

Sample complexity of purity and inner product estimation

Weiyuan Gong^{1,*}, Jonas Haferkamp^{1,†}, Qi Ye^{2,‡} and Zhihan Zhang^{2,§}¹*School of Engineering and Applied Sciences, Harvard University, Allston, Massachusetts 02134, USA*²*Center for Quantum Information, IIIS, Tsinghua University, Beijing 100084, China*

(Received 25 February 2025; accepted 8 January 2026; published 20 October 2026)

We study the sample complexity of the prototypical tasks *quantum purity estimation* and *quantum inner product estimation*. The former has been a key subroutine for experimental benchmarking and entropy estimation, while the latter is a core task for cross-platform validation that compares the unknown quantum states produced by different physical platforms. In purity estimation, we are to estimate $\text{tr}(\rho^2)$ of an unknown quantum state ρ to additive error ε . Meanwhile, for quantum inner product estimation, Alice and Bob are to estimate $\text{tr}(\rho\sigma)$ to additive error ε , given copies of the unknown quantum states ρ and σ using classical communication and restricted quantum communication. We propose a protocol that solves quantum inner product estimation with k -qubit *one-way* quantum communication and unentangled local measurements using $O(\text{median}\{1/\varepsilon^2, 2^{n/2}/\varepsilon, 2^{n-k}/\varepsilon^2\})$ copies of ρ and σ . Our protocol can be modified to estimate the purity of an unknown quantum state ρ using k -qubit quantum memory with the same sample complexity. We also prove that arbitrary protocols with k -qubit quantum memory that estimate purity to error ε require $\Omega(\text{median}\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\})$ copies of ρ . This indicates the same sample complexity lower bound for quantum inner product estimation with *one-way* k -qubit quantum communication and classical communication, and unentangled local measurements. We further improve the sample complexity lower bound for any purity estimation protocols using an identical single-copy projection-valued measurement. Last, we investigate a decisional variant of quantum distributed inner product estimation without quantum communication for mixed states and provide a sample complexity lower bound.

DOI: [10.1103/kh2j-jkn6](https://doi.org/10.1103/kh2j-jkn6)

I. INTRODUCTION

In recent years, notable progress has been made in the development of noisy intermediate-scale quantum devices utilizing different physical platforms [1]. Benchmarking the physical systems on these quantum devices is of fundamental importance for verifying the computational outputs [2]. However, the increasing size of quantum devices has pushed the standard validation methods, such as state tomography [3,4] and state certification [5] to the limit of their capabilities due to the exponential scaling of resource requirement in system size n . These certification algorithms focus on the task of comparing the noisy quantum state (process) with a known target. It is therefore desirable to develop protocols with feasible resource requirements for cross-platform validation [6–8]. Such a protocol should directly compare unknown quantum states generated by different devices, which is an essential task in quantum benchmarking.

In this work, we focus on the task of *quantum inner product estimation*. Here, Alice and Bob are given copies of possibly different quantum states ρ and σ , respectively. Their goal is to estimate the inner product $\text{tr}(\rho\sigma)$ up to additive error ε with a high probability. When Alice and Bob are able to share arbitrary quantum communication channels, a joint SWAP test suffices to efficiently estimate $\text{tr}(\rho\sigma)$. However, quantum communication between Alice and Bob is limited in practice due to hardware constraints. In the extreme case when Alice and Bob are restricted to local quantum operations and classical communication (LOCC), this task is known as *distributed quantum inner product estimation* [9]. It was shown in the same work that $\Theta(2^{n/2})$ copies of ρ and σ are sufficient and necessary to solve distributed inner product estimation to constant accuracy ε even if joint local quantum measurements are allowed. An efficient algorithm for inner product estimation based on Pauli sampling has been provided for quantum states with low magic and entanglement [10]. An experimental demonstration of this protocol among different existing quantum platforms is provided in Ref. [11]. An alternative cross-platform protocol using deep learning is proposed in Ref. [12].

II. CONNECTIONS BETWEEN INNER PRODUCT AND PURITY ESTIMATION

Recent experimental advances in realizing quantum local area networks [13–15] enable cross-platform validation schemes with quantum communication channels [16]. The

*Contact author: wgong@g.harvard.edu

†Contact author: jhaferkamp42@gmail.com

‡Contact author: yeq22@mails.tsinghua.edu.cn

§Contact author: zhihan-z21@mails.tsinghua.edu.cn

capacities of these quantum wires between different platforms are restricted away from the n -qubit quantum communication channel that is required for joint SWAP tests. However, there is no rigorous theoretical understanding of the complexity for cross-platform validation with bounded quantum communications so far. This motivates the characterization of the sample complexity in terms of quantum communication [17]. Formally, we investigate the complexity of the following task:

Problem 1 (Inner product estimation with k -qubit one-way quantum communication). Alice and Bob are given N copies of the unknown quantum states ρ and σ , respectively. In each round of measurement, Alice and Bob each get access to one copy of ρ and σ , and are able to perform k -qubit one-way quantum communication and any one-way classical communication. The goal is to estimate the quantum inner product $\text{tr}(\rho\sigma)$ within additive error ε .

Here, we focus on the scheme where Alice and Bob are only allowed to perform unentangled (i.e., single-copy) measurements locally as it is enough to provide the optimal sample complexity of $O(2^{n/2})$ when $k = 0$ for distributed quantum inner product estimation [9] and local joint measurements are relatively impractical in experiments. In addition, we assume that the communication between Alice and Bob is restricted to a one-way communication from Alice to Bob. The setting where Alice and Bob are allowed to have arbitrary k -qubit quantum communication and classical communication is beyond the scope of this work and remains an open question.

While testing properties among multiple quantum systems in a distributed setting is very different from testing a single system, we demonstrate strong connections between the sample complexity of *purity estimation* and quantum inner product estimation in different measurement schemes. Purity estimation, which estimates the purity $\text{tr}(\rho^2)$ to additive error ε given copies of ρ , emerges as a key subroutine in many quantum information problems such as quantum benchmarking [2] and experimental estimation of quantum Rényi entropy on near-term devices [18]. As pointed out by Ref. [9], a single-copy protocol for distributed inner product estimation, which corresponds to Problem 1 with $k = 0$, indicates a single-copy protocol for purity estimation. Furthermore, by setting $\rho = \sigma$, we can observe that any algorithm for Problem 1 indicates an algorithm for distributed estimation of $\text{tr}(\rho^2)$ with bounded one-way quantum communication between Alice and Bob. It is thus natural to ask whether we can show connections of sample complexity between Problem 1 and the purity estimation setting without communications.

We answer this question affirmatively by showing the same sample complexity scaling for quantum inner product estimation with k -qubit one-way quantum communication (Problem 1) and purity estimation with k -qubit quantum memory (as shown in Fig. 1), which is defined as the following:

Problem 2 (Purity estimation with k -qubit quantum memory, informal). Given N copies of the unknown quantum state ρ , the goal is to estimate $\text{tr}(\rho^2)$ within additive error ε . The algorithm can take two copies of ρ at a time, and maintain n working qubits and a k -qubit quantum memory to store the data. In each round, the algorithm first makes a quantum measurement on the first copy and stores k qubits of the post-measurement state in the quantum memory. The algorithm

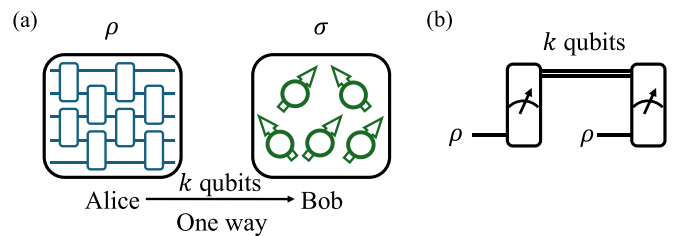


FIG. 1. Comparison of the models for (a) quantum inner product estimation with k -qubit one-way quantum communication and any classical one-way communication (Problem 1) and (b) purity estimation with k -qubit quantum memory (Problem 2).

then performs a quantum measurement on the $n + k$ qubits on the second copy and the quantum memory.

Regarding the existing results for purity estimation of unknown quantum states, well-known algorithms include randomized benchmarking based on single-copy measurements of sample complexity $O(\max\{2^{n/2}/\varepsilon, 1/\varepsilon^2\})$ and SWAP test based on two-copy measurements of sample complexity $O(1/\varepsilon^2)$. An $\Omega(2^{n/2} + 1/\varepsilon^2)$ sample complexity lower bound for single-copy protocols of purity estimation is also provided in Ref. [19]. However, there is still a gap in the optimal ε dependence for ancilla-free protocols. It also remains an open question to understand the quantum memory trade-offs for purity estimation with and without bounded quantum memory between single-copy protocols and two-copy protocols, which has been widely investigated by a long line of recent works to understand the relationship between near-term constraints on quantum devices and algorithms for various quantum learning tasks, and the underlying statistical complexity. A polynomial sample complexity gap for mixedness testing with and without quantum memory was first demonstrated in Ref. [20]. Subsequent works [21–26] further demonstrated exponential separations for a quantum memory for various tasks, including shadow tomography, unitary learning, purity testing, and quantum dynamics learning. In particular, the relationship between the number of qubits in quantum memory and estimating Pauli observables is fully characterized in Refs. [24,27]. Memory-sample trade-offs are also revealed for other tasks such as purity testing [27] and Pauli channel eigenvalue estimation [28,29]. In addition, a classical learning problem that exhibits a quantum memory trade-off was recently proposed [30].

We note that, in contrast to the definition of learning with k -qubit quantum memory in Refs. [24,27] that can take N oracle access to ρ and store data in the quantum memory before the final measurement on $\rho^{\otimes N}$, the definition here follows the definition of learning with two-copy measurements and k -qubit quantum memory in Ref. [27], which only allows the usage of at most *two copies* of ρ in each round (see formal definition in Definition 5). This measurement setting is more feasible in experiments and possesses the connection to the task of inner production estimation with unentangled local measurements between *two* platforms.

III. THE UPPER BOUND

As mentioned above (and rigorously in Appendix B), we show that any protocol for inner product estimation where

Alice and Bob share k qubits of one-way quantum communication and any one-way classical communication (Problem 1) indicates a protocol for purity estimation with k -qubit quantum memory (Problem 2) with the same sample complexity scaling. Therefore, inversely, the lower bound for purity estimation with k -qubit quantum memory indicates a lower bound of the same scaling for inner product estimation where Alice and Bob share k qubits of one-way quantum communication and any one-way classical communication. Using this connection, our first main result gives an algorithm to Problem 1 and thus Problem 2:

Theorem 1 (Informal). Given unknown quantum states ρ and σ , and error parameter ε , there exists a nonadaptive algorithm that solves inner product estimation of $\text{tr}(\rho\sigma)$ with k -qubit one-way quantum communication (Problem 1) using $O(\text{median}\{1/\varepsilon^2, 2^{n/2}/\varepsilon, 2^{n-k}/\varepsilon^2\})$ copies of ρ and σ .

Our protocol is a combination of the distributed inner product estimation protocol with single-copy measurements [9], which gives the upper bound of $O(\max\{1/\varepsilon^2, 2^{n/2}/\varepsilon\})$, and a partial SWAP test algorithm (see Algorithm 2 and Appendix D 1 for details) on the first k qubits of ρ and σ , which gives the upper bound of $O(2^{n-k}/\varepsilon^2)$. In the regime where $\varepsilon \geq 1/\text{poly}(n)$ is at least inverse polynomial in system size, the upper bound can be further reduced to $O(\min\{2^{n/2}/\varepsilon, 2^{n-k}/\varepsilon^2\})$. Combining purity estimation with inner product estimation using the algorithm in Theorem 1, we can estimate fidelitylike measures [6,31] and the squared Hilbert-Schmidt distance $D_{\text{HS}}^2(\rho, \sigma) = \text{tr}(\rho^2) + \text{tr}(\sigma^2) - 2\text{tr}(\rho\sigma)$ with restricted quantum communication and memory.

IV. LOWER BOUNDS

Given the algorithm in Theorem 1, it is natural to ask about the optimality of the algorithm. Our next result provides a lower bound for Problem 2 and thus a lower bound for Problem 1:

Theorem 2 (Informal). For any $0 \leq \varepsilon < 1$, any (possibly adaptive) protocols for estimating purity with k -qubit memory (inner product estimation with one-way k -qubit quantum communication) requires at least $\Omega(\text{median}\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\})$ copies of ρ (ρ and σ).

The detailed proof is technically involved and thus left to Appendixes C 1 and C 2. The high-level strategy of our proof is to combine the learning tree framework [24,25] with the martingale trick [27,29,32]. We model a learning protocol as a decision tree with the depth representing the number of samples, and any choice of unknown quantum state ρ corresponds to a distribution on the leaves. We then consider distinguishing between the maximally mixed state ρ_{mm} with purity 2^{-n} from a state ρ chosen from an ensemble such that $\text{tr}(\rho^2) > 2^{-n} + \varepsilon$ with high probability. We show that the induced probability distributions on the leaves corresponding to the two cases are statistically indistinguishable unless the decision tree is deep enough. The ensemble we employed to prove the purity estimation sample complexity lower bound for strategies with bounded quantum memory in Theorem 2 is significantly different from the previous work [20,24]. When proving lower bounds using the learning tree technique in these previous works, the distinguishing task considered is comparing the

maximally mixed state and an ensemble of quantum states constructed by adding a tiny random perturbation O_{perturb} to the maximally mixed state in the form $(I + O_{\text{perturb}})/2^n$. To ensure that the perturbed state has at least an $O(\varepsilon)$ deviation from the maximally mixed state, the amplitude of the perturbation has to be at least $O(\sqrt{\varepsilon})$. However, this distinguishing task will yield an $\Omega(\max\{2^{n/2}/\varepsilon^{1/4}, 1/\varepsilon^2\})$ sample complexity lower bound using the learning tree framework, which is looser than our result. In contrast, our random ensemble is constructed by adding a random perturbation in the enlarged Hilbert space and partially tracing out the ancillary system.

We note that at $k = 0$, the sample complexity lower bound for purity testing using single-copy measurement provided by Theorem 2 is $\Omega(\max\{2^{n/2}/\varepsilon^{1/2}, 1/\varepsilon^2\})$. This lower bound is weaker than the existing lower bound of $\Omega(\max\{2^{n/2}/\varepsilon, 1/\varepsilon^2\})$ for distributed inner product estimation with only LOCC [9]. However, the hard instance used for inner product estimation cannot be extended to purity estimation as inner product estimation involves in different states ρ and σ , indicating that the lower bound for the latter is more difficult to obtain than the former.

It is also natural to wonder whether the hardness of purity estimation may vary for different purity regimes like quantum state certification [21]. To this end, we provide negative evidence by showing that given an arbitrary purity value, we can construct a hard instance such that estimating the purity to epsilon accuracy around this value requires $\Omega(\max\{2^{n/2}/\varepsilon^{1/2}, 1/\varepsilon^2\})$ copies for any strategies using single-copy measurements in Appendix C 3. In other words, the hardness of purity estimation is *universal* around any *constant* purity value for strategies using single-copy measurements.

Proposition 1 (Informal). For any $0 \leq \varepsilon < 1$ and constant $0 < a < 1 - \varepsilon$, estimating purity close to a^2 requires $\Omega(\max\{2^{n/2}/\sqrt{\varepsilon}, 1/\varepsilon^2\})$ samples for single-copy strategies.

It is helpful to first interpret this in the regime where $\varepsilon \geq 1/\text{poly}(n)$. Within this regime for both purity estimation and inner product estimation, the lower bound of sample complexity is further reduced to $\Omega(\min\{2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\})$. Our result indicates that with $k \leq n/2$ qubits of quantum memory and quantum communication, no protocol can improve upon the sample complexity achieved by single-copy measurements and distributed inner product estimation with single-copy local measurements, respectively. This sample-memory trade-off is similar to the sample-memory trade-off for purity testing [24,27,33], which can be solved by purity estimation with $\varepsilon \simeq 1/2$. However, the proof of the sample complexity lower bound in Theorem 2 is significantly different from obtaining the ε dependence. For general ε , the lower bound, which can also be written as $\Omega(\max\{1/\varepsilon^2, \min\{2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\}\})$, can be understood approximately as a combination of the bound $\Omega(1/\varepsilon^2)$ for the tiny accuracy regime where $\varepsilon = O(2^{-n})$, and $\Omega(\max\{2^{n-k}/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}\})$ for the trade-off of estimation with the k quantum memory qubit (one-way quantum communication).

