

Nearly Query-Optimal Classical Shadow Estimation of Unitary Channels

Zihao Li^{1,2,3,*} Changhao Yi^{1,2,3} You Zhou^{4,5,†} and Huangjun Zhu^{1,2,3,5,‡}

¹State Key Laboratory of Surface Physics, Department of Physics, and Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China

²Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China

³Shanghai Research Center for Quantum Sciences, Shanghai 201315, China

⁴Key Laboratory for Information Science of Electromagnetic Waves (Ministry of Education), Fudan University, Shanghai 200433, China

⁵Hefei National Laboratory, Hefei 230088, China

 (Received 9 December 2024; revised 5 May 2025; accepted 1 August 2025; published 30 September 2025)

Classical shadow estimation (CSE) is a powerful tool for learning the properties of quantum states and quantum processes. Here we consider the CSE task for quantum unitary channels. By querying an unknown unitary channel $\mathcal{U}$ multiple times in quantum experiments, the goal is to learn a classical description from which one can accurately predict many different linear properties of the channel, i.e., the expectation values of arbitrary observables measured on the output of $\mathcal{U}$ upon arbitrary input states. Based on collective measurements on multiple systems, we propose a query-efficient protocol for this task, whose query complexity has a quadratic advantage over the previous best approach for this problem, and almost saturates the information-theoretic lower bound. To further enhance practicality, we also present a variant protocol using only single-copy measurements, which still offers much better query performance than previous protocols that do not use quantum memory, and can serve as a key subroutine for learning an arbitrary unknown Hamiltonian from dynamics. In addition to linear properties of unitary channels, our protocol can also be applied to simultaneously predict many nonlinear properties, such as out-of-time-ordered correlators.

DOI: [10.1103/wccm-zys6](https://doi.org/10.1103/wccm-zys6)

I. INTRODUCTION

In quantum mechanics, the time dynamics of a closed d -dimensional system is governed by a unitary channel acting on quantum states. Learning complex quantum dynamics is a fundamental problem in physics and plays a crucial role in many applications in quantum information processing [1–5]. The traditional learning framework, known as process tomography, aims to give a full characterization of the unknown unitary channel $\mathcal{U}$ with high precision in the diamond distance. As shown by a recent work [6], at least (order) d^2 applications of the channel are necessary to achieve this task in any quantum experiment, even with access to quantum memory.

Actually, requiring the full classical description of the unknown unitary channel is excessive for many applications. Instead, it is usually sufficient to predict certain properties of $\mathcal{U}$. This is similar to the case of quantum states, where the framework of *shadow tomography* [7,8] and *classical shadow estimation* (CSE) [9–19] has led to results on predicting useful properties of a quantum system by using much fewer samples than full tomography. CSE has a notable feature that the measurement protocol is independent of the specific properties to be predicted. In other words, the measurements should produce some classical data from which the properties of interest can be calculated later [9,20]. This feature enables one to simultaneously predict multiple different properties of the system without losing high efficiency.

Recently, the CSE formalism has been extended from quantum states to general quantum channels [21,22]. Given query access to an unknown qudit quantum channel $\mathcal{C}$ that represents a physical process happening in a laboratory, these works considered the task of predicting a collection of M linear properties of the form

$$\text{Tr}(O_l \mathcal{C}(\rho_l)), \quad l = 1, 2, \dots, M, \quad (1)$$

*Contact author: zihao120@fudan.edu.cn

†Contact author: you_zhou@fudan.edu.cn

‡Contact author: zhuhuangjun@fudan.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

for arbitrary quantum states ρ_l and observables O_l . Inheriting the mindset of CSE, the learning protocols of Refs. [21, 22] do not depend on the specific input states and output observables. Provided that the (squared) Frobenius norms of the observables are bounded by $B \geq \max_l \|O_l\|_2^2$, they require $\mathcal{O}(d^2 B \log M)$ applications of the channel $\mathcal{C}$ to achieve this learning task (see Sec. VI for details). Although many works are beginning to emerge to improve the query efficiency (i.e., reduce the number of applications) [23–29], they impose stringent restrictions on the channel or on the input states and output observables, for example, assuming that the channel can be generated in polynomial time (see Sec. VI).

In this work, we re-examine the above significant learning problem under an alternative but natural and common assumption—the channel to be learned is unitary. We denote the learning task in this scenario as *CSE of unitary channels* (CSEU), which is of fundamental and practical interest in quantum physics and quantum information processing. For instance, the quantum gates and circuits used in quantum computing are in general unitary. As shown in Fig. 1, given access to an unknown unitary channel $\mathcal{U}$, CSEU aims to accurately predict properties of the form specified in Eq. (1), with channel $\mathcal{C}$ replaced by $\mathcal{U}$. Such properties play crucial roles in various research areas, including quantum machine learning [5, 29, 31, 32] and variational quantum algorithms [33–35].

Based on collective measurements on multiple systems, we propose a protocol for CSEU that uses $\mathcal{O}(d\sqrt{B} \log M)$ applications of $\mathcal{U}$. Compared with existing methods like *quantum process tomography of unitary channels* (QPTU) [6] or *CSE of general channels* (CSEC) [21, 22], our protocol achieves a square-root reduction in the scaling behavior with the dimension d (see Table I). In addition, by embedding the problem of learning an unknown Boolean function into the task of CSEU, we establish a complementary lower bound, indicating that our protocol has the asymptotically optimal query complexity with respect to the dimension (up to a $\log d$ factor). The lower bound is built in a worst-case sense, namely, the learner is required to

make an accurate prediction for any input state and output observable. We further show that it can be relaxed when considering the average-case prediction over random input states, thereby highlighting the separation in query complexity between the worst- and average-case considerations for the CSEU task.

To address the limitation that collective measurements may be challenging to implement on current devices, we further propose a variant protocol for CSEU, which requires neither quantum memory nor ancillary systems. This variant is appealing for practical applications because it requires only the preparation of a computational-basis state, the application of a random Clifford unitary to a single system, and a measurement in the computational basis. Compared with existing quantum memory-free learning protocols with bounded ancilla, our protocol can significantly enhance the efficiency (see Table I). As a key application, we combine this protocol with the polynomial interpolation technique [24, 36, 37] to learn an arbitrary Hamiltonian H without any prior information on the structure of H . Remarkably, even without using quantum memory, our protocol can achieve much better performance than previous approaches to this problem, in terms of both the number of queries to H and total evolution time under H .

The rest of this paper is organized as follows. In Sec. II, we introduce the basic notation and necessary preliminaries for this work. In Sec. III, we present our protocol for CSEU and analyze its query complexity. In Sec. IV, we establish an information-theoretic lower bound for the query number required to complete the CSEU task. In Sec. V, we study the performance of our protocol under the average-case scenario over random input states. In Sec. VI, we compare our CSEU protocol with a number of previous protocols. In Sec. VII, we apply our CSEU protocol to learn an arbitrary Hamiltonian. In Sec. VIII, we generalize our protocol to predict nonlinear properties of unitary channels and discuss an application in estimating out-of-time-ordered correlators (OTOCs). In Sec. IX we summarize our results and discuss potential future

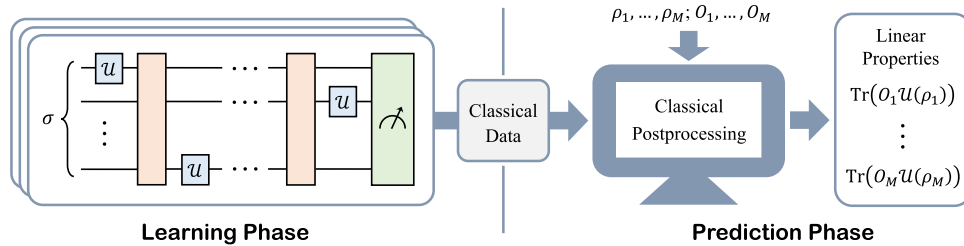


FIG. 1. Illustration of the CSEU task. The task contains two separate phases. In the learning phase, we apply the unknown unitary channel $\mathcal{U}$ multiple times in some quantum experiments and obtain some classical data. In the prediction phase, we are given a collection of quantum states $\rho_1, \dots, \rho_M$ and observables $O_1, \dots, O_M$. By performing classical postprocessing on the data collected from the learning phase, the goal is to accurately predict linear properties $\text{Tr}(O_l \mathcal{U}(\rho_l))$ for all $l = 1, 2, \dots, M$.

TABLE I. Comparison of various protocols for the CSEU task (see Sec. VI for details). Here, d is the system dimension; B is an upper bound on the squared Frobenius norm of the target observables; M is the number of linear properties to be predicted; and ϵ is the target additive error. The parameter s in our protocol represents the number of systems collectively measured in the learning phase. Each protocol is classified based on whether the protocol requires quantum memory, and whether its query complexity achieves the optimal scaling behavior $\tilde{\mathcal{O}}(d)$ with respect to d . The performance of protocols in Refs. [23–29] are not shown here because they have additional restrictions or assumptions on the input states, output observables, or the unitary channel to be learned.

Protocol	Quantum memory?	Query complexity	Is scaling optimal in d ?
Ancilla-assisted CSEC [21,22] (see Sec. VIA)	No	$\mathcal{O}(d^2 B \epsilon^{-2} \log M)$	No
Ancilla-free CSEC [21,22] (see Sec. VIA)	No	$\mathcal{O}(d^3 B \epsilon^{-2} \log M)$	No
QPTU protocol in Ref. [30] (see Sec. VIC)	No	$\mathcal{O}(d^4 \epsilon^{-2})$	No
Our protocol with $s = 1$ (this work)	No	$\mathcal{O}((d \epsilon^{-2} + d^2 \sqrt{B} \epsilon^{-1}) \log M)$	No
Protocol based on Ref. [11] (see Sec. VIB)	Yes	$\mathcal{O}(d^2 \epsilon^{-2} \log M)$	No
QPTU protocol in Ref. [6] (see Sec. VIC)	Yes	$\mathcal{O}(d^2 \epsilon^{-1})$	No
Our protocol with $s = \Theta(d)$ (this work)	Yes	$\mathcal{O}(d(\epsilon^{-2} + \sqrt{B} \epsilon^{-1}) \log M)$	Yes

directions. To streamline the presentation, most technical proofs are relegated to the Appendices.

II. PRELIMINARIES

In this section, we define the key notation that will be frequently used throughout the paper, and review different models of learning quantum channels.

A. Notation and definitions

Consider an n -qubit Hilbert space $\mathcal{H}$ with dimension $d = 2^n$. We denote by $\mathcal{D}(\mathcal{H})$ the set of all density operators on $\mathcal{H}$, $\mathcal{L}(\mathcal{H})$ the set of linear operators on $\mathcal{H}$, and $\mathcal{L}_H(\mathcal{H})$ the set of Hermitian operators on $\mathcal{H}$. Let I be the identity operator on $\mathcal{H}$, and $|\Phi\rangle := \sum_{i=1}^d |ii\rangle \in \mathcal{H}^{\otimes 2}$ be an unnormalized maximally entangled state, where $\{|i\rangle\}_i$ represents the computational basis on $\mathcal{H}$. We denote by $U \in \mathcal{L}(\mathcal{H})$ an unknown unitary operator that we aim to learn, and by $\mathcal{U}(\cdot) = U(\cdot)U^\dagger$ the corresponding unitary channel on $\mathcal{L}(\mathcal{H})$. For any quantum channel $\mathcal{C} : \mathcal{L}(\mathcal{H}) \rightarrow \mathcal{L}(\mathcal{H})$, let

$$\Upsilon_{\mathcal{C}} := (\mathcal{C} \otimes \mathcal{I})(|\Phi\rangle\langle\Phi|) \quad (2)$$

be the corresponding Choi operator, where $\mathcal{I}$ is the identity channel. For any $A \in \mathcal{L}(\mathcal{H})$, we denote by $\|A\|$ its operator norm and $\|A\|_2$ its Frobenius norm. We write $\psi := |\psi\rangle\langle\psi|$ for any pure state $|\psi\rangle \in \mathcal{H}$. For $B \in [1, d]$, we define the set of observables

$$\text{Obs}(B) := \{O \in \mathcal{L}_H(\mathcal{H}) \mid \text{Tr}(O^2) \leq B, \|O\| \leq 1\}. \quad (3)$$

For positive integer t , we denote by $\Pi_{\text{sym}}^{(t)}$ the projector onto the symmetric subspace of $\mathcal{H}^{\otimes t}$, and $\kappa_t = \binom{t+d-1}{t}$ the dimension of the symmetric subspace. An ensemble $\mathcal{E}$ of pure states in $\mathcal{H}$ is said to form a state t -design if

and only if its t th moment agrees with that of the Haar distribution, i.e.,

$$\mathbb{E}_{|\psi\rangle \sim \mathcal{E}} [(|\psi\rangle\langle\psi|)^{\otimes t}] = \mathbb{E}_{|\psi\rangle \sim \mu_H} [(|\psi\rangle\langle\psi|)^{\otimes t}] \equiv \frac{\Pi_{\text{sym}}^{(t)}}{\kappa_t}, \quad (4)$$

where the second expectation is over the Haar measure μ_H on quantum states. For example, for any $|\phi\rangle \in \mathcal{H}$, the ensemble $\{V|\phi\rangle\}_{V \in \text{Cl}(n)}$ is a state 3-design if the unitary operator V is sampled uniformly from the n -qubit Clifford group $\text{Cl}(n)$ [38,39]. A more detailed introduction to state designs is given in Appendix B 1.

We will frequently use the big- $\mathcal{O}$, big- Ω , and big- Θ notation for function asymptotics. In particular, for two functions $f, g : [0, \infty) \rightarrow [0, \infty)$, $f = \mathcal{O}(g)$ if there exists a $c > 0$ such that $f(x) \leq cg(x)$ as $x \rightarrow \infty$, $f = \Omega(g)$ if $g = \mathcal{O}(f)$, and $f = \Theta(g)$ if both $f = \mathcal{O}(g)$ and $f = \Omega(g)$. We will also use $\tilde{\mathcal{O}}$ and $\tilde{\Omega}$ to hide factors that are polylogarithmic in the leading-order term.

B. Learning quantum channels or Hamiltonians with and without quantum memory

Following the classifications in Refs. [24,29,40], here we describe several different models of learning an unknown quantum channel $\mathcal{C}$ or Hamiltonian H , depending on the type of access to the quantum system and the amount of additional resources utilized.

The first type of learning protocols work *without quantum memory* [41]. That is, the learner can use the channel $\mathcal{C}$ (or the time evolution under the Hamiltonian H) only once in each step. To be more explicit, when learning $\mathcal{C}$, in the k th round of experiments, the learner can first prepare an arbitrary quantum state σ_k , then evolve it under $\mathcal{C}$ to yield the output state $(\mathcal{I} \otimes \mathcal{C})(\sigma_k)$, and finally perform an arbitrary quantum measurement on the output

state, which is described by a *positive operator-valued measure* (POVM); here, $\mathcal{I}$ is the identity channel acting on the ancillary system. When learning the Hamiltonian H , the channel $\mathcal{C}$ used in the above procedure is replaced by $\mathcal{U}_{t_k}(\cdot) = e^{-iHt_k}(\cdot)e^{iHt_k}$ for describing the time evolution under H , with $t_k > 0$ being an arbitrary chosen time. We call a protocol *ancilla-assisted* if it uses an ancillary system, and *ancilla-free* otherwise. If an ancillary system is allowed but its dimension is bounded by a constant, then the protocol is *ancilla-bounded*.

The second type of protocols work *with quantum memory*, that is, they have the capability to store quantum data during experiments. When learning an unknown channel $\mathcal{C}$, in each round of experiments, the learner is allowed to access $\mathcal{C}$ multiple times before measuring, possibly interleaving them with the application of other quantum channels. The resulting output state before the measurement has the form

$$(\mathcal{C} \otimes \mathcal{I}) \circ \mathcal{C}_{L-1} \circ \cdots \circ \mathcal{C}_2 \circ (\mathcal{C} \otimes \mathcal{I}) \circ \mathcal{C}_1 \circ (\mathcal{C} \otimes \mathcal{I})(\sigma), \quad (5)$$

where σ is an arbitrary initial state, and $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_{L-1}$ are arbitrary quantum channels. Finally, the learner can perform an arbitrary POVM on the output state. When learning an Hamiltonian H , the channels $\mathcal{C}$ in the above procedure are replaced with time evolutions under H , and the output state in Eq. (5) is replaced by

$$(\mathcal{U}_{t_L} \otimes \mathcal{I}) \circ \mathcal{C}_{L-1} \circ \cdots \circ (\mathcal{U}_{t_2} \otimes \mathcal{I}) \circ \mathcal{C}_1 \circ (\mathcal{U}_{t_1} \otimes \mathcal{I})(\sigma), \quad (6)$$

where $t_1, t_2, \dots, t_L > 0$ are arbitrary chosen times.

III. PROTOCOL FOR CSEU

In this section, we present our protocol for CSEU and analyze its query complexity. The protocol is separated into the learning phase, in which the unknown unitary channel $\mathcal{U}$ is applied several times in quantum experiments, and the prediction phase, in which we predict desired linear properties of $\mathcal{U}$ by using the classical data obtained from the learning phase.

A. Learning phase

In the learning phase, we use a measurement applied on quantum states in $\mathcal{D}(\mathcal{H}^{\otimes s})$. This measurement, denoted by $\mathcal{M}_s$, is called the *symmetric collective measurement*. Suppose $\{|\phi_j\rangle\} \in \mathcal{H}_{j=1}^K$ forms a state $(s+2)$ -design. Then $\mathcal{M}_s$ can be constructed as the following POVM:

$$\mathcal{M}_s = \left\{ A_{\phi_1}, A_{\phi_2}, \dots, A_{\phi_K}, \mathbb{I}_s - \Pi_{\text{sym}}^{(s)} \right\}, \quad (7)$$

where $A_{\phi_j} := (\kappa_s/K) \phi_j^{\otimes s}$, and $\mathbb{I}_s$ is the identity operator on $\mathcal{H}^{\otimes s}$. Thanks to Eq. (4), the sum of the operators A_{ϕ_j}

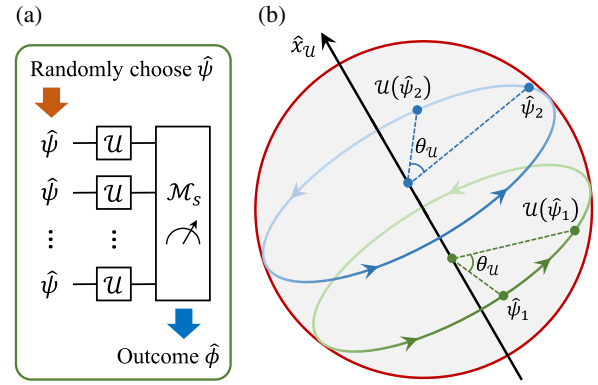


FIG. 2. Learning phase of our protocol for CSEU: procedure and intuition. (a) Schematic view of our learning algorithm $\mathcal{A}_{\text{learn}}$. In each round, we first prepare s copies of a pure state $\hat{\psi}$ chosen from a 4-design ensemble, and rotate them with the unitary channel $\mathcal{U}$, then measure the resulting states using the symmetric collective measurement $\mathcal{M}_s$. (b) Geometric intuition of our protocol illustrated on the single-qubit Bloch sphere.

equals the projector $\Pi_{\text{sym}}^{(s)}$ onto the symmetric subspace. This measurement has been studied before and is known to be optimal for many learning tasks [11, 42–47]. In general, the implementation of $\mathcal{M}_s$ on large systems with current technology remains challenging. Nevertheless, for systems containing only a few qubits, such measurements with a small s have been successfully demonstrated in superconducting, trapped-ion, and photonic quantum systems [48–51].

We denote by $\mathcal{A}_{\text{learn}}$ our quantum algorithm working in the learning phase. To extract meaningful information about $\mathcal{U}$, as shown in Fig. 2(a), $\mathcal{A}_{\text{learn}}$ runs the following learning procedure m times independently:

- (1) Randomly choose a pure state $|\hat{\psi}\rangle \in \mathcal{H}$ from a 4-design state ensemble $\{|\psi_i\rangle\}_{i=1}^L$; then prepare s copies of the state $\hat{\psi}$.
- (2) Apply the unitary channel $\mathcal{U}$ on each copy of $\hat{\psi}$.
- (3) Measure the rotated state $\mathcal{U}(\hat{\psi})^{\otimes s}$ with the symmetric collective measurement $\mathcal{M}_s$. If the measurement outcome $A_{\hat{\phi}}$ is obtained ($\hat{\phi} \in \{\phi_1, \dots, \phi_K\}$), we then record the classical description of the corresponding pure state $\hat{\phi}$.

It is worth noting that we will never see the measurement outcome $(\mathbb{I}_s - \Pi_{\text{sym}}^{(s)})$ in Step 3 of the above procedure, because $\{A_{\phi_1}, \dots, A_{\phi_K}\}$ forms a POVM on the symmetric subspace of $\mathcal{H}^{\otimes s}$, and the state $\mathcal{U}(\hat{\psi})^{\otimes s}$ to be measured is supported in this subspace.

Before providing a detailed explanation of $\mathcal{A}_{\text{learn}}$, we offer some additional remarks. First, if we choose parameter $s \geq 2$ in the learning procedure, then our protocol relies on collective measurements and requires quantum memory. In contrast, when choosing $s = 1$, our protocol relies

solely on single-copy measurements and does not need quantum memory. In this special case, we use at most the 3-design property of the measurement $\mathcal{M}_{s=1}$. Since any orbit of the multiqubit Clifford group forms a state 3-design [38,39], the *random global Clifford measurement* (RGCM) used in the original CSE protocol [9] can be applied in place of $\mathcal{M}_{s=1}$ in our learning procedure, and the analysis will be identical. RGCM only requires implementation of Clifford unitaries and measurements in the computational basis, and is thus appealing to practical applications. Second, recall that the initial state $\hat{\psi}$ is chosen from a 4-design state ensemble in the learning procedure. In fact, this 4-design property is utilized only in the case of $s \geq 2$ throughout the paper. When $s = 1$, the 3-design property is sufficient for proving our results. Therefore, in the $s = 1$ case, we can prepare the initial state $\hat{\psi}$ by simply applying a random Clifford unitary to the computational-basis state $|0^n\rangle\langle 0^n|$.

After each round of the learning experiments, the description of the prepared state $\hat{\psi}$ and the outcome $\hat{\phi}$ are stored in a classical memory. It is instructive to look at the expectation of their tensor product, which is clarified in the following lemma; see Appendix C for a proof.

