Maximal information leakage from quantum encoding of classical data

Farhad Farokhi *

Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia

(Received 24 July 2023; accepted 5 January 2024; published 16 February 2024)

An alternative measure of information leakage for quantum encoding of classical data is defined. An adversary can access a single copy of the state of a quantum system that encodes some classical data and is interested in correctly guessing a general randomized or deterministic function of the data (e.g., a specific feature or attribute of the data in quantum machine learning) that is unknown to the security analyst. The resulting measure of information leakage, referred to as maximal quantum leakage, is the multiplicative increase of the probability of correctly guessing any function of the classical data upon observing measurements of the quantum state. Maximal quantum leakage is shown to satisfy the postprocessing inequality (i.e., applying a quantum channel reduces information leakage) and independence property (i.e., leakage is zero if the quantum state is independent of the classical data), which are fundamental properties required for privacy and security analysis. It also bounds accessible information. Effects of global and local depolarizing noise models on the maximal quantum leakage are established.

DOI: 10.1103/PhysRevA.109.022608

I. INTRODUCTION

This paper deals with quantifying the "amount" of classical information that can be leaked from a quantum system whose state encodes the said classical information. This is a basic question that arises in security and privacy analysis of quantum computing systems. In a setting where classical information is encoded into the state of a quantum system, e.g., quantum machine learning [1], an adversary may access the system either legitimately or maliciously (by hacking), perform measurements on the state of the system, and extract private or secret information. This has, in part, motivated development of privacy-preserving quantum computing [2–6]. This measure of information leakage can also be used to investigate security in communication or transmission of classical data over insecure quantum channels, where an eavesdropper may attempt to extract some classical information by performing measurements on the communicated qubits. This setup is akin to quantum wiretap channels [7]; see Sec. VI for more information.

Any useful measure of information leakage must satisfy a few requirements [8,9]. First, and foremost, the measure should possess an operational interpretation. As stated in [10] (p. 313), "the ultimate test for whether we truly understand an information measure is if it is the answer to some operational task." This will enable a designer or analyst to explain what guarantees can be extracted from minimizing or bounding the measure of information leakage. Second, assumptions regarding the adversary must be minimal so that a large family of adversaries can be modeled and analyzed. For instance, it is customary to assume that the adversary seeks to estimate the entire data accurately while, in practice, the adversary might only be seeking to extract as much information as possible[1] or might only be interested in estimating subsets of the data or specific features that may be unknown to the security analyst. Third, the measure should satisfy certain properties, such as the postprocessing inequality (i.e., further processing of the quantum system by an arbitrary quantum channel must reduce information leakage) and independence property (i.e., leakage is zero if the quantum state is independent of the classical data). The former enables the analyst to make statements that are independent of the computational power of the adversary, i.e., postprocessing by advanced computing techniques or powerful machines should not increase information leakage, while the latter implies that the measure of information leakage is not conservative, i.e., assigning a risk of information leakage to situations where the adversary is guaranteed to not gain any insight. Finally, the measure of information leakage should align with intuition, e.g., noisy quantum circuits must reduce the information leakage.

Common measures of information leakage, while satisfying some of these properties, often fail to meet all requirements. For instance, in quantum computing and information theory, accessible information and its upper bound, Holevo information, do not meet the requirement on minimal assumptions on the adversary. They can be only used in the context that the adversary is interested in estimating the entirety of the classical data. This is because accessible information and Holevo information are formulated to study reliable information transmission [11], not information leakage in security analysis. Also, it is well understood that mutual information is not a good measure of information leakage

*farhad.farokhi@unimelb.edu.au

[1]Akin to the so-called fishing expedition, which refers to nonspecific search for information.

 ©2024 American Physical Society

in security and privacy [8]. Again, the operational interpretation of mutual information stems from communication and compression, which differ from security and privacy. In compression, for instance, we must be able to decode the entire information without any loss. However, in security, an adversary may not be interested in extracting the entire classical information. It may merely want to extract some or any private information. In the classical setting, these observations have motivated moving away from mutual information for measuring information leakage in privacy and security frameworks [8,9,12,13].

This paper presents a metric or measure for information leakage from quantum systems referred to as *maximal quantum leakage*. The measure is built upon a similar classical notion of information leakage known as maximal leakage [8]. The adversary has access to a single copy of the state ρ_A^X of quantum system A that encodes some classical data X, which is assumed to be private and must be kept secure. The adversary is interested in guessing or estimating a general, possibly randomized function of the original classical data X, called Z. The adversary's intention or target, i.e., the underlying randomized function of the classical data, is not known to the designer or analyst. Therefore, the measure of information leakage must be maximized over all possible choices of this function. This motivates the use of the term *maximal* in maximal quantum leakage. This threat model captures a large family of potential adversaries and thus minimizes the assumptions made regarding the adversary's intent. The adversary can perform measurements on the state of the quantum system to observe a random variable Y, i.e., the outcome of the measurements. The adversary then attempts to guess Z based on Y and verify whether the choice is correct. For instance, in the security framework, this could model guessing an individual's password and attempting to log in using the guess [8]. However, other interpretations can be provided, e.g., this could capture guessing someone's private information to use for phishing attacks against them. The adversary's goal is to maximize the probability of correctly guessing Z. That is, the adversary attempts to extract some information regarding X, modeled by Z, with high certainty. The information leakage measures the worst-case (i.e., maximal) ratio of the probability of correctly guessing Z with access to Y and without access to Y. Therefore, the measure investigates cases where the probability of correctly guessing Z increases considerably based on access to Y, i.e., Y leaks considerable information about Z, which is in turn a function of the private data X. This provides a natural interpretation for the information leakage: The multiplicative increase in the probability of correctly guessing any general random or deterministic function of the private data upon access to the quantum encoding of the data is upper bounded by the maximal quantum leakage.

Important properties for the maximal quantum leakage are established. First, maximal quantum leakage is zero if the quantum encoding is indistinguishable, i.e., if the quantum state is independent of the classical data. This is a natural property as otherwise the measure of leakage acts conservatively, i.e., it assigns a nonzero leakage to a scenario that possesses no risk. Second, maximal quantum leakage admits the postprocessing inequality, i.e., maximal leakage reduces if the state of the quantum system is manipulated by an arbitrary quantum channel. This is a useful property in privacy and security analysis because it implies that we only need to compute the information leakage at the beginning of the data analysis chain to establish the risk of data breach, and further computation cannot increase the risk. Finally, the effect of quantum noise models, such as global and local depolarizing channels, on the maximal quantum leakage is investigated. As expected, quantum noise reduces information leakage. This is not surprising given previous observations on the effect of noise in quantum devices on data privacy [2,3]. However, establishing such results is important in ensuring that the proposed notion of information leakage accords with intuition.

Before moving on to the technical content of the paper, a few remarks must be stated regarding the relationship between the framework of this paper, its classical counterpart [8], and the relevant literature on accessible information [14–16]. The main difference with the classical results in [8] stems from the fact that the conditional probability of measurements given private classical data can be written explicitly in terms of quantum states (i.e., the quantum encoding of the classical data) and the positive operator-valued measure (POVM) modeling the measurement process using Born's rule. This implies a degree of freedom that is missing in the classical counterpart. That is, in [8], the conditional probabilities are fixed, but, in this paper, the adversary can changes the conditional probabilities by varying the POVM, i.e., the adversary can select the optimal measurement process for extracting as much information as possible. This implies an additional optimization over the POVMs, which can be potentially unbounded. To alleviate this difficulty, we use methods developed for accessible information [14–16] to show that the number of elements in the POVM is bounded and can be computed iteratively.