For the ancilla-free case ($k = 0$), one may be tempted to employ the hard instance for the mixedness testing task [21], which tests whether a state is maximally mixed or epsilon away from maximally mixed in terms of trace distance. Given that the mixedness testing sample complexity lower bound for

any single-copy strategy is $\Omega(2^{3n/2}/\varepsilon)$, we fix the accuracy demand $\varepsilon' = \varepsilon/2^n$ for purity estimation due to the d factor between trace distance and Hilbert-Schmidt norm and obtain an $\Omega(2^{n/2}/\varepsilon)$ sample complexity lower bound. However, we have to point out that the accuracy required using this reduction is always smaller than $1/2^n$. In this regime, the lower bound of sample complexity $\Omega(2^{n/2}/\varepsilon)$ obtained by reduction to mixedness tests is not optimal because it is always smaller than our result, giving $\Omega(1/\varepsilon^2)$.

We note that compared to Theorem 1 that gives an $O(\text{median}\{1/\varepsilon^2, 2^{n/2}/\varepsilon, 2^{n-k}/\varepsilon^2\})$ sample complexity upper bound, the lower bound in Theorem 2 exhibits a $\sqrt{\varepsilon}$ gap from the $2^{n/2}/\varepsilon$ term in the upper bound. While for general quantum protocols closing this gap remains open, we prove an improved lower bound of $\Omega(2^{n/2}/\varepsilon)$ for purity estimation (and thus distributed inner product estimation) using an identical projection-valued measurement (PVM) in every iteration:

Theorem 3 (Informal). For any $0 \leq \varepsilon < 1$, any protocols for estimating the purity and distributed inner product estimation using an identical single-copy PVM in every round require at least $\Omega(\max\{1/\varepsilon^2, 2^{n/2}/\varepsilon\})$ copies of ρ .

Although the assumption that the algorithm implements only one identical PVM throughout the protocol looks too restrictive, it captures the known algorithms for distributed inner product estimation and purity estimation [6,9], and indicates that these protocols are optimal under this measurement restriction.

While our method of proving lower bounds for purity estimation also implies lower bounds for inner product estimation, it says little about the decision version of this problem. In particular, this leaves the possibility of deciding efficiently whether two devices show the same state or not. The latter is the key objective for cross-platform validation. We also investigate a decisional variant of quantum distributed inner product estimation, mixed state decisional inner product estimation (DIPE) [34]. We obtain the strongest lower bounds for the following version:

Problem 3 (DIPE: Convex mixture of Haar random states). Alice and Bob are given copies of states, promised that one of the following two cases holds:

(1) Alice and Bob both have state $\rho = \frac{1}{r} \sum_{i=1}^r |\psi_i\rangle\langle\psi_i|$, where the states $|\psi_i\rangle$ are drawn independent and identically distributed (iid) from μ (the Haar measure).

(2) Alice has state $\rho = \frac{1}{r} \sum_{i=1}^r |\psi_i\rangle\langle\psi_i|$ and Bob has state $\sigma = \frac{1}{r} \sum_{i=1}^r |\phi_i\rangle\langle\phi_i|$, where the states $|\phi_i\rangle$ and $|\psi_i\rangle$ are drawn iid from μ .

Their goal is to decide which case they are in with high probability using only LOCC.

When $r = 1$, and ρ and σ are pure states, Problem 3 reduces to *decisional inner product estimation*, for which $\Theta(2^{n/2})$ copies are shown to be sufficient and necessary when Alice and Bob can only use LOCC [9]. In Problem 3, the inner product between the states of Alice and Bob is $\mathbb{E} \text{tr}(\rho^2) \sim 1/r$ in the first case, while in the second case, the inner product is approximately 2^{-n} . Assuming $\text{tr}(\rho^2) = \omega(2^{-n})$, we can solve it using a distributed quantum inner product estimation algorithm with accuracy $\text{tr}(\rho^2)$, which uses $O(\max\{2^{n/2}/\text{tr}(\rho^2)\})$ copies. We prove the following theorem in Appendix D 2.

Theorem 4. Any protocol for mixed state DIPE requires at least $\Omega(\sqrt{d}/\mathbb{E} \text{tr}(\rho^2))$ copies even if Alice and Bob are

allowed to perform arbitrary interactive protocols (or arbitrary LOCC operations).

Additionally, we consider another version of mixed DIPE in Appendix D 2, where Alice and Bob either have the same state $U\rho U^\dagger$ or two independent states $U\rho U^\dagger$ and $V\rho V^\dagger$ for two independent Haar random unitaries U and V . For this variant, we restrict to one-way classical communication between Alice and Bob, but we obtain a particularly direct proof using the concentration of measure phenomenon.

V. OUTLOOK

In this work, we propose protocols for quantum inner product estimation with limited quantum communication and unentangled measurements and quantum purity estimation with limited quantum memory, and prove that they are near optimal to at most a $\sqrt{\varepsilon}$ factor. Here, we would like to point out the following differences and improvements from Ref. [9]. First, while the previous work focused on the task of distributed inner product estimation with classical communications, our work focuses on both the tasks of purity estimation with bounded quantum memory and distributed inner product estimation with bounded quantum communications, and reveals the connection between the two models. The algorithms and the hard instances we employed for the upper and lower bounds are significantly different from the previous work. In addition, we also investigate a decisional variant of quantum distributed inner product estimation without quantum communication for mixed states and provide a sample complexity lower bound.

Several open problems remain to be answered. In Problem 1, we consider quantum inner product estimation when Alice and Bob have k -qubit one-way communication and one-way classical communication. It is thus natural to ask whether the lower bound in this work holds for the case when Alice and Bob have k -qubit arbitrary quantum communication and arbitrary LOCC. While our protocols yield sample complexity upper bound $O(\text{median}\{1/\varepsilon^2, 2^{n/2}/\varepsilon, 2^{n-k}/\varepsilon^2\})$ for inner product estimation with k -qubit quantum communication and unentangled measurements and quantum purity estimation with k -qubit quantum memory, there is a $\sqrt{\varepsilon}$ gap from the lower bound of $\Omega(\text{median}\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\})$ we obtained in this work. A similar $\sqrt{\text{tr}(\rho^2)}$ gap appears for the mixed state DIPE problem. An immediate open question is to find the true ε and $\text{tr}(\rho^2)$ dependence for the sample complexity of these two tasks. In addition, we consider the setting in which Alice and Bob are only allowed to perform unentangled local measurements and share k qubits of quantum communication for each pair of ρ and σ . An alternative model for restricted quantum communication is when Alice and Bob can perform local joint measurements but can only send k qubits of quantum communication in total. It is interesting to see the sample complexity under this model. Finally, it is an interesting future direction to explore whether our techniques can yield protocols or lower bounds for other distributed quantum property estimation problems of the form $\text{tr}(\rho \otimes \sigma \cdot W)$ for some two-body operators W with only classical communication or restricted quantum communication.

ACKNOWLEDGMENTS

We thank Anurag Anshu and Sitan Chen for illuminating suggestions and comments throughout this research. We thank Aayush Karan, Yunchao Liu, and Yihui Quek for the helpful discussions on inner product estimation and representation theory. We thank Rui Sun for helpful discussions on purity estimation. We thank Zhenhuan Liu for helpful discussions on random induced states. Arunachalam and Schatzki [35] have related results concurrent to our work, and we thank them for the helpful exchange. J.H. acknowledges funding from the Harvard Quantum Initiative postdoctoral fellowship. Q.Y. was supported in part by the National Natural Science Foundation of China Grant No. T2225008, Tsinghua University Dushi Program, and Shanghai Qi Zhi Institute.

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

APPENDIX A: PRELIMINARIES

In this Appendix, we provide the basic concepts and results required throughout this paper. We use $\|A\|$, $\|A\|_2$, and $\|A\|_1$ to represent the operator norm, Euclidean norm, and trace norm of matrix A , respectively. For vectors, we use $\|v\|_p$ to denote the L_p norm of vector v . The total variation distance between probability distributions p and q is denoted as $\|p - q\|_{\text{TV}}$. We will abbreviate $\{1, 2, \dots, m\}$ as $[m]$ and $(x_{i_1}, x_{i_1+1}, \dots, x_{i_2})$ as $x_{[i_1:i_2]}$. We use the standard big- O notation (O , Ω) throughout. When we say “with high probability” without specification, we mean with probability at least $2/3$. We denote the n -qubit identity operator by I_n and ignore n when it is clear from the context.

1. Basic results in quantum information

We start with the standard definitions and folklore in quantum information. In general, an n -qubit quantum state can be represented as a positive semidefinite matrix $\rho \in \mathbb{C}^{2^n \times 2^n}$ with $\text{tr}(\rho) = 1$. In the following, we also use $d = 2^n$ to denote the dimension of an n -qubit quantum state. A pure state, which is usually represented as $|\psi\rangle$, $|\phi\rangle$, or $|\varphi\rangle$ throughout this paper, is a rank-1 state and thus $\text{tr}(\rho^2) = 1$. Given an (possibly unnormalized) n -qubit quantum state ρ and a subset $S \subseteq [n]$, we use the partial trace $\text{tr}_S(\rho)$ to denote the remaining state after tracing out the qubits in S . A SWAP_n gate is a $2n$ -qubit operator that SWAPS the first and last n qubits. We omit n when it is clear from the context.

a. Quantum measurements

Quantum measurements are represented as positive operator-valued measures (POVMs) in the general case. An n -qubit POVM is defined by a set of positive-semidefinite matrices $\{F_s\}_s$ with $\sum_s F_s = I$, where each F_s is a *POVM element* corresponding to a *measurement outcome* s . The probability of observing outcome s when we measure a given quantum state ρ using the POVM $\{F_s\}_s$ is $\text{tr}(F_s \rho)$.

When F_s is rank 1 for all s , we call $\{F_s\}_s$ a *rank-1 POVM*. Lemma 4.8 in Ref. [24] proves that any POVM can be simulated by a rank-1 POVM and classical postprocessing information theoretically. We will thus assume when we prove our lower bounds in the following parts of the paper that all measurements in the learning protocols only use rank-1 POVMs. We can thus parametrize any rank-1 POVMs as

$$\{w_s 2^n |\psi_s\rangle \langle \psi_s|\}_s,$$

for pure states $\{|\psi_s\rangle\}$ and non-negative weights w_s with $\sum_s w_s = 1$. When $w_s = 2^{-n}$ for all s , the POVM elements satisfy $F_s = |\psi_s\rangle \langle \psi_s|$ and the POVM then reduces to a PVM.

We also consider the postmeasurement quantum state after the POVM $\{F_s\}$. The postmeasurement state is not determined by the POVM $\{F_s\}_s$ itself, but by the physical realization of it. In particular, we need to specify a decomposition for every POVM element F_s in $\{F_s = M_s^\dagger M_s\}_s$, where M_s is a $2^m \times 2^n$ matrix. The postmeasurement state of ρ corresponding to the outcome s is an m -qubit state given by

$$\rho \rightarrow \frac{M_s \rho M_s^\dagger}{\text{tr}(M_s^\dagger M_s \rho)}.$$

We will refer to such a (realization of a) POVM as having *n -qubit input and m -qubit output* or simply as being an $(n \rightarrow m)$ *POVM* to emphasize the dimension of the postmeasurement state. We say “POVM” in this paper for a realization with 0-qubit output unless we specify the number of qubits in the output.

b. Measurements with classical communications

In this paper, we usually have to consider the measurement $\{F_s\}_s = \{M_{A,B}^s\}_s$ that is jointly performed by two parties, Alice and Bob. We ignore the outcome s when it does not matter in the context. Suppose Alice and Bob can only have classical communications. If the classical is restricted to one-way communication, we can always depict the measurement scheme for two-case distinguishing tasks as follows: Alice performs a POVM $\{M_i\}$ on her copies and sends the result i to Bob. After receiving i , Bob performs a two-outcome POVM $\{N_i, I - N_i\}$ on his copies to decide which case they are in. We thus have every $M_{A,B}$ is in the form of

$$M_{A,B} = \sum_i M_i \otimes N_i, \quad (\text{A1})$$

with $\sum_i M_i = I$.

If Alice and Bob can have arbitrary classical communication, $\{M_{A,B}^s\}_s$ can be described as quantum measurements with LOCC. A rigorous mathematical representation for measurements with LOCC is complicated. Instead, we consider separable measurements, which are defined as follows.

Definition 1 (Separable measurements). A two-outcome POVM $\{M_{A,B}, I - M_{A,B}\}$ is separable if and only if

$$M_{A,B} = \sum_t A_t \otimes B_t,$$

$$I - M_{A,B} = \sum_t A'_t \otimes B'_t,$$

$$A_t, A'_t, B_t, B'_t \geq 0.$$

It is proved that any quantum measurements with LOCC are included in the class of separable measurements [36,37].

2. A brief review of Haar random unitaries

A common tool to prove lower bounds on the sample complexity of learning problems is to consider random instances. The random states we consider here are based on Haar random unitaries. The Haar measure μ on the unitary group $U(d)$ is the unique probability measure that is invariant under left and right multiplication. That is,

$$\mathbb{E}_{U \sim \mu} f(UV) = \mathbb{E}_{U \sim \mu} f(VU) = \mathbb{E}_{U \sim \mu} f(U),$$

for any unitary $V \in U(d)$. We can also define a unique rotation invariant measure on states by $U|\psi\rangle$ with $U \sim \mu$. Here, $|\psi\rangle \in \mathbb{C}^d$ can be any quantum state. Slightly abusing notation, we will write $\psi \sim \mu$.

The Haar measure has multiple desirable properties for proving lower bounds, most noticeably *concentration of measure* in the form of Levy's lemma: A function $f : U(d) \rightarrow \mathbb{R}$ is called Lipschitz continuous with Lipschitz constant L if

$$|f(U) - f(V)| \leq L \|U - V\|_2,$$

where $\|\cdot\|_2$ denotes the Frobenius or Schatten 2-norm.

Theorem 5 (Levy's lemma for Haar-random pure states). Let $f : U(d) \rightarrow \mathbb{R}$ be a function that is Lipschitz with Lipschitz constant L . Then,

$$\Pr_{U \sim \mu} [|f(U) - \mathbb{E}_{U \sim \mu} [f(U)]| > \tau] \leq 4 \exp \left(-\frac{2d\tau^2}{9\pi^3 L^2} \right)$$

for any $\tau > 0$.

We will apply this to the outcome probabilities of measurements using k copies of $U \sim \mu$. We can then use Theorem 5 to show that, for small k , the outcomes of such measurements do not depend on U and therefore reveal only little information.

In some settings, we will need to explicitly compute expectation values over the Haar measure. A key tool is the following folklore formula:

Lemma 1 (See e.g. Ref. [38]).

$$\mathbb{E}_{\psi \sim \mu} (|\psi\rangle\langle\psi|)^{\otimes k} = \frac{S_k}{\binom{d+k-1}{k}} = \frac{1}{d \cdots (d+k-1)} \sum_{\pi \in S_k} \pi^d,$$

where S_k denotes the projector onto the symmetric subspace $\text{Sym}^k(\mathbb{C}^d)$, S_k is the set of permutation over k elements, and π^d acts on $(\mathbb{C}^d)^{\otimes k}$ via

$$\pi^d |i_1, \dots, i_k\rangle = |i_{\pi^{-1}(1)}, \dots, i_{\pi^{-1}(k)}\rangle.$$

3. Tree representations and the Le Cam's two-point method

In this paper, we focus on adaptive protocols, which can be naturally modeled as learning trees [20,23,24,27,39]. In each node, we select c copies of the unknown state ρ , perform some POVM on $\rho^{\otimes c}$, and step to its child node corresponding to the outcome. Adapting Ref. [27], we call such protocol a $(c, \mathcal{M})$ learning protocol, where $\mathcal{M}$ is the set of feasible POVMs on $\rho^{\otimes c}$. Throughout the paper, we only consider two specific kinds of protocols:

(1) $c = 1$ and $\mathcal{M}$ is the set of all single-copy POVMs. This characterizes all protocols that only use single-copy measurements.

(2) $c = 2$ and $\mathcal{M}$ is the set of all rank-1 two-copy POVMs where there are only k qubits of quantum memory (see Definition 5 for the formal definition).

Definition 2 (Tree representation for $(c, \mathcal{M})$ learning protocols [27]). Given an unknown n -qubit quantum state ρ , a $(c, \mathcal{M})$ learning protocol can be represented as a rooted tree $\mathcal{T}$ of depth T with each node on the tree recording the measurement outcomes of the algorithm so far. In particular, it has the following properties:

(1) We assign a probability $p^\rho(u)$ for any node u on the tree $\mathcal{T}$.

(2) The probability assigned to the root r of the tree is $p^\rho(r) = 1$.