Lemma 1. The expectation of $\hat{\phi} \otimes \hat{\psi}$ (over both random state preparation and measurement outcomes) is

$$\mathbb{E}[\hat{\phi} \otimes \hat{\psi}] = \frac{(d+1+s)(I \otimes I) + s(U \otimes U^\dagger)T_{(1,2)}}{d(d+1)(d+s)}, \quad (8)$$

where $T_{(1,2)}$ is the swap operator.

Therefore, after obtaining the description of $\hat{\psi}$ and $\hat{\phi}$ in a single run, we can apply a transformation to $\hat{\psi} \otimes \hat{\phi}$ to generate a classical snapshot $\hat{X}$, whose expectation contains all information of the unknown unitary channel $\mathcal{U}$. Concretely, let

$$\hat{X} := \frac{d(d+1)(d+s)(\hat{\phi} \otimes \hat{\psi}^\top) - (d+1+s)(I \otimes I)}{s}, \quad (9)$$

where $(\cdot)^\top$ denotes the transpose operation (with respect to the computational basis). According to Lemma 1, $\hat{X}$ is an unbiased estimator for the Choi operator $\Upsilon_{\mathcal{U}}$ of $\mathcal{U}$:

$$\mathbb{E}[\hat{X}] = [(U \otimes U^\dagger)T_{(1,2)}]^\top = \Upsilon_{\mathcal{U}}, \quad (10)$$

where $\mathbb{E}$ denotes the average over the randomness of state preparation and quantum measurement in the learning phase, $(\cdot)^\top$ is the transpose operation on the second subsystem, and the second equality can easily be derived via the tensor-network diagrams introduced in Appendix B 2.

From the quantum experiments we obtain a collection of m independent classical snapshots of $\mathcal{U}$:

$$\text{UShadow}(\mathcal{U}, m) = \{\hat{X}_1, \dots, \hat{X}_m\}. \quad (11)$$

This collection, as the output of the learning phase, is called the *classical shadow data*. As we shall see in the following sections, many properties of $\mathcal{U}$ can be extracted efficiently from $\text{UShadow}(\mathcal{U}, m)$.

The reason why $\mathcal{A}_{\text{learn}}$ learns all information about $\mathcal{U}$ efficiently can be intuitively explained by a toy model illustrated in Fig. 2(b), where an unknown single-qubit unitary channel $\mathcal{U}$ can rotate the Bloch sphere around an axis $\hat{x}_{\mathcal{U}}$ by an angle $\theta_{\mathcal{U}}$. To recover the information about $\mathcal{U}$, one needs to determine both the value of $\theta_{\mathcal{U}}$ and the direction of $\hat{x}_{\mathcal{U}}$. Mathematically, this goal can be achieved by randomly choosing a few points on the sphere (two points are enough in the usual case), and observe the positions to which they are rotated. Physically, these operations correspond to first randomly selecting a few pure states $\hat{\psi}_1, \dots, \hat{\psi}_m$ on $\mathcal{H}$ (Step 1 of $\mathcal{A}_{\text{learn}}$), and then learning information about the rotated states $\mathcal{U}(\hat{\psi}_i)$. We expect that the latter goal can be efficiently achieved by performing $\mathcal{M}_s$ on multiple copies of $\mathcal{U}(\hat{\psi}_i)$ (Step 3 of $\mathcal{A}_{\text{learn}}$), because this measurement is optimal for pure states for many estimation problems [11,43–47].

B. Prediction phase

In the prediction phase, we are given classical descriptions of M quantum states $\rho_1, \dots, \rho_M \in \mathcal{D}(\mathcal{H})$ and M observables $O_1, \dots, O_M \in \text{Obs}(B)$, where $\text{Obs}(B)$ is

ALGORITHM 1. Algorithm for $\mathcal{A}_{\text{pred}}$ in the prediction phase.

Input: $\text{UShadow}(\mathcal{U}, m)$, $\rho_1, \dots, \rho_M$, $O_1, \dots, O_M$

Output: $\hat{E}(O_1, \rho_1), \dots, \hat{E}(O_M, \rho_M)$

1: Split the shadow data $\text{UShadow}(\mathcal{U}, m)$ into R batches of equal size, where each batch contains $q = m/R$ snapshots (we may assume that m is a multiple of R for simplicity), and set

$$\hat{Y}_{(r)} := \frac{1}{q(q-1)d} \sum_{i,j=(r-1)q+1|i \neq j}^{rq} \hat{X}_i \hat{X}_j, \quad r = 1, \dots, R. \quad (13)$$

2: **for** $l = 1, 2, \dots, M$, **do**

3: Construct R independent sample mean estimators:

$$\hat{Z}_{(r)}(O_l, \rho_l) := \text{Tr}[(O_l \otimes \rho_l^\top) \hat{Y}_{(r)}], \quad r = 1, \dots, R. \quad (14)$$

4: Compute the median of means:

$$\hat{E}(O_l, \rho_l) := \text{median}\{\hat{Z}_{(1)}(O_l, \rho_l), \dots, \hat{Z}_{(R)}(O_l, \rho_l)\}. \quad (15)$$

5: **end for**

6: **return** $\hat{E}(O_l, \rho_l)$ as the estimate for $\text{Tr}(O_l \mathcal{U}(\rho_l))$.

defined in Eq. (3). The goal is to accurately predict linear properties $\text{Tr}(O_l \mathcal{U}(\rho_l))$ for all $l = 1, 2, \dots, M$ using the classical shadow data $\text{UShadow}(\mathcal{U}, m) = \{\hat{X}_1, \dots, \hat{X}_m\}$.

Recall that each snapshot $\hat{X}_i$ exactly reproduces the Choi operator $\Upsilon_{\mathcal{U}}$ of the unitary channel $\mathcal{U}$ in expectation. As a consequence, we have

$$\mathbb{E}[\text{Tr}((O \otimes \rho^\top) \hat{X})] = \text{Tr}[(O \otimes \rho^\top) \Upsilon_{\mathcal{U}}] = \text{Tr}(O \mathcal{U}(\rho)) \quad (12)$$

for any state ρ and observable O , where the second equality holds by channel-state duality [52]. Hence, a natural way to predict $\text{Tr}(O \mathcal{U}(\rho))$ is to calculate $\text{Tr}[(O \otimes \rho^\top) \hat{X}_i]$ for many independent snapshots $\hat{X}_i$ and then take their average. However, as shown in Appendix E, this method requires $\mathcal{O}(d^2 B / \epsilon^2)$ applications of $\mathcal{U}$ to predict a single $\text{Tr}(O \mathcal{U}(\rho))$ within additive error ϵ , even when collective measurements on multiple systems are employed. If only single-copy measurements are accessible, then the query complexity further increases, reaching up to $\mathcal{O}(d^3 B / \epsilon^2)$. Such high query costs are unsatisfactory.

To improve the efficiency, we need to construct an estimator for the unitary channel in a different way. Inspired by the CSE protocol for pure states in Ref. [11], we note that the product of two independent snapshots is still proportional to the Choi operator in expectation,

$$\mathbb{E}[\hat{X}_i \hat{X}_j] = \mathbb{E}[\hat{X}_i] \mathbb{E}[\hat{X}_j] = \Upsilon_{\mathcal{U}}^2 = d \Upsilon_{\mathcal{U}}, \quad \forall i \neq j. \quad (16)$$

In light of this, one can predict $\text{Tr}(O \mathcal{U}(\rho))$ by calculating the quadratic estimators

$$\tilde{\Lambda}_{i,j} := d^{-1} \text{Tr}[(O \otimes \rho^\top) \hat{X}_i \hat{X}_j] \quad (17)$$

for all index pairs $\{(i,j) | i \neq j\}$ and taking their average. Since the number of different $\tilde{\Lambda}_{i,j}$ is much larger than that of $\text{Tr}[(O \otimes \rho^\top) \hat{X}_i]$, this method is expected to enhance the estimation precision and thereby reduce the query cost compared with the former method. Indeed, as we shall see in Sec. III C, it can achieve nearly a square-root reduction in query complexity.

To be concrete, the procedure of our classical algorithm $\mathcal{A}_{\text{pred}}$ working in the prediction phase is summarized in Algorithm 1. Thanks to Eq. (16), the operator $\hat{Y}_{(r)}$ defined in Eq. (13) equals the Choi operator $\Upsilon_{\mathcal{U}}$ in expectation. This fact implies that $\hat{Z}_{(r)}(O_l, \rho_l)$ constructed in Eq. (14) is an unbiased estimator for the desired linear property $\text{Tr}(O_l \mathcal{U}(\rho_l))$. In Step 4 of Algorithm 1, we use the *median-of-means* method [9] to suppress the failure probability of accurate prediction.

C. Rigorous performance guarantees

Now we investigate the query complexity of our protocol for CSEU. That is, in order to achieve

$$\Pr\left\{|\hat{E}(O_l, \rho_l) - \text{Tr}(O_l \mathcal{U}(\rho_l))| < \epsilon \quad \forall 1 \leq l \leq M\right\} \geq 1 - \delta \quad (18)$$

for some $0 < \delta, \epsilon < 1$, how many applications of $\mathcal{U}$ are required by our protocol? To answer this question, we first analyze the statistical fluctuation of each sample mean estimator $\hat{Z}_{(r)}(O_l, \rho_l)$. Using the definition of $\hat{Y}_{(r)}$ in Eq. (13), for any quantum state ρ and observable O on $\mathcal{H}$, we can expand the variance as

$$\text{Var}[\hat{Z}_{(r)}(O, \rho)] = \frac{1}{q^2(q-1)^2 d^2} \sum_{i \neq j} \sum_{k \neq \ell} \text{Cov}(\Lambda_{i,j}, \Lambda_{k,\ell}), \quad (19)$$

where q is the number of snapshots in each batch (see Step 1 of Algorithm 1), $\text{Cov}(\cdot, \cdot)$ is the covariance between two variables, and

$$\Lambda_{i,j} := d \tilde{\Lambda}_{i,j} = \text{Tr}[(O \otimes \rho^\top) \hat{X}_i \hat{X}_j]. \quad (20)$$

The sum in Eq. (19) consists of five different cases: when all indices i, j, k, ℓ are distinct, the covariance is 0 by independence; the other four cases are summarized in the following lemma, which is proved in Appendix F. For simplicity, we denote by $\wp = \text{Tr}(\rho^2)$ the purity of ρ .

Lemma 2. The covariance term $\text{Cov}(\Lambda_{i,j}, \Lambda_{k,\ell})$ has the following upper bound for each combination (i, j, k, ℓ) with $i \neq j$ and $k \neq \ell$:

(1) One index matches ($|\{i, j\} \cap \{k, \ell\}| = 1$)

(a) Match in different positions ($i = \ell$ or $j = k$):

$$\mathcal{O}(d^2 \min\{1, B \wp\}). \quad (21)$$

(b) Match in the same position ($i = k$ or $j = \ell$):

$$\mathcal{O}\left(\frac{d^3 \wp}{s} + d^2 \min\{1, B \wp\}\right), \quad (22)$$

where s is the number of systems collectively measured in the learning phase.

(2) Both indices match ($|\{i, j\} \cap \{k, \ell\}| = 2$)

(a) Order swapped ($i = \ell$ and $j = k$):

$$\mathcal{O}\left(\left(\frac{d^3}{s^4} + \frac{d^2}{s^3} + \frac{d\sqrt{d\wp}}{s^2} + \frac{\sqrt{d\wp}}{s} + \wp\right) d^2 B\right). \quad (23)$$

(b) Same order ($i = k$ and $j = \ell$):

$$\mathcal{O}\left(\left(\frac{d^4}{s^4} + \frac{d^3}{s^3} + \frac{d^2}{s^2} + \frac{d}{s} + 1\right)d^2 B\wp\right). \quad (24)$$

For all combinations of i, j, k, ℓ in the r th batch with $i \neq j$ and $k \neq \ell$, there are $\mathcal{O}(q^4)$ combinations where all four indices are distinct; $\mathcal{O}(q^3)$ combinations where exactly one index matches; and $\mathcal{O}(q^2)$ combinations where both indices match. These facts, together with Eq. (19) and Lemma 2, immediately imply the following proposition, which clarifies the variance of $\hat{Z}_{(r)}(O, \rho)$.

Proposition 1. Given $O \in \text{Obs}(B)$ and $\rho \in \mathcal{D}(\mathcal{H})$, the variance of $\hat{Z}_{(r)}(O, \rho)$ is upper bounded by

$$\mathcal{O}\left[\frac{1}{q}\left(\frac{d\wp}{s} + \min\{1, B\wp\}\right) + \frac{1}{q^2}\left(\frac{d^4}{s^4} + 1\right)B\wp\right]. \quad (25)$$

Having obtained the variance, we can apply Chebyshev's inequality to bound the failure probability of each sample mean estimator $\hat{Z}_{(r)}(O, \rho)$:

$$\Pr\left\{\left|\hat{Z}_{(r)}(O, \rho) - \text{Tr}(OU(\rho))\right| \geq \epsilon\right\} \leq \frac{\text{Var}[\hat{Z}_{(r)}(O, \rho)]}{\epsilon^2}. \quad (26)$$

The query cost of our protocol can then be derived from this bound.

Theorem 1. Suppose $s \geq 1$ is the number of systems collectively measured in our learning phase, $\rho_1, \dots, \rho_M \in \mathcal{D}(\mathcal{H})$, and $O_1, \dots, O_M \in \text{Obs}(B)$. To estimate all M properties $\text{Tr}(O_i \mathcal{U}(\rho_i))$ within additive error ϵ and failure probability at most δ , the number of queries required by our protocol for CSEU is upper bounded by

$$\mathcal{O}\left(\left(\frac{\max\{d, s\}}{\epsilon^2} + \frac{\max\{d^2, s^2\}\sqrt{B}}{s\epsilon}\right) \log\left(\frac{M}{\delta}\right)\right). \quad (27)$$

Theorem 1 is a simplified version of Theorem G1 in Appendix G. Note that the query complexity is closely related to the value of s ; see Fig. 3 for an illustration. In particular, when $s = 1$, i.e., single-copy measurements are used, the query number reduces to

$$\mathcal{O}\left(\left(\frac{d}{\epsilon^2} + \frac{d^2\sqrt{B}}{\epsilon}\right) \log\left(\frac{M}{\delta}\right)\right). \quad (28)$$

As s increases from 1, the query number decreases and reaches the following minimum value when $s = \Theta(d)$,

$$\mathcal{O}\left(d\left(\frac{1}{\epsilon^2} + \frac{\sqrt{B}}{\epsilon}\right) \log\left(\frac{M}{\delta}\right)\right). \quad (29)$$

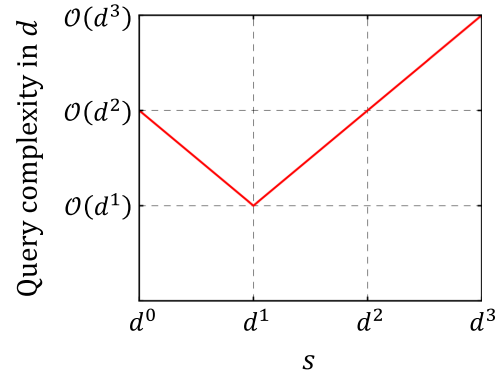


FIG. 3. The query complexity of our protocol [see Eq. (27)] with respect to the dimension d . Here s denotes the number of systems collectively measured in our learning phase. The optimal complexity $\mathcal{O}(d)$ is attained when $s = \Theta(d)$.

When s exceeds $\Theta(d)$, however, the query number starts to increase. Hence, if we are allowed to perform collective measurements on multiple systems, the best choice of s is $s = \Theta(d)$. Equations (28) and (29) show that our protocol significantly outperforms previous approaches for the CSEU task (see Sec. VI for details). This exceptional query performance constitutes a key advantage of our protocol.

It is worth noting that while the symmetric collective measurement $\mathcal{M}_s$ with $s = \Theta(d)$ can achieve maximal reduction in query complexity, it demands both a state $\Theta(d)$ -design and joint operations on $\Theta(d)$ systems. These requirements make its experimental implementation extremely challenging with current technology. By contrast, choosing $s = 1$ only requires performing Clifford unitaries and measurements in the computational basis, which are significantly easier to realize. In practice, one may adjust the parameter s to strike a balance between query cost and measurement feasibility.

IV. LOWER BOUND FOR THE QUERY COMPLEXITY

Although our protocol has a much higher query efficiency than previous approaches for the CSEU problem, it still requires order d applications of the unknown unitary channel $\mathcal{U}$. One may further ask if a better protocol exists or if there is a fundamental limitation on the query cost. In this section, we show that indeed a $\tilde{\Omega}(d)$ query complexity lower bound exists for the CSEU task, which stems from information-theoretic reasons.

The CSEU task is formally defined as follows; we refer to Fig. 1 for a visualization.

Definition 1 (CSEU task). The CSEU task involves developing a protocol $\mathcal{B}$ that contains two separate algorithms $\mathcal{B}_{\text{learn}}$ and $\mathcal{B}_{\text{pred}}$. The “learning” algorithm $\mathcal{B}_{\text{learn}}$ is given access to an unknown n -qubit unitary channel $\mathcal{U}$; after applying it multiple times in some quantum

experiments, $\mathcal{B}_{\text{learn}}$ should output some classical data. The “prediction” algorithm $\mathcal{B}_{\text{pred}}$ takes this classical data, a collection of states $\rho_1, \dots, \rho_M$ in $\mathcal{D}(\mathcal{H})$ and observables $O_1, \dots, O_M$ in $\text{Obs}(B)$ as inputs, and should output M estimates $E(O_l, \rho_l)$. The goal of the task is to reach

$$\Pr\{|E(O_l, \rho_l) - \text{Tr}(O_l \mathcal{U}(\rho_l))| < \epsilon \ \forall 1 \leq l \leq M\} \geq 1 - \delta, \quad (30)$$

where $0 < \delta, \epsilon < 1$ are accuracy parameters.

Theorem 2. Suppose $1 \leq B \leq d$, $0 < \delta, \epsilon \leq 1/3$, and $M \geq 1$. Then any protocol has to query the channel $\mathcal{U}$ at least $\tilde{\Omega}(d)$ times to complete the CSEU task.

By this theorem, an exponential query complexity (with respect to the qubit number) for the CSEU task is inevitable, even if highly complex adaptive or collective operations are used. In addition, the scaling of the lower bound in dimension d matches (up to a $\log d$ factor) the query complexity of our protocol constructed in Sec. III [see Eq. (29)]. In other words, our protocol for CSEU is nearly query-optimal.

A. Proof of Theorem 2

In this subsection, we prove Theorem 2 via a reduction from a problem of learning Boolean functions to the CSEU task. That is, any protocol for the CSEU task can be modified to complete another task defined below, whose query complexity lower bound is known in the literature. This existing lower bound will imply that the CSEU task must require a significant amount of queries.

Definition 2. The *exact learning task of Boolean functions* involves developing a quantum protocol $\mathcal{P}$ for learning an unknown Boolean function $c : \{0, 1\}^n \rightarrow \{0, 1\}$. Here $\mathcal{P}$ is given access to an oracle $\text{MQ}(c)$, which is a quantum unitary satisfying $\text{MQ}(c)|x, b\rangle = |x, b \oplus c(x)\rangle$ for all $x \in \{0, 1\}^n$ and $b \in \{0, 1\}$. The goal of the task is to achieve: with probability at least $2/3$, $\mathcal{P}$ outputs a classical function h such that $h(x) = c(x)$ for all $x \in \{0, 1\}^n$.

Lemma 3 ([53, 54]). Any learning protocol has to query the oracle $\text{MQ}(c)$ at least $\Omega(2^n)$ times to achieve the exact learning task of Boolean functions.

Now we are ready to prove Theorem 2. Suppose the unitary channel $\mathcal{U}$ is acting on $n + 1$ qubits, and there is a protocol $\mathcal{B} = (\mathcal{B}_{\text{learn}}, \mathcal{B}_{\text{pred}})$ that can use N queries of $\mathcal{U}$ to successfully achieve the CSEU task, with parameters $B = M = 1$ and $\delta, \epsilon = 1/3$. In the following, we shall prove that necessarily $N \geq \Omega(2^n/n)$, which implies Theorem 2. Note that here we assume that the qubit number is $n + 1$ instead of n for the convenience of discussion. The proof below is divided into two parts.

1. Part 1: Predicting many expectation values

In this part, we show that there exists a new protocol $\mathcal{B}^* = (\mathcal{B}_{\text{learn}}^*, \mathcal{B}_{\text{pred}}^*)$ that can use $N\mathcal{O}(\log(M/\delta))$ queries of $\mathcal{U}$ to complete the CSEU task with parameters $B = 1$, $\epsilon = 1/3$, any $0 < \delta < 1$, and any integer $M \geq 1$.

Based on $\mathcal{B}$, a new protocol $\mathcal{B}^*$ is constructed as follows using the median-of-means technique: in the learning phase, we run the algorithm $\mathcal{B}_{\text{learn}}$ in total k times independently to obtain k pieces of classical data; in the prediction phase, we classically process these data using $\mathcal{B}_{\text{pred}}$ to obtain k estimates $E_1(O_l, \rho_l), \dots, E_k(O_l, \rho_l)$ for each $1 \leq l \leq M$, and we take the median of them as our final estimate for $\text{Tr}(O_l \mathcal{U}(\rho_l))$:

$$\tilde{E}(O_l, \rho_l) := \text{median}\{E_1(O_l, \rho_l), \dots, E_k(O_l, \rho_l)\}. \quad (31)$$

To estimate all $\text{Tr}(O_l \mathcal{U}(\rho_l))$ accurately with failure probability at most δ , it suffices to choose $k = \mathcal{O}(\log(M/\delta))$ by an analysis identical to that used in the proof of Theorem 1. Thus, the total number of queries required by $\mathcal{B}^*$ is $N \cdot k = N\mathcal{O}(\log(M/\delta))$. In particular, when $\delta = 1/3$ and $M = 2^n$, the query number reads

$$N\mathcal{O}(\log(M/\delta)) = N\mathcal{O}(\log(3 \times 2^n)) = N\mathcal{O}(n). \quad (32)$$

2. Part 2: Reduction to the learning task of Boolean functions

In this part, we show that the protocol $\mathcal{B}^*$ constructed in Part 1 can be used to complete the exact learning task of Boolean functions given in Definition 2. Consequently, the lower bound on the query complexity for this exact learning task can be turned into a lower bound for the CSEU task.

