mathsf{D}(\mathcal{A}(\omega^x)\|\mathcal{A}(\omega)) \leq \min\left\{\varepsilon,\frac{\varepsilon^2}{2}\right\}. && (375)
\end{aligned}$$

The last inequality holds because $\mathcal{A}$ satisfies ε-QDP, along with Proposition 9. Since this inequality holds for all possible $z^{n\setminus i}$ such that $\{X^{n\setminus i}=z^{n\setminus i}\}$, the desired relation holds.

For the second part, consider that

$$\begin{aligned}
&\mathsf{I}(X_i;B|X^{n\setminus i}=z^{n\setminus i})_\sigma \\
&\quad= \mathsf{S}(\mathcal{A}(\omega)) - \sum_{x\in\mathcal{X}} P_{X_i}(x)\mathsf{S}(\mathcal{A}(\omega^x)) && (376)\\
&\quad\leq \delta'\ln(d-1)+\mathsf{h}(\delta'), && (377)
\end{aligned}$$

where the last inequality holds because $\|\mathcal{A}(\omega)-\mathcal{A}(\omega^x)\|_1 \leq 2\delta'$ with $\mathcal{A}$ being (ε,δ)-QDP for the pair $\sigma\sim\sigma^x$, and again applying the continuity result for quantum entropy, as in the proof of Proposition 15. ∎

By the data processing inequality for mutual information, we have $\mathsf{I}(X_i;Y|X^{n\setminus i}) \leq \varepsilon' \coloneqq \min\{\varepsilon,\varepsilon^2/2\}$. This showcases that the setup of Proposition 16 satisfies ε'-mutual information differential privacy, as proposed in [11], where a randomized mechanism $A:\mathcal{X}^{n\times k}\to\mathcal{Y}$ is defined to be ε-mutual information differentially private if

$$\sup_{\substack{P_X\in\mathcal{P}(\mathcal{X}^{n\times k}),\\ i\in\{1,\ldots,n\}}} \mathsf{I}(X_i;A(X)|X^{n\setminus i}) \leq \varepsilon. \tag{378}$$

ACKNOWLEDGMENT

The authors would like to thank Rochisha Agarwal, Hansadi Jayamaha, Kaiyuan Ji, Margarite LaBorde, Hemant Mishra, Dhrumil Patel, Aby Philip, Soorya Rethinasamy, and Vishal Singh for helpful discussions, and they are indebted to Prof. Çişik Balyk for sharing his valuable insights.

REFERENCES

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptogr. Conf.*, 2006, pp. 265–284.

[2] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[3] D. Kifer and A. Machanavajjhala, "Pufferfish: A framework for mathematical privacy definitions," *ACM Trans. Database Syst.*, vol. 39, no. 1, pp. 1–36, Jan. 2014.

[4] S. Song, Y. Wang, and K. Chaudhuri, "Pufferfish privacy mechanisms for correlated data," in *Proc. ACM Int. Conf. Manag. Data*, May 2017, pp. 1291–1306.

[5] W. Zhang, O. Ohrimenko, and R. Cummings, "Attribute privacy: Framework and mechanisms," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Jun. 2022, pp. 757–766.

[6] S. Kessler, E. Buchmann, and K. Böhm, "Deploying and evaluating pufferfish privacy for smart meter data," in *Proc. IEEE 12th Int. Conf. Ubiquitous Intell. Comput. IEEE 12th Int. Conf. Autonomic Trusted Comput. IEEE 15th Int. Conf. Scalable Comput. Commun. Associated Workshops (UIC-ATC-ScalCom)*, Aug. 2015, pp. 229–238.

[7] R. M. Cardell-Oliver and B. Ke, "Towards an activity-aware pufferfish framework for local privacy of household smart water meter data," in *Proc. 10th ACM Int. Conf. Syst. Energy-Efficient Buildings, Cities, Transp.*, Nov. 2023, pp. 328–332.

[8] W. Liang, H. Chen, R. Liu, Y. Wu, and C. Li, "A pufferfish privacy mechanism for monitoring web browsing behavior under temporal correlations," *Comput. Secur.*, vol. 92, May 2020, Art. no. 101754.

[9] Y. Zeng, Y. Sang, S. Luo, and M. Song, "A pufferfish privacy mechanism for the trajectory clustering task," in *Parallel Architectures, Algorithms and Programming*. Singapore: Springer, 2021, pp. 307–317.

[10] T. Nuradha and Z. Goldfeld, "Pufferfish privacy: An information-theoretic study," *IEEE Trans. Inf. Theory*, vol. 69, no. 11, pp. 7336–7356, Nov. 2023.