(3) At each nonleaf node u , we measure $\rho^{\otimes c}$ using an adaptively chosen POVM $M_u = \{F_s^u\}_s \in \mathcal{M}$, which results in a classical outcome s . Each child node v corresponding to the classical outcome s of the node u is connected through the edge $e_{u,s}$.

(4) If v is the child of u through the edge $e_{u,s}$, the probability to traverse this edge is $p^\rho(s|u) = \text{tr}(F_s^u \rho^{\otimes c})$. Then,

$$p^\rho(v) = p^\rho(u) p^\rho(s|u) = p^\rho(u) \text{tr}(F_s^u \rho^{\otimes c}).$$

(5) Each root-to-leaf path is of length T . Note that for a leaf node ℓ , $p^\rho(\ell)$ is the probability for the classical memory to be in state ℓ at the end of the learning protocol. Denote the set of leaves of $\mathcal{T}$ by $\text{leaf}(\mathcal{T})$.

At the end of the protocol, a classical algorithm $\mathcal{A}$ maps any leaf node ℓ to a desired output of the protocol. $\mathcal{A}$ is called the postprocessing part of the protocol.

To prove the lower bound of purity estimation, we consider the reduction from the following distinguishing problem:

Problem 4 (Distinguishing the maximally mixed state and an ensemble $\mathcal{D}$). Given access to copies of an unknown state ρ of dimension $d = 2^n$, distinguish the following two cases:

(1) ρ is the maximally mixed state $\rho_m = I/2^n$.

(2) ρ is sampled from a known distribution $\mathcal{D}$ over n -qubit mixed states.

We will explicitly construct an ensemble $\mathcal{D}_{n,\varepsilon}$ (Definition 6) such that the distinguishing problem can be reduced to a purity test. There are mature tools for proving lower bounds of many-versus-one distinguishing problems, including Le Cam's two-point method [40], one-sided likelihood ratio [24], and the martingale technique [32]. Here, we adapt notations and statements in Ref. [27].

Definition 3 (Likelihood ratio). Let $\mathcal{T}$ be the tree representation of a learning protocol for the many-versus-one distinguishing task in Problem 4. For any $\ell \in \text{leaf}(\mathcal{T})$, define the likelihood ratio

$$L(\ell) := \frac{\mathbb{E}_{\psi \sim \mathcal{D}} p^\psi(\ell)}{p^{\rho_m}(\ell)}.$$

We also define the likelihood ratio for each edge $e_{u,s}$ and each state ψ :

$$L_\psi(u, s) := \frac{p^\psi(s|u)}{p^{\rho_m}(s|u)}, \quad L_\psi(\ell) := \frac{p^\psi(\ell)}{p^{\rho_m}(\ell)}.$$

Lemma 2 (Toolbox for proving lower bounds). Suppose $\mathcal{T}$ is a learning tree with depth T that solves the distinguishing problem in Problem 4 with probability p_{succ} .

(a) (Le Cam's two-point method [40])

$$p_{\text{succ}} \leq d_{\text{TV}}(\mathbb{E}_{\psi \sim \mathcal{D}} p^\psi, p^{\rho_m}) \\ := \frac{1}{2} \sum_{\ell \in \text{leaf}(\mathcal{T})} |\mathbb{E}_{\psi \sim \mathcal{D}} p^\psi(\ell) - p^{\rho_m}(\ell)|.$$

(b) (One-sided likelihood ratio [24]; see also Ref. [27], Lemma 6). For any $\beta > 0$,

$$d_{\text{TV}}(\mathbb{E}_{\psi \sim \mathcal{D}} p^\psi, p^{\rho_m}) \leq \Pr_{\ell \sim p^{\rho_m}(\ell), \psi \sim \mathcal{D}} [L_\psi(\ell) \leq \beta] + 1 - \beta,$$

$$d_{\text{TV}}(\mathbb{E}_{\psi \sim \mathcal{D}} p^\psi, p^{\rho_m}) \leq \Pr_{\ell \sim p^{\rho_m}(\ell)} [L(\ell) \leq \beta] + 1 - \beta.$$

(c) (Martingale technique [32]; see also Ref. [27], Lemma 7). If there exists a $\delta > 0$ such that for every node u we have

$$\mathbb{E}_{\psi \sim \mathcal{D}} \mathbb{E}_{s \sim p^{\rho_m}(s|u)} [(L_\psi(u, s) - 1)^2] \leq \delta,$$

then

$$\Pr_{\ell \sim p^{\rho_m}(\ell), \psi \sim \mathcal{D}} [L_\psi(\ell) \leq 0.9] \leq 0.1 + \alpha \delta T,$$

where α is a universal constant.

These tools upper bound the success probability from different perspectives. Lemma 2(a) is tree-based: We upper bound the statistical distance between the leaf distributions under the two cases. Lemma 2(b) is path based: we have to prove that for most of the paths from root to leaves, the likelihood ratios are not too small. Lemma 2(c) is edge based: We only have to focus on each edge and show that the likelihood ratio over the edge concentrates around 1.

APPENDIX B: FROM PURITY TO INNER PRODUCT ESTIMATION

In this Appendix, we consider the models for quantum inner product estimation with k -qubit one-way quantum communication (Problem 1) and purity estimation with k -qubit quantum memory (Problem 2), and show the connection between the sample complexity of the two tasks (see Fig. 1 for illustration).

For quantum inner product estimation defined in Problem 1, Alice and Bob are allowed to perform k -qubit one-way quantum communication, any one-way classical communication, and unentangled local measurements. As shown in Fig. 1(a), we can always describe such a protocol sequentially, as the quantum communication is one way:

Definition 4 (Sequential description of Problem 1). The sequential description of a protocol for Problem 1.

(1) Alice takes a copy of ρ and selects an $(n \rightarrow k)$ POVM $\{M_s^\dagger M_s\}_s$ (see Appendix A 1 for definitions) to measure ρ .

(1) Alice obtains the classical outcome s with probability $\text{tr}(\rho \cdot M_s^\dagger M_s)$, sends arbitrary classical message (might depend on s), and sends the k -qubit postmeasurement $\varrho_{s,k} = M_s \rho M_s^\dagger / \text{tr}(\rho \cdot M_s^\dagger M_s)$ to Bob.

(1) Bob receives the classical message and $\varrho_{s,k}$, and performs an $(n + k \rightarrow 0)$ POVM $\{M_{s'}^\dagger M_{s'}\}_{s'}$ on $\varrho_{s,k} \otimes \sigma$.

For purity estimation with k -qubit quantum memory defined in Problem 2 [as shown in Fig. 1(b)], we can also provide a sequential definition as below adapting Ref. [27]:

Definition 5 (Purity estimation with k qubits of quantum memory: see also Definition 4 of Ref. [27]). The algorithm consists of $N/2$ rounds, and each round involves two copies of ρ . In each round, the algorithm maintains a k -qubit quantum memory ϱ (initialized to a fixed quantum state at first). Within the round, we get a fresh copy of ρ , select an $(n \rightarrow k)$ POVM $\{M_s^\dagger M_s\}_s$ to measure ρ , obtain the outcome s with probability $\text{tr}(\rho \cdot M_s^\dagger M_s)$, and store the k -qubit postmeasurement state $\varrho_{s,k} = M_s \rho M_s^\dagger / \text{tr}(\rho M_s^\dagger M_s)$ into the quantum memory. The algorithm then takes a fresh copy of ρ , selects an $(n + k \rightarrow 0)$ POVM $\{M_{s'}^\dagger M_{s'}\}_{s'}$ to measure $\varrho_{s,k} \otimes \rho$, and obtains the outcome s' with probability $\text{tr}((\varrho_{s,k} \otimes \rho) \cdot M_{s'}^\dagger M_{s'})$. Finally, the quantum memory is reset to empty, and the next round begins. After $N/2$ rounds, the algorithm predicts purity $\text{tr}(\rho^2)$ of ρ . The total sample complexity of the protocol is N .

Under this learning model with k qubit of quantum memory, purity testing, which distinguishes between ρ being a Haar random pure state from the maximally mixed state, requires $\Theta(\min\{2^{n/2}, 2^{n-k}\})$ copies of ρ for algorithms with k -qubit quantum memory [27]. However, for purity estimation to additive error ε , the dependence on ε for sample complexity is unclear.

Given the sequential description of quantum inner product estimation with bounded one-way quantum communication in Problem 1 and purity estimation with bounded memory in Definition 5 (Problem 2), we can observe the following connection between the sample complexity of Problems 1 and 2:

Claim 1. Consider quantum inner product estimation with k -qubit one-way quantum communication (Problem 1) and purity estimation with k -qubit quantum memory (Problem 2). Suppose we have an algorithm $\mathcal{A}$ that can solve Problem 1 with N copies of ρ and σ , we have an algorithm $\mathcal{A}'$ that solves Problem 2 using $2N$ copies of ρ .

The intuition for this claim is straightforward given the sequential descriptions of two tasks. Suppose we have an algorithm $\mathcal{A}$ for Problem 1 using N copies of ρ and σ (i.e., N rounds in Definition 4), we can describe the algorithm $\mathcal{A}'$ in a similar pattern for Problem 2:

(1) We take a copy of ρ and select the same $(n \rightarrow k)$ POVM $\{M_s^\dagger M_s\}_s$ in $\mathcal{A}$ to measure ρ .

(2) We obtain the classical outcome s with probability $\text{tr}(\rho \cdot M_s^\dagger M_s)$ and store the k -qubit postmeasurement $\varrho_{s,k} = M_s \rho M_s^\dagger / \text{tr}(\rho \cdot M_s^\dagger M_s)$ into the quantum memory.

(3) We take a fresh copy of ρ and perform the same $(n + k \rightarrow 0)$ POVM $\{M_{s'}^\dagger M_{s'}\}_{s'}$ in $\mathcal{A}$ on $\varrho_{s,k} \otimes \rho$.

One can verify that the above sequence satisfies Definition 5 and $\mathcal{A}'$ is a protocol for Problem 2 using $2N$ copies of ρ .

APPENDIX C: LOWER BOUNDS FOR PURITY ESTIMATION

1. Universal lower bound for purity estimation

In this Appendix, we prove a lower bound for purity estimation when arbitrary entangled measurements are allowed. Since entangled measurements are the most general form of

quantum algorithms, the lower bound also applies to the restrictive settings that we will consider later.

Lemma 3. Any algorithm $\mathcal{A}$ that estimates the purity within additive error $\varepsilon \leq 1/36$ using entangled measurements has sample complexity at least $\Omega(1/\varepsilon^2)$.

Proof. Then, proof is elementary and similar ideas have been employed in Refs. [9,19]. Let $\delta = \frac{1}{6} - \sqrt{\frac{1}{36} - \varepsilon}$. Consider $\rho_0 = \frac{1}{3}|0\rangle\langle 0| + \frac{2}{3}|1\rangle\langle 1|$ and $\rho_1 = (\frac{1}{3} + \delta)|0\rangle\langle 0| + (\frac{2}{3} - \delta)|1\rangle\langle 1|$. We have $\text{tr}(\rho_0^2) = \frac{5}{9}$ and $\text{tr}(\rho_1^2) = \frac{5}{9} - \frac{2}{3}\delta + 2\delta^2$, so the purity of the two states differs by $\frac{2}{3}\delta - 2\delta^2 = 2\varepsilon$. Thus, we can use $\mathcal{A}$ to distinguish the two states based on whether the estimated purity is larger than $\frac{5}{9} - \varepsilon$. However, by Helstrom theorem [41], the probability of distinguishing t copies of ρ_0 and t copies of ρ_1 is at most $\frac{1}{2} + \frac{1}{4}\|\rho_0^{\otimes t} - \rho_1^{\otimes t}\|_1$. Therefore, $\|\rho_0^{\otimes t} - \rho_1^{\otimes t}\|_1 \geq \frac{2}{3}$. Defining the fidelity between ρ and σ as $F(\rho, \sigma) = (\text{tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}))^2$, we have both

$$\begin{aligned} F(\rho_0^{\otimes t}, \rho_1^{\otimes t}) &\leq 1 - \frac{1}{4}\|\rho_0^{\otimes t} - \rho_1^{\otimes t}\|_2^2 \leq \frac{8}{9}, \\ F(\rho_0^{\otimes t}, \rho_1^{\otimes t}) &= F(\rho_0, \rho_1)^t \\ &= \left(\sqrt{\frac{1}{3}\left(\frac{1}{3} + \delta\right)} + \sqrt{\frac{2}{3}\left(\frac{2}{3} - \delta\right)} \right)^{2t} \\ &\geq (1 - 40\varepsilon^2)^t \geq 1 - 40\varepsilon^2 t, \end{aligned}$$

where the second inequality uses the well-known relation that fidelity is upper bounded by 1 minus the squared trace distance. Thus, the number of copies t needed for $\mathcal{A}$ is $\Omega(1/\varepsilon^2)$. ■

2. Lower bound for purity estimation with bounded quantum memory

In this Appendix, we prove a lower bound of $\Omega(\min\{\sqrt{d/\varepsilon}, 2^{n-k}/\varepsilon^2\})$ for protocols with k qubits of memory. This together with the $\Omega(1/\varepsilon^2)$ lower bound from Appendix C 1 yields a lower bound of

$$\begin{aligned} &\Omega(\max\{1/\varepsilon^2, \min\{\sqrt{d/\varepsilon}, 2^{n-k}/\varepsilon^2\}\}) \\ &= \Omega(\text{median}\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\}). \end{aligned}$$

When $k = 0$, the protocols considered are degenerate to protocols that only use single-copy measurements. This implies an $\Omega(\max\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}\})$ lower bound for single copy measurements.

Our lower bound is proved via a reduction from Problem 4. To this end, we need an appropriate ensemble $\mathcal{D}$ of n -qubit mixed states satisfying the following conditions:

(1) With high probability, a state ψ sample from $\mathcal{D}$ has high fidelity, i.e., $\text{tr}(\psi^2) = \text{tr}(\rho_m^2) + \Theta(\varepsilon)$, so that an algorithm for purity estimation implies an algorithm for distinguishing ρ_m and $\mathcal{D}$.

(2) There is a neat expression of $\mathbb{E}_{\psi \sim \mathcal{D}} \psi^{\otimes T}$ that allows us to prove lower bound using tools from Lemma 2.

We achieve the two conditions by tracing out part of the qubits from a large Haar random state. This is called a random induced state [42] that has been considered extensively in the literature of quantum entanglement [43–45]. To the best of

our knowledge, this is the first time it has been considered for quantum hypothesis testing.

Definition 6 (Ensemble of random induced states). Define $\mathcal{D}_{n,\varepsilon}$ as the distribution of $\psi = \text{tr}_{>n}(|H\rangle\langle H|)$, where $|H\rangle$ is a $(D = 2^{n+\log(1/\varepsilon)})$ -dimensional Haar random state, and $\text{tr}_{>n}$ means tracing out the last $\log(1/\varepsilon)$ qubits.

We now instantiate Problem 4 into the following problem and show that the problem can be reduced to purity estimation.

Problem 5. Distinguish the maximally mixed state ρ and the ensemble $\mathcal{D}_{n,\varepsilon}$ of random induced states (in the sense of Problem 4) with probability at least $7/12$.

Lemma 4. An algorithm $\mathcal{A}$ that estimates the purity up to within the additive error $\varepsilon/3$ can be used to construct an algorithm that solves Problem 5 with the same sample complexity.