For $x \in \{0, 1\}^n$, define quantum state $\rho_x = |x, 0\rangle\langle x, 0|$ and observable $O_x = |x, 1\rangle\langle x, 1| \in \text{Obs}(1)$. Then for any Boolean function $c : \{0, 1\}^n \rightarrow \{0, 1\}$, its corresponding $\text{MQ}(c)$ oracle (see Definition 2) satisfies

$$\text{Tr}[O_x \text{MQ}(c) \rho_x \text{MQ}(c)^\dagger] = |\langle x, 1 | x, c(x) \rangle|^2 = c(x). \quad (33)$$

Recall that by using $N\mathcal{O}(n)$ queries of the $\text{MQ}(c)$ oracle, the protocol $\mathcal{B}^*$ can estimate all $\text{Tr}[O_x \text{MQ}(c) \rho_x \text{MQ}(c)^\dagger]$ up to precision $\epsilon = 1/3$, with failure probability at most $1/3$. Let $\tilde{E}_x$ be the estimate for $\text{Tr}[O_x \text{MQ}(c) \rho_x \text{MQ}(c)^\dagger]$ and define the function

$$h(x) := \begin{cases} 0 & \text{when } \tilde{E}_x \leq 1/2, \\ 1 & \text{when } \tilde{E}_x > 1/2. \end{cases} \quad (34)$$

According to Eq. (33), with probability at least $2/3$, we have $h(x) = c(x)$ for all $x \in \{0, 1\}^n$. Therefore, $\mathcal{B}^*$ can successfully achieve the exact learning task of Boolean functions by using $N\mathcal{O}(n)$ queries. This query number cannot be smaller than the lower bound in Lemma 3; that is, $N\mathcal{O}(n) \geq \Omega(2^n)$. It follows that $N \geq \Omega(2^n/n)$, which completes the proof of Theorem 2.

V. CSEU IN THE AVERAGE CASE

We have so far considered CSEU in the worst-case scenario, where the learner is asked to make an accurate prediction for any input state ρ and observable O . As shown in Sec. IV, at least $\tilde{\Omega}(d)$ queries of the unitary channel $\mathcal{U}$ are required to achieve this goal. In this section, we turn to consider CSEU in the average-case scenario, and show that the query complexity lower bound $\tilde{\Omega}(d)$ can be broken in many parameter regimes.

In the average-case scenario, we still consider that an arbitrary observable $O \in \text{Obs}(B)$ is input into the prediction phase. Using the classical data obtained from the learning phase, the goal is to find a real-valued function $E(O, \rho)$ that can predict $\text{Tr}(O\mathcal{U}(\rho))$ within a small mean absolute error

$$\Xi(\mathcal{T}, O) := \text{Ave}_{\rho \sim \mathcal{T}} |E(O, \rho) - \text{Tr}(O\mathcal{U}(\rho))|, \quad (35)$$

where the average is over quantum states ρ that obey some specific distribution $\mathcal{T}$ on $\mathcal{D}(\mathcal{H})$. In this section we use the notation “Ave” instead of “ $\mathbb{E}$ ” to avoid confusion with the meaning of the expectation $\mathbb{E}$ in Sec. III.

The above task can be solved via our learning protocol developed in Sec. III: in the prediction phase, we set the number of batches $R = 1$, and use $\hat{E}(O, \rho)$ in Eq. (15) as the function for estimating $\text{Tr}(O\mathcal{U}(\rho))$. As clarified in the following lemma, the performance of our protocol in the average-case scenario is closely related to the average variance of $\hat{E}(O, \rho)$, see Appendix H for a proof.

Lemma 4. Suppose $\mathcal{T}$ is a state distribution on $\mathcal{D}(\mathcal{H})$, $O \in \text{Obs}(B)$, accuracy parameters $0 < \delta, \epsilon < 1$, and

$$\text{Ave}_{\rho \sim \mathcal{T}} \text{Var}[\hat{E}(O, \rho)] \leq \frac{\delta \epsilon^2}{4}. \quad (36)$$

Then, with failure probability at most δ , the mean absolute error $\Xi(\mathcal{T}, O) < \epsilon$.

When $R = 1$, the function $\hat{E}(O, \rho)$ reduces to the mean estimate $\hat{Z}_{(r)}(O, \rho)$ in Eq. (14), whose variance has been clarified in Proposition 1. So we can use Proposition 1 and Lemma 4 to determine the number of queries needed by our protocol in the average-case scenario for any state distribution $\mathcal{T}$.

In the following, we assume that $\mathcal{T} = \pi_{d,\lambda}$ is a distribution of random density operators in $\mathcal{D}(\mathcal{H})$ constructed by taking partial trace of Haar-random pure states on a composite system [55–57].

Definition 3. The λ -induced distribution $\pi_{d,\lambda}$ of density matrix is the distribution on $\mathcal{D}(\mathcal{H})$ induced by the uniform distribution of pure states in $\mathcal{H} \otimes \mathcal{H}_\lambda$, where the dimensions of $\mathcal{H}$ and $\mathcal{H}_\lambda$ are d and λ , respectively. A

state ρ following the distribution $\pi_{d,\lambda}$ can be generated by $\rho = \text{Tr}_\lambda(|\varphi\rangle\langle\varphi|)$, where $|\varphi\rangle \in \mathcal{H} \otimes \mathcal{H}_\lambda$ is a pure state drawn from the Haar measure.

The average purity of quantum states $\rho \sim \pi_{d,\lambda}$ reads $(d + \lambda)/(\lambda d + 1) \approx 1/d + 1/\lambda$ [58]. This fact together with Proposition 1 immediately implies the following corollary.

Corollary 1. Suppose $1 \leq s \leq d$ is the number of systems collectively measured in the learning phase, $R = 1$ is the number of batches we set in the prediction phase, dimension $\lambda \leq d$, and observable $O \in \text{Obs}(B)$. Then

$$\begin{aligned} \text{Ave}_{\rho \sim \pi_{d,\lambda}} \text{Var}[\hat{E}(O, \rho)] \\ = \mathcal{O}\left[\frac{1}{m} \left(\frac{d}{s\lambda} + \min\left\{1, \frac{B}{\lambda}\right\} \right) + \frac{d^4 B}{m^2 s^4 \lambda}\right], \end{aligned} \quad (37)$$

where m is the number of snapshots obtained from the learning phase. In particular, when $s = 1$ we have

$$\text{Ave}_{\rho \sim \pi_{d,\lambda}} \text{Var}[\hat{E}(O, \rho)] = \mathcal{O}\left(\frac{d}{m\lambda} + \frac{d^4 B}{m^2 \lambda}\right). \quad (38)$$

From Corollary 1 and Lemma 4, we can easily determine the query number $m \cdot s$ of our protocol required to carry out the CSEU task in the average-case scenario with distribution $\pi_{d,\lambda}$. First, if we are limited to single-copy measurements, i.e., $s = 1$, then

$$m \cdot s = \mathcal{O}\left(\frac{d}{\lambda \epsilon^2} + \frac{d^2}{\epsilon} \sqrt{\frac{B}{\lambda}}\right) \quad (39)$$

queries are sufficient to ensure that the mean absolute error satisfies $\Xi(\pi_{d,\lambda}, O) < \epsilon$ with high probability. This query complexity is lower than that of our protocol in the worst-case scenario [see Eq. (28)].

Second, if collective measurements on multiple systems are accessible, then the conditions in Lemma 4 can be satisfied by choosing $s = \lfloor d\lambda^{-1/4} \rfloor$ and

$$m = \mathcal{O}\left[\frac{1}{\epsilon^2} \left(\frac{1}{\lambda^{3/4}} + \min\left\{1, \frac{B}{\lambda}\right\} \right) + \frac{\sqrt{B}}{\epsilon}\right], \quad (40)$$

in which case the query number reads

$$m \cdot s = \mathcal{O}\left[\frac{d}{\epsilon^2} \left(\frac{1}{\lambda} + \frac{\min\{\lambda, B\}}{\lambda^{5/4}} \right) + \frac{d\sqrt{B}}{\lambda^{1/4} \epsilon}\right]. \quad (41)$$

When B and ϵ are independent of the system size, the query complexity reduces to $m \cdot s = \mathcal{O}(d\lambda^{-1/4})$. Notably, $\mathcal{O}(d^{3/4})$ queries of $\mathcal{U}$ are sufficient when $\lambda = \Theta(d)$. Hence,

the query lower bound in Theorem 2 for the worst-case scenario does not apply to the average-case scenario. Previously, such remarkable separation in query complexity between the worst- and average-case considerations has been demonstrated in certain channel learning tasks [25,27].

VI. COMPARISON WITH PREVIOUS WORKS

In this section, we survey existing works related to CSEU and compare them with our results. The performance of these protocols on the CSEU task are summarized in Table I in Sec. I.

A. CSE of general quantum channels

By applying the concepts of classical shadow formalism to quantum states, two recent works [21,22] considered CSE of general quantum channels (CSEC). They proposed two protocols for CSEC that do not utilize any quantum memory. The first protocol is ancilla-assisted: the learner first prepares the Choi state $(1/d)\Upsilon_C$ of the unknown channel $\mathcal{C}$ by applying $\mathcal{C}$ to the first system of the maximally entangled state $|\Phi\rangle\langle\Phi|/d$ on $\mathcal{H}^{\otimes 2}$, and then performs CSE on this Choi state. For any state ρ and observable O , the property $\text{Tr}(O\mathcal{C}(\rho))$ can be recovered by virtue of channel-state duality [52]:

$$\text{Tr}(O\mathcal{C}(\rho)) = d \text{Tr}\left[(O \otimes \rho^\top) \frac{\Upsilon_C}{d}\right]. \quad (42)$$

If the M observables O_l to be estimated are chosen from $\text{Obs}(B)$, this protocol requires $\mathcal{O}(d^2 B \epsilon^{-2} \log M)$ applications of $\mathcal{C}$ to estimate the properties in Eq. (1) all within additive error ϵ . However, this protocol requires an ancillary system with dimension d , which may limit its practical applicability.

The second protocol for CSEC is ancilla-free [21,22]. In each round of experiments, one prepares a random pure state on $\mathcal{H}$, evolves it under $\mathcal{C}$, and then performs a randomized measurement on the output state. This procedure is identical to that of our protocol based on single-copy measurements. Using the experimental results, one can construct an unbiased estimator for the Choi operator Υ_C , from which any desired linear property can be recovered. To predict M properties in Eq. (1) all within additive error ϵ , this protocol requires

$$\mathcal{O}\left(\frac{d^2 \log M}{\epsilon^2} \max_{1 \leq l \leq M} \|S(O_l)\|\right) \quad (43)$$

applications of $\mathcal{C}$ (see Theorem 2 in Ref. [21]), where

$$S(O) := 4[\text{Tr}(O)]^2 I + 2 \text{Tr}(O^2) I + 4 \text{Tr}(O) O + 4 O^2. \quad (44)$$

Notice that $\|S(O)\|$ can be very large for some specific observables in $\text{Obs}(B)$. In particular, due to the first term

in Eq. (44), it attains its maximum scaling $\Theta(dB)$ when $O = \sqrt{d^{-1}B} I \in \text{Obs}(B)$. As a result, if all observables O_l to be predicted are chosen from $\text{Obs}(B)$, this protocol has a $\mathcal{O}(d^3 B \epsilon^{-2} \log M)$ query complexity.

In contrast to the two protocols mentioned above, our protocol is designed for predicting properties of unitary channels rather than general channels. However, our protocol based on single-copy measurements not only works without quantum memory and ancilla, but also has much higher query efficiency than the above protocols (see Table I). To better visualize this advantage, let us consider the case in which $\{O_l\}$ is the collection of 4^n Pauli observables, and high-precision estimation with error $\epsilon = 1/d$ is desired. In this case, our CSEU protocol that does not use quantum memory needs $\tilde{\mathcal{O}}(d^{3.5})$ queries, whereas the CSEC protocols mentioned above [21,22] need $\tilde{\mathcal{O}}(d^5)$ queries. The intrinsic reason behind this advantage is that our protocol employs the property of the snapshots $\hat{X}_i$ presented in Eq. (16) to construct the quadratic estimator $d^{-1} \text{Tr}[(O \otimes \rho^\top) \hat{X}_i \hat{X}_j]$ (see Sec. III B). For a general quantum channel, Eq. (16) may fail and such an unbiased quadratic estimator cannot be constructed.

B. Learning protocol based on Ref. [11]

A recent work [11] found that when collective measurements are used, the task of CSE for pure states can be solved with a lower sample complexity than the original shadow protocol in Ref. [9]. More specifically, suppose we are given multiple copies of an unknown pure state σ , and $Q_1, \dots, Q_M$ are arbitrary observables in $\text{Obs}(B)$. To predict M expectation values $\text{Tr}(Q_i \sigma)$ all within additive error ϵ , Ref. [11] showed that it suffices to use the following number of samples:

$$\mathcal{O}\left(\left(\frac{\sqrt{B}}{\epsilon} + \frac{1}{\epsilon^2}\right) \log M\right). \quad (45)$$

Based on this result, we can naively construct the following protocol for CSEU: one first prepares the (pure) Choi state $(1/d)\Upsilon_U$ of the unitary channel $\mathcal{U}$, then performs CSE on the Choi state using the protocol of Ref. [11]. Suppose $\rho_1, \rho_2, \dots, \rho_M$ are quantum states in $\mathcal{D}(\mathcal{H})$ and $O_1, O_2, \dots, O_M$ are observables in $\text{Obs}(B)$. According to Eq. (45), to estimate $\text{Tr}[(O_l \otimes \rho_l^\top)(1/d)\Upsilon_U]$ within error $0 < \eta < 1$ for all l , it suffices to query the channel

$$\mathcal{O}\left(\left(\frac{\sqrt{B}}{\eta} + \frac{1}{\eta^2}\right) \log M\right) \quad (46)$$

times. In addition, by Eq. (42), to estimate $\text{Tr}(O_l \mathcal{U}(\rho_l))$ within ϵ precision, one needs to keep the additive error η in estimating $\text{Tr}[(O_l \otimes \rho_l^\top)(1/d)\Upsilon_U]$ smaller than $d^{-1}\epsilon$. As a

result, the query complexity of this protocol reads

$$\mathcal{O}\left(\left(\frac{\sqrt{B}}{d^{-1}\epsilon} + \frac{1}{[d^{-1}\epsilon]^2}\right) \log M\right) = \mathcal{O}\left(\frac{d^2}{\epsilon^2} \log M\right), \quad (47)$$

where we use the relation $B \leq d$. By contrast, the query complexity of our protocol based on collective measurements is $\mathcal{O}(d(\epsilon^{-2} + \sqrt{B}\epsilon^{-1}) \log M)$, which improves the scaling behavior in the dimension d quadratically.

C. Process tomography of unitary channels

Another approach for completing the CSEU task is to first use quantum process tomography of unitary channels (QPTU) [6,30,59,60] to construct a classical description of the unknown channel $\mathcal{U}$, and then use this description to estimate $\text{Tr}(O\mathcal{U}(\rho))$. When quantum memory is accessible, Ref. [6] showed that $\Theta(d^2\epsilon^{-1})$ queries of $\mathcal{U}$ are necessary and sufficient to output a classical description that is ϵ -close to $\mathcal{U}$ in the diamond norm. Consequently, any linear property $\text{Tr}(O\mathcal{U}(\rho))$ can be estimated within additive error ϵ from $\mathcal{O}(d^2\epsilon^{-1})$ queries. In the tomography protocol of Ref. [6], the unknown channel $\mathcal{U}$ is applied in series on one system of dimension d , so quantum memory is needed but ancillary systems are not. Compared with our protocol based on collective measurements, this protocol has a smaller space overhead, but a larger query complexity for completing the CSEU task when $d > \epsilon^{-1}$, which is the usual case.

When quantum memory is not accessible, the protocol in Ref. [30] has the highest efficiency among all QPTU protocols as far as we are aware. Using $\mathcal{O}(d^2\eta^{-2})$ queries of the unknown channel $\mathcal{U}$, this protocol can learn a classical description of the Choi state $(1/d)\Upsilon_{\mathcal{U}}$ within $0 < \eta < 1$ trace distance (see Theorem 1 in Ref. [30]). From this description, one can then estimate $\text{Tr}[(O \otimes \rho^{\top})(1/d)\Upsilon_{\mathcal{U}}]$ within additive error η for any state ρ and observable O . According to Eq. (42), this implies a $\mathcal{O}(d^4\epsilon^{-2})$ query complexity for estimating $\text{Tr}(O\mathcal{U}(\rho))$ within ϵ error—nearly a quadratic overhead compared with our CSEU protocol that does not use quantum memory.

D. Other works

Recently, Ref. [23] proposed an efficient protocol without quantum memory for learning to predict $\text{Tr}(O\mathcal{C}(\rho))$ for an arbitrary unknown channel $\mathcal{C}$, achieving a small average error over input states ρ drawn from any locally flat distribution. Here, O is restricted to a *bounded-degree observable* that can be expressed as a sum of local observables, each has support on a constant number of qubits. In contrast, our protocol for CSEU focuses on unitary channels, but can deal with the worst-case input states and general observables.

Using quantum memory, Ref. [24] also gave an efficient procedure for predicting linear properties $\text{Tr}(O\mathcal{C}(\rho))$

for an arbitrary unknown channel $\mathcal{C}$. However, here the input state ρ is restricted to Pauli-sparse quantum states, and the observable O is restricted to bounded Pauli-sparse observables.

Instead of general quantum channels, recent works [25–29] considered quantum channels or unitaries $\mathcal{C}$ that can be generated in polynomial time, and showed that a polynomial amount of queries are sufficient to estimate $\text{Tr}(O\mathcal{C}(\rho))$ in this restricted class. By contrast, our protocol for CSEU applies to arbitrary unitary channels, in which case an exponential number of queries is inevitable according to Theorem 2.

VII. APPLICATION IN HAMILTONIAN LEARNING

A. Background and our result

Learning an unknown Hamiltonian is a central task in many-body quantum physics. In an isolated quantum system, the Hamiltonian H fully characterizes the system's time evolution through the dynamic operator e^{-iHt} , so all information about the future state of the system is determined by its initial state and H . Beyond its fundamental significance in understanding quantum dynamics, Hamiltonian learning has broad applications across various fields, including quantum metrology and sensing [61,62], and quantum device engineering [63,64].

Here we consider the problem of learning an unknown n -qubit Hamiltonian H with respect to the *normalized Frobenius norm* (NFN). More explicitly, given time evolution access to H , our goal is to learn a classical description $\hat{H} \in \mathcal{L}_{\text{H}}(\mathcal{H})$ of the Hamiltonian, such that

$$\Pr\left\{\frac{1}{\sqrt{2^n}} \|\hat{H} - H\|_2 \leq \epsilon\right\} \geq 1 - \delta \quad (48)$$

for accuracy $0 < \delta, \epsilon < 1/3$. Here we do not make prior assumptions on the structure of H , but only have two natural constraints $\text{Tr}(H) = 0$ and $\|H\| = \mathcal{O}(\text{poly}(n))$. According to Theorem 1.4 of Ref. [65], even with access to quantum memory, any protocol has to make at least $\tilde{\Omega}(d^2)$ applications of H to solve this problem.

We choose NFN as the error metric because it reflects how well we can predict the system on average based on the learned description $\hat{H}$ of the Hamiltonian [66]. Suppose we are given a known initial state ρ_0 , and we want to predict observation expectation values of the system at a given time t by using $\hat{H}$. Starting from ρ_0 , the true state $\rho(t)$ at time t and the state $\hat{\rho}(t)$ we predict will be

$$\rho(t) = e^{-iHt} \rho_0 e^{iHt}, \quad \hat{\rho}(t) = e^{-i\hat{H}t} \rho_0 e^{i\hat{H}t}. \quad (49)$$

Accordingly, the true and predicted expectation values of a given observable O will be

$$\langle O \rangle_{\rho(t)} = \text{Tr}[O\rho(t)], \quad \langle O \rangle_{\hat{\rho}(t)} = \text{Tr}[O\hat{\rho}(t)], \quad (50)$$

respectively. The absolute value of their difference represents our prediction error based on $\hat{H}$. If the initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$ is a random pure state drawn from the Haar distribution, then the mean squared error can be bounded as [66]

$$\mathbb{E}_{|\psi_0\rangle \sim \mu_H} \left[|\langle O \rangle_{\rho(t)} - \langle O \rangle_{\hat{\rho}(t)}|^2 \right] \leq 2 \left(t \|O\|_2 \frac{\|\hat{H} - H\|_2}{\sqrt{2^n}} \right)^2. \quad (51)$$

Therefore, a small NFN-error bound guarantees a high average accuracy in predicting the expectation values of observables under a random initial state [66,67].

In recent years, the task of learning an unknown Hamiltonian H through its unitary dynamics has been extensively studied in the literature [24,36,37,64–66,68–76]. Most existing protocols use repeated applications of the Hamiltonian's time evolution in each experiment, thereby demanding the resource of quantum memory. In addition, these protocols generally rely on strong assumptions about the structure of H —for instance, assuming that H exhibits sparse interactions or contains Pauli terms with constant weights. However, in many practical scenarios, the Hamiltonian may have an unknown structure and involve arbitrary interaction couplings, which restrict the applicability of these protocols.

Furthermore, it is worth mentioning that learning a Hamiltonian H with respect to NFN is much harder than simply estimating its Pauli coefficients, although the latter is the focus of most previous Hamiltonian learning protocols. More precisely, estimating all Pauli coefficients of H within additive error ϵ only implies learning H within error $d\epsilon$ in NFN (see Appendix I), where $d = 2^n$ is the dimension of the system. Due to this reason, when attempting to extend existing protocols to learn an arbitrary Hamiltonian in NFN, the number of required experimental runs and total evolution time under H would become prohibitively large, as shown in Appendix I.

Here, to demonstrate the power of our CSEU protocol, we show that it can be combined with the polynomial interpolation technique to learn an arbitrary Hamiltonian H in NFN. The learning protocol is nonadaptive, and does not need quantum memory or ancillary qubits. In each round of experiments, it requires only the preparation of a computational-basis state, the application of a random Clifford unitary to a single system, a measurement in the computational basis, and a real $\mathcal{O}(1/\|H\|)$ -time evolution under H . The protocol's query complexity and evolution time complexity are summarized in the following theorem, which is a simplified version of Theorem I1 in Appendix I.