The remainder of the paper is organized as follows. We finish this section with some preliminary definitions and useful notations for quantum systems. Maximal quantum leakage is formally defined and a semiexplicit formula for its computation is presented in Sec. II. Properties of the maximal quantum leakage, i.e., the independence property, postprocessing inequality, and upper and lower bounds for the leakage, are established in Sec. III. The effect of depolarizing noises inherent to quantum devices on maximal quantum leakage is investigated in Sec. IV. To minimize interruptions to the flow of the paper and to focus on definitions and properties without getting bogged down in the mathematics, the proofs of all the results in Secs. II–IV are presented across different subsections in the Appendix. An iterative algorithm using subgradient ascent for computing maximal quantum leakage is presented in Sec. V. Finally, Sec. VI presents some concluding remarks and avenues for future research.

Preliminary definitions and notations

The basic definitions and useful notations presented in this review section are mostly borrowed from [10].

The state space of a quantum system is modeled by complex Hilbert space $\mathcal{H}$. Dirac's notation is used to denote pure quantum states. A *pure quantum state* is an element of Hilbert space $\mathcal{H}$ with unit norm, e.g., $|\psi\rangle \in \mathcal{H}$ with $\||\psi\rangle\|_2 = \sqrt{\langle\psi|\psi\rangle} = 1$. The smallest nontrivial quantum system is a *qubit* corresponding to a two-dimensional Hilbert space. Combination of any two

quantum states $|\phi\rangle$ and $|\psi\rangle$ is denoted by their tensor product $|\phi\rangle \otimes |\psi\rangle$. A *mixed quantum state* is characterized by ensemble $\{(p_1, |\psi_1\rangle), \ldots, (p_k, |\psi_k\rangle)\}$, where $p_i \geqslant 0$ for all $i \in \{1, \ldots, k\}$ and $\sum_i p_i = 1$. The mixed state signifies that the quantum system is in pure state $|\psi_i\rangle$ with probability p_i. The density operator for the mixed quantum state $\{(p_1, |\psi_1\rangle), \ldots, (p_k, |\psi_k\rangle)\}$ is $\rho := \sum_i p_i |\psi_i\rangle\langle\psi_k|$. Thus, $\mathrm{tr}(\rho) = 1$. Any pure quantum state $|\phi\rangle$ can be modeled using rank-1 density operator $\rho = |\phi\rangle\langle\phi|$. Therefore, without loss of generality, the density operator can denote the state of a quantum system. Combination of any two density operators ρ and σ is denoted by their tensor product $\rho \otimes \sigma$.

When the postmeasurement state of the quantum system is of no interest (e.g., the quantum system is discarded after measurement), a measurement for a quantum system can be modeled using a POVM, which is a set of positive semidefinite matrices $F = \{F_i\}_i$ such that $\sum_i F_i = I$. In this case, the probability of obtaining output i when taking a measurement on a system with quantum state ρ is $\mathrm{tr}(\rho F_i) = \mathrm{tr}(F_i \rho)$.

A quantum channel is a mapping from the space of density operators to potentially another space of density operators that is both completely positive and trace preserving. Following the Choi-Kraus the Choi-Kraus theorem (Theorem 4.4.1) of [10], for each quantum channel $\mathcal{E}$, there exists a family of linear operators $\{E_j\}_j$ satisfying $\sum_j E_j^\dagger E_j = I$ such that $\mathcal{E}(\rho) = \sum_j E_j \rho E_j^\dagger$ for all density operators ρ. For any matrix A, $A^\dagger$ denotes its conjugate transpose or Hermitian. The tensor product of quantum channels $\mathcal{E}_1$ and $\mathcal{E}_2$ is defined as $\mathcal{E}_1 \otimes \mathcal{E}_2(\rho_1 \otimes \rho_2) := \mathcal{E}_1(\rho_1) \otimes \mathcal{E}_2(\rho_2)$ for all density operators ρ_1 and ρ_2.

II. DEFINITION AND COMPUTATION OF MAXIMAL QUANTUM LEAKAGE

The classical data that must be kept private or secure are modeled by discrete random variable X with finite support set $\mathbb{X}$. Assume that $p_X(x) := \mathbb{P}\{X = x\} > 0$ for all $x \in \mathbb{X}$. This assumption is without loss of generality as the set $\mathbb{X}$ can always be trimmed so that this assumption holds. Knowledge of the support set of secret variables is referred to as *domain knowledge* and is postulated to be required for developing privacy-preserving mechanisms [5]. As shown later, the maximal quantum leakage is not a function of p_X and is thus robust to the choice of the secret prior. This aligns with the requirement to keep the assumptions on the measure of information leakage minimal.

For each realization of discrete random variable $X = x \in \mathbb{X}$, a quantum system A in mixed state $\rho_A^x \in \mathcal{D}(\mathcal{H}_A)$ is prepared, i.e., ensemble $\mathcal{E} := \{p_X(x), \rho_A^x\}_{x\in\mathbb{X}}$ is prepared. The quantum state A is handed over to an adversary without revealing the realization of the classical random variable X. The expected density operator is then $\rho_A = \mathbb{E}\{\rho_A^X\} = \sum_{x\in\mathbb{X}} p_X(x)\rho_A^x$. This is the state of the quantum system from the perspective of someone who does not know the realization of X, i.e., the adversary.

The objective of the adversary is to estimate or guess a possibly randomized discrete function of the random variable X, denoted by random variable Z, by performing measurements on the quantum system A. The adversary performs a POVM $F = \{F_y\}_y$ on the quantum system A. Random variable Y with finite support set $\mathbb{Y} = \{1, \ldots, |F|\}$ denotes the outcome of the measurement. The probability of obtaining output $Y = y \in \mathbb{Y}$ when taking a measurement on quantum state ρ_A^X is given by $\mathrm{tr}(\rho_A^X F_i)$. Therefore,

$$\mathbb{P}\{Y = y \,|\, X = x\} = \mathrm{tr}(\rho_A^x F_y), \quad \forall x \in \mathbb{X}, y \in \mathbb{Y}. \tag{1}$$

Upon observing the measurement outcome Y, the adversary takes a one-shot guess of the random variable Z denoted by the random variable $\widehat{Z}$. The adversary then attempts to verify whether $\widehat{Z}$ is correct or not.

Definition 1 (maximal quantum leakage). The maximal quantum leakage from random variable X through quantum encoding of the data via ensemble $\{p_X(x), \rho_A^x\}_{x\in\mathbb{X}}$ is

$$\mathcal{Q}(X \to A)_{\rho_A} := \sup_{\{F_y\}_y} \sup_{Z,\widehat{Z}} \log_2 \left(\frac{\mathbb{P}\{Z = \widehat{Z}\}}{\max\limits_{z\in\mathbb{Z}} \mathbb{P}\{Z = z\}} \right), \tag{2}$$

where the inner supremum is taken over all random variables Z and $\widehat{Z}$ with arbitrary support set $\mathbb{Z}$ and the outer supremum is taken over all POVMs $F = \{F_y\}_y$.

The maximal quantum leakage, as characterized in Definition 1, captures the multiplicative increase in the probability of correctly guessing any general random or deterministic function of the private data upon accessing the quantum encoding of the data. The probability of correctly guessing the realization of random variable Z with access to measurement Y is $\mathbb{P}\{Z = \widehat{Z}\}$. Without access to any measurements, the adversary's best guess of the realization of random variable Z would have been the most likely or probable realization $\widetilde{Z} := \arg\max_{z\in\mathbb{Z}} \mathbb{P}\{Z = z\}$. Therefore, the probability of correctly guessing the realization of random variable Z without access to any measurements is $\mathbb{P}\{Z = \widetilde{Z}\} = \max_{z\in\mathbb{Z}} \mathbb{P}\{Z = z\}$. A large maximal quantum leakage implies that there exist features of the private data that can be guessed more reliably by accessing the quantum state. This demonstrates information leakage along those features. Noting that those features can be potentially exploited by the adversary, an analyst, who is not aware of the target of the adversary, has to investigate and mitigate the weak point.