Proof. We need some basic properties of random induced states, which can be obtained from direct calculations using Lemma 1:

$$\mathbb{E}_{\psi \sim \mathcal{D}_{n,\varepsilon}} \text{tr}(\psi^2) = \frac{2^n \varepsilon + 1}{2^n + \varepsilon},$$

$$\text{Var}_{\psi \sim \mathcal{D}_{n,\varepsilon}} \text{tr}(\psi^2) = \frac{2(2^{2n} - 1)(1/\varepsilon^2 - 1)}{(2^n/\varepsilon + 1)^2(2^n/\varepsilon + 2)(2^n/\varepsilon + 3)}.$$

By Chebyshev's inequality,

$$\begin{aligned} p &= \Pr_{\psi \sim \mathcal{D}_{n,\varepsilon}} \left[\text{tr}(\psi^2) \leq 2^{-n} + \frac{2\varepsilon}{3} \right] \\ &\leq \frac{\text{Var}_{\psi \sim \mathcal{D}_{n,\varepsilon}} \text{tr}(\psi^2)}{(\mathbb{E}_{\psi \sim \mathcal{D}_{n,\varepsilon}} \text{tr}(\psi^2) - (2^{-n} + 2\varepsilon/3))^2} = O(2^{-2n}). \end{aligned}$$

To solve Problem 5, use $\mathcal{A}$ to estimate the purity and output ρ_m if the estimated purity is smaller than $2^{-n} + \varepsilon/3$. With probability $1 - p$, ρ in the random induced state case has purity greater than $2^{-n} + 2\varepsilon/3$. Then, with probability at least $\frac{2}{3}$, $\mathcal{A}$ estimates purity within error $\varepsilon/3$, in which case the distinguishing task in Problem 5 is solved correctly. Thus, the probability of solving the distinguishing task correctly is at least $(1 - p)2/3 \geq \frac{7}{12}$. ■

Hence, it suffices to prove lower bounds for Problem 5. The techniques are similar to the ones used in Ref. [27]. We include some lemmas there:

Lemma 5 (See Lemma 16 of Ref. [27]). Let x, y be two positive integers, ρ_x be a mixed state on x qudits (each qudit has local dimension d), and ρ_y be a mixed state on y qudits. Then,

$$\text{tr} \left(\rho_x \otimes \rho_y \sum_{\pi \in S_{x+y}} \pi^d \right) \geq \text{tr} \left(\rho_x \sum_{\pi \in S_x} \pi^d \right) \text{tr} \left(\rho_y \sum_{\pi \in S_y} \pi^d \right).$$

Lemma 6. Let $F_\ell = \bigotimes_{i=1}^T F_{s_i}^{u_i}$ be a product of T non-negative d -dimensional Hermitian operators, and $B_1 \subsetneq B_2 \subseteq [T]$. Then,

$$\text{tr} \left(F_\ell \sum_{\pi \in S_{B_2}} \pi^d \otimes I^{\otimes \overline{B_2}} \right) \geq \text{tr} \left(F_\ell \sum_{\pi \in S_{B_1}} \pi^d \otimes I^{\otimes \overline{B_2}} \right).$$

Proof. Without loss of generality, we assume $B_1 = \{1, 2, \dots, k\}$ and $B_2 = \{1, 2, \dots, k'\}$ with $k < k'$. We also assume $F_{s_i}^{u_i}$ is of trace 1. Then, we can recursively apply

Lemma 5 to obtain

$$\begin{aligned} \text{tr} \left(F_\ell \sum_{\pi \in S_{B_2}} \pi^d \otimes I^{\otimes \overline{B_2}} \right) &= \text{tr} \left(\bigotimes_{t=1}^{k'-1} F_{s_t}^{u_t} \sum_{\pi \in S_{k'}} \pi^d \right) \geq \text{tr} \left(\bigotimes_{t=1}^{k'-1} F_{s_t}^{u_t} \sum_{\pi \in S_{k-1}} \pi^d \right) \text{tr}(F_{s_{k'}}^{u_{k'}}) \\ &= \text{tr} \left(\bigotimes_{t=1}^{k'-1} F_{s_t}^{u_t} \sum_{\pi \in S_{k-1}} \pi^d \right) \geq \dots \geq \text{tr} \left(\bigotimes_{t=1}^k F_{s_t}^{u_t} \sum_{\pi \in S_k} \pi^d \right) = \text{tr} \left(F_\ell \sum_{\pi \in S_{B_1}} \pi^d \otimes I^{\otimes \overline{B_1}} \right). \quad \blacksquare \end{aligned}$$

Lemma 7 (See Lemma 17 of Ref. [27]).

$$\max_{\{F_s\}_s \in \mathcal{M}_{2,n}^k} \sum_s \frac{\text{tr}(F_s \text{SWAP})^2}{\text{tr}(F_s)} \leq 2^{k+n}.$$

Now we are ready for the lower bound.

Theorem 6. Any algorithm $\mathcal{A}$ that estimates the purity within additive error ε with two-copy protocols with k qubits of quantum memory requires $\Omega(\text{median}\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\})$ copies.

By Claim 1, Theorem 6 also indicates an $\Omega(\text{median}\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\})$ sample complexity for quantum inner product estimation using k -qubit one-way communication as in Problem 1.

Proof. As argued before, it suffices to prove an $\Omega(\min\{2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\})$ lower bound for Problem 5. Using the tree representation, consider a path $\{e_{u_t, s_t}\}_{t=1}^T$ from root leaf ℓ . We have the likelihood ratio for ψ :

$$L_\psi(\ell) = \prod_{t=1}^T \frac{\text{tr}(F_{s_t}^{u_t} \psi^{\otimes 2})}{\text{tr}(F_{s_t}^{u_t} \rho_m^{\otimes 2})} = \frac{\text{tr}(\bigotimes_{t=1}^T F_{s_t}^{u_t} \text{tr}_{>n}^{\otimes 2T}(|H\rangle\langle H|^{\otimes 2T}))}{\text{tr}(\bigotimes_{t=1}^T F_{s_t}^{u_t} \rho_m^{\otimes 2T})} = \frac{\text{tr}(\bigotimes_{t=1}^T \text{tr}_{>n}^{\otimes 2\ddagger}(F_{s_t}^{u_t}) |H\rangle\langle H|^{\otimes 2T})}{\text{tr}(\bigotimes_{t=1}^T F_{s_t}^{u_t} \rho_m^{\otimes 2T})},$$

where $\text{tr}_{>n}^{\otimes x}$ means tracing out the last $\log(1/\varepsilon)$ qubits for all x copies and $\text{tr}_{>n}^{\otimes x\ddagger}$ is its adjoint. Then, the likelihood ratio is

$$\begin{aligned} L(\ell) &= \mathbb{E}_{\psi \sim \mathcal{D}_{n,\varepsilon}} [L_\psi(\ell)] = \mathbb{E}_{|H\rangle \sim \mu} \frac{\text{tr}(\bigotimes_{t=1}^T \text{tr}_{>n}^{\otimes 2\ddagger}(F_{s_t}^{u_t}) |H\rangle\langle H|^{\otimes 2T})}{\text{tr}(\bigotimes_{t=1}^T F_{s_t}^{u_t} \rho_m^{\otimes 2T})} \\ &= \frac{d^{2T}}{D(D+1) \cdots (D+2T-1)} \cdot \frac{\text{tr}(\bigotimes_{t=1}^T \text{tr}_{>n}^{\otimes 2\ddagger}(F_{s_t}^{u_t}) \sum_{\pi \in S_{2T}} \pi^D)}{\text{tr}(\bigotimes_{t=1}^T F_{s_t}^{u_t})} \\ &\geq \frac{d^{2T}}{D(D+1) \cdots (D+2T-1)} \prod_{t=1}^T \frac{\text{tr}(\text{tr}_{>n}^{\otimes 2\ddagger}(F_{s_t}^{u_t}) \sum_{\pi \in S_2} \pi^D)}{\text{tr}(F_{s_t}^{u_t})} \\ &= \frac{d^{2T}}{D(D+1) \cdots (D+2T-1)} \prod_{t=1}^T \frac{\text{tr}(F_{s_t}^{u_t} \text{tr}_{>n}^{\otimes 2}(\sum_{\pi \in S_2} \pi^D))}{\text{tr}(F_{s_t}^{u_t})} \\ &= \frac{D^{2T}}{D(D+1) \cdots (D+2T-1)} \prod_{t=1}^T \frac{\text{tr}(F_{s_t}^{u_t} (I + \frac{d}{D} \text{SWAP}))}{\text{tr}(F_{s_t}^{u_t})} \\ &\geq \left(1 - \frac{2T}{D}\right)^{2T} \left(1 + \frac{1}{D}\right)^T \prod_{t=1}^T \frac{\text{tr}(F_{s_t}^{u_t} \rho_2)}{\text{tr}(F_{s_t}^{u_t} \rho_m^{\otimes 2})} \geq \left(1 - \frac{4T^2}{D}\right) \prod_{t=1}^T \frac{\text{tr}(F_{s_t}^{u_t} \rho_2)}{\text{tr}(F_{s_t}^{u_t} \rho_m^{\otimes 2})}. \end{aligned}$$

In the third line, we used Lemma 1. The fourth line uses Lemma 5 recursively. We defined the $2n$ -qubit state $\rho_2 = \frac{I+d/D\text{SWAP}}{d^2(1+1/D)}$ in the seventh line. Note that $\prod_{t=1}^T \frac{\text{tr}(F_{s_t}^{u_t} \rho_2)}{\text{tr}(F_{s_t}^{u_t} \rho_m^{\otimes 2})}$ is the likelihood ratio of distinguishing $\rho_m^{\otimes 2}$ and ρ_2 when using $(1, \mathcal{M}_{2,n}^k)$ protocols. We can bound it using Lemma 2(c). To apply Lemma 2(c), we need to bound the likelihood ratio at each step.

$$\begin{aligned} &\max_{\{F_s\}_s \in \mathcal{M}_{2,n}^k} \mathbb{E}_{s \sim \text{tr}(F_s \rho_m^{\otimes 2})} \left[\left(\frac{\text{tr}(F_s \rho_2)}{\text{tr}(F_s \rho_m^{\otimes 2})} - 1 \right)^2 \right] \\ &= \max_{\{F_s\}_s \in \mathcal{M}_{2,n}^k} \sum_s \frac{\text{tr}(F_s)}{d^2} \left(\frac{d}{D+1} \frac{\text{tr}(F_s \text{SWAP})}{\text{tr}(F_s)} - \frac{1}{D+1} \right)^2 \\ &\leq \frac{2}{(D+1)^2} + 2 \max_{\{F_s\}_s \in \mathcal{M}_{2,n}^k} \frac{\text{tr}(F_s \text{SWAP})^2}{(D+1)^2 \text{tr}(F_s)} \leq \frac{2}{(D+1)^2} + \frac{2^{k+n+1}}{(D+1)^2} \leq 4\varepsilon^2 2^{k-n}. \end{aligned}$$

In the second line, we used the inequality $(x - y)^2 \leq 2x^2 + 2y^2$. Thus, by Lemma 2(c) we have

$$\Pr_{\ell \sim p_m^{\otimes 2}(\ell)} \left[\prod_{t=1}^T \frac{\text{tr}(F_{s_t}^{u_t} \rho_2)}{\text{tr}(F_{s_t}^{u_t} \rho_m^{\otimes 2})} > 0.9 \right] \geq 0.9 - cT(4\varepsilon^2 2^{k-n}),$$

for a constant c . This means that

$$\Pr_{\ell \sim p_m^{\otimes 2}(\ell)} \left[L(\ell) > 0.9 \left(1 - \frac{4T^2}{D} \right) \right] \geq 0.9 - cT(4\varepsilon^2 2^{k-n}).$$

Assuming $T \leq \min\{\frac{2^{n/2}}{20\sqrt{\varepsilon}}, \frac{2^{n-k}}{400c\varepsilon^2}\}$, we have $0.9(1 - 4T^2/D) \geq 0.9 \times 0.99 \geq 0.89$ and $cT(4\varepsilon^2 2^{k-n}) \leq 0.01$. Thus, in this case

$$\Pr_{\ell \sim p_m^{\otimes 2}(\ell)} [L(\ell) > 0.89] \geq 0.89.$$

By Lemmas 2(a) and 2(b), the success probability is at most $0.11 + 0.11 = 0.22 < \frac{7}{12}$. We conclude that to solve Problem 5 with probability $\frac{7}{12}$, $T \geq \min\{\frac{2^{n/2}}{20\sqrt{\varepsilon}}, \frac{2^{n-k}}{400c\varepsilon^2}\} = \Omega(2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2)$. ■

Theorem 7 (The exponential hardness of purity estimation is universal). The hard instances in Definition 6 (used in the lower bound in Theorem 6) are all close to the maximally mixed state. Here, we show that estimating purity close to any constant value also requires $\Omega(2^{n/2}/\sqrt{\varepsilon})$ samples for single-copy strategies. This indicates that purity estimation is also hard for states far from maximally mixed (even for pure states). By contrast, in quantum state certification, certifying states that are far away from maximally mixed can be significantly easier [21].

Proof. For any fixed constant $0 < a < 1$, we consider a distinguishing task between the following ensembles:

- (1) The n -qubit quantum state is of the form $\rho_1 = a|\psi\rangle\langle\psi| + (1-a) \cdot I/d$, where $|\psi\rangle\langle\psi|$ is sampled Haar randomly and $d = 2^n$.
- (2) The n -qubit quantum state is of the form $\rho_2 = (a + \varepsilon)|\psi\rangle\langle\psi| + (1-a-\varepsilon) \cdot I/2^n$, where $|\psi\rangle\langle\psi|$ is sampled Haar randomly.

By calculating the expectation and variance, we see that the purity is exponentially concentrated around $a^2 + (1-a^2)/d$ ($(a + \varepsilon)^2 + (1 - (a + \varepsilon)^2)/d$) in the first (second) case. Therefore, an algorithm $\mathcal{A}$ that estimates purity within additive error $2a\varepsilon/3$ can solve the above distinguishing task. Now we show that the sample complexity of this distinguishing task is at least $\Omega(\sqrt{d}/\sqrt{\varepsilon})$ for single-copy strategies.

In the following, we consider a strategy using T single-copy measurements. We consider the learning tree framework in Definition 2. Similar to Lemma 2(a), the success probability of distinguishing between the two cases is given by [40]

$$p_{\text{succ}} \leq d_{\text{TV}}(\mathbb{E}_{\psi} p^{\rho_1}, \mathbb{E}_{\psi} p^{\rho_2}) := \frac{1}{2} \sum_{\ell \in \text{leaf}(\mathcal{T})} |\mathbb{E}_{\psi} p^{\rho_1}(\ell) - \mathbb{E}_{\psi} p^{\rho_2}(\ell)|. \quad (\text{C1})$$

By the definition, the probability $p^{\rho}(\ell) = \text{tr}(F_{\ell} \rho^{\otimes T})$, where $F_{\ell} = \otimes_{t=1}^T F_{s_t}^{u_t}$ is the tensor product of POVM elements along path ℓ . We first prove the following fact $\forall B_1 \subsetneq B_2 \subseteq [T]$:

$$\sum_{\ell} |\text{tr}(F_{\ell} \mathbb{E}_{\psi} \psi^{\otimes B_1} \otimes (I/d)^{\otimes \overline{B_1}}) - \text{tr}(F_{\ell} \mathbb{E}_{\psi} \psi^{\otimes B_2} \otimes (I/d)^{\otimes \overline{B_2}})| \leq \frac{|B_2|^2 - |B_1|^2}{d}. \quad (\text{C2})$$

Write $|B_1| = b_1$ and $|B_2| = b_2$. By Lemma 2(b), we only need to bound the one-sided likelihood ratio given by

$$\begin{aligned} \frac{\text{tr}(F_{\ell} \mathbb{E}_{\psi} \psi^{\otimes B_2} \otimes (I/d)^{\otimes \overline{B_2}})}{\text{tr}(F_{\ell} \mathbb{E}_{\psi} \psi^{\otimes B_1} \otimes (I/d)^{\otimes \overline{B_1}})} &= \frac{d^{b_2-b_1}}{(d+b_1) \cdots (d+b_2-1)} \frac{\text{tr}(F_{\ell} \sum_{\pi \in S_{B_2}} \pi^d \otimes I^{\otimes \overline{B_2}})}{\text{tr}(F_{\ell} \sum_{\pi \in S_{B_1}} \pi^d \otimes I^{\otimes \overline{B_1}})} \\ &\geq \frac{d^{b_2-b_1}}{(d+b_1) \cdots (d+b_2-1)} = \prod_{i=b_1}^{b_2-1} \left(1 - \frac{i}{d+i} \right) \geq \prod_{i=b_1}^{b_2-1} \left(1 - \frac{i}{d} \right) \\ &\geq 1 - \sum_{i=b_1}^{b_2-1} \frac{i}{d} = 1 - \frac{b_2(b_2-1) - b_1(b_1-1)}{2d} \geq 1 - \frac{b_2^2 - b_1^2}{2d}. \end{aligned}$$

Here, we use Haar integral in the first line, Lemma 6 in the second line, and the Bernoulli inequality $\prod_{i=1}^n (1 - x_i) \geq 1 - \sum_{i=1}^n x_i$, $x_i \in (0, 1)$ in the last line. Hence, Eq. (C2) holds by Lemma 2(b).