Theorem 3. Given time evolution access to an arbitrary n -qubit traceless Hamiltonian H with $\|H\| = \mathcal{O}(\text{poly}(n))$. There is an ancilla-free protocol without quantum memory

that outputs a classical description $\hat{H} \in \mathcal{L}_H(\mathcal{H})$ such that Eq. (48) holds. Both the number of queries to H and the total evolution time required by the protocol scale as

$$\tilde{\mathcal{O}}\left(\left(\frac{d^2}{\epsilon^2} + \frac{d^3}{\epsilon}\right) \log(\delta^{-1})\right). \quad (52)$$

To the best of our knowledge, this is the first protocol based on $\tilde{\mathcal{O}}(d^3)$ queries and evolution time for learning an arbitrary Hamiltonian with respect to NFN. Notably, even without using quantum memory, the query complexity is still close to the known theoretical lower bound $\tilde{\Omega}(d^2)$ [65], demonstrating both high efficiency and practical applicability. In contrast, the most efficient protocol in the literature requires $\tilde{\mathcal{O}}(d^4)$ Hamiltonian queries and total evolution time to achieve this task even if quantum memory is accessible (see Appendix I).

B. Schematic overview of our Hamiltonian learning protocol

In this section, we formalize the basic principles and methodology of our Hamiltonian learning protocol. Additional technical details regarding the protocol can be found in Appendix I.

Our goal is to learn an arbitrary unknown traceless Hamiltonian H , which can be expanded in the form:

$$H = \sum_{\mathbf{q} \in \{0,1,2,3\}^n} \mu(\mathbf{q}) \sigma_{\mathbf{q}}. \quad (53)$$

Here, $\sigma_{\mathbf{q}}$ are Pauli operators, and $\mu(\mathbf{q}) \in \mathbb{R}$ are Pauli coefficients with $\mu(0^n) = 0$. For a system with an initial state ρ evolving under H , the expectation value of an operator O at time t can be written as [24,36,37]

$$\langle O \rangle_{\rho(t)} = \text{Tr}[O \mathcal{U}_t(\rho)] = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \text{Tr}[g_k(H, O) \rho], \quad (54)$$

where $\mathcal{U}_t(\cdot) := e^{-iHt}(\cdot)e^{iHt}$ denotes the unitary channel describing the time evolution, and

$$g_k(H, O) := \underbrace{[H, [H, [\dots, [H, O] \dots]]}_{k \text{ times}}, \quad g_{k=0}(H, O) := O. \quad (55)$$

As shown in Refs. [24,36,37], for any Pauli operator $\sigma_{\mathbf{q}} \neq I$, there is a quantum state $\rho_{\mathbf{q}}$ and another Pauli operator $\sigma_{\mathbf{q}'}$ that satisfy $\text{Tr}(i[H, \sigma_{\mathbf{q}'}]\rho_{\mathbf{q}}) = 2\mu(\mathbf{q})$. This fact together with Eq. (54) leads to the following key observation. For any nonidentity Pauli operator $\sigma_{\mathbf{q}}$, if the initial state ρ and observable O are properly chosen, then the first-order time derivative of $\langle O \rangle_{\rho(t)}$ at $t = 0$ tells us the Pauli coefficient $\mu(\mathbf{q})$. In addition, this derivative can be extracted

by performing polynomial interpolation on the function $t \mapsto \langle O \rangle_{\rho(t)}$ [24,36,37].

Building on the above observations, our protocol integrates the CSEU method developed in Sec. III and the derivative estimation procedure of Ref. [24] to learn the unknown Hamiltonian. In the first step, for several different time points t , we treat $\mathcal{U}_t$ as an unknown unitary channel, and use our CSEU protocol to accurately estimate the expectation value $\langle O \rangle_{\rho(t)}$ corresponding to each $\sigma_{\mathbf{q}} \neq I$. Due to the high efficiency of our CSEU method, this step requires much fewer Hamiltonian queries (and thus less total evolution time) than existing approaches. The second step of our protocol is classical postprocessing, in which we estimate each Pauli coefficient $\mu(\mathbf{q})$ within additive error ϵ/d by using polynomial interpolation, following the same procedure as in Ref. [24]. Let $\hat{\mu}(\mathbf{q})$ be the estimate for $\mu(\mathbf{q})$; then our protocol outputs $\hat{H} = \sum_{\mathbf{q}} \hat{\mu}(\mathbf{q}) \sigma_{\mathbf{q}}$ as the learned description of H , which satisfies

$$\begin{aligned} \frac{\|\hat{H} - H\|_2}{\sqrt{d}} &= \sqrt{\frac{1}{d} \text{Tr}[(\hat{H} - H)^2]} \\ &= \sqrt{\sum_{\mathbf{q} \in \{0,1,2,3\}^n} |\hat{\mu}(\mathbf{q}) - \mu(\mathbf{q})|^2} \\ &\leq \sqrt{\sum_{\mathbf{q} \in \{0,1,2,3\}^n} \left(\frac{\epsilon}{d}\right)^2} = \epsilon. \end{aligned} \quad (56)$$

In this way, the unknown Hamiltonian H is learned within error ϵ with respect to NFN.

VIII. APPLICATION IN ESTIMATING OTOCS

So far we have considered predicting linear functions of the unknown unitary channel $\mathcal{U}$. Actually, the classical shadow data $\text{UShadow}(\mathcal{U}, m)$ obtained in our learning phase may also be used to predict nonlinear functions of $\mathcal{U}$. One particularly crucial nonlinear function is the out-of-time-ordered correlators (OTOCs). As a tool to quantify quantum information scrambling, OTOC plays critical roles in both quantum many-body physics [77,78] and quantum information [79,80]. Suppose $U(t)$ is a unitary time-evolution operator and W is a local Hermitian operator. For chaotic dynamics, the Heisenberg operator $W_{U(t)} = U(t)^\dagger W U(t)$ rapidly becomes nonlocal and noncommutative with an initially nonoverlapping local operator V . The degree of this noncommutativity can be quantified by the OTOC

$$C(U(t)) = \text{Tr}(\rho W_{U(t)} V^\dagger W_{U(t)} V), \quad (57)$$

with ρ being the initial quantum state. To simplify the notation, in the following we often omit “(t)” and use U to denote the unitary time-evolution operator.

For given W and V , below we consider the task of predicting OTOC by querying the unknown time-evolution operator U several times. Using the tensor-network diagram (see Appendix B 2), the OTOC can be written as

$$C(U) = \text{Tr}[(\Upsilon_U \otimes \Upsilon_U)(W \otimes (V^\dagger)^\top \otimes W \otimes (V\rho)^\top) T_{(1,3)}]. \quad (58)$$

In this diagram, the order in which operators are multiplied is from top to bottom, “ Υ ” and “ λ ” means the periodic boundary condition for top and bottom legs, and $T_{(1,3)}$ is the swap operator on the first and third systems. Recall that each snapshot $\hat{X}$ in the classical shadow data $\text{UShadow}(\mathcal{U}, m)$ equals the Choi operator Υ_U in expectation. So one can predict $C(U)$ by using the following estimator constructed from pairs of distinct snapshots:

$$\begin{aligned} \hat{C}(U) &:= \frac{1}{m(m-1)} \sum_{i \neq j} \text{Tr}[(\hat{X}_i \otimes \hat{X}_j) \\ &\quad \times (W \otimes (V^\dagger)^\top \otimes W \otimes (V\rho)^\top) T_{(1,3)}]. \end{aligned} \quad (59)$$

Here the summation is over all $m(m-1)$ quadratic terms with $i, j = 1, \dots, m$ and $i \neq j$. Due to the independence between $\hat{X}_i$ and $\hat{X}_j$, this estimator correctly predicts the OTOC in expectation.

To study the query efficiency of this approach, following Refs. [81,82], here we focus on the infinite temperature thermal state $\rho = I/d$, and assume that W and V are both unitary, traceless, and Hermitian operators. In this case, as proved in Appendix J, the variance of our estimator reads

$$\text{Var}[\hat{C}(U)] = \mathcal{O}\left[\frac{1}{m} \left(\frac{d^2}{s^2} + 1\right) + \frac{d^2}{m^2} \left(\frac{d^4}{s^4} + 1\right)\right]. \quad (60)$$

By combining this result with Chebyshev’s inequality, we immediately arrive at the following proposition, which clarifies the query complexity for estimating OTOC.

Proposition 2. In order to estimate $C(U)$ within additive error $0 < \epsilon < 1$ and some high (constant) success

probability, the number of queries used by our protocol is

$$m \cdot s = \mathcal{O}\left(\frac{d^2}{\epsilon^2} + \frac{d^3}{\epsilon}\right) \quad (61)$$

when we use only single-copy measurements, i.e., $s = 1$; and is

$$m \cdot s = \mathcal{O}\left(\frac{d}{\epsilon^2} + \frac{d^2}{\epsilon}\right) \quad (62)$$

when we use collective measurements on d systems, i.e., $s = d$.

Due to its shadow nature, our protocol can estimate many OTOCs for different initial operators W and V simultaneously by classical postprocessing. In contrast, previous methods [81,82] need to encode the information on W and V into the initial state preparation, unitary evolution, and final measurements. As a result, the measurement outcomes can only be used to estimate the chosen OTOC. In addition, our protocol based on collective measurements can significantly enhance the query efficiency compared with Ref. [82], as shown in Eq. (62). Furthermore, it is straightforward to generalize our protocol to predict higher-order OTOCs [82,83].

IX. CONCLUSION

In this work, we initiated a comprehensive investigation of the CSEU problem, which is of fundamental interest to both theoretical studies and practical applications. By utilizing the power of collective measurements, we proposed a novel protocol for the CSEU task, whose query complexity is nearly asymptotically optimal for prediction in the worst case, and can be further reduced for prediction in the average case. To further enhance practicality, we also introduced an alternative ancilla-free protocol based on single-copy measurements, which achieves much higher query efficiency compared with previous protocols that do not use quantum memory. This protocol can serve as a key subroutine for learning an arbitrary unknown Hamiltonian with respect to NFN, significantly outperforming existing approaches for this task. Furthermore, our protocols can be extended to predict nonlinear properties of unitary channels, and have demonstrated high efficiency in estimating OTOC, a crucial quantity in quantum many-body physics. This work represents a significant advance in the study of learning quantum channels, and is expected to trigger a cascade of future research works on this subject.

There are some intriguing problems that merit further investigation. First, although our protocol based on collective measurements achieves a nearly optimal query complexity, it applies the unknown unitary channel $\mathcal{U}$ in parallel, resulting in a large space overhead. It is natural to ask whether this overhead can be avoided by considering sequential applications of $\mathcal{U}$ [6], which would make

the protocol more practical. Second, our protocol applies quadratic estimators in the prediction phase to estimate linear properties (see Sec. III B). It would be interesting to study the potential benefits of higher-order estimators constructed from more snapshots, which may further reduce the query complexity with respect to the parameter B . The third question is whether our protocol based on single-copy measurements can achieve the optimal query complexity with respect to the system dimension d over all protocols that do not rely on quantum memory. We leave these lines of research to future study.

ACKNOWLEDGMENTS

The authors thank Penghui Yao, Zhenhuan Liu, Datong Chen, Fuchuan Wei, Qi Ye, and Masahito Hayashi for inspiring discussions and valuable comments. Z.L., C.Y., and H.Z. are supported by the Shanghai Science and Technology Innovation Action Plan (Grant No. 24LZ1400200), the Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01), the National Key Research and Development Program of China (Grant No. 2022YFA1404204), and the National Natural Science Foundation of China (Grant No. 92165109). H.Z. is also supported by the Innovation Program for Quantum Science and Technology (Grant No. 2024ZD0300101). Y.Z. is supported by the Innovation Program for Quantum Science and Technology (Grants No. 2024ZD0301900 and No. 2021ZD0302000), the National Natural Science Foundation of China (Grant No. 12205048), the Shanghai Science and Technology Innovation Action Plan (Grant No. 24LZ1400200), and the start-up funding of Fudan University.

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX A: FREQUENTLY USED SYMBOLS

In Table II, we list the frequently used symbols throughout this paper for readers' reference.

APPENDIX B: PRELIMINARIES

For our proofs, it will be useful to have some familiarity with quantum state designs and tensor-network diagrams. A more detailed review can be found in Ref. [84].

1. Quantum state designs

A *Haar-random state* $|\psi\rangle$ in the Hilbert space $\mathcal{H}$ is a pure state drawn randomly according to the Haar measure μ_H on quantum states. For integer $t \geq 1$, the expectation of

TABLE II. Frequently used symbols and their meanings.

Symbol	Meaning
$\mathcal{U}$	the unknown unitary channel we aim to learn
$\Upsilon_{\mathcal{U}}$	the Choi operator of $\mathcal{U}$
$\mathcal{H}$	the n -qubit Hilbert space that $\mathcal{U}$ acts on
d	$d = 2^n$ is the dimension of $\mathcal{H}$
$\mathcal{D}(\mathcal{H})$	the set of all density operators on $\mathcal{H}$
$\text{Obs}(B)$	the set of observables on $\mathcal{H}$ with bounded norms; see Eq. (3)
$\mathcal{A}_{\text{learn}}$	our algorithm working in the learning phase; see Sec. III A
$\mathcal{A}_{\text{pred}}$	our algorithm working in the prediction phase; see Sec. III B
$\hat{X}_i$	a classical snapshot of $\mathcal{U}$ obtained by $\mathcal{A}_{\text{learn}}$; see Eq. (9)
$\mathcal{M}_s$	the symmetric collective measurement used in $\mathcal{A}_{\text{learn}}$; see Eq. (7)
s	the number of copies jointly measured by $\mathcal{M}_s$ in the learning phase
m	the number of snapshots obtained by $\mathcal{A}_{\text{learn}}$
R	the number of batches used in $\mathcal{A}_{\text{pred}}$
q	$q = m/R$ is the number of snapshots in each batch
$\hat{Z}_{(r)}$	the mean estimator constructed from the r th batch; see Eq. (14)
$\hat{E}(O, \rho)$	our protocol's final estimate for $\text{Tr}(O\mathcal{U}(\rho))$; see Eq. (15)
ϵ	the error that we can tolerate
δ	the failure probability that we can tolerate
$\Xi(T, O)$	the mean absolute error for predicting $\text{Tr}(O\mathcal{U}(\rho))$; see Eq. (35)
$\pi_{d,\lambda}$	the λ -induced distribution of density matrices; see Definition 3
H	the unknown Hamiltonian that we aim to learn
$\Pi_{\text{sym}}^{(t)}$	the projector onto the symmetric subspace of $\mathcal{H}^{\otimes t}$
S_t	the permutation group of order t
T_π	the permutation operator on $\mathcal{H}_d^{\otimes t}$ corresponding to $\pi \in S_t$

$(|\psi\rangle\langle\psi|)^{\otimes t}$ with respect to the Haar measure reads

$$\mathbb{E}_{|\psi\rangle\sim\mu_{\text{H}}}[(|\psi\rangle\langle\psi|)^{\otimes t}] = \frac{1}{\kappa_t} \Pi_{\text{sym}}^{(t)}, \quad \kappa_t = \binom{t+d-1}{t}, \quad (\text{B1})$$

where $\Pi_{\text{sym}}^{(t)}$ is the projector onto the symmetric subspace of $\mathcal{H}^{\otimes t}$, and κ_t is the dimension of the symmetric subspace.

An ensemble $\mathcal{E}$ of pure states in $\mathcal{H}$ is said to be a state t -design if and only if its t th moment agrees with that of the Haar distribution, i.e.,

$$\mathbb{E}_{|\psi\rangle\sim\mathcal{E}}[(|\psi\rangle\langle\psi|)^{\otimes t}] = \mathbb{E}_{|\psi\rangle\sim\mu_{\text{H}}}[(|\psi\rangle\langle\psi|)^{\otimes t}] = \frac{1}{\kappa_t} \Pi_{\text{sym}}^{(t)}. \quad (\text{B2})$$

If $\mathcal{E}$ is a state t -design, then it is also a state $(t-1)$ -design by definition.

Let S_t be the permutation group of order t . For any element $\pi \in S_t$, we define its corresponding permutation operator to be the unitary operator T_π on $\mathcal{H}^{\otimes t}$ that satisfies

$$T_\pi(|\varphi_1\rangle \otimes \cdots \otimes |\varphi_t\rangle) = |\varphi_{\pi^{-1}(1)}\rangle \otimes \cdots \otimes |\varphi_{\pi^{-1}(t)}\rangle \quad (\text{B3})$$

for $|\varphi_1\rangle, |\varphi_2\rangle, \dots, |\varphi_t\rangle \in \mathcal{H}$. For example, for $\pi = (1, 2) \in S_2$, the corresponding permutation operator $T_{(1,2)}$ is the swap operator. An important property of the projector $\Pi_{\text{sym}}^{(t)}$ is that it can be decomposed into a sum of permutation operators:

$$\Pi_{\text{sym}}^{(t)} = \frac{1}{t!} \sum_{\pi \in S_t} T_\pi. \quad (\text{B4})$$

2. Tensor-network diagrams

Tensor networks provide a graphical method that simplifies the analysis of tensor operations. In tensor-network diagrams, a linear operator $A \in \mathcal{L}(\mathcal{H})$ on the space $\mathcal{H}$ is represented as a box with open legs:

$$A = \begin{array}{c} \text{---} \\ | \\ \boxed{A} \\ | \\ \text{---} \end{array}, \quad (\text{B5})$$

where the top leg stands for row index of A , and the bottom leg stands for column index of A . The trace of A and the

transpose of A are represented as

$$\text{Tr}(A) = \begin{array}{c} \text{---} \\ | \\ \boxed{A} \\ | \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ | \\ \boxed{A} \\ | \\ \text{---} \end{array} \quad \text{and} \quad A^\top = \begin{array}{c} \text{---} \\ | \\ \boxed{A} \\ | \\ \text{---} \end{array}, \quad (\text{B6})$$

respectively, where “ --- ” and “ --- ” means the periodic boundary condition for the top and bottom legs. In particular, the identity operator on $\mathcal{H}$ and its trace are represented as

$$I = \begin{array}{c} | \\ | \\ \boxed{I} \\ | \\ | \end{array} \quad \text{and} \quad \text{Tr}(I) = \bigcirc = \dim(\mathcal{H}). \quad (\text{B7})$$

Given operators $A, B \in \mathcal{L}(\mathcal{H})$, their product AB is represented as a connection (a contraction of the indices) between two boxes, and their tensor product $A \otimes B$ is represented as putting A and B together with no leg connection:

$$AB = \begin{array}{c} \boxed{A} \\ \boxed{B} \end{array}, \quad A \otimes B = \begin{array}{cc} \boxed{A} & \boxed{B} \end{array}. \quad (\text{B8})$$

An operator $A \in \mathcal{L}(\mathcal{H}^{\otimes 2})$ is represented as a box that has two legs on the top and two on the bottom, and its partial transpose is represented as changing the directions of legs on one side:

$$A = \begin{array}{c} | \\ | \\ \boxed{A} \\ | \\ | \end{array}, \quad A^{\top_2} = \begin{array}{c} | \\ | \\ \boxed{A} \\ | \\ | \end{array}. \quad (\text{B9})$$

In tensor-network diagrams, permutation operators are represented as changing the position of legs. For example, the swap operator is represented as exchanging two legs, and cyclic permutation operators are represented as sequentially moving each leg to its neighboring position:

$$T_{(1,2)} = \begin{array}{c} \diagup \quad \diagdown \\ \diagdown \quad \diagup \end{array}, \quad T_{(1,3,2)} = \begin{array}{c} \diagup \quad \diagdown \quad \diagup \\ \diagdown \quad \diagup \quad \diagdown \end{array}. \quad (\text{B10})$$

A particularly important diagram is that of the unnormalized maximally entangled state:

$$|\Phi\rangle\langle\Phi| = \begin{array}{c} \diagup \quad \diagdown \\ \diagdown \quad \diagup \end{array}, \quad (\text{B11})$$

from which we can easily deduce that $(|\Phi\rangle\langle\Phi|)^{\top_2} = T_{(1,2)}$.

APPENDIX C: PROOF OF LEMMA 1

The expectation of $\hat{\phi} \otimes \hat{\psi}$ (over both random state preparation and measurement outcomes) is

$$\begin{aligned} \mathbb{E}[\hat{\phi} \otimes \hat{\psi}] &= \sum_{i=1}^L \sum_{j=1}^K (\phi_j \otimes \psi_i) \Pr\{\hat{\psi} = \psi_i, \hat{\phi} = \phi_j\} \\ &= \sum_{i=1}^L \sum_{j=1}^K (\phi_j \otimes \psi_i) \frac{1}{L} \text{Tr}[A_{\phi_j} (U \psi_i U^\dagger)^{\otimes s}] \\ &= \frac{1}{L} \sum_{i=1}^L \left(\frac{\kappa_s}{K} \sum_{j=1}^K \phi_j \text{Tr}[\phi_j^{\otimes s} (U \psi_i U^\dagger)^{\otimes s}] \right) \otimes \psi_i. \end{aligned} \quad (\text{C1})$$

According to Lemma C1 below and the fact that $\{|\phi_j\rangle\}_{j=1}^K$ forms a state $(s+1)$ -design (see Sec. III A), the term in the parentheses in the last line can be rewritten as

$$\frac{\kappa_s}{K} \sum_{j=1}^K \phi_j \text{Tr}[\phi_j^{\otimes s} (U \psi_i U^\dagger)^{\otimes s}] = \frac{I + s U \psi_i U^\dagger}{d + s}. \quad (\text{C2})$$

So we have

$$\begin{aligned} \mathbb{E}[\hat{\phi} \otimes \hat{\psi}] &= \frac{1}{L} \sum_{i=1}^L \left(\frac{I + s U \psi_i U^\dagger}{d + s} \right) \otimes \psi_i \\ &= \frac{I}{d + s} \otimes \left(\frac{1}{L} \sum_{i=1}^L \psi_i \right) \\ &\quad + \frac{s}{d + s} (U \otimes I) \left(\frac{1}{L} \sum_{i=1}^L \psi_i \otimes \psi_i \right) (U^\dagger \otimes I) \\ &\stackrel{(a)}{=} \frac{I}{d + s} \otimes \frac{I}{\kappa_1} + \frac{s}{d + s} (U \otimes I) \frac{\Pi_{\text{sym}}^{(2)}}{\kappa_2} (U^\dagger \otimes I) \\ &\stackrel{(b)}{=} \frac{I \otimes I}{\kappa_1(d + s)} + \frac{s}{2\kappa_2(d + s)} \sum_{\pi \in S_2} (U \otimes I) T_\pi (U^\dagger \otimes I) \\ &= \frac{(d + 1 + s)(I \otimes I) + s(U \otimes I) T_{(1,2)} (U^\dagger \otimes I)}{d(d + 1)(d + s)} \\ &\stackrel{(c)}{=} \frac{(d + 1 + s)(I \otimes I) + s(U \otimes U^\dagger) T_{(1,2)}}{d(d + 1)(d + s)}, \end{aligned} \quad (\text{C3})$$

which confirms Lemma 1 in the main text. Here (a) follows from Eq. (4) in the main text and the fact that $\{|\psi_i\rangle\}_{i=1}^L$ is a state 2-design (see Sec. III A), (b) follows from Eq. (B4), and (c) holds because $T_{(1,2)}(U^\dagger \otimes I) = (I \otimes U^\dagger) T_{(1,2)}$.