Theorem 1. The maximal quantum leakage is given by

$$\mathcal{Q}(X \to A)_{\rho_A} = \sup_{\{F_y\}_y} \log_2 \left(\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \mathrm{tr}(\rho_A^x F_y) \right). \tag{3}$$

Proof. See Appendix 1. ■

Theorem 1 provides a semiexplicit formula for computing maximal quantum leakage. This is done by removing the inner maximization over random variables Z and $\widehat{Z}$. Interestingly, Theorem 1 shows that $\mathcal{Q}(X \to A)_{\rho_A}$ is independent of the prior for the secret p_X. Therefore, maximal quantum leakage is immune from or robust to incorporating a wrong assumption on the secret random variable X. The same cannot be said about accessible information, quantum mutual information, or Holevo information. In Theorem 1, however, the outer maximization on $\{F_y\}$ still remains. This is a particularly troubling problem as the number of outcomes in POVM $\{F_y\}$ is not bounded, i.e., it can range to infinity. The next theorem shows that we can restrict our search to POVMs that have

at most $\dim(\mathcal{H}_A)^2$ outcomes. Note that, because we cannot exchange the sum over y and the maximization over x in (3), the information leakage is nonzero in general.

Theorem 2. Let $\mathcal{H}_A$ have finite dimension d_A. The maximal quantum leakage is attained by POVM $F = \{F_y\}_{y=1}^m$ with $m \leqslant d_A^2$, such that F_y are rank-1 operators.

Proof. See Appendix 2 ■.

Maximal quantum leakage in Definition 1 considers a scenario where the adversary only makes and verifies a single guess. This might not be entirely realistic in practice. An adversary might be able to make several guesses. For instance, the adversary might devise multiple privacy or security attacks based on various plausible guesses of the random variable Z. Therefore, we may assume that, upon observing the measurement outcome Y, the adversary makes k guesses of the random variable Z denoted by the random variable $\widehat{Z}_1, \ldots, \widehat{Z}_k$ and then attempts to verify them. In this scenario, we can modify maximal quantum leakage to compute k-maximal quantum leakage, defined below.

Definition 2 (k-maximal quantum leakage). The k-guess maximal quantum leakage from random variable X through quantum encoding of the data via ensemble $\{p_X(x), \rho_A^x\}_{x\in\mathbb{X}}$ is

$$\mathcal{Q}^{(k)}(X \to A)_{\rho_A} := \sup_{\{F_y\}_y} \sup_{Z,\widehat{Z}_1,\ldots,\widehat{Z}_k} \log_2 \left(\frac{\mathbb{P}\{\exists j : Z = \widehat{Z}_j\}}{\max_{\mathcal{Z}\subset\mathbb{Z},|\mathcal{Z}|\leqslant k} \mathbb{P}\{Z \in \mathcal{Z}\}} \right), \quad (4)$$

where the inner supremum is taken over all random variables $Z, \widehat{Z}_1, \ldots, \widehat{Z}_k$ with arbitrary support set $\mathbb{Z}$ and the outer supremum is taken over all POVMs $F = \{F_y\}_y$.

Theorem 3. $\mathcal{Q}^{(k)}(X \to A)_{\rho_A} = \mathcal{Q}(X \to A)_{\rho_A}$.

Proof. The proof follows from Theorem 4 of [8] and the proof of Theorem 1. ■

Theorem 3 implies that the number of guesses that the adversary can make is immaterial in measuring information leakage. Therefore, the choice of one-shot guesses in maximal quantum leakage is without loss of generality.

III. PROPERTIES OF MAXIMAL QUANTUM LEAKAGE

In this section, properties of maximal quantum leakage are established. Maximal quantum leakage satisfies the independence property (i.e., leakage is zero if the quantum state is independent of the classical data) and postprocessing inequality (i.e., applying a quantum channel reduces information leakage). We can also bound maximal quantum leakage based on the dimension of the quantum system and the cardinality of the support set of the secret random variable. We start with the independence property. To do so, we must define indistinguishably to establish when the quantum state is independent of the classical data.

Definition 3 (indistinguishability). $(\rho_A^x)_{x\in\mathbb{X}}$ is indistinguishable if $\rho_A^x = \rho_A^{x'}$ for all $x, x' \in \mathbb{X}$.

Indistinguishability implies that, for various realizations of the classical data X, the quantum state remains the same. Therefore, an adversary cannot obtain any measurements from the quantum states that correlate with the classical data. Therefore, there is no leakage of classical data. This is established in the next result.

Proposition 1. $\mathcal{Q}(X \to A)_{\rho_A} \geqslant 0$ with equality if and only if $(\rho_A^x)_{x\in\mathbb{X}}$ is indistinguishable.

Proof. See Appendix 3. ■

In the next proposition, we provide an upper bound for maximal quantum leakage based on the dimension of the quantum system $d_A = \dim(\mathcal{H}_A)$ and the cardinality of the support set of the secret random variable $|\mathbb{X}|$. A discrete random variable $\mathbb{X}$ with a support set of size $|\mathbb{X}|$ has no more than $\log_2(|\mathbb{X}|)$ bits of information to be leaked. Therefore, $\mathcal{Q}(X \to A)_{\rho_A}$ can never be larger than $\log_2(|\mathbb{X}|)$.

Proposition 2. Let $\mathcal{H}_A$ have finite dimension d_A. Then, $\mathcal{Q}(X \to A)_{\rho_A} \leqslant \min\{\log_2(|\mathbb{X}|), \log_2(d_A^2)\}$.

Proof. See Appendix 4. ■

Another important property is the data-processing inequality stating that quantum maximal leakage can be only reduced by application of an arbitrary quantum channel. This implies that we only need to compute the information leakage at the beginning of the data analysis chain to establish the risk of data breach, and further computation cannot increase the risk. The data-processing inequality is proved in the next proposition.

Proposition 3. For any quantum channel $\mathcal{E}$, $\mathcal{Q}(X \to A)_{\mathcal{E}(\rho_A)} \leqslant \mathcal{Q}(X \to A)_{\rho_A}$.

Proof. See Appendix 5. ■

An important notion of information in quantum information theory is accessible information [10] (p. 298). For ensemble $\mathcal{E} = \{p_X(x), \rho_A^x\}_{x\in\mathbb{X}}$, defined in Sec. II, the accessible information is

$$I_{\text{acc}}(\mathcal{E}) := \sup_{\{F_y\}_y} I(X;Y),$$

where $I(X;Y)$ is the classical mutual information between the random variable X denoting the secret or private information and the random variable Y denoting the measurement outcome. The next proposition provides a relationship between accessible information and maximal quantum leakage.

Proposition 4. $I_{\text{acc}}(\mathcal{E}) \leqslant \mathcal{Q}(X \to A)_{\rho_A}$.

Proof. See Appendix 6. ■

The inequality in Proposition 4 is rather intuitive. The accessible information, in the context of security analysis, deals with an adversary that seeks to estimate the entire secret data. However, in defining maximal quantum leakage, we let the adversary extract as much information as possible by estimating any general possibly randomized function of the data. The adversary in the maximal quantum leakage setting is stronger and more general in comparison with the adversary in the context of the accessible information. This inequality is a direct consequence of requiring minimal assumptions on the adversary.