[11] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 43–54.

[12] T. Nuradha and Z. Goldfeld, "An information-theoretic characterization of pufferfish privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2022, pp. 2005–2010.

[13] L. Zhou and M. Ying, "Differential privacy in quantum computation," in *Proc. IEEE 30th Comput. Secur. Found. Symp. (CSF)*, Aug. 2017, pp. 249–262.

[14] S. Aaronson and G. N. Rothblum, "Gentle measurement of quantum states and differential privacy," in *Proc. 51st Annu. ACM SIGACT Symp. Theory Comput.*, Jun. 2019, pp. 322–333.

[15] C. Hirche, C. Rouzé, and D. S. França, "Quantum differential privacy: An information theory perspective," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5771–5787, Sep. 2023.

[16] Y. Quek, S. Arunachalam, and J. A. Smolin, "Private learning implies quantum stability," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 20503–20515.

[17] Y. Du, M.-H. Hsieh, T. Liu, D. Tao, and N. Liu, "Quantum noise protects quantum classifiers against adversaries," *Phys. Rev. Res.*, vol. 3, no. 2, May 2021, Art. no. 023153.

[18] M. Senekane, M. Mafu, and B. M. Taele, "Privacy-preserving quantum machine learning using differential privacy," in *Proc. IEEE AFRICON*, Sep. 2017, pp. 1432–1435.

[19] A. Angrisani and E. Kashefi, "Quantum local differential privacy and quantum statistical query model," 2022, *arXiv:2203.03591*.

[20] Y. Du, M.-H. Hsieh, T. Liu, S. You, and D. Tao, "Quantum differentially private sparse regression learning," *IEEE Trans. Inf. Theory*, vol. 68, no. 8, pp. 5217–5233, Aug. 2022.

[21] W. M. Watkins, S. Y.-C. Chen, and S. Yoo, "Quantum machine learning with differential privacy," *Sci. Rep.*, vol. 13, no. 1, p. 2453, Feb. 2023.

[22] J.-C. Huang et al., "Certified robustness of quantum classifiers against adversarial examples through quantum noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[23] A. Angrisani, M. Doosti, and E. Kashefi, "Differential privacy amplification in quantum and quantum-inspired algorithms," 2022, *arXiv:2203.03604*.

[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] R. Bassily, A. Groce, J. Katz, and A. Smith, "Coupled-worlds privacy: Exploiting adversarial uncertainty in statistical data privacy," in *Proc. IEEE 54th Annu. Symp. Found. Comput. Sci.*, Oct. 2013, pp. 439–448.

[26] B. M. Terhal, D. P. DiVincenzo, and D. W. Leung, "Hiding bits in Bell states," *Phys. Rev. Lett.*, vol. 86, no. 25, pp. 5807–5810, Jun. 2001.

[27] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal, "Quantum data hiding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 580–598, Nov. 2002.

[28] T. Eggeling and R. F. Werner, "Hiding classical data in multipartite quantum states," *Phys. Rev. Lett.*, vol. 89, no. 9, Aug. 2002, Art. no. 097905.

[29] P. Hayden, D. Leung, P. W. Shor, and A. Winter, "Randomizing quantum states: Constructions and applications," *Commun. Math. Phys.*, vol. 250, no. 2, pp. 371–391, Sep. 2004.

[30] P. Hayden, D. Leung, and G. Smith, "Multiparty data hiding of quantum information," *Phys. Rev. A, Gen. Phys.*, vol. 71, no. 6, Jun. 2005, Art. no. 062339.

[31] C. Lupo, M. M. Wilde, and S. Lloyd, "Quantum data hiding in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3745–3756, Jun. 2016.

[32] L. Lami, C. Palazuelos, and A. Winter, "Ultimate data hiding in quantum mechanics and beyond," *Commun. Math. Phys.*, vol. 361, no. 2, pp. 661–708, Jul. 2018.

[33] T. Murakami and Y. Kawamoto, "Utility-optimized local differential privacy mechanisms for distribution estimation," in *Proc. 28th USENIX Secur. Symp. (USENIX Secur.)*, 2019, pp. 1877–1894.

[34] J. Guan, W. Fang, and M. Ying, "Verifying fairness in quantum machine learning," in *Proc. Int. Conf. Comput. Aided Verification*. Cham, Switzerland: Springer, 2022, pp. 408–429.

[35] E. Perrier, "Quantum fair machine learning," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, Jul. 2021, pp. 843–853.