In this proof, we use at most the $(s+1)$ -design property of the ensemble $\{|\phi_j\rangle\}_{j=1}^K$ and the 2-design property of the ensemble $\{|\psi_i\rangle\}_{i=1}^L$. The $(s+2)$ -design property of $\{|\phi_j\rangle\}_{j=1}^K$ and the 4-design property of $\{|\psi_i\rangle\}_{i=1}^L$ will be used later in the proof of Lemma D1.

The following auxiliary lemma follows directly from Lemma 13 in Ref. [11].

Lemma C1. Suppose integer $s \geq 1$, pure state $|\varphi\rangle \in \mathcal{H}$, and $\{|\phi_j\rangle\}_{j=1}^K$ forms a state $(s+1)$ -design on $\mathcal{H}$. Then

$$\frac{\kappa_s}{K} \sum_j \phi_j \text{Tr}(\phi_j^{\otimes s} \varphi^{\otimes s}) = \frac{I + s\varphi}{d + s}. \quad (\text{C4})$$

APPENDIX D: AUXILIARY LEMMAS

To establish our results in the main text, here we prepare several auxiliary lemmas.

Recall that our algorithm $\mathcal{A}_{\text{learn}}$ working in the learning phase outputs a classical description of the prepared state $\hat{\psi}$ and the measurement outcome $\hat{\phi}$ in each round (see Sec. III A). In the main text, we have clarified the expectation of their tensor product $\hat{\phi} \otimes \hat{\psi}$ in Lemma 1. Here, we calculate the expectation of $\hat{\phi}^{\otimes 2} \otimes \hat{\psi}^{\otimes 2}$, which is useful for proving Lemma 2 in the main text (see Appendix F).

Lemma D1. The expectation of $\hat{\phi}^{\otimes 2} \otimes \hat{\psi}^{\otimes 2}$ (over both random state preparation and measurement outcomes) is

$$\mathbb{E}[\hat{\phi}^{\otimes 2} \otimes \hat{\psi}^{\otimes 2}] = \frac{2}{(d+s)(d+s+1)} \left(\frac{1}{\kappa_2} \Delta_1 + \frac{s}{\kappa_3} \Delta_2 + \frac{s}{\kappa_3} \Delta_3 + \frac{s(s-1)}{2\kappa_4} \Delta_4 \right), \quad (\text{D1})$$

where

$$\Delta_1 := \Pi_{\text{sym}}^{(2)} \otimes \Pi_{\text{sym}}^{(2)}, \quad (\text{D2})$$

$$\Delta_2 := (I \otimes U \otimes I \otimes I) \left[I_1 \otimes \left(\Pi_{\text{sym}}^{(3)} \right)_{2,3,4} \right] (I \otimes U^\dagger \otimes I \otimes I) \left(\Pi_{\text{sym}}^{(2)} \otimes I \otimes I \right), \quad (\text{D3})$$

$$\Delta_3 := (U \otimes I \otimes I \otimes I) \left[I_2 \otimes \left(\Pi_{\text{sym}}^{(3)} \right)_{1,3,4} \right] (U^\dagger \otimes I \otimes I \otimes I) \left(\Pi_{\text{sym}}^{(2)} \otimes I \otimes I \right), \quad (\text{D4})$$

$$\Delta_4 := (U \otimes U \otimes I \otimes I) \Pi_{\text{sym}}^{(4)} (U^\dagger \otimes U^\dagger \otimes I \otimes I), \quad (\text{D5})$$

I_i denotes the identity operator acting on the i th system, and $(\cdot)_{i,j,k}$ denotes the operator acting on the i th, j th, and k th systems.

Proof of Lemma D1. We have

$$\begin{aligned} \mathbb{E}[\hat{\phi}^{\otimes 2} \otimes \hat{\psi}^{\otimes 2}] &= \sum_{i=1}^L \sum_{j=1}^K \left(\phi_j^{\otimes 2} \otimes \psi_i^{\otimes 2} \right) \Pr\{\hat{\psi} = \psi_i, \hat{\phi} = \phi_j\} \\ &= \sum_{i=1}^L \sum_{j=1}^K \left(\phi_j^{\otimes 2} \otimes \psi_i^{\otimes 2} \right) \frac{1}{L} \text{Tr}[A_{\phi_j} (U \psi_i U^\dagger)^{\otimes s}] \\ &= \frac{1}{L} \sum_{i=1}^L \left(\frac{\kappa_s}{K} \sum_{j=1}^K \phi_j^{\otimes 2} \text{Tr}[\phi_j^{\otimes s} (U \psi_i U^\dagger)^{\otimes s}] \right) \otimes \psi_i^{\otimes 2}. \end{aligned} \quad (\text{D6})$$

According to Lemma D2 below and the fact that $\{|\phi_j\rangle\}_{j=1}^K$ forms a state $(s+2)$ -design (see Sec. III A),

$$\frac{\kappa_s}{K} \sum_{j=1}^K \phi_j^{\otimes 2} \text{Tr}[\phi_j^{\otimes s} (U \psi_i U^\dagger)^{\otimes s}] = \frac{2}{(d+s)(d+s+1)} \left[(I + s U \psi_i U^\dagger)^{\otimes 2} - \frac{s(s+1)}{2} (U \psi_i U^\dagger)^{\otimes 2} \right] \Pi_{\text{sym}}^{(2)}. \quad (\text{D7})$$

So we have

$$\begin{aligned} \mathbb{E}[\hat{\phi}^{\otimes 2} \otimes \hat{\psi}^{\otimes 2}] &= \frac{2}{(d+s)(d+s+1)} \cdot \frac{1}{L} \sum_{i=1}^L \left[(I + s U \psi_i U^\dagger)^{\otimes 2} - \frac{s(s+1)}{2} (U \psi_i U^\dagger)^{\otimes 2} \right] \Pi_{\text{sym}}^{(2)} \otimes \psi_i^{\otimes 2} \\ &= \frac{2}{(d+s)(d+s+1)} \cdot \frac{1}{L} \sum_{i=1}^L \left[I^{\otimes 2} + s I \otimes (U \psi_i U^\dagger) + s (U \psi_i U^\dagger) \otimes I + \frac{s(s-1)}{2} (U \psi_i U^\dagger)^{\otimes 2} \right] \Pi_{\text{sym}}^{(2)} \otimes \psi_i^{\otimes 2}. \end{aligned} \quad (\text{D8})$$

Since $\{|\psi_i\rangle\}_{i=1}^L$ is a state 4-design (see Sec. III A), we can apply Eq. (4) in the main text to calculate the sum in Eq. (D8) term-by-term. The first term is calculated as

$$\frac{1}{L} \sum_{i=1}^L \left(I^{\otimes 2} \Pi_{\text{sym}}^{(2)} \right) \otimes \psi_i^{\otimes 2} = \frac{1}{\kappa_2} \Pi_{\text{sym}}^{(2)} \otimes \Pi_{\text{sym}}^{(2)} = \frac{\Delta_1}{\kappa_2}, \quad (\text{D9})$$

where we have used the 2-design property of $\{|\psi_i\rangle\}_{i=1}^L$. The second term is calculated as

$$\begin{aligned} \frac{1}{L} \sum_{i=1}^L [sI \otimes (U\psi_i U^\dagger)] \Pi_{\text{sym}}^{(2)} \otimes \psi_i^{\otimes 2} &= \frac{s}{L} \sum_{i=1}^L (I \otimes U \otimes I \otimes I) (I \otimes \psi_i^{\otimes 3}) (I \otimes U^\dagger \otimes I \otimes I) (\Pi_{\text{sym}}^{(2)} \otimes I \otimes I) \\ &= s(I \otimes U \otimes I \otimes I) \left(I \otimes \frac{\Pi_{\text{sym}}^{(3)}}{\kappa_3} \right) (I \otimes U^\dagger \otimes I \otimes I) (\Pi_{\text{sym}}^{(2)} \otimes I \otimes I) = \frac{s}{\kappa_3} \Delta_2, \end{aligned} \quad (\text{D10})$$

where we have used the 3-design property of $\{|\psi_i\rangle\}_{i=1}^L$. By a similar reasoning, the third term is given by

$$\frac{1}{L} \sum_{i=1}^L [s(U\psi_i U^\dagger) \otimes I] \Pi_{\text{sym}}^{(2)} \otimes \psi_i^{\otimes 2} = \frac{s}{\kappa_3} \Delta_3. \quad (\text{D11})$$

The last term is calculated as

$$\begin{aligned} \frac{1}{L} \sum_{i=1}^L \left[\frac{s(s-1)}{2} (U\psi_i U^\dagger)^{\otimes 2} \Pi_{\text{sym}}^{(2)} \right] \otimes \psi_i^{\otimes 2} &\stackrel{(a)}{=} \frac{s(s-1)}{2} \cdot \frac{1}{L} \sum_{i=1}^L (U\psi_i U^\dagger)^{\otimes 2} \otimes \psi_i^{\otimes 2} \\ &= \frac{s(s-1)}{2} \cdot \frac{1}{L} \sum_{i=1}^L (U \otimes U \otimes I \otimes I) (\psi_i^{\otimes 4}) (U^\dagger \otimes U^\dagger \otimes I \otimes I) \\ &\stackrel{(b)}{=} \frac{s(s-1)}{2} (U \otimes U \otimes I \otimes I) \frac{\Pi_{\text{sym}}^{(4)}}{\kappa_4} (U^\dagger \otimes U^\dagger \otimes I \otimes I) = \frac{s(s-1)}{2\kappa_4} \Delta_4, \end{aligned} \quad (\text{D12})$$

where (a) holds because $(U\psi_i U^\dagger)^{\otimes 2} \Pi_{\text{sym}}^{(2)} = (U\psi_i U^\dagger)^{\otimes 2}$, and (b) holds because $\{|\psi_i\rangle\}_{i=1}^L$ is a 4-design. Equations (D8)–(D12) together confirm Lemma D1. $\blacksquare$

Note that, in the above proof, the 4-design property of $\{|\psi_i\rangle\}_{i=1}^L$ is only utilized in the case of $s \geq 2$. When $s = 1$, the last term $[s(s-1)/2] (U\psi_i U^\dagger)^{\otimes 2}$ in the bracket of Eq. (D8) vanishes and thus the 3-design property of $\{|\psi_i\rangle\}_{i=1}^L$ is sufficient to prove Lemma D1.

The following auxiliary lemma follows directly from Lemma 14 in Ref. [11].

Lemma D2. Suppose integer $s \geq 1$, pure state $|\varphi\rangle \in \mathcal{H}$, and $\{|\phi_j\rangle\}_{j=1}^K$ forms a state $(s+2)$ -design on $\mathcal{H}$. Then

$$\frac{\kappa_s}{K} \sum_j \phi_j^{\otimes 2} \text{Tr}(\phi_j^{\otimes s} \varphi^{\otimes s}) = \frac{2}{(d+s)(d+s+1)} \left[(I + s\varphi)^{\otimes 2} - \frac{s(s+1)}{2} \varphi^{\otimes 2} \right] \Pi_{\text{sym}}^{(2)}. \quad (\text{D13})$$

Lemma D3 (Hölder's inequality). Suppose A, B are operators on $\mathcal{H}$. Then

$$|\text{Tr}(AB)| \leq \|A\|_1 \|B\|, \quad |\text{Tr}(AB)| \leq \|A\|_2 \|B\|_2. \quad (\text{D14})$$

Lemma D4. Suppose U is a unitary operator on $\mathcal{H}$, $\rho \in \mathcal{D}(\mathcal{H})$, and $O \in \text{Obs}(B)$. Let $O_U := U^\dagger O U$ and $\wp := \text{Tr}(\rho^2)$. Then

$$|\text{Tr}(O)| \leq \sqrt{dB}. \quad (\text{D15})$$

$$|\text{Tr}(O_U \rho)| \leq \min\{\sqrt{B\wp}, 1\}, \quad (\text{D16})$$

$$|\text{Tr}(O_U^2 \rho)| \leq \min\{\sqrt{B\wp}, 1\}, \quad (\text{D17})$$

$$|\text{Tr}(O_U \rho^2)| \leq \wp, \quad (\text{D18})$$

$$|\text{Tr}(O_U^2 \rho^2)| \leq \wp, \quad (\text{D19})$$

$$|\text{Tr}(O_U \rho O_U \rho)| \leq \wp. \quad (\text{D20})$$

Proof of Lemma D4. Equation (D15) can be derived as

$$|\text{Tr}(O)| \leq \|I\|_2 \|O\|_2 = \sqrt{d \text{Tr}(O^2)} \leq \sqrt{dB}, \quad (\text{D21})$$

where the first inequality follows from Lemma D3, and the second inequality holds because $O \in \text{Obs}(B)$.

Similarly, Eq. (D16) can be derived as

$$\begin{aligned} |\text{Tr}(O_U \rho)| &\leq \|O_U\|_2 \|\rho\|_2 = \|O\|_2 \|\rho\|_2 \\ &= \sqrt{\text{Tr}(O^2) \text{Tr}(\rho^2)} \leq \sqrt{B\wp}, \end{aligned} \quad (\text{D22})$$

$$|\text{Tr}(O_U \rho)| \leq \|O_U\| \|\rho\|_1 = \|O\| \text{Tr}(\rho) \leq 1, \quad (\text{D23})$$

given that $\|O\| \leq 1$ when $O \in \text{Obs}(B)$.

Equation (D17) follows from Eq. (D16) and the fact that $O^2 \in \text{Obs}(B)$ when $O \in \text{Obs}(B)$.

Equation (D18) can be derived as

$$|\text{Tr}(O_U \rho^2)| \leq \|O_U\| \|\rho^2\|_1 = \|O\| \text{Tr}(\rho^2) \leq \wp. \quad (\text{D24})$$

Equation (D19) follows from Eq. (D18) and the fact that $O^2 \in \text{Obs}(B)$ when $O \in \text{Obs}(B)$.

Equation (D20) can be derived as

$$\begin{aligned} |\text{Tr}(O_U \rho O_U \rho)| &\leq \|O_U \rho\|_2^2 = \text{Tr}[(O_U \rho)^\dagger (O_U \rho)] \\ &= \text{Tr}(O_U^2 \rho^2) \leq \wp, \end{aligned} \quad (\text{D25})$$

where the last inequality follows from Eq. (D19). $\blacksquare$

APPENDIX E: PREDICTING $\text{Tr}(OU(\rho))$ BY AVERAGING $\text{Tr}[(O \otimes \rho^\top) \hat{X}_i]$

Recall that our protocol for CSEU outputs the classical shadow data $\text{UShadow}(\mathcal{U}, m) = \{\hat{X}_1, \hat{X}_2, \dots, \hat{X}_m\}$ in the learning phase, which contains m independent classical snapshots. Given any quantum state $\rho \in \mathcal{D}(\mathcal{H})$ and observable $O \in \text{Obs}(B)$ in the prediction phase, as mentioned in Sec. III B of the main text, a simple way to predict

$\text{Tr}(OU(\rho))$ is to calculate $\text{Tr}[(O \otimes \rho^\top) \hat{X}_i]$ and take their average:

$$\tilde{Z}_m(O, \rho) := \frac{1}{m} \sum_{i=1}^m \text{Tr}[(O \otimes \rho^\top) \hat{X}_i]. \quad (\text{E1})$$

In this Appendix we shall show that, with this method, the number of queries required to estimate $\text{Tr}(OU(\rho))$ within additive error ϵ scales as $\mathcal{O}(d^2 B / \epsilon^2)$ even when collective measurements on multiple systems are used. If only single-copy measurements are accessible, then the query complexity further increases, reaching up to $\mathcal{O}(d^3 B / \epsilon^2)$.

The fluctuations of $\tilde{Z}_m(O, \rho)$ around its expectation $\text{Tr}(OU(\rho))$ are controlled by the variance. By Lemma E1 below, we have

$$\begin{aligned} \text{Var}[\tilde{Z}_m(O, \rho)] &= \frac{1}{m} \text{Var}[\text{Tr}((O \otimes \rho^\top) \hat{X})] \\ &\leq \frac{1}{m} \Theta\left[\left(1 + \frac{d^2}{s^2}\right) dB\right], \end{aligned} \quad (\text{E2})$$

where the bound can be saturated when ρ is a pure state and $O = \sqrt{B/d} I \in \text{Obs}(B)$.

By Chebyshev's inequality, in order to suppress the additive error of $\tilde{Z}_m(O, \rho)$ to ϵ with some constant success probability $1 - \delta$, it suffices to choose m such that $\text{Var}[\tilde{Z}_m(O, \rho)] \leq \delta \epsilon^2$; that is, to choose

$$m \geq C \left(1 + \frac{d^2}{s^2}\right) dB \epsilon^{-2} \quad (\text{E3})$$

for some constant $C > 0$. So the total number of required queries reads

$$s \cdot m = C \left(1 + \frac{d^2}{s^2}\right) \frac{s dB}{\epsilon^2}. \quad (\text{E4})$$

In the special case $s = 1$, i.e., only single-copy measurements are used, we have $s \cdot m = \mathcal{O}(d^3 B / \epsilon^2)$. This query complexity agrees with that of the protocol for CSEC developed in Ref. [21], which is ancilla-free and does not use quantum memory (see Sec. VI in the main text). If collective measurements on multiple systems are accessible, the query number can be further reduced by increasing s . When $s = \Theta(d)$, it can be reduced to the greatest extent, in which case the query complexity reads $s \cdot m = \mathcal{O}(d^2 B / \epsilon^2)$.

In contrast to the above method, our method for estimating $\text{Tr}(OU(\rho))$ in the main text is based on the quadratic estimator $d^{-1} \text{Tr}[(O \otimes \rho^\top) \hat{X}_i \hat{X}_j]$ instead of $\text{Tr}[(O \otimes \rho^\top) \hat{X}_i]$. As shown in Table III below, in both single-copy and collective measurement scenarios, the quadratic estimator demonstrates strictly better query efficiency.

TABLE III. The query complexity of two different methods for estimating $\text{Tr}(OU(\rho))$.

Method	Single-copy measurements with $s = 1$	Collective measurements with $s = \Theta(d)$
Averaging $\text{Tr}[(O \otimes \rho^\top) \hat{X}_i]$	$\mathcal{O}(d^3 B \epsilon^{-2})$	$\mathcal{O}(d^2 B \epsilon^{-2})$
Averaging $d^{-1} \text{Tr}[(O \otimes \rho^\top) \hat{X}_i \hat{X}_j]$	$\mathcal{O}(d \epsilon^{-2} + d^2 \sqrt{B} \epsilon^{-1})$ [see Eq. (28)]	$\mathcal{O}(d(\epsilon^{-2} + \sqrt{B} \epsilon^{-1}))$ [see Eq. (29)]

Lemma E1. Suppose $\rho \in \mathcal{D}(\mathcal{H})$ and $O \in \text{Obs}(B)$. Then

$$\text{Var}\left[\text{Tr}\left((O \otimes \rho^\top) \hat{X}\right)\right] \leq \Theta\left[\left(1 + \frac{d^2}{s^2}\right) dB\right], \quad (\text{E5})$$

where the inequality can be saturated when ρ is a pure state and $O = \sqrt{d^{-1}B}I$.

Note that when the observable O is proportional to the identity, i.e., $O = cI$, the single-shot estimator

$$\begin{aligned} \text{Tr}\left[(O \otimes \rho^\top) \hat{X}\right] &= \text{Tr}\left[(cI \otimes \rho^\top) \frac{d(d+1)(d+s)(\hat{\phi} \otimes \hat{\psi}^\top) - (d+1+s)(I \otimes I)}{s}\right] \\ &= \frac{c d(d+1)(d+s)}{s} \text{Tr}(\rho \hat{\psi}) - \frac{c d(d+1+s)}{s} \end{aligned} \quad (\text{E6})$$

is not a constant, and has fluctuation due to the randomness of $\hat{\psi}$. Hence, the variance of $\text{Tr}[(O \otimes \rho^\top) \hat{X}]$ does not vanish in this case.