IV. EFFECT OF QUANTUM NOISE ON MAXIMAL LEAKAGE

A common noise model in quantum systems is the (global) depolarizing channel defined as

$$\mathcal{D}_{p,d_A}(\rho) := \frac{p}{d_A} I + (1-p)\rho, \quad (5)$$

where d_A is the dimension of the Hilbert space $\mathcal{H}_A$ to which the system belongs and $p \in [0, 1]$ is a probability parameter.

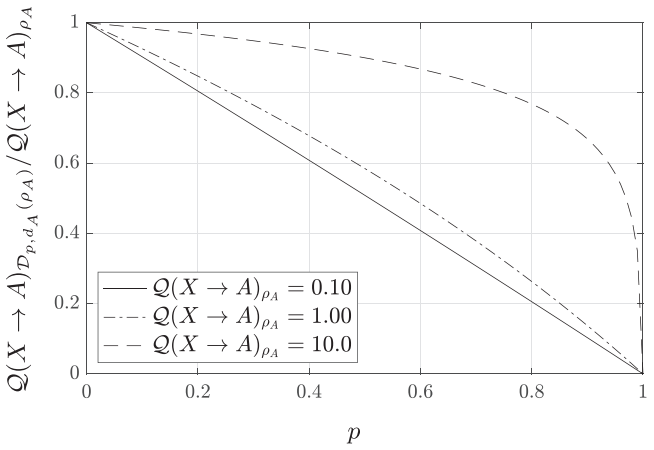

FIG. 1. Ratio of information leakage without and with the global depolarizing channel vs the probability parameter p. As expected, the noisier the channel is, i.e., the higher the probability parameter is, the smaller maximal quantum leakage is.

In the next proposition, it is shown that the depolarizing channel results in reduction of the maximal quantum leakage. This means that maximal quantum leakage accords with intuition and similar results on privacy-preserving quantum systems [2–6].

Proposition 5. For global depolarizing channel $\mathcal{D}_{p,d_A}$,

$$\mathcal{Q}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)} = \log_2[p + (1-p)2^{\mathcal{Q}(X\to A)_{\rho_A}}].$$

Particularly,

$$\frac{d}{dp}\mathcal{Q}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)} < 0 \text{ if } \mathcal{Q}(X \to A)_{\rho} > 0.$$

Therefore, $\mathcal{Q}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)}$ is a decreasing function of the probability parameter p.

Proof. See Appendix 7. ∎

Figure 1 illustrates the ratio of information leakage without and with the global depolarizing channel versus the probability parameter. When the probability parameter rises, and therefore the global depolarizing channel becomes noisier, the maximal quantum leakage drops continuously.

The previous proposition demonstrated how maximal quantum leakage is affected by global depolarizing noise. However, in quantum computing devices, each qubit can be affected by local noise. Consider the case where the Hilbert space $\mathcal{H}_A$ is composed of k qubits, i.e., $d_A = 2^k$. In this case, we consider local depolarizing noise channel $\mathcal{D}_{p,2}^{\otimes k} := \mathcal{D}_{p,2} \otimes \cdots \otimes \mathcal{D}_{p,2}$, where a depolarizing channel $\mathcal{D}_{p,2}$ acts on each qubit separately. The effect of the local depolarizing channel on the maximal quantum leakage is investigated in the next proposition.

Proposition 6. For local depolarizing channel $\mathcal{D}_{p,2}^{\otimes k}$,

$$\mathcal{Q}(X \to A)_{\mathcal{D}_{p,2}^{\otimes k}(\rho_A)} \leqslant \log_2[p^k + (1-p^k)2^{\mathcal{Q}(X\to A)_{\rho_A}}].$$

Proof. See Appendix 8. ∎

Propositions 5 and 6 show that noisy intermediate-scale quantum (NISQ) devices inherently provide security and privacy. This is of course not surprising. The noise in NISQ devices has been shown to ensure quantum differential privacy [2,3] and privacy against hypothesis testing adversaries [6]. Noisy devices can also improve security of quantum machine learning models against adversarial attacks [17,18], albeit these guarantees go hand in hand with performance degradation [19].

Algorithm 1: Subgradient ascent algorithm for computing maximal quantum leakage.

Require: $\{\rho_A^x\}_{x\in\mathbb{X}}$, $\mathbb{Y} = \{1, \ldots, d_A^2\}$, $\mu > 0$, and $\epsilon > 0$
Ensure: $\{F_y\}$ and $\mathcal{Q}(X \to A)_{\rho_A}$
OldCost $\leftarrow \infty$
Random initialization $F_y \succeq 0$, $\forall y \in \mathbb{Y}$
NewCost $\leftarrow \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \mathrm{tr}(\rho_A^x F_y)$
while |OldCost − NewCost| $\geqslant \epsilon$ **do**
  OldCost $\leftarrow$ NewCost
  $x^*(y) \leftarrow \arg\max_{x\in\mathbb{X}} \mathrm{tr}(\rho_A^x F_y)$, $\forall y \in \mathbb{Y}$
  $G_y \leftarrow I + \mu(\rho_A^{x^*(y)} - \sum_{z\in\mathbb{Y}} \rho_A^{x^*(z)} F_z)$, $\forall y \in \mathbb{Y}$
  $\tilde{F}_y \leftarrow G_y^\dagger F_y G_y$
  $S \leftarrow \sum_{y\in\mathbb{Y}} \tilde{F}_y$
  $F_y \leftarrow S^{-1/2} \tilde{F}_y S^{-1/2}$
  NewCost $\leftarrow \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \mathrm{tr}(\rho_A^x F_y)$
end while
$\mathcal{Q}(X \to A)_{\rho_A} = \log_2(\text{NewCost})$
return $\{F_y\}$ and $\mathcal{Q}(X \to A)_{\rho_A}$

V. ITERATIVE ALGORITHM FOR COMPUTING MAXIMAL QUANTUM LEAKAGE

In this section, we follow the approach of [16] to compute the maximal quantum leakage using an iterative algorithm. The main difference here is the use of subgradient (as opposed to gradient) ascent. This is due to nondifferentiability of the cost function in maximal quantum leakage with respect to the POVMs (due to the inner maximization on $x \in \mathbb{X}$); see Secs. 14.2 and 14.3 of [20] for more information on subgradients and nonsmooth optimization. Note that

$$2^{\mathcal{Q}(X\to A)_{\rho_A}} = \sup_{\{F_y\}} \sum_{y\in\mathbb{Y}} \mathrm{tr}\big(\rho_A^{x^*(y)} F_y\big), \tag{6a}$$

$$\text{s.t. } 0 \preceq F_y, y \in \mathbb{Y}, \sum_{y\in\mathbb{Y}} F_y = I, \tag{6b}$$

where $\mathbb{Y} = \{1, \ldots, d_A^2\}$, $x^*(y) \in \arg\max_{x\in\mathbb{X}} \mathrm{tr}(\rho_A^x F_y)$, and d_A is the dimension of $\mathcal{H}_A$. If $\arg\max_{x\in\mathbb{X}} \mathrm{tr}(\rho_A^x F_y)$ admits a unique solution for all $y \in \mathbb{Y}$, the cost function is differentiable at that point and the subgradient and the gradient are equal to each other. If $\arg\max_{x\in\mathbb{X}} \mathrm{tr}(\rho_A^x F_y)$ does not admit a unique solution, selecting each solution results in a different subgradient. In fact, any convex combination of said subgradients will be also a subgradient. Note that after fixing $x^*(y)$, the cost function and constraints of this optimization problem in (6) have the same form as in [16] and, therefore, a similar approach can be used to compute the subgradients (instead of gradients). Algorithm 1 summarizes an iterative subgradient ascent method for computing the maximal leakage. In this algorithm, $\mu > 0$ is the step size and can be selected adaptively to keep the cost function increasing or can be set *a priori* small enough to ensure convergence. Furthermore, threshold $\epsilon > 0$ is selected to determine termination of the algorithm, i.e., the algorithm is terminated if the improvement in the cost function is not larger