Now we can bound the total variation distance in Eq. (C1):

$$\begin{aligned}
d_{\text{TV}}(\mathbb{E}_\psi p^{\rho_1} \mathbb{E}_\psi p^{\rho_2}) &= \frac{1}{2} \sum_{\ell} |\mathbb{E}_\psi p^{\rho_1}(\ell) - \mathbb{E}_\psi p^{\rho_2}(\ell)| \\
&= \frac{1}{2} \sum_{\ell} |\text{tr}(F_{\ell}(\mathbb{E}_\psi(a\psi + a(I/d) + (1-a-\varepsilon)(I/d))^{\otimes T} - \mathbb{E}_\psi(a\psi + \varepsilon\psi + (1-a-\varepsilon)(I/d))^{\otimes T})| \\
&= \frac{1}{2} \sum_{B_1, B_2, B_3 \text{ partition}[T]} a^{|B_1|} \varepsilon^{|B_2|} (1-a-\varepsilon)^{|B_3|} \sum_{\ell} |\text{tr}(F_{\ell}(\mathbb{E}_\psi \psi^{\otimes B_1 \cup B_2} \otimes (I/d)^{\otimes B_3} - \mathbb{E}_\psi \psi^{\otimes B_1} \otimes (I/d)^{\otimes B_2 \cup B_3}))| \\
&\leq \frac{1}{2} \sum_{B_1, B_2, B_3 \text{ partition}[T]} a^{|B_1|} \varepsilon^{|B_2|} (1-a-\varepsilon)^{|B_3|} \frac{|B_1 \cup B_2|^2 - |B_1|^2}{d} \\
&= \frac{1}{2d} \sum_{b_1+b_2+b_3=T} \binom{T}{b_1, b_2, b_3} a^{b_1} \varepsilon^{b_2} (1-a-\varepsilon)^{b_3} ((b_1+b_2)^2 - b_1^2) \\
&= \frac{1}{2d} \sum_{b_1+b_2+b_3=T} a^{b_1} \varepsilon^{b_2} (1-a-\varepsilon)^{b_3} (2b_1b_2 + b_2(b_2-1) + b_2) \binom{T}{b_1, b_2, b_3} \\
&= \frac{1}{2d} \sum_{b_1+b_2+b_3=T} a^{b_1} \varepsilon^{b_2} (1-a-\varepsilon)^{b_3} \left(2T(T-1) \binom{T-2}{b_1-1, b_2-1, b_3} \right. \\
&\quad \left. + T(T-1) \binom{T-2}{b_1, b_2-2, b_3} + T \binom{T-1}{b_1, b_2-1, b_3} \right) \\
&= \frac{a\varepsilon T(T-1)}{d} + \frac{\varepsilon^2 T(T-1)}{2d} + \frac{T\varepsilon}{2d}.
\end{aligned}$$

Therefore, when $T \ll \sqrt{d}/\sqrt{\varepsilon}$, the TV distance is much less than $2/3$, and the algorithm cannot distinguish the two cases with high probability. $\blacksquare$

3. Improved lower bound restricted to protocols with fixed measurements

The analysis above gives an $\Omega(\text{median}\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}, 2^{n-k}/\varepsilon^2\})$ lower bound. There is a $\sqrt{\varepsilon}$ gap between this lower bound and the $O(\text{median}\{1/\varepsilon^2, 2^{n/2}/\varepsilon, 2^{n-k}/\varepsilon^2\})$ upper bound that we will show in the next section. We believe that the gap is due to some technical inadequacy in the lower bound proof. Here, we provide a piece of evidence by improving the lower bound in a restricted setting: estimating purity with a randomly fixed PVM.

Definition 7 (Estimating purity with a randomly fixed single-copy PVM). We say an algorithm $\mathcal{A}$ estimates the purity within additive error ε with a randomly fixed single-copy PVM, if it proceeds as the following: Given access to an unknown state ρ , $\mathcal{A}$ samples a unitary U from a well-designed distribution $\mathcal{U}$ and performs the single-copy PVM $M_U := \{U|x\rangle\langle x|U^\dagger\}_{x=0}^{d-1}$ on T copies of ρ . It then classically processes all outcomes and outputs an estimator $\hat{E}$ such that with probability at least $2/3$, $|\hat{E} - \text{tr}(\rho^2)| \leq \varepsilon$. T is called the sample complexity of $\mathcal{A}$.

In the single-copy scenario (i.e., $k=0$), one can verify that the aforementioned lower and upper bounds simplify to $\Omega(\max\{1/\varepsilon^2, 2^{n/2}/\sqrt{\varepsilon}\})$ and $O(\max\{1/\varepsilon^2, 2^{n/2}/\varepsilon\})$, respectively. Now we prove that in the setting of estimating with a randomly fixed PVM, the lower bound can be improved to $\Omega(\max\{1/\varepsilon^2, 2^{n/2}/\varepsilon\})$, matching the upper bound.

Theorem 8. Any algorithm $\mathcal{A}$ that estimates the purity within additive error ε with a randomly fixed single-copy PVM (as in Definition 7) has sample complexity at least $\Omega(\max\{1/\varepsilon^2, 2^{n/2}/\varepsilon\})$.

Proof. The $\Omega(1/\varepsilon^2)$ lower bound comes directly from Lemma 3. Here, we only need to prove the $\Omega(2^{n/2}/\varepsilon)$ lower bound when $1/\varepsilon \leq 2^{n/2}$. By Lemma 4, it suffices to prove a lower bound for Problem 5. Fix a unitary U sampled from $\mathcal{U}$. We index a leaf $\ell \in \text{Leaf}(\mathcal{T})$ by outcomes $(x_1, x_2, \dots, x_T) \in \{0, 1, \dots, d-1\}^T$ and define $|\ell\rangle := |x_1, \dots, x_T\rangle$. We calculate the probabilities of ℓ in both cases.

$$\begin{aligned}
p(\ell) &:= p^{\rho_m}(\ell) \\
&= \langle \ell | (U^\dagger)^{\otimes T} \rho_m^{\otimes T} U^{\otimes T} | \ell \rangle \langle \ell | (U^\dagger)^{\otimes T} \rho_m^{\otimes T} U^{\otimes T} | \ell \rangle \\
&= d^{-T}.
\end{aligned} \tag{C3}$$

$$\begin{aligned}
q(\ell) &:= \mathbb{E}_{\psi \sim \mathcal{D}_{n,\varepsilon}} p^\psi(\ell) \\
&= \langle \ell | (U^\dagger)^{\otimes T} \mathbb{E}_\psi \psi^{\otimes T} U^{\otimes T} | \ell \rangle \langle \ell | (U^\dagger)^{\otimes T} \mathbb{E}_\psi \psi^{\otimes T} U^{\otimes T} | \ell \rangle \\
&= \langle \ell | \mathbb{E}_\psi \psi^{\otimes T} | \ell \rangle \langle \ell | \mathbb{E}_\psi \psi^{\otimes T} | \ell \rangle.
\end{aligned} \tag{C4}$$

Here, we use the fact that the distribution of $U^\dagger \psi U$ is the same as the distribution of ψ (i.e., Haar measure is invariant under unitary multiplication). Therefore, the distributions p and q are indeed independent of U . The goal is to upper bound the success probability by Lemma 2(b) when the depth of the learning tree $T \ll \sqrt{d}/\varepsilon$.

We now explicitly calculate $q(\ell)$ using Lemma 1. Here, we explicitly specify the local dimension of a permutation operator. For a permutation $\pi \in S_T$, let π^D be the permutation operator on T copies, π^d be the permutation operator on the first n qubits of T copies, and $\pi^{1/\varepsilon}$ be the permutation operator on the last $\log(1/\varepsilon)$ qubits of T copies. Then, we have $\pi^D = \pi^d \otimes \pi^{1/\varepsilon}$. Let $c(\pi)$ be the number of cycles in π . By Lemma 1,

$$\begin{aligned} \mathbb{E}_{\psi \sim \mathcal{D}_{n,\varepsilon}} \psi^{\otimes T} &= \frac{\text{tr}_{>n}^{\otimes T}(\sum_{\pi \in S_T} \pi^D)}{D(D+1)\dots(D+T-1)} = \frac{\sum_{\pi \in S_T} \pi^d \text{tr}(\pi^{1/\varepsilon})}{D(D+1)\dots(D+T-1)} \\ &= \frac{\sum_{\pi \in S_T} \pi \varepsilon^{-c(\pi)}}{D(D+1)\dots(D+T-1)} = d^{-T} \frac{\sum_{\pi \in S_T} \pi \varepsilon^{T-c(\pi)}}{(1+\varepsilon/d)\dots(1+(T-1)\varepsilon/d)}. \end{aligned} \quad (\text{C5})$$

Here, $\text{tr}_{>n}^{\otimes T}$ means tracing out the last $\log(1/\varepsilon)$ qubits for all copies. We use Lemma 1 in the first line, $\pi^D = \pi^d \otimes \pi^{1/\varepsilon}$ in the second line, and $\text{tr}(\pi^{1/\varepsilon}) = \varepsilon^{-c(\pi)}$ in the third line. We omit the superscript d and write π^d as π from now on. Notice that for any term $\langle \ell | \pi | \ell | \ell | \pi | \ell \rangle$ to be nonzero, π must connect copies with the same outcomes, that is, $x_i = x_{\pi(i)}$ for any i . Define $B_y^\ell := \{i \in \{1, 2, \dots, T\} : x_i = y\}$ for $y \in \{0, 1, \dots, d-1\}$, $b_y^\ell := |B_y^\ell|$, and let $S_{B_y^\ell}$ be the set of permutations over copies in B_y^ℓ . $\langle \ell | \sum_{\pi \in S_T} \pi \varepsilon^{T-c(\pi)} | \ell \rangle$ has a block structure as follows:

$$\left\langle \ell \left| \sum_{\pi \in S_T} \pi \varepsilon^{T-c(\pi)} \right| \ell \right\rangle = \prod_{y=0}^{d-1} \left\langle y^{\otimes b_y^\ell} \left| \sum_{\pi \in B_y^\ell} \pi \varepsilon^{b_y^\ell - c(\pi)} \right| y^{\otimes b_y^\ell} \right\rangle = \prod_{y=0}^{d-1} \sum_{\pi \in B_y^\ell} \varepsilon^{b_y^\ell - c(\pi)}. \quad (\text{C6})$$

Recall that the coefficient of x^k in $x(x+1)\dots(x+t-1)$ is the unsigned Stirling number of the first kind, which is the number of permutations of t elements with k cycles (see, e.g., Ref. [46]). So, $x(x+1)\dots(x+t-1) = \sum_{\pi \in S_t} x^{c(\pi)}$. Setting $x = 1/\varepsilon$, we obtain that $\sum_{\pi \in B_y^\ell} \varepsilon^{b_y^\ell - c(\pi)} = (1+\varepsilon)(1+2\varepsilon)\dots(1+(b_y^\ell-1)\varepsilon)$. Combining with Eqs. (C3)–(C6), we finally obtain the likelihood ratio

$$L(\ell) := \frac{q(\ell)}{p(\ell)} = \frac{\prod_{y=0}^{d-1} (1+\varepsilon)\dots(1+(b_y^\ell-1)\varepsilon)}{(1+\varepsilon/d)\dots(1+(T-1)\varepsilon/d)}.$$

Since $x - \frac{x^2}{2} \leq \ln(1+x) \leq x$ for $x \geq 0$,

$$\begin{aligned} \ln(L(\ell)) &= \sum_{y=0}^{d-1} \sum_{i=1}^{b_y^\ell-1} \ln(1+i\varepsilon) - \sum_{i=1}^{T-1} \ln(1+i\varepsilon/d) \geq \sum_{y=0}^{d-1} \sum_{i=1}^{b_y^\ell-1} \left(i\varepsilon - \frac{(i\varepsilon)^2}{2} \right) - \sum_{i=1}^{T-1} \frac{i\varepsilon}{d} \\ &= \varepsilon \sum_{y=0}^{d-1} \binom{b_y^\ell}{2} - \frac{\varepsilon}{d} \binom{T}{2} - \varepsilon^2 \sum_{y=0}^{d-1} \frac{(b_y^\ell-1)b_y^\ell(2b_y^\ell-1)}{12} \\ &= \left(\varepsilon - \frac{1}{2}\varepsilon^2 \right) \sum_{y=0}^{d-1} \binom{b_y^\ell}{2} - \frac{\varepsilon}{d} \binom{T}{2} - \varepsilon^2 \sum_{y=0}^{d-1} \binom{b_y^\ell}{3} \\ &= \left(\varepsilon - \frac{1}{2}\varepsilon^2 \right) \sum_{i < j} 1[x_i = x_j] - \frac{\varepsilon}{d} \binom{T}{2} - \varepsilon^2 \sum_{i < j < k} 1[x_i = x_j = x_k]. \end{aligned} \quad (\text{C7})$$

We show in Lemma 8 that, assuming $T < \sqrt{d}/(100\varepsilon)$ and $1/\varepsilon < \sqrt{d}$, the right-hand side of Eq. (C7) is larger than -0.13 with probability at least 0.97. In other words, $\Pr_{\ell \sim p(\ell)}[L(\ell) \leq e^{-0.115}] \leq 0.03$. By Lemma 2(b), the success probability is at most $0.03 + (1 - e^{-0.115}) < 2/3$. Therefore, to achieve a $2/3$ success probability, $T = \Omega(\max\{1/\varepsilon^2, \sqrt{d}/\varepsilon\})$. ■

Lemma 8. Assume $T < \sqrt{d}/(100\varepsilon)$ and $1/\varepsilon < \sqrt{d}$. Suppose $x_1, x_2, \dots, x_T$ are uniformly drawn from $\{0, 1, \dots, d-1\}$. With probability at least 0.97, the right-hand side of Eq. (C7) is at least -0.13 .

Proof. Define two random variables $X := \sum_{i < j} 1[x_i = x_j]$, $Y := \sum_{i < j < k} 1[x_i = x_j = x_k]$. It is easy to see $\mathbb{E}[X] = \frac{1}{d} \binom{T}{2} < \frac{1}{10000\varepsilon^2}$ and $\mathbb{E}[Y] = \frac{1}{d^2} \binom{T}{3} < \frac{1}{10000\varepsilon^3\sqrt{d}} < \frac{1}{10000\varepsilon^2}$.

Furthermore, for $i_1 < j_1, i_2 < j_2$,

$$\begin{aligned} &\mathbb{E}(1[x_{i_1} = x_{j_1}]1[x_{i_2} = x_{j_2}]) - \mathbb{E}(1[x_{i_1} = x_{j_1}])\mathbb{E}(1[x_{i_2} = x_{j_2}]) \\ &= \begin{cases} \frac{1}{d} - \frac{1}{d^2}, & i_1 = i_2, j_1 = j_2, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

So, $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = (\frac{1}{d} - \frac{1}{d^2}) \binom{T}{2} < \frac{1}{d} \binom{T}{2} < \frac{1}{10000\varepsilon^2}$.

By Chebyshev's inequality, with probability at least 0.99, $|X - \mathbb{E}[X]| \leq 10\sqrt{\text{Var}[X]} < \frac{1}{10\varepsilon}$. By Markov's inequality, with probability at least 0.99, $X \leq 100\mathbb{E}[X] < \frac{1}{100\varepsilon^2}$. With probability at least 0.99, $Y \leq 100\mathbb{E}[Y] < \frac{1}{100\varepsilon^2}$. Therefore, with probability at least 0.97, all events happen, and the right-hand side of Eq. (C7) is

$$\varepsilon(X - \mathbb{E}[X]) - \frac{1}{2}\varepsilon^2 X - \varepsilon^2 Y > -0.115. \quad \blacksquare$$

ALGORITHM 1. Distributed inner product estimation using unentangled local measurement [9].

Input : Accuracy demand ε , N copies of unknown ρ and σ .

Output : An estimation of $\text{tr}(\rho\sigma)$ within additive error ε .

- 1: Divide the N copies of ρ and σ into N_b groups, each containing m copies of ρ and σ .
- 2: **for** $i = 1, \dots, N_b$ **do**
- 3: Alice and Bob sample a ($d = 2^n$)-dimensional Haar random unitary U_i , and load a fresh batch of m copies of ρ and σ , respectively.
- 4: Alice measures m copies of ρ in the basis $\{U_i^\dagger |b\rangle \langle b| U_i\}_{b=0}^{d-1}$ to obtain outputs $x_1, \dots, x_m$. Alice then sends these classical outputs to Bob.
- 5: Bob measures m copies of σ in the basis $\{U_i^\dagger |b\rangle \langle b| U_i\}_{b=0}^{d-1}$ to obtain outputs $y_1, \dots, y_m$.
- 6: Bob computes the collision estimator

$$\tilde{g}_i = \frac{1}{m^2} \sum_{j,k=1}^m \mathbb{I}[x_j = y_k].$$

- 7: Let $w_i = (2^n + 1)\tilde{g}_i - 1$.

8: **end for**

- 9: Bob outputs $w = \frac{1}{N_b} \sum_{i=1}^{N_b} w_i$.