Proof of Lemma E1. Using the definition of $\hat{X}$ in Eq. (9) of the main text, we can expand the variance as

$$\begin{aligned} \text{Var}\left[\text{Tr}\left((O \otimes \rho^\top) \hat{X}\right)\right] &= \text{Var}\left[\frac{d(d+1)(d+s)}{s} \text{Tr}\left[(O \otimes \rho^\top)(\hat{\phi} \otimes \hat{\psi}^\top)\right] - \frac{d+1+s}{s} \text{Tr}(O \otimes \rho^\top)\right] \\ &= \left[\frac{d(d+1)(d+s)}{s}\right]^2 \text{Var}\left[\text{Tr}\left((O \otimes \rho)(\hat{\phi} \otimes \hat{\psi})\right)\right] \\ &= \left[\frac{d(d+1)(d+s)}{s}\right]^2 \left\{ \mathbb{E}\left[\left(\text{Tr}[(O \otimes \rho)(\hat{\phi} \otimes \hat{\psi})]\right)^2\right] - \left(\mathbb{E}\left[\text{Tr}[(O \otimes \rho)(\hat{\phi} \otimes \hat{\psi})]\right)\right]^2 \right\} \\ &\leq \left[\frac{d(d+1)(d+s)}{s}\right]^2 \text{Tr}\left[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \mathbb{E}[\hat{\phi}^{\otimes 2} \otimes \hat{\psi}^{\otimes 2}]\right]. \end{aligned} \quad (\text{E7})$$

According to Lemma D1, the trace term in the last line of Eq. (E7) can be expanded as

$$\text{Tr}\left[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \mathbb{E}[\hat{\phi}^{\otimes 2} \otimes \hat{\psi}^{\otimes 2}]\right] = \frac{2}{(d+s)(d+s+1)} \text{Tr}\left[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \left(\frac{\Delta_1}{\kappa_2} + \frac{s\Delta_2}{\kappa_3} + \frac{s\Delta_3}{\kappa_3} + \frac{s(s-1)\Delta_4}{2\kappa_4}\right)\right]. \quad (\text{E8})$$

Next, we calculate $\text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \Delta_j]$ for $j = 1, 2, 3, 4$, respectively. First,

$$\begin{aligned} \text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \Delta_1] &= \text{Tr}(O^{\otimes 2} \Pi_{\text{sym}}^{(2)}) \text{Tr}(\rho^{\otimes 2} \Pi_{\text{sym}}^{(2)}) = \frac{1}{4} \text{Tr}[O^{\otimes 2} (I + T_{(12)})] \text{Tr}[\rho^{\otimes 2} (I + T_{(12)})] \\ &= \frac{1}{4} [\text{Tr}(O^2) + (\text{Tr}(O))^2] [\text{Tr}(\rho^2) + 1] = \mathcal{O}(Bd), \end{aligned} \quad (\text{E9})$$

where the last equality follows from Lemma D4 in Appendix D. Second, $\text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \Delta_2]$ can be calculated via tensor-network diagrams:

$$\begin{aligned}
 \text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \Delta_2] &= \text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2})(I \otimes U \otimes I \otimes I)(I \otimes \Pi_{\text{sym}}^{(3)})(I \otimes U^\dagger \otimes I \otimes I)(\Pi_{\text{sym}}^{(2)} \otimes I \otimes I)] \\
 &= \text{Diagram 1} = \frac{1}{3! \cdot 2!} \sum_{\pi_1 \in S_3} \sum_{\pi_2 \in S_2} \text{Diagram 2} \\
 &= \frac{1}{12} \text{Tr}(O) [\text{Tr}(O) + \text{Tr}(O) \text{Tr}(\rho^2) + 2 \text{Tr}(OU\rho U^\dagger) + 2 \text{Tr}(OU\rho^2 U^\dagger)] \\
 &\quad + \frac{1}{12} [\text{Tr}(O^2) + \text{Tr}(O^2) \text{Tr}(\rho^2) + 2 \text{Tr}(O^2 U \rho U^\dagger) + 2 \text{Tr}(O^2 U \rho^2 U^\dagger)] \\
 &= \mathcal{O}(Bd),
 \end{aligned} \tag{E10}$$

where we use Lemma D4 to derive the last equality. Third, from the symmetry between Δ_2 and Δ_3 we have

$$\text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \Delta_3] = \text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \Delta_2] = \mathcal{O}(Bd). \tag{E11}$$

Fourth,

$$\begin{aligned}
 \text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \Delta_4] &= \text{Tr}[(U^\dagger \otimes U^\dagger \otimes I \otimes I)(O^{\otimes 2} \otimes \rho^{\otimes 2})(U \otimes U \otimes I \otimes I) \Pi_{\text{sym}}^{(4)}] \\
 &= \frac{1}{4!} \sum_{\pi \in S_4} \text{Tr}[(O_U^{\otimes 2} \otimes \rho^{\otimes 2}) T_\pi] \\
 &= \frac{1}{24} [(\text{Tr}(O_U))^2 + \text{Tr}(O_U^2) + 4 \text{Tr}(O_U) \text{Tr}(O_U \rho) + (\text{Tr}(O_U))^2 \text{Tr}(\rho^2) + 4 \text{Tr}(O_U) \text{Tr}(O_U \rho^2) \\
 &\quad + 4 \text{Tr}(O_U^2 \rho) + \text{Tr}(O_U^2) \text{Tr}(\rho^2) + 2(\text{Tr}(O_U \rho))^2 + 4 \text{Tr}(O_U^2 \rho^2) + 2 \text{Tr}(O_U \rho O_U \rho)] \\
 &= \mathcal{O}(Bd),
 \end{aligned} \tag{E12}$$

where $O_U = U^\dagger O U$, and the last equality follows from Lemma D4 again.

Therefore, Eq. (E8) can be rewritten as

$$\text{Tr}[(O^{\otimes 2} \otimes \rho^{\otimes 2}) \mathbb{E}[\hat{\phi}^{\otimes 2} \otimes \hat{\psi}^{\otimes 2}]] = \frac{2}{(d+s)(d+s+1)} \left(\frac{1}{\kappa_2} + \frac{2s}{\kappa_3} + \frac{s(s-1)}{2\kappa_4} \right) \mathcal{O}(dB). \tag{E13}$$

By plugging Eq. (E13) into Eq. (E7), we have

$$\text{Var}[\text{Tr}((O \otimes \rho^\top) \hat{X})] \leq \left[\frac{d(d+1)(d+s)}{s} \right]^2 \frac{\mathcal{O}(dB)}{(d+s)(d+s+1)} \left(\frac{1}{\kappa_2} + \frac{2s}{\kappa_3} + \frac{s(s-1)}{2\kappa_4} \right) = \mathcal{O} \left[\left(1 + \frac{d^2}{s^2} \right) dB \right], \tag{E14}$$

which confirms the inequality Eq. (E5).

At last, we show that the inequality Eq. (E5) can be saturated when ρ is a pure state and $O = \sqrt{B/d}I \in \text{Obs}(B)$. In this case,

$$\begin{aligned}
\text{Var}\left[\text{Tr}\left((O \otimes \rho^\top)\hat{X}\right)\right] &= \text{Var}\left[\frac{d(d+1)(d+s)}{s} \text{Tr}\left[(O \otimes \rho^\top)(\hat{\phi} \otimes \hat{\psi}^\top)\right] - \frac{d+1+s}{s} \text{Tr}(O \otimes \rho^\top)\right] \\
&= \text{Var}\left[\frac{d(d+1)(d+s)}{s} \sqrt{\frac{B}{d}} \text{Tr}(\rho \hat{\psi})\right] \\
&= \frac{B}{d} \left[\frac{d(d+1)(d+s)}{s}\right]^2 \left[\text{Tr}\left((\rho \otimes \rho)\mathbb{E}[\hat{\psi} \otimes \hat{\psi}]\right) - \left(\text{Tr}(\rho \mathbb{E}[\hat{\psi}])\right)^2\right] \\
&\stackrel{(a)}{=} \frac{B}{d} \left[\frac{d(d+1)(d+s)}{s}\right]^2 \left[\text{Tr}\left((\rho \otimes \rho) \frac{\Pi_{\text{sym}}^{(2)}}{\kappa_2}\right) - \left(\text{Tr}(\rho I / \kappa_1)\right)^2\right] \\
&= \frac{B}{d} \left[\frac{d(d+1)(d+s)}{s}\right]^2 \left[\frac{2}{d(d+1)} - \frac{1}{d^2}\right] \\
&= \Theta\left[\left(1 + \frac{d^2}{s^2}\right)dB\right], \tag{E15}
\end{aligned}$$

where (a) follows from Eq. (4) in the main text and the fact that the state $\hat{\psi}$ is chosen randomly from a 4-design state ensemble. This completes the proof of Lemma E1. $\blacksquare$

APPENDIX F: PROOF OF LEMMA 2

In preparation for proving Lemma 2, it will be convenient to first calculate the second moment of the snapshot $\hat{X}$, which is defined in Eq. (9) of the main text.

Lemma F1. The expectation of $\hat{X} \otimes \hat{X}$ (over both random state preparation and measurement outcomes) is

$$\mathbb{E}[\hat{X} \otimes \hat{X}] = \frac{2d^2(d+1)^2(d+s)}{s^2(d+s+1)} \left(\frac{1}{\kappa_2} \tilde{\Delta}_1 + \frac{s}{\kappa_3} \tilde{\Delta}_2 + \frac{s}{\kappa_3} \tilde{\Delta}_3 + \frac{s(s-1)}{2\kappa_4} \tilde{\Delta}_4 \right) - \frac{d+1+s}{s} (\tilde{\Delta}_5 + \tilde{\Delta}_6) - \frac{(d+1+s)^2}{s^2} \tilde{\Delta}_7, \tag{F1}$$

where

$$\tilde{\Delta}_1 := \left(\Pi_{\text{sym}}^{(2)}\right)_{1,3} \otimes \left(\Pi_{\text{sym}}^{(2)}\right)_{2,4}, \tag{F2}$$

$$\tilde{\Delta}_2 := \left((I \otimes I \otimes U \otimes I) \left[I_1 \otimes \left(\Pi_{\text{sym}}^{(3)}\right)_{2,3,4} \right] (I \otimes I \otimes U^\dagger \otimes I) \left[\left(\Pi_{\text{sym}}^{(2)}\right)_{1,3} \otimes I_2 \otimes I_4 \right] \right)^{\top_{2,4}}, \tag{F3}$$

$$\tilde{\Delta}_3 := \left((U \otimes I \otimes I \otimes I) \left[I_3 \otimes \left(\Pi_{\text{sym}}^{(3)}\right)_{1,2,4} \right] (U^\dagger \otimes I \otimes I \otimes I) \left[\left(\Pi_{\text{sym}}^{(2)}\right)_{1,3} \otimes I_2 \otimes I_4 \right] \right)^{\top_{2,4}}, \tag{F4}$$

$$\tilde{\Delta}_4 := \left((U \otimes I \otimes U \otimes I) \Pi_{\text{sym}}^{(4)} (U^\dagger \otimes I \otimes U^\dagger \otimes I) \right)^{\top_{2,4}}, \tag{F5}$$

$$\tilde{\Delta}_5 := \Upsilon_U \otimes I \otimes I, \quad \tilde{\Delta}_6 := I \otimes I \otimes \Upsilon_U, \quad \tilde{\Delta}_7 := I \otimes I \otimes I \otimes I, \tag{F6}$$

and $\top_{2,4}$ denotes the partial transpose on the second and fourth subsystems.

Proof of Lemma F1. We have

$$\begin{aligned}
\mathbb{E}[\hat{X} \otimes \hat{X}] &= \mathbb{E} \left[\left(\frac{d(d+1)(d+s)(\hat{\phi} \otimes \hat{\psi}^\top) - (d+1+s)(I \otimes I)}{s} \right)^{\otimes 2} \right] \\
&= \frac{d^2(d+1)^2(d+s)^2}{s^2} \mathbb{E} [\hat{\phi} \otimes \hat{\psi} \otimes \hat{\phi} \otimes \hat{\psi}]^{\top_{2,4}} - \frac{d(d+1)(d+s)(d+1+s)}{s^2} \mathbb{E} [\hat{\phi} \otimes \hat{\psi} \otimes I \otimes I]^{\top_2} \\
&\quad - \frac{d(d+1)(d+s)(d+1+s)}{s^2} \mathbb{E} [I \otimes I \otimes \hat{\phi} \otimes \hat{\psi}]^{\top_4} + \frac{(d+1+s)^2}{s^2} (I \otimes I \otimes I \otimes I) \\
&\stackrel{(a)}{=} \frac{2d^2(d+1)^2(d+s)}{s^2(d+s+1)} \left(\frac{1}{\kappa_2} \tilde{\Delta}_1 + \frac{s}{\kappa_3} \tilde{\Delta}_2 + \frac{s}{\kappa_3} \tilde{\Delta}_3 + \frac{s(s-1)}{2\kappa_4} \tilde{\Delta}_4 \right) + \frac{(d+1+s)^2}{s^2} (I \otimes I \otimes I \otimes I) \\
&\quad - \frac{d(d+1)(d+s)(d+1+s)}{s^2} \left[\frac{(d+1+s)(I \otimes I) + s\Upsilon_{\mathcal{U}}}{d(d+1)(d+s)} \otimes I \otimes I \right] \\
&\quad - \frac{d(d+1)(d+s)(d+1+s)}{s^2} \left[I \otimes I \otimes \frac{(d+1+s)(I \otimes I) + s\Upsilon_{\mathcal{U}}}{d(d+1)(d+s)} \right] \\
&= \frac{2d^2(d+1)^2(d+s)}{s^2(d+s+1)} \left(\frac{1}{\kappa_2} \tilde{\Delta}_1 + \frac{s}{\kappa_3} \tilde{\Delta}_2 + \frac{s}{\kappa_3} \tilde{\Delta}_3 + \frac{s(s-1)}{2\kappa_4} \tilde{\Delta}_4 \right) \\
&\quad - \frac{d+1+s}{s} (I \otimes I \otimes \Upsilon_{\mathcal{U}} + \Upsilon_{\mathcal{U}} \otimes I \otimes I) - \frac{(d+1+s)^2}{s^2} (I \otimes I \otimes I \otimes I), \tag{F7}
\end{aligned}$$

which confirms Lemma F1. Here the equality (a) follows from Lemma 1 in the main text, Lemma D1 in Appendix D, and the relation $[(U \otimes U^\dagger)T_{(1,2)}]^{\top_2} = \Upsilon_{\mathcal{U}}$. ■

Now we are ready to bound the covariance term $\text{Cov}(\Lambda_{ij}, \Lambda_{k\ell})$ for each combination of the four indices (i, j, k, ℓ) with $i \neq j$ and $k \neq \ell$. In the rest of this appendix, Appendix F1 deals with the case in which exactly one index matches in different positions ($i = \ell$ or $j = k$); Appendix F2 deals with the case in which exactly one index matches in the same position ($i = k$ or $j = \ell$); Appendix F3 deals with the case in which both indices match with order swapped ($i = \ell$ and $j = k$); and Appendix F4 deals with the case in which both indices match with the same order ($i = k$ and $j = \ell$). Lemmas F2–F5 in the following subsections together will confirm Lemma 2 in the main text.

1. Exactly one index matches in different positions

Lemma F2. For all distinct i, j, k , we have $\text{Cov}(\Lambda_{ij}, \Lambda_{j,k}) = \mathcal{O}(d^2 \min\{1, B_{\mathcal{D}}\})$, where $\mathcal{D} = \text{Tr}(\rho^2)$.

Proof of Lemma F2. First, we have

$$\begin{aligned}
\text{Cov}(\Lambda_{ij}, \Lambda_{j,k}) &= \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top) \hat{X}_i \hat{X}_j) \text{Tr}((O \otimes \rho^\top) \hat{X}_j \hat{X}_k)^* \right] - \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top) \hat{X}_i \hat{X}_j) \right] \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top) \hat{X}_j \hat{X}_k)^* \right] \\
&\stackrel{(a)}{=} \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top) \hat{X}_i \hat{X}_j) \text{Tr}((O \otimes \rho^\top) \hat{X}_j \hat{X}_k) \right] - \text{Tr}[(O \otimes \rho^\top) d \Upsilon_{\mathcal{U}}] \text{Tr}[(O \otimes \rho^\top) d \Upsilon_{\mathcal{U}}]^* \\
&= \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\hat{X}_i \hat{X}_j \otimes \hat{X}_j \hat{X}_k)) \right] - d^2 [\text{Tr}(O \mathcal{U}(\rho))]^2, \tag{F8}
\end{aligned}$$

where (a) holds because $O \otimes \rho^\top, \hat{X}_j, \hat{X}_k$ are Hermitian, and

$$\mathbb{E}[\hat{X}_i \hat{X}_j] = \mathbb{E}[\hat{X}_j \hat{X}_k] = \mathbb{E}[\hat{X}]^2 = \Upsilon_{\mathcal{U}}^2 = d \Upsilon_{\mathcal{U}}. \tag{F9}$$

The second moment term in the last line of Eq. (F8) can be further calculated as

$$\begin{aligned}
\mathbb{E} \left[\text{Tr}((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\hat{X}_i \hat{X}_j \otimes \hat{X}_j \hat{X}_k)) \right] &= \text{Tr}((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\mathbb{E}[\hat{X}_i] \otimes \mathbb{E}[\hat{X}_k]) \mathbb{E}[\hat{X}_j \otimes \hat{X}_j]) \\
&= \text{Tr}((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_{\mathcal{U}} \otimes \Upsilon_{\mathcal{U}}) \mathbb{E}[\hat{X} \otimes \hat{X}]). \tag{F10}
\end{aligned}$$

The expression of $\mathbb{E}[\hat{X} \otimes \hat{X}]$ in Lemma F1 has 7 terms. Plugging them into Eq. (F10), respectively, we get

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \tilde{\Delta}_5\right] = d [\text{Tr}(OU(\rho))]^2 \geq 0, \quad (\text{F11})$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \tilde{\Delta}_6\right] = d [\text{Tr}(OU(\rho))]^2 \geq 0, \quad (\text{F12})$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \tilde{\Delta}_7\right] = [\text{Tr}(OU(\rho))]^2 \geq 0; \quad (\text{F13})$$

and

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \tilde{\Delta}_1\right] = \mathcal{O}(\min\{1, B\varphi\}), \quad (\text{F14})$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \tilde{\Delta}_2\right] = \mathcal{O}(d \min\{1, B\varphi\}), \quad (\text{F15})$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \tilde{\Delta}_3\right] = \mathcal{O}(d \min\{1, B\varphi\}), \quad (\text{F16})$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \tilde{\Delta}_4\right] = \mathcal{O}(d^2 \min\{1, B\varphi\}). \quad (\text{F17})$$

Here Eqs. (F14)–(F17) can be derived by using tensor-network diagrams. For example,

$$\begin{aligned} & \text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \tilde{\Delta}_2\right] \\ &= \text{Diagram 1} = \frac{1}{3! \cdot 2!} \sum_{\pi_1 \in S_3} \sum_{\pi_2 \in S_2} \text{Diagram 2} \\ &= \frac{1}{12} [4(\text{Tr}(OU\rho U^\dagger))^2 + 2d(\text{Tr}(OU\rho U^\dagger))^2 + 4\text{Tr}(OU\rho U^\dagger OU\rho U^\dagger) + 2d\text{Tr}(OU\rho U^\dagger OU\rho U^\dagger)] \\ &= \mathcal{O}(d \min\{1, B\varphi\}), \end{aligned} \quad (\text{F18})$$

The diagrammatic representation shows the contraction of indices for the trace expression. The left side (Diagram 1) shows the full expression with indices $\Upsilon_U \otimes \Upsilon_U$ and $\tilde{\Delta}_2$ highlighted. The right side (Diagram 2) shows the expression after applying the permutation operators T_{π_1} and T_{π_2} .

where the last equality follows from Lemma D4.

Therefore, by Eqs. (F8) and (F10),

$$\begin{aligned} \text{Cov}(\Lambda_{i,j}, \Lambda_{j,k}) &\leq \text{Tr}\left((O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \mathbb{E}[\hat{X} \otimes \hat{X}]\right) \\ &\stackrel{(a)}{\leq} \text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes \Upsilon_U) \frac{2d^2(d+1)^2(d+s)}{s^2(d+s+1)} \left(\frac{1}{\kappa_2} \tilde{\Delta}_1 + \frac{s}{\kappa_3} \tilde{\Delta}_2 + \frac{s}{\kappa_3} \tilde{\Delta}_3 + \frac{s(s-1)}{2\kappa_4} \tilde{\Delta}_4\right)\right] \end{aligned}$$

$$\begin{aligned}
&\stackrel{(b)}{=} \mathcal{O}\left[\frac{d^4}{s^2} \left(\frac{\min\{1, B_{\wp}\}}{\kappa_2} + \frac{sd \min\{1, B_{\wp}\}}{\kappa_3} + \frac{s(s-1)d^2 \min\{1, B_{\wp}\}}{2\kappa_4} \right)\right] \\
&= \mathcal{O}(d^2 \min\{1, B_{\wp}\}),
\end{aligned} \tag{F19}$$

where (a) follows from Lemma F1 and Eqs. (F11)–(F13), and (b) follows from Eqs. (F14)–(F17). ■

2. Exactly one index matches in the same position

Lemma F3. For all distinct i, j, k , we have $\text{Cov}(\Lambda_{i,j}, \Lambda_{k,j}) = \mathcal{O}\left(\frac{d^3}{s} \wp + d^2 \min\{1, B_{\wp}\}\right)$, where $\wp = \text{Tr}(\rho^2)$.

Proof of Lemma F3. First, we have

$$\begin{aligned}
\text{Cov}(\Lambda_{i,j}, \Lambda_{k,j}) &= \mathbb{E}\left[\text{Tr}\left((O \otimes \rho^\top) \hat{X}_i \hat{X}_j\right) \text{Tr}\left((O \otimes \rho^\top) \hat{X}_k \hat{X}_j\right)^*\right] - \mathbb{E}\left[\text{Tr}\left((O \otimes \rho^\top) \hat{X}_i \hat{X}_j\right)\right] \mathbb{E}\left[\text{Tr}\left((O \otimes \rho^\top) \hat{X}_k \hat{X}_j\right)^*\right] \\
&\stackrel{(a)}{=} \mathbb{E}\left[\text{Tr}\left((O \otimes \rho^\top) \hat{X}_i \hat{X}_j\right) \text{Tr}\left((O \otimes \rho^\top) \hat{X}_k \hat{X}_j\right)\right] - \text{Tr}[(O \otimes \rho^\top) d \Upsilon_U] \text{Tr}[(O \otimes \rho^\top) d \Upsilon_U]^* \\
&= \mathbb{E}\left[\text{Tr}\left((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\hat{X}_i \hat{X}_j \otimes \hat{X}_k \hat{X}_j)\right)\right] - d^2 [\text{Tr}(OU(\rho))]^2,
\end{aligned} \tag{F20}$$

where (a) holds because $O \otimes \rho^\top$, $\hat{X}_j$, $\hat{X}_k$ are Hermitian, and $\mathbb{E}[\hat{X}_i \hat{X}_j] = \mathbb{E}[\hat{X}_j \hat{X}_k] = \mathbb{E}[\hat{X}]^2 = \Upsilon_U^2 = d \Upsilon_U$. The second moment term in the last line can be further calculated as

$$\begin{aligned}
&\mathbb{E}\left[\text{Tr}\left((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\hat{X}_i \hat{X}_j \otimes \hat{X}_k \hat{X}_j)\right)\right] \\
&= \text{Tr}\left((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\mathbb{E}[\hat{X}_i] \otimes I \otimes I) \mathbb{E}[\hat{X}_j \otimes \hat{X}_j] (I \otimes I \otimes \mathbb{E}[\hat{X}_k])\right) \\
&= \text{Tr}\left((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_U \otimes I \otimes I) \mathbb{E}[\hat{X} \otimes \hat{X}] (I \otimes I \otimes \Upsilon_U)\right).
\end{aligned} \tag{F21}$$

The expression of $\mathbb{E}[\hat{X} \otimes \hat{X}]$ in Lemma F1 has 7 terms. Plugging them into Eq. (F21), respectively, we get

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_U \otimes I \otimes I) \tilde{\Delta}_5 (I \otimes I \otimes \Upsilon_U)\right] = d [\text{Tr}(OU(\rho))]^2 \geq 0, \tag{F22}$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_U \otimes I \otimes I) \tilde{\Delta}_6 (I \otimes I \otimes \Upsilon_U)\right] = d [\text{Tr}(OU(\rho))]^2 \geq 0, \tag{F23}$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_U \otimes I \otimes I) \tilde{\Delta}_7 (I \otimes I \otimes \Upsilon_U)\right] = [\text{Tr}(OU(\rho))]^2 \geq 0, \tag{F24}$$

and

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_U \otimes I \otimes I) \tilde{\Delta}_1 (I \otimes I \otimes \Upsilon_U)\right] = \mathcal{O}(d \wp), \tag{F25}$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_U \otimes I \otimes I) \tilde{\Delta}_2 (I \otimes I \otimes \Upsilon_U)\right] = \mathcal{O}(d \min\{1, B_{\wp}\}), \tag{F26}$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_U \otimes I \otimes I) \tilde{\Delta}_3 (I \otimes I \otimes \Upsilon_U)\right] = \mathcal{O}(d^2 \wp), \tag{F27}$$

$$\text{Tr}\left[(O \otimes \rho^\top \otimes O \otimes \rho^\top) (\Upsilon_U \otimes I \otimes I) \tilde{\Delta}_4 (I \otimes I \otimes \Upsilon_U)\right] = \mathcal{O}(d^2 \min\{1, B_{\wp}\}). \tag{F28}$$

Here Eqs. (F25)–(F28) can be derived by using tensor-network diagrams. For example,

$$\begin{aligned}
& \text{Tr}[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Omega \otimes I \otimes I) \tilde{\Delta}_4(I \otimes I \otimes \Omega)] \\
&= \text{Diagram 1} = \frac{1}{4!} \sum_{\pi \in S_4} \text{Diagram 2} \\
&= \frac{1}{24} [(d^2 + 5d + 6)(\text{Tr}(OU\rho U^\dagger))^2 + (d^2 + 5d + 6) \text{Tr}(O^2 U \rho^2 U^\dagger)] \\
&= \mathcal{O}(d^2 \min\{1, B_\wp\}),
\end{aligned} \tag{F29}$$

The diagrammatic representation shows the trace of a product of operators. The first diagram (Diagram 1) consists of four vertical lines representing indices. The top two lines are labeled O and $\rho^\top$, and the bottom two are labeled O and $\rho^\top$. Between the top and bottom pairs, there are green boxes labeled U and $U^\dagger$. A red dashed box labeled Υ_U encloses the U and $U^\dagger$ boxes on the left. A red dashed box labeled $\tilde{\Delta}_4$ encloses the central part of the diagram. A red dashed box labeled Υ_U encloses the U and $U^\dagger$ boxes on the right. The second diagram (Diagram 2) shows a similar structure but with a red box labeled T_π in the center, representing a permutation π over the four indices.

where the last inequality follows from Lemma D4.