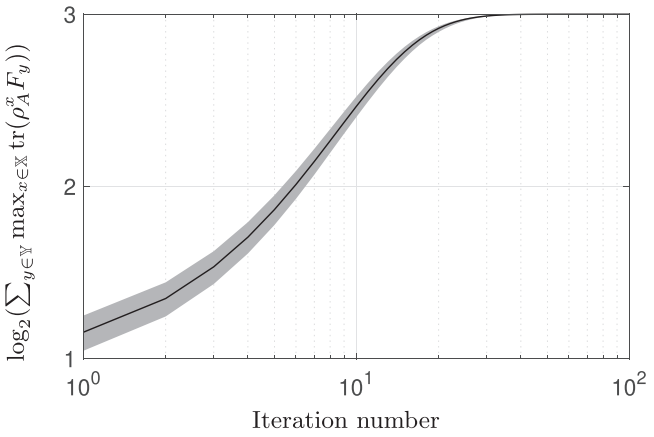

FIG. 2. Information leakage vs iterations of Algorithm 1 for the index encoding example.

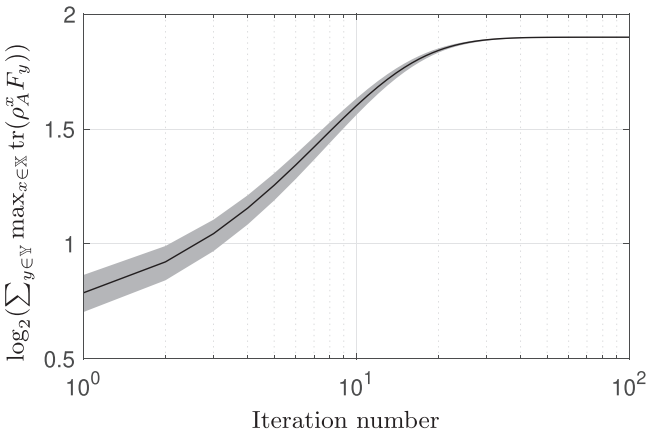

FIG. 3. Information leakage vs iterations of Algorithm 1 for the amplitude encoding example.

than ϵ (so not enough headway towards an optimal solution is made).

To demonstrate the validity of the proposed algorithm, we start by index encoding of classical data in quantum states for which maximal quantum leakage can be computed easily. Consider random variable X with support set $\mathbb{X} = \{1, \ldots, d_A\}$, where $d_A = 8$ denotes the dimension of $\mathcal{H}_A$. Let $\rho_A^x = |x\rangle\langle x|$ for all $x \in \mathbb{X}$. In this case, by selecting $F_y = |y\rangle\langle y|$ for $y \in \{1, \ldots, d_A\}$, we get $\log_2(\sum_{y\in\mathbb{Y}} \mathrm{tr}(\rho_A^{x^*(y)} F_y)) = \log_2(d_A) = 3$. This is the maximum attainable leakage according to Proposition 2. Therefore, $\mathcal{Q}(X \to A)_{\rho_A} = 3$ bits. Figure 2 illustrates the information leakage for iterations of Algorithm 1 starting from a random POVM. Following Theorem 2, we select $d_A^2 = 64$ as the number of elements in the starting POVM. Furthermore, we use $\mu = 10^{-1}$. The gray area demonstrates the maximum and minimum in each iteration (note the randomness in the initialization) and the solid black line shows the mean in each iteration. As expected, the algorithm rapidly converges to $\mathcal{Q}(X \to A)_{\rho_A} = 3$.

Now, we expand our attention to a more complex encoding strategy. Let $X = (X_1, X_2, X_3) \in \{0, 1\}^3$. Assume that $\mathcal{H}_A$ is a Hilbert space of dimension $d_A = 8$ and $\rho_A^x = |\psi^x\rangle\langle\psi^x|$, where $\psi^x = x_1|0\rangle + (1 - x_1)|1\rangle + x_2|2\rangle + (1 - x_2)|3\rangle + x_3|4\rangle + (1 - x_3)|5\rangle$. This is more similar to amplitude encoding that is often utilized in quantum machine learning. Figure 3 shows the information leakage versus iterations of Algorithm 1 starting from a random POVM. Similarly, we use $\mu = 10^{-1}$ and select $d_A^2 = 64$ as the number of elements in the starting POVM. The gray area demonstrates the maximum and minimum in each iteration and the solid black line shows the mean in each iteration. The algorithm rapidly converges to $\mathcal{Q}(X \to A)_{\rho_A} = 1.9$ bits. Interestingly, amplitude encoding seems to leak less information in comparison with index encoding.

VI. CONCLUSIONS AND FUTURE WORK

We considered an adversary that is interested in correctly guessing a potentially randomized function of a secret or private data with access to a single copy of the state of a quantum system encoding it. We proposed the notion of maximal quantum leakage, which captures the multiplicative increase in the probability of correctly guessing any function of the data upon observing measurements of the quantum state. We proved that maximal quantum leakage satisfies the postprocessing inequality and independence property and bounds accessible information. Future work can focus on the following topics.

(1) *Quantum wiretap*: As stated in the introduction, accessible information (related to mutual information) is an appropriate measure of information when considering lossless communication [11] while maximal quantum leakage generalizes the assumptions on the adversary and is perfect for investigating eavesdroppers. This motivates using a combination of accessible information and maximal quantum leakage in wiretap or obfuscation channels. Figure 4 illustrates a wiretap channel, where Alice wants to communicate effectively with Bob while minimizing the leaked information to Eve. Here, communication channels can be any completely positive trace-preserving mappings. Alice's strategy in the wiretap channel can be computed by finding an encoding policy $x \mapsto \rho_A^x$ that maximizes $I_{\mathrm{acc}}(\{p_X(x), \rho_B^x\}_{x\in\mathbb{X}})$ (i.e., the rate of information transfer to Bob) subject to $\mathcal{Q}(X \to E)_{\rho_E} \leqslant \varepsilon$ for small constant $\varepsilon \geqslant 0$ (i.e., restricts the amount of leaked information to Eve).

(2) *Generalization in quantum machine learning*: Classical maximal leakage has been already utilized to better understand generalization of machine learning models [21,22]. Therefore, we expect to be able to use maximal quantum leakage to analyze generalization of various quantum machine learning models.

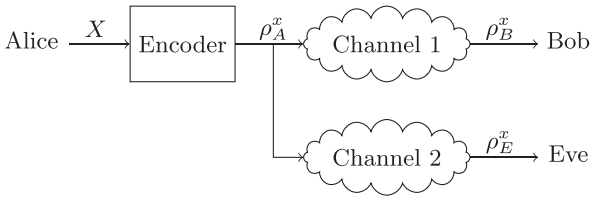

FIG. 4. A quantum wiretap channel, where Alice wants to communicate effectively with Bob while minimizing the leaked information to Eve.