APPENDIX D: SAMPLE COMPLEXITY OF INNER PRODUCT ESTIMATION

1. Inner product estimation with bounded quantum communication

In this Appendix, we provide an algorithm for inner product estimation with k -qubit one-way quantum communication (Problem 1), and thus an algorithm for purity estimation with k -qubit quantum memory (Problem 2) by Claim 1.

When $k = 0$ in Problem 1, the task reduces to distributed inner product estimation with unentangled measurements where Alice and Bob estimate $\text{tr}(\rho\sigma)$ using single-copy local measurements and classical communication (i.e., LOCC). In this case, Ref. [9] proposes a collision-based algorithm using single-copy local measurements and classical communications between Alice and Bob as in Algorithm 1.

Here, we write the protocol in a one-way communication version. The following lemma proves the performance of Algorithm 1:

Lemma 9 (Performance of Algorithm 1, [9]). There exists a choice of N_b and m such that

$$N = N_b m = O\left(\max\left\{\frac{1}{\varepsilon^2}, \frac{2^{n/2}}{\varepsilon}\right\}\right),$$

and the estimator w output from Algorithm 1 is an approximation of $\text{tr}(\rho\sigma)$ within additive error ε with a high probability.

In Problem 1, Alice and Bob are allowed to have k -qubit one-way quantum communication for each copy pair ρ and σ . To this end, we propose the following partial SWAP test algorithm making use of one-way quantum communication in Algorithm 2.

Now, we analyze the performance of Algorithm 2. From Line 1 to Line 3, Alice and Bob are to perform the folklore SWAP test on the first k qubits of ρ and σ . By choosing $N_k = O(1/\varepsilon^2)$, one can guarantee that Alice and Bob can estimate $\tilde{f}_k$

ALGORITHM 2. Partial SWAP test using k -qubit one-way communication unentangled local measurement.

Input: Accuracy demand ε , $N + N_k$ copies of unknown ρ and σ where $N_k = O(1/\varepsilon^2)$.

Output: An estimation of $\text{tr}(\rho\sigma)$ within additive error ε .

- 1: Alice and Bob load N_k copies of ρ and σ , respectively.
- 2: Alice sends the first k qubits of each copy to Bob.
- 3: Bob applies a partial SWAP test, i.e. POVM $\{\frac{I_{2k} + \text{SWAP}_k}{2}, \frac{I_{2k} - \text{SWAP}_k}{2}\}$, on the first k qubits of ρ received and σ for each copy. He can then estimate $\tilde{f}_k$ within $\varepsilon/2$ of the inner product $f_k = \text{tr}(\text{tr}_{>k}(\rho)\text{tr}_{>k}(\sigma))$.
- 4: Divided the N copies of ρ and σ into N_b groups, each containing m copies of ρ and σ .
- 5: **for** $i = 1, \dots, N_b$ **do**
- 6: Alice and Bob sample a 2^{n-k} -dimensional Haar random unitary U_i , and load a fresh batch of m copies of ρ and σ , respectively.
- 7: Alice sends the first k qubits of each copy to Bob.
- 8: Bob applies a partial SWAP test, i.e. POVM $\{\frac{I_{2k} + \text{SWAP}_k}{2}, \frac{I_{2k} - \text{SWAP}_k}{2}\}$, on the first k qubits of ρ received and σ for each copy, and obtains outcome $z_1, \dots, z_m$.
- 9: Alice measures m copies of the remaining $n - k$ qubits of ρ each in the basis $\{U_i^\dagger |b\rangle \langle b| U_i\}_{b=0}^{2^{n-k}-1}$ to obtain outputs $x_1, \dots, x_m$. Alice then sends these classical outputs to Bob.
- 10: Bob measures m copies of the remaining $n - k$ qubits of σ each in the basis $\{U_i^\dagger |b\rangle \langle b| U_i\}_{b=0}^{2^{n-k}-1}$ to obtain outputs $y_1, \dots, y_m$.
- 11: Bob computes the value

$$\tilde{g}_i = \frac{1}{m} \sum_{i=1}^m (\mathbb{I}[z_i = 1, x_i = y_i] - \mathbb{I}[z_i = -1, x_i = y_i])$$

- 12: Let $w_i = (2^{n-k} + 1)\tilde{g}_i - \tilde{f}_k$

13: **end for**

- 14: Bob outputs $w = \frac{1}{N_b} \sum_{i=1}^{N_b} w_i$

within $\varepsilon/2$ of the inner product $f_k = \text{tr}(\text{tr}_{>k}(\rho)\text{tr}_{>k}(\sigma))$ with high probability.

We then consider the remaining part of the algorithm. In each iteration $i = 1, \dots, N_b$, we denote $X = \{x_1, \dots, x_m\}$, $Y = \{y_1, \dots, y_m\}$, $Z = \{z_1, \dots, z_m\}$, and $S = \{X, Y, Z\}$. We can thus regard $\tilde{g}_i$ as a random function $g(U, S)$. Note that for fixed U , we have

$$\Pr[z_i = 1, x_i = y_i]$$

$$= \sum_{b=0}^{d-1} \text{tr}\left(\frac{I_{2k} + \text{SWAP}_k}{2} \otimes (U^\dagger |b\rangle \langle b| U)^{\otimes 2} \cdot \rho \otimes \sigma\right),$$

$$\Pr[z_i = -1, x_i = y_i]$$

$$= \sum_{b=0}^{d-1} \text{tr}\left(\frac{I_{2k} - \text{SWAP}_k}{2} \otimes (U^\dagger |b\rangle \langle b| U)^{\otimes 2} \cdot \rho \otimes \sigma\right).$$

The expectation value of $g(U, S)$ is thus

$$\mathbb{E}_{U,S}[g(U, S)]$$

$$= \mathbb{E}_U \left[\sum_{b=0}^{d-1} \text{tr}(\text{SWAP}_k \otimes (U^\dagger |b\rangle \langle b| U)^{\otimes 2} \cdot \rho \otimes \sigma) \right]$$

$$\begin{aligned}
&= \frac{\text{tr}(\text{SWAP}_k \otimes (I_{n-k} \otimes I_{n-k} + \text{SWAP}_{n-k}) \cdot \rho \otimes \sigma)}{2^{n-k} + 1} \\
&= \frac{f + f_k}{2^{n-k} + 1},
\end{aligned}$$

where $f_k = \text{tr}(\text{tr}_{>k}(\rho)\text{tr}_{>k}(\sigma))$ and $f = \text{tr}(\rho\sigma)$. This implies the following:

Lemma 10. The value w output from Bob in Algorithm 2 is an unbiased estimation for $\text{tr}(\rho\sigma)$.

Note that since Alice and Bob can estimate $\tilde{f}_k$ within $\varepsilon/2$ of the inner product $f_k = \text{tr}(\text{tr}_{>k}(\rho)\text{tr}_{>k}(\sigma))$ with high probability, we only need to compute the parameters N , N_b , and m such that the remaining part of w deviates from $\text{tr}(\rho\sigma) + f_k$ for at most $\varepsilon/2$. To achieve this, we consider the variance of w . Note that for

$$\text{Var}(w) = \frac{1}{N_b^2} \sum_{i=1}^{N_b} \text{Var}(w_i) = \frac{(2^{n-k} + 1)^2}{N_b} \text{Var}(g(U, S)), \quad (\text{D1})$$

it is sufficient to compute the variance of $g(U, S)$. The variance of $\text{Var}(g(U, S))$ can be divided into the variance of the random unitary U and the variance of the intrinsic randomness of quantum measurement. This indicates that

$$\text{Var}(g(U, S)) = \text{Var}_U(\mathbb{E}[g(U, S)|U]) + \mathbb{E}_U[\text{Var}(g(U, S)|U)]. \quad (\text{D2})$$

We now compute the two terms. We have the following lemma for the first term:

Lemma 11. The first term in Eq. (D2) is bounded by

$$\text{Var}_U(\mathbb{E}[g(U, S)|U]) = O\left(\frac{1}{2^{3(n-k)}}\right).$$

Lemma 12 (See Lemma 18 of Ref. [9]).

$$\mathbb{E}_{\psi' \sim \mu_{\perp\psi}}[|\psi'\rangle\langle\psi'|^{\otimes 2}] = \frac{1}{2^{n-k}(2^{n-k} - 1)}((I - |\psi\rangle\langle\psi|)^{\otimes 2} + (I - |\psi\rangle\langle\psi|)^{\otimes 2} \text{SWAP}_{n-k}(I - |\psi\rangle\langle\psi|)^{\otimes 2}).$$

By plugging this lemma into the case when $b_1 \neq b_2$, it can be written as

$$\begin{aligned}
&2^{n-k}(2^{n-k} - 1) \mathbb{E}_{\substack{\psi \sim \mu \\ \psi' \sim \mu_{\perp\psi}}} \text{tr}((\text{SWAP}_k)^{\otimes 2} \otimes (|\psi\rangle\langle\psi|)^{\otimes 2} \otimes (|\psi'\rangle\langle\psi'|)^{\otimes 2} \cdot (\rho \otimes \sigma)^{\otimes 2}) \\
&= \mathbb{E}_{\psi \sim \mu} \text{tr}((\text{SWAP}_k)^{\otimes 2} \otimes (|\psi\rangle\langle\psi|)^{\otimes 2} \otimes (I \otimes I + \text{SWAP}_{n-k}) \cdot (\rho \otimes \sigma)^{\otimes 2}) + O(2^{-3(n-k)}) \\
&= \frac{\text{tr}((\text{SWAP}_k)^{\otimes 2} \otimes (I \otimes I + \text{SWAP}_{n-k}) \otimes (I \otimes I + \text{SWAP}_{n-k}) \cdot (\rho \otimes \sigma)^{\otimes 2})}{2^{n-k}(2^{n-k} + 1)} + O(2^{-3(n-k)}) \\
&= \frac{(f + f_k)^2}{2^{n-k}(2^{n-k} + 1)} + O(2^{-3(n-k)}). \quad (\text{D5})
\end{aligned}$$

By inserting Eqs. (D4) and (D5) into Eq. (D3), we arrive at

$$\text{Var}_U(\mathbb{E}[g(U, S)|U]) = \frac{(f + f_k)^2}{2^{n-k}(2^{n-k} + 1)} + O(2^{-3(n-k)}) - \left(\frac{f + f_k}{2^{n-k} + 1}\right)^2 = O(2^{-3(n-k)}),$$

which proves Lemma 11. ■

Proof. The first term in Eq. (D2) can be computed as

$$\begin{aligned}
&\text{Var}_U(\mathbb{E}[g(U, S)|U]) \\
&= \mathbb{E}_U[g(U)^2] - \mathbb{E}_U[g(U)]^2 \\
&= \sum_{b_1, b_2=0}^{2^{n-k}-1} \mathbb{E}_U[\text{tr}((\text{SWAP}_k)^{\otimes 2} \otimes (U^\dagger |b_1\rangle\langle b_1| U)^{\otimes 2} \\
&\quad \otimes (U^\dagger |b_2\rangle\langle b_2| U)^{\otimes 2} \cdot (\rho \otimes \sigma)^{\otimes 2})] - \mathbb{E}_U[g(U)]^2. \quad (\text{D3})
\end{aligned}$$

While the last term above is $\mathbb{E}_U[g(U)]^2 = (\frac{f+f_k}{2^{n-k}+1})^2$, the first term can be further decomposed into the case when $b_1 = b_2$ and $b_1 \neq b_2$. For $b_1 = b_2$, the total contribution can be computed as

$$\begin{aligned}
&2^{n-k} \mathbb{E}_{\psi \sim \mu} [\text{tr}((\text{SWAP}_k)^{\otimes 2} \otimes (|\psi\rangle\langle\psi|)^{\otimes 4} \cdot (\rho \otimes \sigma)^{\otimes 2})] \\
&= \frac{1}{(2^{n-k} + 1)(2^{n-k} + 2)(2^{n-k} + 3)} \sum_{\pi \in S_4} \text{tr}((\text{SWAP}_k)^{\otimes 2} \\
&\quad \otimes \pi^{2^{n-k}} \cdot (\rho \otimes \sigma)^{\otimes 2}) \\
&= O\left(\frac{1}{2^{3(n-k)}}\right), \quad (\text{D4})
\end{aligned}$$

where the second line follows from Lemma 1 and the last line follows from the fact that the summation is over 24 elements each of at most 1. When $b_1 \neq b_2$, the contribution can be computed as

$$\begin{aligned}
&2^{n-k}(2^{n-k} - 1) \mathbb{E}_{\substack{\psi \sim \mu \\ \psi' \sim \mu_{\perp\psi}}} \text{tr}((\text{SWAP}_k)^{\otimes 2} \otimes (|\psi\rangle\langle\psi|)^{\otimes 2} \\
&\quad \otimes (|\psi'\rangle\langle\psi'|)^{\otimes 2} \cdot (\rho \otimes \sigma)^{\otimes 2}).
\end{aligned}$$

Here, $\mu_{\perp\psi}$ is the Haar measure over states perpendicular to $|\psi\rangle$. We have the following lemma to deal with the above equation:

Regarding the second term, we have the following lemma:
Lemma 13. The second term in Eq. (D2) is bounded by

$$\mathbb{E}_U[\text{Var}(g(U, S)|U)] = O\left(\frac{1}{m2^{n-k}}\right).$$

Proof. We first note that when we fix a U ,

$$\begin{aligned} \text{Var}(g) &= \mathbb{E}[g^2] - \mathbb{E}[g]^2 \\ &= \frac{1}{m^2} \sum_{i,j=1}^m (\mathbb{1}[z_i = 1, x_i = y_i] - \mathbb{1}[z_i = -1, x_i = y_i]) \\ &\quad (\mathbb{1}[z_j = 1, x_j = y_j] - \mathbb{1}[z_j = -1, x_j = y_j]) - \mathbb{E}[g]^2 \\ &= \frac{\mathbb{E}[g']}{m} + \left(\frac{m-1}{m} - 1\right) \mathbb{E}[g]^2 \\ &\leq \frac{\mathbb{E}[g']}{m}, \end{aligned}$$

Here,

$$g' = \tilde{g}_i = \frac{1}{m} \sum_{i=1}^m (\mathbb{1}[z_i = 1, x_i = y_i] + \mathbb{1}[z_i = -1, x_i = y_i])$$

and the third line follows from considering $i = j$ and $i \neq j$. Averaging over U , we have

$$\mathbb{E}_U[\text{Var}(g(U, S)|U)] = \frac{\mathbb{E}_{U,S}[g']}{m} = O\left(\frac{1}{m2^{n-k}}\right),$$

where the second step follows from the fact that

$$\begin{aligned} \mathbb{E}_{U,S}[g'] &= \mathbb{E}_U \left[\sum_{b=0}^{d-1} \text{tr}(I_k \otimes I_k \otimes (U^\dagger |b\rangle \langle b| U)^{\otimes 2} \cdot \rho \otimes \sigma) \right] \\ &= \frac{1}{2^{n-k} + 1} \text{tr}(I_k \otimes I_k \otimes (I_{n-k} \otimes I_{n-k} + \text{SWAP}_{n-k}) \cdot \rho \otimes \sigma) \\ &\leq \frac{2}{2^{n-k} + 1}. \end{aligned}$$

Combining Lemmas 10, 11, and 13, we have the following lemma regarding the performance of Algorithm 2:

Lemma 14 (Performance of Algorithm 2). There exists a choice of N_b and m such that

$$N = N_b m = O\left(\frac{2^{n-k}}{\varepsilon^2}\right),$$

and the estimator w output from Algorithm 2 is an approximation of $\text{tr}(\rho)$ within additive error ε with a high probability.