Therefore,

$$\begin{aligned}
\text{Cov}(\Lambda_{i,j}, \Lambda_{k,j}) &\leq \text{Tr}[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes I \otimes I) \mathbb{E}[\hat{X} \otimes \hat{X}](I \otimes I \otimes \Upsilon_U)] \\
&\stackrel{(a)}{\leq} \text{Tr} \left[(O \otimes \rho^\top \otimes O \otimes \rho^\top)(\Upsilon_U \otimes I \otimes I) \right. \\
&\quad \times \frac{2d^2(d+1)^2(d+s)}{s^2(d+s+1)} \left(\frac{1}{\kappa_2} \tilde{\Delta}_1 + \frac{s}{\kappa_3} \tilde{\Delta}_2 + \frac{s}{\kappa_3} \tilde{\Delta}_3 + \frac{s(s-1)}{2\kappa_4} \tilde{\Delta}_4 \right) (I \otimes I \otimes \Upsilon_U) \Big] \\
&\stackrel{(b)}{=} \mathcal{O} \left[\frac{d^4}{s^2} \left(\frac{d\wp}{\kappa_2} + \frac{sd \min\{1, B_\wp\}}{\kappa_3} + \frac{sd^2 \wp}{\kappa_3} + \frac{s(s-1)d^2 \min\{1, B_\wp\}}{2\kappa_4} \right) \right] \\
&= \mathcal{O} \left(\frac{d^3}{s} \wp + d^2 \min\{1, B_\wp\} \right),
\end{aligned} \tag{F30}$$

where (a) follows from Lemma F1 and Eqs. (F22)–(F24), and (b) follows from Eqs. (F25)–(F28). ■

3. Both indices match with order swapped

Lemma F4. For distinct i, j , we have

$$\text{Cov}(\Lambda_{i,j}, \Lambda_{j,i}) = \mathcal{O} \left[\left(\frac{d^3}{s^4} + \frac{d^2}{s^3} + \frac{d\sqrt{d\wp}}{s^2} + \frac{\sqrt{d\wp}}{s} + \wp \right) d^2 B \right]. \tag{F31}$$

Proof of Lemma F4. For distinct i, j , we have

$$\begin{aligned} \text{Cov}(\Lambda_{ij}, \Lambda_{ji}) &= \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top) \hat{X}_i \hat{X}_j) \text{Tr}((O \otimes \rho^\top) \hat{X}_i \hat{X}_j) \right] - \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top) \hat{X}_i \hat{X}_j) \right] \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top) \hat{X}_j \hat{X}_i)^* \right] \\ &= \mathbb{E} \left[\text{Tr}((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\hat{X}_i \hat{X}_j \otimes \hat{X}_i \hat{X}_j)) \right] - d^2 [\text{Tr}(OU(\rho))]^2, \\ &\leq \text{Tr}((O \otimes \rho^\top \otimes O \otimes \rho^\top) \mathbb{E}[\hat{X} \otimes \hat{X}] \mathbb{E}[\hat{X} \otimes \hat{X}]). \end{aligned} \quad (\text{F32})$$

Define

$$\xi(\mu, \nu) := \left| \text{Tr}((O \otimes \rho^\top \otimes O \otimes \rho^\top) \tilde{\Delta}_\mu \tilde{\Delta}_\nu) \right|, \quad \mu, \nu \in \{1, 2, \dots, 7\}. \quad (\text{F33})$$

Then, according to Lemma F1, we can expand the last line of Eq. (F32) as

$$\begin{aligned} &\text{Tr}((O \otimes \rho^\top \otimes O \otimes \rho^\top) \mathbb{E}[\hat{X} \otimes \hat{X}] \mathbb{E}[\hat{X} \otimes \hat{X}]) \\ &= \mathcal{O} \left(\frac{d^4}{s^4} \xi(1, 1) + \frac{d^3}{s^3} [\xi(1, 2) + \xi(1, 3) + \xi(2, 1) + \xi(3, 1)] + \frac{d}{s} [\xi(2, 4) + \xi(3, 4) + \xi(4, 2) + \xi(4, 3)] \right. \\ &\quad + \frac{d^2}{s^2} [\xi(2, 2) + \xi(2, 3) + \xi(1, 4) + \xi(3, 2) + \xi(3, 3) + \xi(4, 1)] + \xi(4, 4) + \frac{(d+s)^4}{s^4} \xi(7, 7) \\ &\quad + \frac{d(d+s)}{s^2} [\xi(3, 5) + \xi(3, 6) + \xi(2, 5) + \xi(2, 6) + \xi(5, 2) + \xi(5, 3) + \xi(6, 2) + \xi(6, 3)] \\ &\quad + \frac{d^2(d+s)}{s^3} [\xi(1, 5) + \xi(1, 6) + \xi(5, 1) + \xi(6, 1)] + \frac{(d+s)}{s} [\xi(4, 5) + \xi(4, 6) + \xi(5, 4) + \xi(6, 4)] \\ &\quad + \frac{(d+s)^3}{s^3} [\xi(5, 7) + \xi(6, 7) + \xi(7, 5) + \xi(7, 6)] + \frac{d(d+s)^2}{s^3} [\xi(2, 7) + \xi(3, 7) + \xi(7, 2) + \xi(7, 3)] \\ &\quad \left. + \frac{(d+s)^2}{s^2} [\xi(5, 5) + \xi(5, 6) + \xi(6, 5) + \xi(6, 6) + \xi(4, 7) + \xi(7, 4)] + \frac{d^2(d+s)^2}{s^4} [\xi(1, 7) + \xi(7, 1)] \right). \end{aligned} \quad (\text{F34})$$

To bound the covariance, we derive bounds for all $\xi(\mu, \nu)$ using tensor-network diagrams, as summarized in Table IV. For example,

$$\begin{aligned} &\text{Tr}[(O \otimes \rho^\top \otimes O \otimes \rho^\top) \tilde{\Delta}_1 \tilde{\Delta}_3] \\ &= \text{Diagram 1} = \frac{1}{2! \cdot 3! \cdot 2!} \sum_{\substack{\pi_1, \pi_2 \in S_2 \\ \pi_3 \in S_3}} \text{Diagram 2} \\ &= \frac{1}{12} \left[(\text{Tr}(O))^2 (\text{Tr}(\rho))^2 + 2 \text{Tr}(O) \text{Tr}(\rho) \text{Tr}(OU\rho U^\dagger) + (\text{Tr}(O))^2 \text{Tr}(\rho^2) + 2 \text{Tr}(O) \text{Tr}(OU\rho^2 U^\dagger) \right. \\ &\quad \left. + \text{Tr}(O^2) (\text{Tr}(\rho))^2 + 2 \text{Tr}(\rho) \text{Tr}(O^2 U \rho U^\dagger) + \text{Tr}(O^2) \text{Tr}(\rho^2) + 2 \text{Tr}(O^2 U \rho^2 U^\dagger) \right]. \end{aligned} \quad (\text{F35})$$

TABLE IV. Upper bounds for $\xi(\mu, \nu)$ derived by using tensor-network diagrams. Here $\wp = \text{Tr}(\rho^2)$.

	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$	$\nu = 7$
$\mu = 1$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(\sqrt{d\wp}B)$	$\mathcal{O}(\sqrt{d\wp}B)$	$\mathcal{O}(dB)$
$\mu = 2$	$\mathcal{O}(dB)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(dB\sqrt{\wp})$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(dB)$
$\mu = 3$	$\mathcal{O}(dB)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(dB\sqrt{\wp})$	$\mathcal{O}(dB)$
$\mu = 4$	$\mathcal{O}(dB)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d^2B\wp)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(dB)$
$\mu = 5$	$\mathcal{O}(\sqrt{d\wp}B)$	$\mathcal{O}(\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(B\wp)$	$\mathcal{O}(\sqrt{d\wp}B)$
$\mu = 6$	$\mathcal{O}(\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(\sqrt{d\wp}B)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(B\wp)$	$\mathcal{O}(d\sqrt{d\wp}B)$	$\mathcal{O}(\sqrt{d\wp}B)$
$\mu = 7$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(\sqrt{d\wp}B)$	$\mathcal{O}(\sqrt{d\wp}B)$	$\mathcal{O}(dB)$

This equation and Lemma D4 together imply that

$$\xi(1, 3) = \left| \text{Tr} \left[(O \otimes \rho^\top \otimes O \otimes \rho^\top) \tilde{\Delta}_1 \tilde{\Delta}_3 \right] \right| \leq \frac{dB + 2\sqrt{dB} + B + 2}{6} = \mathcal{O}(dB). \quad (\text{F36})$$

The other bounds in Table IV can be derived in a similar way.

By combining Eqs. (F32), (F34), and Table IV, we have

$$\begin{aligned} & \text{Cov}(\Lambda_{ij}, \Lambda_{ji}) \\ & \leq \mathcal{O} \left(\frac{d^4}{s^4} dB + \frac{d^3}{s^3} dB + \frac{d}{s} dB \sqrt{d\wp} + \frac{d^2}{s^2} dB \sqrt{d\wp} + d^2 B \wp + \frac{(d+s)^4}{s^4} dB + \frac{d(d+s)}{s^2} dB \sqrt{d\wp} + \frac{d^2(d+s)}{s^3} B \sqrt{d\wp} \right. \\ & \quad \left. + \frac{(d+s)}{s} dB \sqrt{d\wp} + \frac{(d+s)^3}{s^3} B \sqrt{d\wp} + \frac{d(d+s)^2}{s^3} dB + \frac{(d+s)^2}{s^2} dB \sqrt{d\wp} + \frac{d^2(d+s)^2}{s^4} dB \right) \\ & = \mathcal{O} \left[\left(\frac{d^3}{s^4} + \frac{d^2}{s^3} + \frac{d\sqrt{d\wp}}{s^2} + \frac{\sqrt{d\wp}}{s} + \wp \right) d^2 B \right], \end{aligned} \quad (\text{F37})$$

which confirms Lemma F4. ■

4. Both indices match with the same order

Lemma F5. For distinct i, j , we have

$$\text{Cov}(\Lambda_{ij}, \Lambda_{ij}) = \mathcal{O} \left[\left(\frac{d^4}{s^4} + \frac{d^3}{s^3} + \frac{d^2}{s^2} + \frac{d}{s} + 1 \right) d^2 B \wp \right]. \quad (\text{F38})$$

Proof of Lemma F5. For distinct i, j , we have

$$\begin{aligned} \text{Cov}(\Lambda_{ij}, \Lambda_{ij}) &= \mathbb{E} \left[\text{Tr} \left((O \otimes \rho^\top) \hat{X}_i \hat{X}_j \right) \text{Tr} \left((O \otimes \rho^\top) \hat{X}_j \hat{X}_i \right) \right] - d^2 [\text{Tr}(O \mathcal{U}(\rho))]^2 \\ &\leq \mathbb{E} \left[\text{Tr} \left((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\hat{X}_i \hat{X}_j \otimes \hat{X}_j \hat{X}_i) \right) \right], \end{aligned} \quad (\text{F39})$$

Note that the trace inside the last line of Eq. (F39) can be further written as

$$\begin{aligned} & \text{Tr} \left((O \otimes \rho^\top \otimes O \otimes \rho^\top) (\hat{X}_i \hat{X}_j \otimes \hat{X}_j \hat{X}_i) \right) \\ &= \text{Tr} \left((O \otimes \rho^\top \otimes O \otimes \rho^\top \otimes I \otimes I) (I \otimes I \otimes \hat{X}_i \otimes \hat{X}_i) (\hat{X}_j \otimes \hat{X}_j \otimes I \otimes I) T_{(1,5)(2,6)(3)(4)} \right). \end{aligned} \quad (\text{F40})$$

So the covariance satisfies

$$\begin{aligned} & \text{Cov} \left[\text{Tr} \left((O \otimes \rho^\top) \hat{X}_i \hat{X}_j \right), \text{Tr} \left((O \otimes \rho^\top) \hat{X}_j \hat{X}_i \right) \right] \\ & \leq \text{Tr} \left[(O \otimes \rho^\top \otimes O \otimes \rho^\top \otimes I \otimes I) (I \otimes I \otimes \mathbb{E}[\hat{X} \otimes \hat{X}]) (\mathbb{E}[\hat{X} \otimes \hat{X}] \otimes I \otimes I) T_{(1,5)(2,6)(3)(4)} \right]. \end{aligned} \quad (\text{F41})$$

Define

$$\zeta(\mu, \nu) := \left| \text{Tr} \left[(O \otimes \rho^\top \otimes O \otimes \rho^\top \otimes I \otimes I) (I \otimes I \otimes \tilde{\Delta}_\mu) (\tilde{\Delta}_\nu \otimes I \otimes I) T_{(1,5)(2,6)(3)(4)} \right] \right| \quad \mu, \nu \in \{1, 2, \dots, 7\}. \quad (\text{F42})$$

According to Lemma F1, we can then expand the right-hand side of Eq. (F41) as

Right-hand side of Eq. (F41)

$$\begin{aligned} &= \mathcal{O} \left(\frac{d^4}{s^4} \zeta(1, 1) + \frac{d^3}{s^3} [\zeta(1, 2) + \zeta(1, 3) + \zeta(2, 1) + \zeta(3, 1)] + \frac{d}{s} [\zeta(2, 4) + \zeta(3, 4) + \zeta(4, 2) + \zeta(4, 3)] \right. \\ &\quad + \frac{d^2}{s^2} [\zeta(2, 2) + \zeta(2, 3) + \zeta(1, 4) + \zeta(3, 2) + \zeta(3, 3) + \zeta(4, 1)] + \zeta(4, 4) + \frac{(d+s)^4}{s^4} \zeta(7, 7) \\ &\quad + \frac{d(d+s)}{s^2} [\zeta(3, 5) + \zeta(3, 6) + \zeta(2, 5) + \zeta(2, 6) + \zeta(5, 2) + \zeta(5, 3) + \zeta(6, 2) + \zeta(6, 3)] \\ &\quad + \frac{d^2(d+s)}{s^3} [\zeta(1, 5) + \zeta(1, 6) + \zeta(5, 1) + \zeta(6, 1)] + \frac{(d+s)}{s} [\zeta(4, 5) + \zeta(4, 6) + \zeta(5, 4) + \zeta(6, 4)] \\ &\quad + \frac{(d+s)^3}{s^3} [\zeta(5, 7) + \zeta(6, 7) + \zeta(7, 5) + \zeta(7, 6)] + \frac{d(d+s)^2}{s^3} [\zeta(2, 7) + \zeta(3, 7) + \zeta(7, 2) + \zeta(7, 3)] \\ &\quad \left. + \frac{(d+s)^2}{s^2} [\zeta(5, 5) + \zeta(5, 6) + \zeta(6, 5) + \zeta(6, 6) + \zeta(4, 7) + \zeta(7, 4)] + \frac{d^2(d+s)^2}{s^4} [\zeta(1, 7) + \zeta(7, 1)] \right). \quad (\text{F43}) \end{aligned}$$

To bound the covariance, we derive bounds for all $\zeta(\mu, \nu)$ using tensor-network diagrams, as summarized in Table V. For example,

$$\begin{aligned} &\text{Tr} \left[(O \otimes \rho^\top \otimes O \otimes \rho^\top \otimes I \otimes I) (I \otimes I \otimes \tilde{\Delta}_4) (\tilde{\Delta}_3 \otimes I \otimes I) T_{(1,5)(2,6)(3)(4)} \right] \\ &= \text{Diagram 1} = \text{Diagram 2} = \frac{1}{2! \cdot 3! \cdot 4!} \sum_{\substack{\pi_1 \in S_3 \\ \pi_2 \in S_4 \\ \pi_3 \in S_2}} \text{Diagram 3} \\ &= (C_1 d^2 + C_2 d + C_3) \text{Tr}(O_U^2) \text{Tr}(\rho^2) + (C_4 d^2 + C_5 d + C_6) (\text{Tr}(O_U \rho))^2 + (C_7 d^2 + C_8 d + C_9) \text{Tr}(O_U^2 \rho^2) \\ &\quad + (C_{10} d + C_{11}) \text{Tr}(O_U) \text{Tr}(O_U \rho) + (C_{12} d + C_{13}) (\text{Tr}(O_U))^2 \text{Tr}(\rho^2) + (C_{14} d + C_{15}) \text{Tr}(O_U) \text{Tr}(O_U \rho^2) \\ &\quad + (C_{16} d + C_{17}) \text{Tr}(O_U^2 \rho) + (C_{18} d + C_{19}) \text{Tr}(O_U \rho O_U \rho) + (C_{20} d + C_{21}) \text{Tr}(O^2) + (C_{22} d + C_{23}) (\text{Tr}(O))^2, \quad (\text{F44}) \end{aligned}$$

TABLE V. Upper bounds for $\zeta(\mu, \nu)$ derived by using tensor-network diagrams. Here $\wp = \text{Tr}(\rho^2)$.

	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$	$\nu = 7$
$\mu = 1$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(\sqrt{d \wp} B)$	$\mathcal{O}(\sqrt{d \wp} B)$	$\mathcal{O}(dB)$
$\mu = 2$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(dB \sqrt{\wp})$	$\mathcal{O}(dB)$
$\mu = 3$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(\sqrt{d \wp} B)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(dB)$
$\mu = 4$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d^2 B \wp)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(dB)$
$\mu = 5$	$\mathcal{O}(\sqrt{d \wp} B)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(\sqrt{d \wp} B)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(B \wp)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(\sqrt{d \wp} B)$
$\mu = 6$	$\mathcal{O}(\sqrt{d \wp} B)$	$\mathcal{O}(dB \sqrt{\wp})$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(d \sqrt{d \wp} B)$	$\mathcal{O}(B \wp)$	$\mathcal{O}(\sqrt{d \wp} B)$
$\mu = 7$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(dB)$	$\mathcal{O}(\sqrt{d \wp} B)$	$\mathcal{O}(\sqrt{d \wp} B)$	$\mathcal{O}(dB)$

where $C_1, C_2, \dots, C_{23}$ are constants, and $O_U = U^\dagger O U$. This equation and Lemma D4 together imply that

$$\zeta(4, 3) = \left| \text{Tr} \left[(O \otimes \rho^\top \otimes O \otimes \rho^\top \otimes I \otimes I) (I \otimes I \otimes \tilde{\Delta}_4) (\tilde{\Delta}_3 \otimes I \otimes I) T_{(1,5)(2,6)(3)(4)} \right] \right| = \mathcal{O}(d^2 B \wp). \quad (\text{F45})$$

The other bounds in Table V can be derived in a similar way.

By combining Eqs. (F41), (F43), and Table V, we have

$$\begin{aligned} & \text{Cov}(\Lambda_{ij}, \Lambda_{ij}) \\ & \leq \mathcal{O} \left(\frac{d^4}{s^4} d^2 B \wp + \frac{d^3}{s^3} d^2 B \wp + \frac{d}{s} d^2 B \wp + \frac{d^2}{s^2} d^2 B \wp + d^2 B \wp + \frac{(d+s)^4}{s^4} dB + \frac{d(d+s)}{s^2} dB \sqrt{d \wp} + \frac{d^2(d+s)}{s^3} B \sqrt{d \wp} \right. \\ & \quad \left. + \frac{(d+s)}{s} dB \sqrt{d \wp} + \frac{(d+s)^3}{s^3} B \sqrt{d \wp} + \frac{d(d+s)^2}{s^3} dB + \frac{(d+s)^2}{s^2} dB \sqrt{d \wp} + \frac{d^2(d+s)^2}{s^4} dB \right) \\ & = \mathcal{O} \left[\left(\frac{d^4}{s^4} + \frac{d^3}{s^3} + \frac{d^2}{s^2} + \frac{d}{s} + 1 \right) d^2 B \wp \right], \end{aligned} \quad (\text{F46})$$

which confirms Lemma F5. ■

APPENDIX G: PROOF OF THEOREM 1

To prove Theorem 1, it suffices to prove the following theorem, which is a stronger version of Theorem 1.