(3) *Privacy-preserving quantum computing*: Privacy analysis in quantum system is relatively new with only recent studies on quantum differential privacy and puffer-fish privacy [2–6]. In many scenarios differential privacy can result in conservative results and bad performance. Information-theoretic privacy [23] can provide a systematic approach to balancing privacy and utility in general settings. The proposed notion of information leakage can provide an operational measure of privacy leakage for balancing utility and privacy in quantum systems.

APPENDIX: PROOFS OF ALL THE PRESENTED RESULTS

The proofs of all the results in Secs. II–IV are presented across the following subsections. The proofs are moved to this section to minimize interruptions to the flow of the paper and to focus on definitions and properties without excessive mathematics.

1. Proof of Theorem 1

The proof follows from the fact that

$$\begin{aligned}
&\sup_{Z,\widehat{Z}} \log_2 \left(\frac{\mathbb{P}\{Z=\widehat{Z}\}}{\max\limits_{z\in\mathbb{Z}} \mathbb{P}\{Z=z\}} \right) \\
&\quad = \log_2 \left(\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \mathbb{P}\{Y=y\,|\,X=x\} \right) \\
&\quad = \log_2 \left(\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}(\rho_A^x F_y) \right),
\end{aligned}$$

where the first equality follows from Theorem 1 of [8] and the second equality is a direct consequence of (1).

2. Proof of Theorem 2

The proof follows the same line of reasoning as the proof of Theorem 1 of [14], which is reformulated in Proposition 5.8 of [15].

Let $\mathcal{F}_k$ denote the set of POVMs with k outcomes, i.e., $\mathcal{F}_k := \{\{F_y\}_{y=1}^k | F_y \succeq 0, \sum_{y=1}^k F_y = I\}$. Define $g : \cup_{k\geqslant 0}\mathcal{F}_k \to \mathbb{R}_{\geqslant 0}$ as

$$g(F) := \sum_{y=1}^{k} \max_{x\in\mathbb{X}} \operatorname{tr}(\rho_A^x F_y)$$

where $F = \{F_y\}_{y=1}^k$. Evidently,

$$\mathcal{Q}(X\to A)_{\rho_A} = \log_2 \left(\sup_{F\in\cup_{k\geqslant 0}\mathcal{F}_k} g(F) \right).$$

The following lemma extends Lemma 2 of [14] to $g(\cdot)$.

Lemma 1. Consider $F = \{F_y\}_{y=1}^k \in \mathcal{F}_k$. Let $F' := \{F'_y\}_{y=1}^{k+1} \in \mathcal{F}_{k+1}$ be such that $F'_y = F_y$ for all $y \in \{1,\dots,k\}\setminus\{y_0\}$ for some $y_0 \in \{1,\dots,k\}$ and $F'_{y_0}, F'_{k+1} \succ 0$ satisfy $F_{y_0} = F'_{y_0} + F'_{k+1}$. Then, $g(F') \geqslant g(F)$.

Proof. Without loss of generality, up to rearranging the order of the elements in the POVMs, we can assume that $y_0 = k$. Note that

$$\begin{aligned}
g(F) &= \sum_{y=1}^{k} \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F_y\big) \\
&= \sum_{y=1}^{k-1} \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F'_y\big) + \max_{x\in\mathbb{X}} \operatorname{tr}\big[\rho_A^x \big(F'_k + F'_{k+1}\big)\big] \\
&\leqslant \sum_{y=1}^{k-1} \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F'_y\big) + \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F'_k\big) \\
&\quad + \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F'_{k+1}\big) \\
&= \sum_{y=1}^{k+1} \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F'_y\big), \\
&= g(F'),
\end{aligned}$$

where the inequality follows from

$$\begin{aligned}
&\max_{x\in\mathbb{X}} \operatorname{tr}\big[\rho_A^x (F'_k + F'_{k+1})\big] \\
&\quad = \max_{x\in\mathbb{X}} \big[\operatorname{tr}\big(\rho_A^x F'_k\big) + \operatorname{tr}\big(\rho_A^x F'_{k+1}\big)\big] \\
&\quad \leqslant \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F'_k\big) + \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F'_{k+1}\big).
\end{aligned}$$

This concludes the proof. ∎

Lemma 2. $F \mapsto g(F)$ is a convex on $\mathcal{F}_k$ for any $k \geqslant 0$.

Proof. Consider $F = \{F_y\}_{y=1}^k$ and $F' = \{F'_y\}_{y=1}^k$ such that $F, F' \in \mathcal{F}_k$. Evidently, $F'' := \{\lambda F_y + (1-\lambda)F'_y\}_{y=1}^k$ also belongs to $\mathcal{F}_k$ for all $\lambda \in [0,1]$ because $\lambda F_y + (1-\lambda)F'_y \succeq 0$ for all $y \in \{0,\dots,k\}$ and

$$\sum_{y=1}^{k} (\lambda F_y + (1-\lambda)F'_y) = \lambda \sum_{y=1}^{k} F_y + (1-\lambda)\sum_{y=1}^{k} F'_y = I.$$

Furthermore,

$$\begin{aligned}
g(F'') &= \sum_{y=1}^{k} \max_{x\in\mathbb{X}} \operatorname{tr}\big\{\rho_A^x [\lambda F_y + (1-\lambda)F'_y]\big\} \\
&= \sum_{y=1}^{k} \max_{x\in\mathbb{X}} \big[\lambda \operatorname{tr}\big(\rho_A^x F_y\big) + (1-\lambda)\operatorname{tr}\big(\rho_A^x F'_y\big)\big] \\
&\leqslant \sum_{y=1}^{k} \big[\max_{x\in\mathbb{X}} \lambda \operatorname{tr}\big(\rho_A^x F_y\big) + \max_{x\in\mathbb{X}} (1-\lambda)\operatorname{tr}\big(\rho_A^x F'_y\big)\big] \\
&= \lambda \sum_{y=1}^{k} \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F_y\big) + (1-\lambda)\sum_{y=1}^{k} \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x F'_y\big) \\
&= \lambda g(F) + (1-\lambda) g(F').
\end{aligned}$$

This concludes the proof. ∎

Fix $k \geqslant 0$. Because $g(F)$ is continuous in F and $\mathcal{F}_k$ is compact, $\sup_{F\in\mathcal{F}_k} g(F)$ is attained on the set $\mathcal{F}_k$. Assume that $\hat{F} = \{\hat{F}_y\}_{y=1}^k$ maximizes g on $\mathcal{F}_k$. By removing zero components, if necessary, we obtain POVM $\tilde{F} = \{\tilde{F}_y\}_{y=1}^{\ell}$ with $k \geqslant \ell$

such that $g(\hat{F}) = g(\tilde{F})$. Using spectral decomposition, we can decompose each element of the POVM $\tilde{F}$ into a sum of rank-1 elements, i.e., $\tilde{F}_y = \sum_{z=1}^{z_y} \bar{F}_z$, where $\bar{F}_z \succ 0$ are rank-1 matrices. Construct POVM $\bar{F} := \{\{\bar{F}_z\}_{z=1}^{z_y}\}_{y=1}^{\ell}$. Let $m = |\bar{F}|$. By repeatedly using Lemma 1, we can see that $g(\bar{F}) \geqslant g(\hat{F})$. Because of Lemma 2, g is convex on $\mathcal{F}_m$. Therefore, the maximizing observable $\check{F}$ is an extreme point of the set $\mathcal{F}_m$. Theorem 2.21 of [15] shows that $\check{F}$ must be linearly independent and the Carathéodory theorem (see, e.g., Theorem 1.6 of [15]) shows that $\check{F}$ can be represented as a convex combination of at most d_A^2 elements, where d_A is the dimension of $\mathcal{H}_A$. Therefore, without loss of generality, by removing some zero components, $\check{F} \in \mathcal{F}_{d_A^2}$. Since k was chosen arbitrary, $\sup_{F \in \cup_{k\geqslant 0}\mathcal{F}_k} g(F)$ is attained on an observable from the compact convex set $\mathcal{F}_{d_A^2}$.