Proof. We first combine Lemmas 11 and 13; we can compute the variance in Eq. (D2) as

$$\begin{aligned} \text{Var}(g(U, S)) &= \text{Var}_U(\mathbb{E}[g(U, S)|U]) + \mathbb{E}_U[\text{Var}(g(U, S)|U)] \\ &= O\left(\frac{1}{m2^{n-k}} + \frac{1}{2^{3(n-k)}}\right). \end{aligned}$$

Plugging the above equation in Eq. (D1), we have

$$\begin{aligned} \text{Var}(w) &= \frac{(2^{n-k} + 1)^2}{N_b} \text{Var}(g(U, S)) \\ &= \frac{1}{N_b} O\left(\frac{2^{n-k}}{m} + \frac{1}{2^{n-k}}\right). \end{aligned}$$

Note that from Lemma 10, w output from Bob is an unbiased estimation for $\text{tr}(\rho\sigma)$. It is enough to satisfy

$$\text{Var}(w) \leq \left(\frac{\varepsilon}{2}\right)^2.$$

Then, we have with high probability that $w + \tilde{f}_k$ deviates from $\text{tr}(\rho\sigma) + f_k$ for at most $\varepsilon/2$. As $\tilde{f}_k$ deviates from f_k for at most $\varepsilon/2$ with high probability, we conclude that w deviates from $\text{tr}(\rho\sigma)$ for at most ε with high probability. Therefore, we require

$$\frac{2^{n-k}}{N_b m} \leq O(\varepsilon^2), \quad \frac{1}{N_b 2^{n-k}} \leq O(\varepsilon^2).$$

Here, we set m to 1, and $N_b = O(2^{n-k}/\varepsilon^2)$ satisfies the requirements. Thus, we have

$$N = N_b m = O\left(\frac{2^{n-k}}{\varepsilon^2}\right). \quad \blacksquare$$

Therefore, the sample complexity of Algorithm 2 is $N + N_k = O(2^{n-k}/\varepsilon^2)$. Combining Algorithms 1 and 2, we reach the following theorem:

Theorem 9. Given unknown quantum states ρ and σ , error parameter ε , and the capacity of one-way communication k in Problem 1, the combination of Algorithms 1 and 2 yields an algorithm that solves Problem 1 with a high probability using

$$\begin{aligned} O\left(\min\left\{\frac{2^{n-k}}{\varepsilon^2}, \max\left\{\frac{1}{\varepsilon^2}, \frac{2^{n/2}}{\varepsilon}\right\}\right\}\right) \\ = O\left(\text{median}\left\{\frac{1}{\varepsilon^2}, \frac{2^{n/2}}{\varepsilon}, \frac{2^{n-k}}{\varepsilon^2}\right\}\right) \end{aligned}$$

copies of ρ and σ .

2. Lower bound for mixed state decisional inner product estimation

In this Appendix, we provide two lower bounds for mixed state versions of the DIPE introduced in Ref. [9]. For both versions, we obtain a sample lower bound of $\sqrt{d/\text{tr}(\rho^2)}$. For Problem 6, our bound holds for protocols with one-way communication between Alice and Bob. The proof follows directly from concentration of measure and, in particular, does not require Chiribella's theorem. For Problem 11, we can show the lower bound for all separable measurements (see Definition 1) by adapting the proof technique in Ref. [9].

a. Lower bounds from concentration of measure

We consider the following mixed variant of the DIPE:

Problem 6 (Mixed state decisional inner product estimation with fixed spectrum, mixed DIPE I). Fixing an underlying

quantum state ρ , Alice and Bob are given copies of states, promised that one of the following two cases holds:

(1) Alice and Bob both have state $U\rho U^\dagger$ for a Haar randomly chosen unitary U .

(2) Alice has state $U\rho U^\dagger$ and Bob has state $V\rho V^\dagger$ for independent Haar randomly chosen unitary U and V .

Their goal is to decide which case they are in with high probability.

Theorem 10. Deciding DIPE I for any fixed ρ requires at least $\Omega(\sqrt{d/\text{tr}(\rho^2)})$ copies if Alice and Bob are allowed to perform local measurements and one-way communication protocols.

We need a standard lemma bounding the Lipschitz constant for k -copy measurements:

Lemma 15. $f_{U,\rho} := \text{tr}[M(U\rho U^\dagger)^{\otimes k}]$ has Lipschitz constant $L \leq 2k\sqrt{\text{tr}[\rho^2]}$ for any operator M with $\|M\|_\infty \leq 1$.

Proof. We calculate

$$\begin{aligned} & |\text{tr}[M(U\rho U^\dagger)^{\otimes k}] - \text{tr}[M(V\rho V^\dagger)^{\otimes k}]| \\ &= |\text{tr}[M((U\rho U^\dagger)^{\otimes k} - (V\rho V^\dagger)^{\otimes k})]| \\ &\leq \|M\|_\infty \|(U\rho U^\dagger)^{\otimes k} - (V\rho V^\dagger)^{\otimes k}\|_1 \\ &\leq \|(U\rho U^\dagger)^{\otimes k} - (V\rho V^\dagger)^{\otimes k}\|_1. \end{aligned}$$

We can now use the following bound: Consider matrices A, B, C , and D with $\|A\|_1 = \|D\|_1 = 1$. Then,

$$\begin{aligned} & \|A \otimes B - C \otimes D\|_1 \\ &\leq \|A \otimes (B - D) + (A - C) \otimes D\|_1 \\ &\leq \|A\|_1 \|B - D\|_1 + \|A - C\|_1 \|D\|_1 \\ &\leq \|B - D\|_1 + \|A - C\|_1. \end{aligned}$$

We apply this bound k times and obtain

$$\begin{aligned} & |\text{tr}[M(U\rho U^\dagger)^{\otimes k}] - \text{tr}[M(V\rho V^\dagger)^{\otimes k}]| \\ &\leq k\|U\rho U^\dagger - V\rho V^\dagger\|_1 \\ &= k\|U\rho U^\dagger - U\rho V^\dagger + U\rho V^\dagger - V\rho V^\dagger\|_1 \\ &\leq k\|U\rho(U^\dagger - V^\dagger)\|_1 + k\|(U - V)\rho V^\dagger\|_1 \\ &\leq 2k\|\rho\|_2\|U - V\|_2 \end{aligned}$$

by Hölder's inequality and Cauchy-Schwarz. $\blacksquare$

We can now prove Theorem 10:

Proof of Theorem 10. For quantum measurements $M_{A,B} = \sum_i M_i \otimes N_i$ performed by Alice and Bob with one-way communications (see ‘‘Measurements with classical communications’’ part in Appendix A 1 for details), we will show that the assumption

$$\begin{aligned} & \mathbb{E}_{U,V \sim \mu} \text{tr}[M_{A,B}(U\rho U^\dagger)^{\otimes k} \otimes (V\rho V^\dagger)^{\otimes k}] \\ &= \mathbb{E}_{U \sim \mu} \sum_i \text{tr}[M_i(U\rho U^\dagger)^{\otimes k}] \mathbb{E}_{V \sim \mu} \text{tr}[N_i(V\rho V^\dagger)^{\otimes k}] \\ &\geq \frac{2}{3} \end{aligned}$$

leads to a contradiction for $k = o(\sqrt{d/\text{tr}(\rho^2)})$. Theorem 5 implies

$$\begin{aligned} & |\text{Pr}[\text{tr}[N_i(V\rho V^\dagger)^{\otimes k}] - \mathbb{E}_V \text{tr}[N_i(V\rho V^\dagger)^{\otimes k}]] \geq \tau| \\ &\leq 4e^{-\frac{2d\tau^2}{36\pi^3 \text{tr}[\rho^2]k^2}}. \end{aligned}$$

With the notation $B_\tau := \{|\text{tr}[N_i(V\rho V^\dagger)^{\otimes k}] - \mathbb{E}_V \text{tr}[N_i(V\rho V^\dagger)^{\otimes k}]| \leq \tau\}$, we have.

$$\begin{aligned} & \mathbb{E}_{U \sim \mu} \text{tr}[M_{A,B}(U\rho U^\dagger)^{\otimes k} \otimes (U\rho U^\dagger)^{\otimes k}] \\ &= \sum_i \mathbb{E}_{U \sim \mu} \text{tr}[M_i(U\rho U^\dagger)^{\otimes k}] \text{tr}[N_i(U\rho U^\dagger)^{\otimes k}] \\ &\geq \sum_i \int_{B_\tau} \text{tr}[M_i(U\rho U^\dagger)^{\otimes k}] \text{tr}[N_i(U\rho U^\dagger)^{\otimes k}] d\mu(U) \\ &\geq \sum_i \int_{B_\tau} \text{tr}[M_i(U\rho U^\dagger)^{\otimes k}] \\ &\quad \times \left(\mathbb{E}_{V \sim \mu} \text{tr}[N_i(V\rho V^\dagger)^{\otimes k}] - \tau \right) d\mu(U) \\ &\geq \mathbb{E}_{U \sim \mu} \sum_i \text{tr}[M_i(U\rho U^\dagger)^{\otimes k}] \mathbb{E}_{V \sim \mu} \text{tr}[N_i(V\rho V^\dagger)^{\otimes k}] - \tau \\ &\quad - \int_{U(d) \setminus B_\tau} \sum_i \text{tr}[M_i(U\rho U^\dagger)^{\otimes k}] \\ &\geq \frac{2}{3} - \tau - 4e^{-\frac{2d\tau^2}{36\pi^3 \text{tr}[\rho^2]k^2}}. \end{aligned}$$

We can now choose $\tau = 1/10$ and find that for $k \leq \sqrt{d}/400\sqrt{\text{tr}[\rho^2]}$ mixed DIPE I cannot be solved with general one-way protocols. $\blacksquare$

b. Lower bound for all separable measurements

We consider a second mixed variant of the DIPE to prove a lower bound even if Alice and Bob are allowed to perform arbitrary measurements. The following is a restatement of Problem 3:

Problem 7 (Convex mixture of Haar random states, mixed DIPE II). Alice and Bob are given copies of states, promised that one of the following two cases holds:

(1) Alice and Bob both have state $\rho = \frac{1}{r} \sum_{i=1}^r |\psi_i\rangle\langle\psi_i|$, where the states ψ_i are drawn iid from μ .

(2) Alice has state $\rho = \frac{1}{r} \sum_{i=1}^r |\psi_i\rangle\langle\psi_i|$ and Bob has state $\sigma = \frac{1}{r} \sum_{i=1}^r |\phi_i\rangle\langle\phi_i|$, where the states $|\phi_i\rangle$ and $|\psi_i\rangle$ are drawn iid from μ .

Their goal is to decide which case they are in with high probability.

We will prove the following theorem:

Theorem 11. Any protocol involving arbitrary interactive protocols between Alice and Bob requires $k = \Omega(\sqrt{d/\text{tr}(\rho^2)})$ copies to decide mixed DIPE II with high probability.

As in Ref. [9], we can use Chiribella's theorem [47] that characterizes the measure-and-prepare channel via the optimal cloning channel. The measure-and-prepare channel [38] is defined as

$$\text{MP}_{a \rightarrow b}(\rho) := \binom{a+d-1}{a} \mathbb{E}_{\psi \sim \mu} \text{tr}(\rho \cdot \psi^{\otimes a}) \psi^{\otimes b}.$$

As we will see, the measure-and-prepare channel will show up naturally when analyzing mixed DIPE II. On the other hand, we have the optimal cloning channel:

$$\text{Clone}_{a \rightarrow b}(\rho) := \binom{d+a-1}{a} \mathcal{S}_b(\rho \otimes \mathbb{1}^{\otimes b-a}) \mathcal{S}_b.$$

Chiribella's theorem [47] then establishes the following connection:

Theorem 12 (Chiribella's theorem). For integers $a, b \geq 0$ and a density matrix ρ , it holds that

$$\text{MP}_{a \rightarrow b}(\rho) = \sum_{s=0}^{\min a, b} \frac{\binom{a}{s} \binom{d+b-1}{b-s}}{\binom{d+a+b-1}{b}} \text{Clone}_{s \rightarrow b}(\text{tr}_{a-s}[\rho]).$$

Theorem 12 allows us to prove the following generalization of Lemma 7 in Ref. [9]:

Lemma 16. For any input state ρ , we have

$$\mathbb{E}_{\psi \sim \mu} \text{tr}[\rho \psi^{\otimes a}] \psi^{\otimes b} \geq e^{-ab/d} \frac{\mathcal{S}_b}{\binom{d+b-1}{b}}.$$

Proof. Applying Theorem 12, we calculate

$$\begin{aligned} & \mathbb{E}_{\psi} \text{tr}[\rho \psi^{\otimes a}] \psi^{\otimes b} \\ &= \frac{1}{\binom{a+d-1}{a}} \text{MP}_{a \rightarrow b}(\rho) \\ &= \frac{1}{\binom{a+d-1}{a}} \sum_{s=0}^{\min a, b} \frac{\binom{a}{s} \binom{d+b-1}{b-s}}{\binom{d+a+b-1}{b}} \text{Clone}_{s \rightarrow b}(\text{tr}_{a-s}[\rho]) \\ &\geq \frac{1}{\binom{a+d-1}{a}} \frac{\binom{a}{0} \binom{d+b-1}{b}}{\binom{d+a+b-1}{b}} \text{Clone}_{0 \rightarrow b}(\text{tr}[\rho]) \\ &= \frac{1}{\binom{a+d-1}{a}} \frac{\binom{d+b-1}{b}}{\binom{d+a+b-1}{b}} \frac{\mathcal{S}_b}{\binom{d+b-1}{b}} \\ &\geq \frac{1}{\binom{a+d-1}{a}} e^{-ab/d} \frac{\mathcal{S}_b}{\binom{d+b-1}{b}}, \end{aligned}$$

where the last line follows from

$$\frac{\prod_{i=0}^{b-1} (d+i)}{\prod_{i=0}^{b-1} (d+a+i)} = \prod_{i=1}^{b-1} \frac{1}{1 + \frac{a}{d+i}} \geq e^{-\sum_{i=0}^{b-1} \frac{a}{d+i}} \geq e^{-\frac{ab}{d}}.$$

Consequently, we find

$$\begin{aligned} \rho_{\text{Bob}} &= \frac{1}{r^{2k}} \sum_{i_1, \dots, i_k, j_1, \dots, j_k} \mathbb{E}_{\psi_1, \dots, \psi_r} \text{tr} \left[A_t \bigotimes_{l=1}^k |\psi_{i_l}\rangle\langle\psi_{i_l}| \right] \bigotimes_{l=1}^k |\psi_{j_l}\rangle\langle\psi_{j_l}| \\ &= \frac{1}{r^{2k}} \sum_{i_1, \dots, i_k, j_1, \dots, j_k} \mathbb{E}_{\psi_1, \dots, \psi_r} \text{tr} \left[\bigotimes_{q=1}^r \mathcal{S}_{\{l:i_l=q\}} A_t \bigotimes_{q=1}^r \mathcal{S}_{\{l:i_l=q\}} \bigotimes_{l=1}^k |\psi_{i_l}\rangle\langle\psi_{i_l}| \right] \bigotimes_{l=1}^k |\psi_{j_l}\rangle\langle\psi_{j_l}| \\ &= \frac{1}{r^{2k}} \sum_{i_1, \dots, i_k, j_1, \dots, j_k} \text{tr} \left[A_t \bigotimes_{q=1}^r \mathcal{S}_{\{l:i_l=q\}} \right] \mathbb{E}_{\psi_1, \dots, \psi_r} \text{tr} \left[\tau_{i_1, \dots, i_k}^t \bigotimes_{l=1}^k |\psi_{i_l}\rangle\langle\psi_{i_l}| \right] \bigotimes_{l=1}^k |\psi_{j_l}\rangle\langle\psi_{j_l}|, \end{aligned} \quad (\text{D6})$$

where we define the state

$$\tau_{i_1, \dots, i_k}^t := \frac{\bigotimes_{q=1}^r \mathcal{S}_{\{l:i_l=q\}} A_t \bigotimes_{q=1}^r \mathcal{S}_{\{l:i_l=q\}}}{\text{tr}[A_t \bigotimes_{q=1}^r \mathcal{S}_{\{l:i_l=q\}}]}.$$

We can now interpret the expression $\mathbb{E}_{\psi_1, \dots, \psi_r} \text{tr}[\tau_{i_1, \dots, i_k}^t \bigotimes_{l=1}^k |\psi_{i_l}\rangle\langle\psi_{i_l}|] \bigotimes_{l=1}^k |\psi_{j_l}\rangle\langle\psi_{j_l}|$ as a tensor product of measure-and-prepare channels applied to the state τ . We apply Lemma 16 and arrive at the following bound:

$$\mathbb{E}_{\psi_1, \dots, \psi_r \sim \mu} \text{tr} \left[\tau_{i_1, \dots, i_k}^t \bigotimes_{l=1}^k |\psi_{i_l}\rangle\langle\psi_{i_l}| \right] \bigotimes_{l=1}^k |\psi_{j_l}\rangle\langle\psi_{j_l}| \geq \prod_{q=1}^r \left(e^{-|\{l:i_l=q\}||\{l:j_l=q\}|/d} \frac{1}{\binom{|\{l:i_l=q\}|+d-1}{d-1}} \right) \cdot \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:j_l=q\}}}{\binom{|\{l:j_l=q\}|+d-1}{d-1}}.$$

Proof of Theorem 11. We will show that a successful separable two-outcome POVM $\{M_{A,B}, \mathbb{I} - M_{A,B}\}$ with $k = o(\sqrt{d/\mathbb{E}\text{tr}(\rho^2)})$ will lead to a contradiction. More precisely, we assume that $\{M_{A,B}, \mathbb{I} - M_{A,B}\}$ satisfies

- (1) $\mathbb{E}_{\rho} \text{tr}(M_{A,B} \rho^{\otimes k} \otimes \rho^{\otimes k}) \leq \frac{1}{3}.$
- (2) $\mathbb{E}_{\rho, \sigma} \text{tr}(M_{A,B} \rho^{\otimes k} \otimes \sigma^{\otimes k}) \geq \frac{2}{3}.$

First, notice that

$$\begin{aligned} \mathbb{E}_{\rho} \text{tr}(\rho^2) &= \frac{1}{r^2} \sum_{i,j=1}^r \mathbb{E}_{\psi_1, \dots, \psi_r \sim \mu} |\langle\psi_i|\psi_j\rangle|^2 \\ &= \frac{1}{r^2} (r + r(r-1)d^{-1}), \end{aligned}$$

where we used the formula $\mathbb{E}_{\psi} |\langle\psi|\phi\rangle|^2 = d^{-1}$ for any fixed state ϕ . Therefore, we can assume that $\mathbb{E}_{\rho} \text{tr}(\rho^2) = \Theta(r^{-1})$.