[36] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[37] S. Khatri and M. M. Wilde, "Principles of quantum communication theory: A modern approach," 2020, *arXiv:2011.04672v1*.

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

[39] A. Uhlmann, "The 'transition probability' in the state space of a*-algebra," *Rep. Math. Phys.*, vol. 9, no. 2, pp. 273–279, Apr. 1976.

[40] A. Y. Kitaev, "Quantum computations: Algorithms and error correction," *Russian Math. Surveys*, vol. 52, no. 6, pp. 1191–1249, Dec. 1997.

[41] 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.

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

[43] 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.

[44] 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.

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

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

[47] 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.

[48] X. Wang and M. M. Wilde, "Resource theory of asymmetric distinguishability," *Phys. Rev. Res.*, vol. 1, no. 3, Dec. 2019, Art. no. 033170.

[49] A. C. Thompson, "On certain contraction mappings in a partially ordered vector space," *Proc. Amer. Math. Soc.*, vol. 14, no. 3, p. 438, Jun. 1963.

[50] R. Salzmann, N. Datta, G. Gour, X. Wang, and M. M. Wilde, "Symmetric distinguishability as a quantum resource," *New J. Phys.*, vol. 23, no. 8, Aug. 2021, Art. no. 083016.

[51] B. Regula, L. Lami, and M. M. Wilde, "Postselected quantum hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 70, no. 5, pp. 3453–3469, May 2024.

[52] B. Pejó and D. Desfontaines, *Guide to Differential Privacy Modifications: A Taxonomy of Variants and Extensions*. (Springer Briefs in Computer Science Serries). Cham, Switzerland: Springer, 2022.

[53] G. De Palma, M. Marvian, D. Trevisan, and S. Lloyd, "The quantum Wasserstein distance of order 1," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6627–6643, Oct. 2021.

[54] D. Kifer and B.-R. Lin, "An axiomatic view of statistical privacy and utility," *J. Privacy Confidentiality*, vol. 4, no. 1, pp. 1–36, Jul. 2012.

[55] W. Matthews, S. Wehner, and A. Winter, "Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding," *Commun. Math. Phys.*, vol. 291, no. 3, pp. 813–843, Nov. 2009.

[56] E. B. Davies and J. T. Lewis, "An operational approach to quantum probability," *Commun. Math. Phys.*, vol. 17, no. 3, pp. 239–260, Sep. 1970.

[57] E. B. Davies, *Quantum Theory of Open Systems*. New York, NY, USA: Academic, 1976.

[58] M. Ozawa, "Quantum measuring processes of continuous observables," *J. Math. Phys.*, vol. 25, no. 1, pp. 79–87, Jan. 1984.

[59] Ú. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2014, pp. 1054–1067.

[60] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proc. 22nd ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst.*, Jun. 2003, pp. 211–222.

[61] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM J. Comput.*, vol. 40, no. 3, pp. 793–826, Jan. 2011.

[62] S. Nietert, Z. Goldfeld, and R. Cummings, "Outlier-robust optimal transport: Duality, structure, and statistical analysis," in *Proc. Int. Conf. Artif. Intell. Statist.*, vol. 151, 2022, pp. 11691–11719.

[63] J. Watrous, "Semidefinite programs for completely bounded norms," *Theory Comput.*, vol. 5, no. 11, pp. 217–238, 2009.

[64] D. Sutter, V. B. Scholz, A. Winter, and R. Renner, "Approximate degradable quantum channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 12, pp. 7832–7844, Dec. 2017.

[65] M. Horodecki, P. Horodecki, and R. Horodecki, "General teleportation channel, singlet fraction, and quasidistillation," *Phys. Rev. A, Gen. Phys.*, vol. 60, no. 3, pp. 1888–1898, Sep. 1999.

[66] Z. Ding, Y. Wang, G. Wang, D. Zhang, and D. Kifer, "Detecting violations of differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2018, pp. 475–489.

[67] M. Jagielski, J. Ullman, and A. Oprea, "Auditing differentially private machine learning: How private is private SGD?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 22205–22216.

[68] C. Domingo-Enrich and Y. Mroueh, "Auditing differential privacy in high dimensions with the kernel quantum Rényi divergence," 2022, *arXiv:2205.13941*.

[69] J. Watrous, "Limits on the power of quantum statistical zero-knowledge," in *Proc. 43rd Annu. IEEE Symp. Found. Comput. Sci.*, Mar. 2002, pp. 459–468.