3. Proof of Proposition 1: Establishing the independence property

Define $I_\infty(X;Y) := \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \mathbb{P}\{Y = y \,|\, X = x\} = \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}(\rho_A^x F_y)$. Note that $I_\infty(X;Y) \geqslant 0$ Lemma 1 of [8] and, as a result, $\mathcal{Q}(X \to A)_{\rho_A} = \sup_{\{F_y\}_y} I_\infty(X;Y) \geqslant 0$.

Now, we prove that $\mathcal{Q}(X \to A)_{\rho_A} = 0$ if $(\rho_A^x)_{x\in\mathbb{X}}$ is indistinguishable. Consider $x, x' \in \mathbb{X}$ such that $x \neq x'$. We have

$$\begin{aligned}\mathbb{P}\{Y = y \,|\, X = x\} &= \operatorname{tr}\big(\rho_A^x F_y\big)\\ &= \operatorname{tr}\big(\rho_A^{x'} F_y\big)\\ &= \mathbb{P}\{Y = y \,|\, X = x'\},\end{aligned}$$

where the first and the third equality follow from (1) and the second equality follows from indistinguishability of $(\rho_A^x)_{x\in\mathbb{X}}$, i.e., $\rho_A^x = \rho_A^{x'}$ for all $x, x' \in \mathbb{X}$. Therefore, random variables X and Y are independent, which implies that $I_\infty(X;Y) = 0$ [8], Lemma 1, which is true irrespective of the choice of $\{F_y\}_y$. This implies that $\mathcal{Q}(X \to A)_{\rho_A} = \sup_{\{F_y\}_y} I_\infty(X;Y) = 0$.

Next, we prove that $(\rho_A^x)_{x\in\mathbb{X}}$ is indistinguishable if $\mathcal{Q}(X \to A)_{\rho_A} = 0$. Because $\mathcal{Q}(X \to A)_{\rho_A} = 0$, it must be that $I_\infty(X;Y) = 0$ for all POVMs $F = \{F_y\}_y$. Following Lemma 1 of [8], it must be that X and Y are independent for all POVMs F. We have

$$\begin{aligned}\operatorname{tr}\big(\rho_A^x F_y\big) &= \mathbb{P}\{Y = y \,|\, X = x\}\\ &= \mathbb{P}\{Y = y \,|\, X = x'\}\\ &= \operatorname{tr}\big(\rho_A^{x'} F_y\big),\end{aligned}$$

where the second equality follows from statistical independence of X and Y. Therefore, $\operatorname{tr}[(\rho_A^x - \rho_A^{x'})F_y] = 0$ for all $0 \preceq F_y \preceq I$. This implies that $\rho_A^x - \rho_A^{x'} = 0$, or equivalently $\rho_A^x = \rho_A^{x'}$, which concludes the proof.

4. Proof of Proposition 2

Let $\{F_y\}_y$ be the maximizing POVM in Theorem 2. Then, $\mathcal{Q}(X \to A) = I_\infty(X;Y)$. Following Lemma 1 of [8], $I_\infty(X;Y) \leqslant \min\{|\mathbb{X}|, |\mathbb{Y}|\}$. Furthermore, $|\mathbb{Y}| = m \leqslant d_A^2$. Therefore, $\mathcal{Q}(X \to A)_{\rho_A} \leqslant \min\{\log_2(|\mathbb{X}|), \log_2(d_A^2)\}$.

5. Proof of Proposition 3: Establishing the data processing inequality

Note that

$$\mathcal{Q}(X \to A)_{\mathcal{E}(\rho_A)} = \sup_{\{F_y\}\in\mathcal{F}} \log_2 \left(\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}(\mathcal{E}\big(\rho_A^x\big)F_y)\right),$$

where $\mathcal{F}$ denotes the set of all POVMs. According to the Choi-Kraus theorem Theorem 4.4.1 of [10], for each quantum channel $\mathcal{E}$, there exists a family of linear operators $\{E_j\}_{j=1}^n$ for some $n \in \mathbb{N}$ such that $\sum_{j=1}^n E_j^\dagger E_j = I$ and $\mathcal{E}(\rho) = \sum_{j=1}^n E_j \rho E_j^\dagger$ for all density operators ρ. Consider any POVM $F = \{F_y\}_{y=1}^m$. We have

$$\begin{aligned}\operatorname{tr}(\mathcal{E}\big(\rho_A^x\big)F_y) &= \operatorname{tr}\left(\sum_{j=1}^n E_j \rho_A^x E_j^\dagger F_y\right)\\ &= \operatorname{tr}\left[\rho_A^x \left(\sum_{j=1}^n E_j^\dagger F_y E_j\right)\right]\\ &= \operatorname{tr}\big(\rho_A^x \bar{F}_y\big),\end{aligned}$$

where

$$\bar{F}_y := \sum_{j=1}^n E_j^\dagger F_y E_j.$$

Define $\bar{\mathcal{F}} := \{\{\bar{F}_y\}_y \,|\, \bar{F}_y = \sum_{j=1}^n E_j^\dagger F_y E_j, \{F_y\}_y \in \mathcal{F}\}$. Evidently, by construct,

$$\mathcal{Q}(X \to A)_{\mathcal{E}(\rho_A)} = \sup_{\{\bar{F}_y\}\in\bar{\mathcal{F}}} \log_2 \left(\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\big(\rho_A^x \bar{F}_y\big)\right).$$

Let us prove that $\bar{\mathcal{F}} \subset \mathcal{F}$. Let $\{\bar{F}_y\}_y \in \bar{\mathcal{F}}$. Then, there must exist $\{F_y\}_y \in \mathcal{F}$ such that $\bar{F}_y = \sum_{j=1}^n E_j^\dagger F_y E_j$. Note that $\bar{F}_y \succeq 0$ because $F_y \succeq 0$. Furthermore, $\bar{F}_y = \sum_{j=1}^n E_j F_y E_j^\dagger \preceq I$ because $F_y \preceq I$. Furthermore,

$$\begin{aligned}\sum_y \bar{F}_y &= \sum_y \sum_{j=1}^n E_j^\dagger F_y E_j\\ &= \sum_{j=1}^n E_j^\dagger \left(\sum_y F_y\right) E_j\\ &= \sum_{j=1}^n E_j^\dagger E_j\\ &= I,\end{aligned}$$

where the third equality follows from the fact that $\sum_y F_y = I$ and the last equality follows from the fact that $\sum_{j=1}^n E_j^\dagger E_j = I$. Therefore, $\{\bar{F}_y\}_y$ must belong to $\mathcal{F}$, which proves that

$\bar{\mathcal{F}} \subset \mathcal{F}$. This implies that

$$\begin{aligned}\mathcal{Q}(X \to A)_{\mathcal{E}(\rho_A)} &= \sup_{\{\bar{F}_y\}\in\bar{\mathcal{F}}} \log_2 \left(\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x \bar{F}_y\right)\right) \\ &\leqslant \sup_{\{\bar{F}_y\}\in\mathcal{F}} \log_2 \left(\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x \bar{F}_y\right)\right) \\ &= \mathcal{Q}(X \to A)_{\rho_A},\end{aligned}$$

where the inequality follows from the fact that taking the supremum over a larger set results in a larger value.