We will consider the (unnormalized) “post measurement” state of Bob in case 1:

$$\rho_{\text{Bob}} = \mathbb{E}_{\rho} \text{tr}(A_t \rho^{\otimes k}) \rho^{\otimes k}.$$

We expand

$$\mathbb{E}_{\rho} \text{tr}(A_t \rho^{\otimes k}) = \sum_{i_1, \dots, i_k=1}^r \text{tr} \left[A_t \mathbb{E}_{\psi_1, \dots, \psi_r \sim \mu} \bigotimes_{l=1}^k |\psi_{i_l}\rangle\langle\psi_{i_l}| \right].$$

The operator $\mathbb{E}_{\psi_1, \dots, \psi_r} \bigotimes_{l=1}^k |\psi_{i_l}\rangle\langle\psi_{i_l}|$ is a tensor product over maximally mixed states on symmetric subspaces. In the following, for a subset $L \in [k]$, we denote by $\mathcal{S}_L$ the projector onto the symmetric subspace of tensor copies in L . If the subset is clear from context, we denote the projector onto the symmetric subspace of a tensor copies by $\mathcal{S}_a$.

$$\mathbb{E}_{\psi_1, \dots, \psi_r \sim \mu} \bigotimes_{l=1}^k |\psi_{i_l}\rangle\langle\psi_{i_l}| = \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:i_l=q\}}}{\binom{|\{l:i_l=q\}|+d-1}{d-1}}.$$

■

In Eq. (D6), we will typically see tuples $i_1, \dots, i_k$, such that every single value $1 \leq q \leq r$ can be expected to appear roughly k/r many times. This will allow us to lower bound the contribution of $\prod_{q=1}^r e^{-|\{l:i_l=q\}||\{l:j_l=q\}|/d}$. We can bound the tails of this distribution via a Hoeffding bound. More precisely, the probability of drawing the number q greater than $m \geq k/r$ times is bounded by

$$\Pr[|\{l:i_l=q\}| \geq m] \leq 2e^{-2k(1/r-m/k)^2}.$$

In particular, setting $m = Ck/r$ for a constant $C > 1$, we find via a union bound

$$\Pr\left[\max_q \{|\{l:i_l=q\}|, |\{l:j_l=q\}|\} \geq Ck/r\right] \leq \sum_{q=1}^r \Pr[|\{l:i_l=q\}| \geq m] \leq 4ke^{-2k(1-C)^2/r^2}. \quad (\text{D7})$$

Splitting the sum in Eq. (D6) into the typical case and the tail, we find

$$\begin{aligned} \rho_{\text{Bob}} &\geq \frac{1}{r^{2k}} \sum_{\substack{i_1, \dots, j_k \\ \max_q \{|\{l:i_l=q\}|, |\{l:j_l=q\}|\} \leq Ck/r}} \text{tr} \left[A_t \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:i_l=q\}}}{\binom{|\{l:i_l=q\}|+d-1}{d-1}} \right] e^{-C^2k^2/rd} \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:j_l=q\}}}{\binom{|\{l:j_l=q\}|+d-1}{d-1}} \\ &= \frac{1}{r^{2k}} \sum_{i_1, \dots, j_k=1}^r \text{tr} \left[A_t \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:i_l=q\}}}{\binom{|\{l:i_l=q\}|+d-1}{d-1}} \right] e^{-C^2k^2/rd} \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:j_l=q\}}}{\binom{|\{l:j_l=q\}|+d-1}{d-1}} - L, \end{aligned}$$

where we defined

$$L := \frac{1}{r^{2k}} \sum_{\substack{i_1, \dots, j_k \\ \max_q \{|\{l:i_l=q\}|, |\{l:j_l=q\}|\} > Ck/r}} \text{tr} \left[A_t \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:i_l=q\}}}{\binom{|\{l:i_l=q\}|+d-1}{d-1}} \right] e^{-C^2k^2/rd} \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:j_l=q\}}}{\binom{|\{l:j_l=q\}|+d-1}{d-1}}.$$

We can now apply this characterization to case 1 of the mixed DIPE (Problem 7):

$$\begin{aligned} \mathbb{E}_{\rho} \text{tr}[M_{A,B} \rho^{\otimes k} \otimes \rho^{\otimes k}] &= \sum_t \text{tr}[\text{tr}[A_t \rho^{\otimes k}] \rho^{\otimes k} B_t] \\ &\geq \sum_t \frac{e^{-C^2k^2/rd}}{r^{2k}} \sum_{i_1, \dots, j_k=1}^r \text{tr} \left[A_t \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:i_l=q\}}}{\binom{|\{l:i_l=q\}|+d-1}{d-1}} \right] \text{tr} \left[\bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:j_l=q\}}}{\binom{|\{l:j_l=q\}|+d-1}{d-1}} B_t \right] - \sum_t \text{tr}[LB_t] \\ &= e^{-C^2k^2/rd} \mathbb{E}_U \text{tr}[M_{AB} \rho^{\otimes k} \otimes \sigma^{\otimes k}] - \sum_t \text{tr}(LB_t) \geq e^{-C^2k^2/rd} \frac{2}{3} - \sum_t \text{tr}(LB_t). \end{aligned}$$

The second term can be bounded by

$$\begin{aligned} \sum_t \text{tr}[B_t L_t] &\leq \frac{1}{r^{2k}} \sum_{\substack{i_1, \dots, j_k \\ \max_q \{|\{l:i_l=q\}|, |\{l:j_l=q\}|\} > Ck/r}} \text{tr} \left[M_{AB} \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:i_l=q\}}}{\binom{|\{l:i_l=q\}|+d-1}{d-1}} \otimes \bigotimes_{q=1}^r \frac{\mathcal{S}_{\{l:j_l=q\}}}{\binom{|\{l:j_l=q\}|+d-1}{d-1}} \right] \\ &\leq \frac{1}{r^{2k}} \sum_{\substack{i_1, \dots, j_k \\ \max_q \{|\{l:i_l=q\}|, |\{l:j_l=q\}|\} > Ck/r}} 1 \\ &= \Pr[\max_q \{|\{l:i_l=q\}|, |\{l:j_l=q\}|\} > Ck/r] \\ &\leq 4ke^{-\Omega(kC^2/r^2)}, \end{aligned}$$

where we used Hölder's inequality and Eq. (D7). We choose C large enough so that $ke^{-\Omega(kC^2/r^2)} \leq \frac{1}{20}$ for all $k \geq c\sqrt{r}$ for a small constant $c > 0$. Assuming that $k < c\sqrt{r} \leq c\sqrt{d}$, we can show that the convex DIPE is strictly harder than the pure version. Indeed,

$$\prod_{q=1}^r e^{-|\{l:i_l=q\}||\{l:j_l=q\}|/d} = e^{-\sum_q |\{l:i_l=q\}||\{l:j_l=q\}|/d} \geq e^{-k^2/d}.$$

Else, we simply observe that $k \leq o(\sqrt{dr}) = o(\sqrt{d/\mathbb{E}[\rho^2]})$ contradicts the assumption $\mathbb{E}_U \text{tr}[M_{A,B} \rho^{\otimes k} \otimes \rho^{\otimes k}] \leq \frac{1}{3}$, which completes the proof. $\blacksquare$

- [1] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [2] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nat. Rev. Phys.* **2**, 382 (2020).
- [3] R. O'Donnell and J. Wright, Efficient quantum tomography, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (ACM, Cambridge, MA, 2016), pp. 899–912.
- [4] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, *IEEE Trans. Inf. Theory* **63**, 5628 (2017).
- [5] C. Buadescu, R. O'Donnell, and J. Wright, Quantum state certification, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (ACM, Phoenix, AZ, 2019), pp. 503–514.
- [6] A. Elben, B. Vermersch, R. van Bijnen, C. Kokail, T. Brydges, C. Maier, M. K. Joshi, R. Blatt, C. F. Roos, and P. Zoller, Cross-platform verification of intermediate scale quantum devices, *Phys. Rev. Lett.* **124**, 010504 (2020).
- [7] J. Carrasco, A. Elben, C. Kokail, B. Kraus, and P. Zoller, Theoretical and experimental perspectives of quantum verification, *PRX Quantum* **2**, 010102 (2021).
- [8] C. Greganti, T. F. Demarie, M. Ringbauer, J. A. Jones, V. Saggio, I. A. Calafell, L. A. Rozema, A. Erhard, M. Meth, L. Postler, *et al.*, Cross-verification of independent quantum devices, *Phys. Rev. X* **11**, 031049 (2021).
- [9] A. Anshu, Z. Landau, and Y. Liu, Distributed quantum inner product estimation, in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* (ACM, Rome, Italy, 2022), pp. 44–51.
- [10] M. Hinsche, M. Ioannou, S. Jerbi, L. Leone, J. Eisert, and J. Carrasco, Efficient distributed inner product estimation via Pauli sampling, *PRX Quantum* **6**, 030354 (2025).
- [11] D. Zhu, Z.-P. Cui, C. Noel, A. Risinger, D. Biswas, L. Egan, Y. Zhu, A. M. Green, C. H. Alderete, N. H. Nguyen, *et al.*, Cross-platform comparison of arbitrary quantum states, *Nat. Commun.* **13**, 6620 (2022).
- [12] Y. Qian, Y. Du, Z. He, M.-h. Hsieh, and D. Tao, Multimodal deep representation learning for quantum cross-platform verification, *Phys. Rev. Lett.* **133**, 130601 (2024).
- [13] P. Kurpiers, P. Magnard, T. Walter, B. Royer, M. Pechal, J. Heinsoo, Y. Salathé, A. Akin, S. Storz, J.-C. Besse, *et al.*, Deterministic quantum state transfer and remote entanglement using microwave photons, *Nature (London)* **558**, 264 (2018).
- [14] P. Magnard, S. Storz, P. Kurpiers, J. Schär, F. Marxer, J. Lütolf, T. Walter, J.-C. Besse, M. Gabureac, K. Reuer, *et al.*, Microwave quantum link between superconducting circuits housed in spatially separated cryogenic systems, *Phys. Rev. Lett.* **125**, 260502 (2020).
- [15] M. Alshowkan, B. P. Williams, P. G. Evans, N. S. Rao, E. M. Simmerman, H.-H. Lu, N. B. Lingaraju, A. M. Weiner, C. E. Marvinney, Y.-Y. Pai, *et al.*, Reconfigurable quantum local area network over deployed fiber, *PRX Quantum* **2**, 040304 (2021).
- [16] J. Knörzer, D. Malz, and J. I. Cirac, Cross-platform verification in quantum networks, *Phys. Rev. A* **107**, 062424 (2023).
- [17] A. Anshu and S. Arunachalam, A survey on the complexity of learning quantum states, *Nat. Rev. Phys.* **6**, 59 (2023).
- [18] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, *Nature (London)* **626**, 58 (2024).
- [19] K. Chen, Q. Wang, P. Long, and M. Ying, Unitarity estimation for quantum channels, *IEEE Trans. Inf. Theory* **69**, 5116 (2023).
- [20] S. Bubeck, S. Chen, and J. Li, Entanglement is necessary for optimal quantum property testing, in *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Durham, NC, 2020), pp. 692–703.
- [21] S. Chen, J. Li, B. Huang, and A. Liu, Tight bounds for quantum state certification with incoherent measurements, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Denver, CO, 2022), pp. 1205–1213.
- [22] S. Chen, J. Li, and R. O'Donnell, Toward instance-optimal state certification with incoherent measurements, in *Proceedings of Thirty-Fifth Conference on Learning Theory* (PMLR, London, UK, 2022), pp. 2541–2596.
- [23] D. Aharonov, J. Cotler, and X.-L. Qi, Quantum algorithmic measurement, *Nat. Commun.* **13**, 887 (2022).
- [24] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, Exponential separations between learning with and without quantum memory, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022), pp. 574–585.
- [25] H.-Y. Huang, R. Kueng, and J. Preskill, Information-theoretic bounds on quantum advantage in machine learning, *Phys. Rev. Lett.* **126**, 190505 (2021).
- [26] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, *et al.*, Quantum advantage in learning from experiments, *Science* **376**, 1182 (2022).
- [27] S. Chen, W. Gong, and Q. Ye, Optimal tradeoffs for estimating Pauli observables, in *Proceedings of 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Chicago, IL, 2024), pp. 1086–1105.
- [28] S. Chen, S. Zhou, A. Seif, and L. Jiang, Quantum advantages for Pauli channel estimation, *Phys. Rev. A* **105**, 032435 (2022).
- [29] S. Chen and W. Gong, Efficient Pauli channel estimation with logarithmic quantum memory, *PRX Quantum* **6**, 020323 (2025).
- [30] Q. Liu, R. Raz, and W. Zhan, Memory-sample lower bounds for learning with classical-quantum hybrid memory, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023* (Association for Computing Machinery, New York, 2023), pp. 1097–1110.
- [31] Y.-C. Liang, Y.-H. Yeh, P. E. Mendonça, R. Y. Teh, M. D. Reid, and P. D. Drummond, Quantum fidelity measures for mixed states, *Rep. Prog. Phys.* **82**, 076001 (2019).
- [32] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, The complexity of NISQ, *Nat. Commun.* **14**, 6001 (2023).
- [33] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Direct estimations of linear and nonlinear functionals of a quantum state, *Phys. Rev. Lett.* **88**, 217901 (2002).
- [34] We will use DIPE specific to decisional inner product estimation throughout this paper.
- [35] S. Arunachalam and L. Schatzki, Distributed inner product estimation with limited quantum communication, in *Proceedings of 42nd International Symposium on Theoretical Aspects of Computer Science (STACS 2025)* (Leibniz International Proceedings

- in Informatics (LIPIcs) series, Jena, Germany, 2025), No. 11, pp. 11:1–11:17.
- [36] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Quantum nonlocality without entanglement, *Phys. Rev. A* **59**, 1070 (1999).
 - [37] E. Chitambar and R. Duan, Nonlocal entanglement transformations achievable by separable operations, *Phys. Rev. Lett.* **103**, 110502 (2009).
 - [38] A. Harrow, The church of the symmetric subspace, [arXiv:1308.6595](https://arxiv.org/abs/1308.6595).
 - [39] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, *Science* **377**, eabk3333 (2022).
 - [40] B. Yu, Assouad, Fano, and Le Cam, in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* edited by D. Pollard, E. Torgersen, and G. L. Yang (Springer, New York, 1997), Vol. 423, pp. 423–435.
 - [41] C. W. Helstrom, Quantum detection and estimation theory, *J. Stat. Phys.* **1**, 231 (1969).
 - [42] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge university press, Cambridge, UK, 2017).
 - [43] G. Aubrun, Partial transposition of random states and non-centered semicircular distributions, *Random Matrices: Theory Appl.* **01**, 1250001 (2012).
 - [44] P. Hayden, D. W. Leung, and A. Winter, Aspects of generic entanglement, *Commun. Math. Phys.* **265**, 95 (2006).
 - [45] G. Aubrun, S. J. Szarek, and D. Ye, Entanglement thresholds for random induced states, *Commun. Pure Appl. Math.* **67**, 129 (2014).
 - [46] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. (Addison-Wesley, Boston, MA, 2017).
 - [47] G. Chiribella, On quantum estimation, quantum cloning and finite quantum de Finetti theorems, in *Conference on Quantum Computation, Communication, and Cryptography* (Springer, New York, 2010), pp. 9–25.