[70] J. Watrous, "Zero-knowledge against quantum attacks," *SIAM J. Comput.*, vol. 39, no. 1, pp. 25–58, Jan. 2009.

[71] S. Rethinasamy, R. Agarwal, K. Sharma, and M. M. Wilde, "Estimating distinguishability measures on quantum computers," *Phys. Rev. A, Gen. Phys.*, vol. 108, no. 1, Jul. 2023, Art. no. 012409.

[72] R. Chen, Z. Song, X. Zhao, and X. Wang, "Variational quantum algorithms for trace distance and fidelity estimation," *Quantum Sci. Technol.*, vol. 7, no. 1, Jan. 2022, Art. no. 015019.

[73] Q. Wang and Z. Zhang, "Fast quantum algorithms for trace distance estimation," *IEEE Trans. Inf. Theory*, vol. 70, no. 4, pp. 2720–2733, Apr. 2024.

[74] M. Cerezo et al., "Variational quantum algorithms," *Nature Rev. Phys.*, vol. 3, pp. 625–644, Aug. 2021.

[75] K. Bharti et al., "Noisy intermediate-scale quantum (NISQ) algorithms," *Rev. Modern Phys.*, vol. 94, no. 1, Feb. 2022, Art. no. 015004.

[76] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, "Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics," in *Proc. 51st Annu. ACM SIGACT Symp. Theory Comput.*, Jun. 2019, pp. 193–204.

[77] D. Aharonov, V. Jones, and Z. Landau, "A polynomial quantum algorithm for approximating the Jones polynomial," *Algorithmica*, vol. 55, no. 3, pp. 395–421, Nov. 2009.

[78] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," *Nature Phys.*, vol. 10, no. 9, pp. 631–633, Jul. 2014.

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

[80] M. Ohya and D. Petz, *Quantum Entropy and Its Use*. Cham, Switzerland: Springer, 1993.

[81] I. Mironov, "Rényi differential privacy," in *Proc. IEEE 30th Comput. Secur. Found. Symp. (CSF)*, Aug. 2017, pp. 263–275.

[82] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," in *Proc. Int. Conf. Theory Cryptogr.* Berlin, Germany: Springer, 2016, pp. 635–658.

[83] Q. Wang, J. Guan, J. Liu, Z. Zhang, and M. Ying, "New quantum algorithms for computing quantum entropies and distances," *IEEE Trans. Inf. Theory*, early access, 2024. [Online]. Available: https://arxiv.org/abs/2203.13522

[84] C. Fuchs, "Distinguishability and accessible information in quantum theory," Ph.D. dissertation, Dept. Phys. Astron., Univ. New Mexico, Albuquerque, NM, USA, Dec. 1996.

[85] K. Matsumoto, "A new quantum version of f-divergence," in *Reality and Measurement in Algebraic Quantum Theory*, vol. 261. Singapore: Springer, 2018, pp. 229–273.

[86] F. Hiai and M. Mosonyi, "Different quantum f-divergences and the reversibility of quantum operations," *Rev. Math. Phys.*, vol. 29, no. 7, 2017, Art. no. 1750023.

[87] G. Lindblad, "Completely positive maps and entropy inequalities," *Commun. Math. Phys.*, vol. 40, no. 2, pp. 147–151, Jun. 1975.

[88] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *J. Math. Phys.*, vol. 14, no. 12, pp. 1938–1941, Dec. 1973.

[89] K. M. R. Audenaert, "A sharp continuity estimate for the von Neumann entropy," *J. Phys. A, Math. Theor.*, vol. 40, no. 28, pp. 8127–8136, Jul. 2007.

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 2018, and the M.Sc. degree from Cornell University, USA, in 2023, where she is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering. During the bachelor's degree, 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.

Ziv Goldfeld (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from Ben-Gurion University, Israel, in 2012, 2014, and 2017, respectively. From 2017 to 2019, he was a Post-Doctoral Fellow with the Laboratory for Information and Decision Systems (LIDS), MIT. He is currently an Assistant Professor in electrical and computer engineering with Cornell University. He was a recipient of several awards, such as the NSF CAREER Award, the IBM Academic Award, and the Rothschild Fellowship.

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 in electrical and computer engineering with Cornell University. His current research interests are in quantum Shannon theory, quantum computation, quantum optical communication, quantum computational complexity theory, and quantum error correction. He is an Outstanding Referee of the American Physical Society. He was a recipient of the National Science Foundation Career Development Award. He was also a co-recipient of the 2018 AHP-Birkhauser Prize, awarded to "the most remarkable contribution" published in the journal *Annales Henri Poincaré*.