6. Proof of Proposition 4: Relationship between accessible information and maximal quantum leakage

Note that $I_\infty(X;Y) := \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \mathbb{P}\{Y = y \,|\, X = x\} = \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}(\rho_A^x F_y)$ Theorem 1 of [8]. From Lemma 2 of [8], we know that $I_\infty(X;Y) \geqslant I(X;Y)$. The rest follows from taking the supremum with respect to POVM $\{F_y\}_y$ on both sides of the inequality $I_\infty(X;Y) \geqslant I(X;Y)$.

7. Proof of Proposition 5: Effect of the global depolarizing channel

Note that

$$\begin{aligned}\operatorname{tr}(\mathcal{D}_{p,d_A}(\rho_A^x)F_y) &= \operatorname{tr}\left[\left(\frac{p}{d_A} I + (1-p)\rho_A^x\right)F_y\right] \\ &= \frac{p}{d_A}\operatorname{tr}(F_y) + (1-p)\operatorname{tr}\left(\rho_A^x F_y\right).\end{aligned}$$

Therefore,

$$\begin{aligned}&\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left[\mathcal{D}_{p,d_A}(\rho_A^x)F_y\right] \\ &= \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \left(\frac{p}{d_A}\operatorname{tr}(F_y) + (1-p)\operatorname{tr}\left(\rho_A^x F_y\right)\right) \\ &= \frac{p}{d_A}\sum_{y\in\mathbb{Y}} \operatorname{tr}(F_y) + (1-p)\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x F_y\right) \\ &= \frac{p}{d_A}\operatorname{tr}\left(\sum_{y\in\mathbb{Y}} F_y\right) + (1-p)\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x F_y\right) \\ &= \frac{p}{d_A}\operatorname{tr}(I) + (1-p)\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x F_y\right) \\ &= p + (1-p)\sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x F_y\right).\end{aligned}$$

This implies that

$$\begin{aligned}&\mathcal{Q}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)} \\ &= \log_2 \left(\sup_{\{F_y\}_y} \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}(\mathcal{D}_{p,d_A}(\rho_A^x)F_y)\right) \\ &= \log_2 \left(p + (1-p)\sup_{\{F_y\}_y} \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x F_y\right)\right) \\ &= \log_2(p + (1-p)2^{\mathcal{Q}(X\to A)_{\rho_A}}).\end{aligned}$$

8. Proof of Proposition 6: Effect of the local depolarizing channel

Following the proof of Lemma IV.4 of [3], a local depolarizing noise channel can be always represented as

$$\mathcal{D}_{p,2}^{\otimes k} = p^k \frac{I}{2^k} + (1-p^k)\mathcal{M}(p),$$

where $\mathcal{M}$ is an appropriately selected quantum channel, i.e., completely positive and trace preserving mapping. Using the same line of reasoning as in the proof of Proposition 5 in Appendix 7, we get

$$\begin{aligned}&\mathcal{Q}(X \to A)_{\mathcal{D}_{p,2}^{\otimes k}(\rho_A)} \\ &= \log_2 \left(p^k + (1-p^k)\sup_{\{F_y\}_y} \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\mathcal{M}(\rho_A^x)F_y\right)\right). \quad \text{(A1)}\end{aligned}$$

From Proposition 3, we know that

$$\sup_{\{F_y\}_y} \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\mathcal{M}(\rho_A^x)F_y\right) \leqslant \sup_{\{F_y\}_y} \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x F_y\right). \quad \text{(A2)}$$

Combining (A1) and (A2) gives

$$\begin{aligned}&\mathcal{Q}(X \to A)_{\mathcal{D}_{p,2}^{\otimes k}(\rho_A)} \\ &\leqslant \log_2 \left(p^k + (1-p^k)\sup_{\{F_y\}_y} \sum_{y\in\mathbb{Y}} \max_{x\in\mathbb{X}} \operatorname{tr}\left(\rho_A^x F_y\right)\right) \\ &= \log_2[p^k + (1-p^k)2^{\mathcal{Q}(X\to A)_{\rho_A}}].\end{aligned}$$

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[2] L. Zhou and M. Ying, Differential privacy in quantum computation, in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)* (IEEE, New York, 2017), pp. 249–262.

[3] C. Hirche, C. Rouzé, and D. S. França, Quantum differential privacy: An information theory perspective, IEEE Trans. Inf. Theory **69**, 5771 (2023).

[4] S. Aaronson and G. N. Rothblum, Gentle measurement of quantum states and differential privacy, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing STOC 2019* (Association for Computing Machinery, New York, NY, 2019), pp. 322–333.

[5] T. Nuradha, Z. Goldfeld, and M. M. Wilde, Quantum pufferfish privacy: A flexible privacy framework for quantum systems, arXiv:2306.13054.

[6] F. Farokhi, Privacy against hypothesis-testing adversaries for quantum computing, arXiv:2302.12405.

[7] N. Cai, A. Winter, and R. W. Yeung, Quantum privacy and quantum wiretap channels, Probl. Inf. Transm. **40**, 318 (2004).

[8] I. Issa, A. B. Wagner, and S. Kamath, An operational approach to information leakage, IEEE Trans. Inf. Theory **66**, 1625 (2019).

[9] F. Farokhi and N. Ding, Measuring information leakage in non-stochastic brute-force guessing, in *2020 IEEE Information Theory Workshop (ITW)* (IEEE, New York, 2021), pp. 1–5.

[10] M. Wilde, *Quantum Information Theory* (Cambridge University, New York, 2013).

[11] A. S. Holevo, Bounds for the quantity of information transmitted by a quantum communication channel, Prob. Peredachi Inf. **9**, 3 (1973).

[12] J. Liao, O. Kosut, L. Sankar, and F. du Pin Calmon, Tunable measures for information leakage and applications to privacy-utility tradeoffs, IEEE Trans. Inf. Theory **65**, 8043 (2019).

[13] M. Diaz, H. Wang, F. P. Calmon, and L. Sankar, On the robustness of information-theoretic privacy measures and mechanisms, IEEE Trans. Inf. Theory **66**, 1949 (2019).

[14] E. Davies, Information and quantum measurement, IEEE Trans. Inf. Theory **24**, 596 (1978).

[15] A. S. Holevo, *Quantum Systems, Channels, Information: A Mathematical Introduction* (De Gruyter, Boston, 2019).

[16] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Iterative procedure for computing accessible information in quantum communication, Phys. Rev. A **71**, 054303 (2005).

[17] Y. Du, M.-H. Hsieh, T. Liu, D. Tao, and N. Liu, Quantum noise protects quantum classifiers against adversaries, Phys. Rev. Res. **3**, 023153 (2021).

[18] M. Weber, N. Liu, B. Li, C. Zhang, and Z. Zhao, Optimal provable robustness of quantum classification via quantum hypothesis testing, npj Quantum Inf. **7**, 76 (2021).

[19] S. Resch and U. R. Karpuzcu, Benchmarking quantum computers and the impact of quantum noise, ACM Comput. Surv. (CSUR) **54**, 1 (2021).

[20] W. Sun and Y. X. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, Springer Optimization and its Applications (Springer, New York, 2006).

[21] A. R. Esposito, M. Gastpar, and I. Issa, Generalization error bounds via Rényi-, f-divergences and maximal leakage, IEEE Trans. Inf. Theory **67**, 4986 (2021).

[22] A. R. Esposito, M. Gastpar, and I. Issa, A new approach to adaptive data analysis and learning via maximal leakage, arXiv:1903.01777.

[23] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, From the information bottleneck to the privacy funnel, in *2014 IEEE Information Theory Workshop (ITW 2014)* (IEEE, New York, 2014), pp. 501–505.