Theorem G1. Suppose $s \geq 1$ is the number of systems collectively measured in our learning phase; accuracy parameters $0 < \delta, \epsilon < 1$; $\rho_1, \dots, \rho_M$ is a collection of quantum states chosen from $\mathcal{D}(\mathcal{H})$, with $\max_l \text{Tr}(\rho_l^2) = \mathcal{P}$; and $O_1, \dots, O_M$ is a collection of observables chosen from $\text{Obs}(B)$, with $1 \leq B \leq d$. To estimate all M properties $\text{Tr}(O_l \mathcal{U}(\rho_l))$ within additive error ϵ and failure probability at most δ , the number of queries required by our protocol for CSEU is upper bounded by

$$\begin{aligned} & \mathcal{O} \left(\left(\frac{d \mathcal{P} + s \min\{1, B \mathcal{P}\}}{\epsilon^2} + \frac{\max\{d^2, s^2\} \sqrt{B \mathcal{P}}}{s \epsilon} \right) \right. \\ & \quad \left. \times \log \left(\frac{M}{\delta} \right) \right). \end{aligned} \quad (\text{G1})$$

Proof of Theorem G1. First, we consider the case of $M = 1$, namely, the goal is to accurately predict $\text{Tr}(O \mathcal{U}(\rho))$ for an observable $O \in \text{Obs}(B)$ and a quantum state $\rho \in \mathcal{D}(\mathcal{H})$ that has purity $\wp = \text{Tr}(\rho^2)$. Suppose we want the probability in Eq. (26) to be at most $1/4$, i.e.,

$$\Pr \left\{ \left| \hat{Z}_{(r)}(O, \rho) - \text{Tr}(O \mathcal{U}(\rho)) \right| \geq \epsilon \right\} \leq \frac{1}{4}. \quad (\text{G2})$$

Thanks to Proposition 1, it is sufficient to have

$$\begin{aligned} & \frac{1}{q} \left(\frac{d \wp}{s} + \min\{1, B \wp\} \right) \leq \epsilon^2 \quad \text{and} \\ & \frac{1}{q^2} \left(\frac{d^4}{s^4} + 1 \right) B \wp \leq \epsilon^2, \end{aligned} \quad (\text{G3})$$

where we ignored constant coefficients. These two conditions can be achieved by choosing

$$q = C \left(\frac{1}{\epsilon^2} \max \left\{ \frac{d\wp}{s}, \min \{1, B\wp\} \right\} + \frac{\sqrt{B\wp}}{\epsilon} \max \left\{ \frac{d^2}{s^2}, 1 \right\} \right) \quad (\text{G4})$$

for some constant C .

Recall that our final estimate is the median value $\hat{E}(O, \rho)$ in Eq. (15). Assume that the number of batches R is odd, then $\hat{E}(O, \rho)$ equals some $\hat{Z}_{(r)}(O, \rho)$. If $\hat{E}(O, \rho)$ is a bad estimate, i.e., $|\hat{E}(O, \rho) - \text{Tr}(OU(\rho))| \geq \epsilon$, then at least half of the R estimates $\hat{Z}_{(r)}(O, \rho)$ are bad. Hence, when Eq. (G2) holds for all $r = 1, 2, \dots, R$, the failure probability of $\hat{E}(O, \rho)$ can be bounded by

$$\begin{aligned} \Pr \left[\left| \hat{E}(O, \rho) - \text{Tr}(OU(\rho)) \right| \geq \epsilon \right] &\leq \Pr \left[\# \left\{ r : \left| \hat{Z}_{(r)}(O, \rho) - \text{Tr}(OU(\rho)) \right| \geq \epsilon \right\} \geq \frac{R}{2} \right] \\ &\stackrel{(a)}{\leq} \sum_{j=0}^{(R-1)/2} \binom{R}{j} \left(\frac{3}{4} \right)^j \left(\frac{1}{4} \right)^{R-j} \stackrel{(b)}{\leq} \exp \left[-RD \left(\frac{(R-1)/2}{R} \parallel 3/4 \right) \right] \stackrel{(c)}{\leq} e^{-RD(1/2 \parallel 3/4)}, \end{aligned} \quad (\text{G5})$$

where (a) holds because the R estimates $\hat{Z}_{(r)}(O, \rho)$ are independent; $D(x \parallel y) := x \ln \frac{x}{y} + (1-x) \ln \frac{1-x}{1-y}$ is the Kullback-Leibler divergence; (b) follows from the Chernoff bound in Lemma G1 below; and (c) holds because $D(x \parallel y)$ is strictly decreasing in x when $0 < x < y < 1$. Therefore, to ensure that the failure probability in Eq. (G5) is at most δ , it suffices to choose

$$R = \left\lceil \frac{\ln \delta^{-1}}{D(1/2 \parallel 3/4)} \right\rceil = \mathcal{O}(\log \delta^{-1}). \quad (\text{G6})$$

By Eqs. (G4) and (G6), the total number of queries used by our protocol is

$$\begin{aligned} s \cdot q \cdot R &= \mathcal{O} \left(\left(\frac{d\wp + s \min \{1, B\wp\}}{\epsilon^2} + \frac{\max \{d^2, s^2\} \sqrt{B\wp}}{s\epsilon} \right) \right. \\ &\quad \left. \times \log(\delta^{-1}) \right), \end{aligned} \quad (\text{G7})$$

which confirms Theorem G1 in the special case $M = 1$.

To complete the proof, now we consider the general case of $M \geq 1$. According to the union bound, if our protocol can accurately predict one expectation value $\text{Tr}(OU(\rho))$ with failure probability at most $\tilde{\delta}$, then it can also accurately predict M expectation values $\text{Tr}(O_l U(\rho_l))$ for $l = 1, 2, \dots, M$ simultaneously, with failure probability at most $M\tilde{\delta}$. This fact and the query number in Eq. (G7) together confirm Theorem G1 in the general case $M \geq 1$. ■

Lemma G1 (Chernoff bound). Suppose $0 < p < 1$, integers $k, z \geq 0$, and $k \leq pz$. Then we have

$$\sum_{j=0}^k \binom{z}{j} p^j (1-p)^{z-j} \leq e^{-zD(\frac{k}{z} \parallel p)}. \quad (\text{G8})$$

APPENDIX H: PROOF OF LEMMA 4

Note that $\Pr\{\Xi(\mathcal{T}, O) \geq \epsilon\} \leq \delta$ is satisfied if the following relations hold:

$$\begin{aligned} \mathbb{E}[\Xi(\mathcal{T}, O)] &\leq \frac{\epsilon}{2}, \\ \Pr \left[\left| \Xi(\mathcal{T}, O) - \mathbb{E}[\Xi(\mathcal{T}, O)] \right| \geq \frac{\epsilon}{2} \right] &\leq \delta. \end{aligned} \quad (\text{H1})$$

Here and afterwards, the expectation $\mathbb{E}$ is over the randomness in our protocol.

First, we have

$$\begin{aligned} \Xi(\mathcal{T}, O) &:= \text{Ave}_{\rho \sim \mathcal{T}} |E(O, \rho) - \text{Tr}(OU(\rho))| \\ &\leq \sqrt{\text{Ave}_{\rho \sim \mathcal{T}} \left(\left| \hat{E}(O, \rho) - \text{Tr}(OU(\rho)) \right|^2 \right)}. \end{aligned} \quad (\text{H2})$$

It follows that

$$\begin{aligned} \mathbb{E}[\Xi(\mathcal{T}, O)] &\leq \mathbb{E} \left[\sqrt{\text{Ave}_{\rho \sim \mathcal{T}} \left(\left| \hat{E}(O, \rho) - \text{Tr}(OU(\rho)) \right|^2 \right)} \right] \\ &\leq \sqrt{\mathbb{E} \left[\text{Ave}_{\rho \sim \mathcal{T}} \left(\left| \hat{E}(O, \rho) - \text{Tr}(OU(\rho)) \right|^2 \right) \right]}. \end{aligned} \quad (\text{H3})$$

Second, Chebyshev's inequality implies that

$$\begin{aligned} & \Pr\left\{|\Xi(\mathcal{T}, O) - \mathbb{E}[\Xi(\mathcal{T}, O)]| \geq \frac{\epsilon}{2}\right\} \\ & \leq \frac{4}{\epsilon^2} \text{Var}[\Xi(\mathcal{T}, O)] \leq \frac{4}{\epsilon^2} \mathbb{E}[\Xi(\mathcal{T}, O)^2] \\ & \leq \frac{4}{\epsilon^2} \mathbb{E}\left[\text{Ave}_{\rho \sim \mathcal{T}} \left(\left|\hat{E}(O, \rho) - \text{Tr}(OU(\rho))\right|^2\right)\right], \quad (\text{H4}) \end{aligned}$$

where the last inequality follows from Eq. (H2) again.

According to Eqs. (H1), (H3), and (H4), the relation $\Pr\{\Xi(\mathcal{T}, O) \geq \epsilon\} \leq \delta$ is satisfied if

$$\mathbb{E}\left[\text{Ave}_{\rho \sim \mathcal{T}} \left(\left|\hat{E}(O, \rho) - \text{Tr}(OU(\rho))\right|^2\right)\right] \leq \frac{\delta \epsilon^2}{4}. \quad (\text{H5})$$

In addition, we have

$$\begin{aligned} & \mathbb{E}\left[\text{Ave}_{\rho \sim \mathcal{T}} \left(\left|\hat{E}(O, \rho) - \text{Tr}(OU(\rho))\right|^2\right)\right] \\ & = \text{Ave}_{\rho \sim \mathcal{T}} \left[\mathbb{E}\left(\left|\hat{E}(O, \rho) - \text{Tr}(OU(\rho))\right|^2\right)\right] \\ & = \text{Ave}_{\rho \sim \mathcal{T}} \left[\mathbb{E}\left(\left|\hat{E}(O, \rho)\right|^2 - \hat{E}(O, \rho)^* \text{Tr}(OU(\rho)) \right. \right. \\ & \quad \left. \left. - \hat{E}(O, \rho) \text{Tr}(OU(\rho)) + [\text{Tr}(OU(\rho))]^2\right)\right] \\ & \stackrel{(a)}{=} \text{Ave}_{\rho \sim \mathcal{T}} \left[\mathbb{E}\left(\left|\hat{E}(O, \rho)\right|^2\right) - \left|\mathbb{E}[\hat{E}(O, \rho)]\right|^2\right] \\ & = \text{Ave}_{\rho \sim \mathcal{T}} \text{Var}[\hat{E}(O, \rho)], \quad (\text{H6}) \end{aligned}$$

where (a) holds because $\mathbb{E}[\hat{E}(O, \rho)] = \text{Tr}(OU(\rho))$. Equations (H5) and (H6) together confirm Lemma 4.

APPENDIX I: ADDITIONAL RESULTS ON HAMILTONIAN LEARNING

1. Performance of previous protocols for learning an arbitrary Hamiltonian

To illustrate the advantage of our Hamiltonian learning protocol, here we provide more details on the performance of previous protocols for learning an n -qubit unknown Hamiltonian H , including those in Refs. [24,36,37,66,71,72,75]. These protocols are highly efficient when H has a sparse or local structure. However, we show that none of them can learn an arbitrarily complex H with respect to the normalized Frobenius norm (NFN) using only $\tilde{O}(d^3)$ Hamiltonian queries or total evolution time, even with access to quantum memory. Actually, most previous works did not consider the problem of learning fully arbitrary Hamiltonians, and only a few existing protocols can be modified to learn a completely unknown Hamiltonian under our framework.

The Hamiltonian H of interest can be expressed as

$$H = \sum_{\mathbf{q} \in \{0,1,2,3\}^n} \mu(\mathbf{q}) \sigma_{\mathbf{q}}, \quad (\text{I1})$$

where $\sigma_{\mathbf{q}}$ are Pauli operators and $\mu(\mathbf{q}) \in \mathbb{R}$ are Pauli coefficients. H is called a k -body Hamiltonian with M terms if $\mu(\mathbf{q}) \neq 0$ only for M Pauli terms $\sigma_{\mathbf{q}}$ with $\text{wt}(\sigma_{\mathbf{q}}) \leq k$, where the weight $\text{wt}(\sigma_{\mathbf{q}})$ denotes the number of qubits on which $\sigma_{\mathbf{q}}$ acts nontrivially. Most previous Hamiltonian learning protocols focus on estimating all nonzero Pauli coefficients of H . That is, they aim to output accurate estimates $\hat{\mu}(\mathbf{q})$ for $\mu(\mathbf{q})$ that satisfy $\|\hat{\mu} - \mu\|_{\infty} \leq \epsilon$ with high success probability. Here $\mu := (\mu(\mathbf{q}))_{\mathbf{q} \in \{0,1,2,3\}^n}$ and $\hat{\mu} := (\hat{\mu}(\mathbf{q}))_{\mathbf{q} \in \{0,1,2,3\}^n}$ are the vectors of true coefficients and estimated coefficients, respectively, and

$$\|\hat{\mu} - \mu\|_p = \left(\sum_{\mathbf{q} \in \{0,1,2,3\}^n} |\hat{\mu}(\mathbf{q}) - \mu(\mathbf{q})|^p \right)^{1/p} \quad (\text{I2})$$

denotes the ℓ^p norm of $\hat{\mu} - \mu$. Let $\hat{H} = \sum_{\mathbf{q}} \hat{\mu}(\mathbf{q}) \sigma_{\mathbf{q}}$ be the learned description of H , then we have

$$\frac{1}{\sqrt{d}} \|\hat{H} - H\|_2 = \|\hat{\mu} - \mu\|_2 \leq d \|\hat{\mu} - \mu\|_{\infty}, \quad (\text{I3})$$

where the inequality is saturated when $|\hat{\mu}(\mathbf{q}) - \mu(\mathbf{q})|$ are equal for all $\mathbf{q} \in \{0,1,2,3\}^n$. Therefore, estimating all Pauli coefficients of H within additive error ϵ only implies learning H within error $d\epsilon$ in NFN.

Reference [24] proposed an efficient protocol for learning the Pauli coefficients of an arbitrary traceless Hamiltonian H . It uses parallel queries to the time evolution under H in experiments, and thus needs quantum memory. To achieve $\|\hat{\mu} - \mu\|_{\infty} \leq \epsilon$ with high probability, this protocol requires $\tilde{O}(n\|H\|^4\epsilon^{-4})$ queries to H and a total evolution time of $\tilde{O}(n\|H\|^3\epsilon^{-4})$. A similar performance is achieved by the protocol in Ref. [72], which uses sequential queries to H . According to the above discussions, if the goal is to learn H within error ϵ in NFN, then the query number and the evolution time required by these protocols are both $\tilde{O}(d^4\epsilon^{-4})$, assuming that $\|H\| = \mathcal{O}(\text{poly}(n))$.

A recent work Ref. [66] proposed a protocol for learning a k -body Hamiltonian H with M terms, which is robust to state preparation and measurement (SPAM) errors. It employs sequential queries to H in each experiment and achieves the following performance: By using $N = \tilde{O}(3^k M)$ independent experiments and a total evolution time $T = \tilde{O}(9^k M^{1/p+1/2} \epsilon^{-1})$, it can output estimate $\hat{\mu}(\mathbf{q})$ for every $\mu(\mathbf{q})$ such that $\|\hat{\mu} - \mu\|_p \leq \epsilon$ with high probability, where $p = 1$ or 2 . This protocol is highly efficient when k is a constant. However, if we do not make prior assumptions about the structure of H , then k and M can take values

up to $k = n$ and $M = d^2$. For this worst-case scenario, learning H within error ϵ in NFN (that is, $\|\hat{\mu} - \mu\|_2 \leq \epsilon$) requires a $T = \tilde{O}(9^n d^2 \epsilon^{-1}) \approx \tilde{O}(d^{5.170} \epsilon^{-1})$ total evolution time and $N = \tilde{O}(3^n d^2) \approx \tilde{O}(d^{3.585})$ experiments. Another work [71] showed that one can achieve $\|\hat{\mu} - \mu\|_\infty \leq \epsilon$ from $T = \tilde{O}(M^2 \epsilon^{-1})$ total evolution time. From the above discussions, this result implies a $T = \tilde{O}(d^5 \epsilon^{-1})$ time complexity for learning an arbitrary Hamiltonian within error ϵ in NFN.

Reference [75] proposed an efficient protocol for learning an unknown Hamiltonian H with M Pauli terms. In order to learn H within error ϵ in NFN, it needs $\tilde{O}(M \|H\|^2 \epsilon^{-2})$ Hamiltonian queries, which reduces to $\tilde{O}(d^2 \epsilon^{-2})$ if H is an arbitrary unknown Hamiltonian satisfying $\|H\| = \mathcal{O}(\text{poly}(n))$. However, this protocol requires not only the time evolution unitary e^{-iHt} , but also the backwards evolution unitary e^{iHt} , which is physically unrealizable.

There are also Hamiltonian learning protocols that do not use quantum memory and look beyond Hamiltonians with sparse interactions [36,37]. To learn all coefficients of a Hamiltonian H within additive error ϵ , the protocol of Ref. [37] requires $\tilde{O}(M^4 \epsilon^{-2})$ queries, and thus is unsatisfactory when H contains exponentially many terms; the protocol of Ref. [36] requires $\tilde{O}(9^k \epsilon^{-2})$ queries, and thus is unsatisfactory when H contains high-weight terms. From the previous discussions, when the goal is to learn an arbitrary Hamiltonian within error ϵ in NFN, the query costs of these protocols are more than $\tilde{O}(d^5 \epsilon^{-2})$.

Compared with these existing protocols, our Hamiltonian learning protocol demonstrates superior performance by simultaneously (i) operating without the requirement of quantum memory, and (ii) achieving a smaller query number and total evolution time (see Theorem 3 in the main text), making it favorable for real-world implementation.

2. Details of our Hamiltonian learning protocol and its performance

Given time evolution access to an arbitrary n -qubit traceless Hamiltonian H with $\|H\| = \mathcal{O}(\text{poly}(n))$. Our goal is to learn a classical description $\hat{H} \in \mathcal{L}_H(\mathcal{H})$ of the Hamiltonian, such that the following condition holds:

$$\Pr \left\{ \frac{1}{\sqrt{2^n}} \|\hat{H} - H\|_2 \leq \epsilon \right\} \geq 1 - \delta \quad (I4)$$

for accuracy $0 < \delta, \epsilon < 1/3$. Next, we present our learning protocol for this task in Appendix 12 a and show its performance in Appendix 12 b.

a. Hamiltonian learning protocol

We follow the Hamiltonian learning protocol sketched in Sec. VII B of the main text with the following parameter

choices: Our evolution time in each experiment is set to be $t_j = \frac{T}{2}(1 + z_j)$ for $j = 1, 2, \dots, L$, where

$$T = \frac{1}{\|H\|}, \quad z_j = -\cos\left(\frac{2j-1}{2L}\pi\right), \quad L = \left\lceil 2 \log \left(\frac{8 \|H\|}{\sqrt{2\pi} \ln(2)\epsilon} \right) \right\rceil, \quad \epsilon = \frac{\epsilon}{d}. \quad (I5)$$

In the first step of our protocol, for each time t_j ($1 \leq j \leq L$), we treat $\mathcal{U}_{t_j}$ as an unknown unitary channel, and implement our CSEU protocol in Sec. III to estimate the expectation values $\langle \sigma_{\mathbf{q}'} \rangle_{\rho_{\mathbf{q}}(t_j)} = \text{Tr}[\sigma_{\mathbf{q}'} \mathcal{U}_{t_j}(\rho_{\mathbf{q}})]$, where the input state $\rho_{\mathbf{q}}$ and the measured observable $\sigma_{\mathbf{q}'}$ are chosen as follows (see also Ref. [24, Lemma 6.1]): Let $1 \leq i \leq n$ be such that $\mathbf{q}_i \neq 0$; let $\mathbf{q}'_i \in \{1, 2, 3\} \setminus \{\mathbf{q}_i\}$ and set $\mathbf{q}'_j = 0$ for all $j \neq i$; let

$$\rho_{\mathbf{q}} := \frac{1}{d} \left(I + \frac{i}{2} [\sigma_{\mathbf{q}}, \sigma_{\mathbf{q}'}] \right). \quad (I6)$$

We denote by $2\hat{\alpha}_{t_j}^{(\mathbf{q})} \in \mathbb{R}$ our estimate of $\text{Tr}[\sigma_{\mathbf{q}'} \mathcal{U}_{t_j}(\rho_{\mathbf{q}})]$, which should achieve the following accuracy:

$$\Pr \left\{ \left| 2\hat{\alpha}_{t_j}^{(\mathbf{q})} - \text{Tr}[\sigma_{\mathbf{q}'} \mathcal{U}_{t_j}(\rho_{\mathbf{q}})] \right| \leq \tilde{\epsilon}, \quad \forall \mathbf{q} \in \{0, 1, 2, 3\}^n \setminus \{0^n\} \right\} \geq 1 - \frac{\delta}{L}, \quad (I7)$$

where

$$\tilde{\epsilon} := \frac{3T\epsilon}{4(L-1)L(2L-1)}. \quad (I8)$$

Let $1 \leq s \leq d$ be the number of systems collectively measured by our CSEU protocol in each experiment. As explained in Sec. III, if $s > 1$, then quantum memory is required. If instead $s = 1$, then the CSEU protocol does not need quantum memory and ancillary qubits; in each round of experiments, it requires only the preparation of a computational-basis state $|0\rangle^{\otimes n}$, the application of a random Clifford unitary to a single system, a measurement in the computational basis, and one application of $\mathcal{U}_{t_j}$ [i.e., one real $\mathcal{O}(1/\|H\|)$ -time evolution under H].

In the second step of our Hamiltonian learning protocol, we perform Chebyshev polynomial interpolation [24,37] to estimate the Pauli coefficients $\mu(\mathbf{q})$ by using $\hat{\alpha}_{t_j}^{(\mathbf{q})}$ learned in the first step. The Chebyshev coefficients are estimated as

$$\hat{b}_0^{(\mathbf{q})} = \frac{1}{L} \sum_{j=1}^L \hat{\alpha}_{t_j}^{(\mathbf{q})}, \quad (I9)$$

$$\hat{b}_l^{(\mathbf{q})} = \frac{2}{L} \sum_{j=1}^L \hat{\alpha}_{t_j}^{(\mathbf{q})} T_l(z_j), \quad 1 \leq l \leq L-1,$$

where T_l is the l th Chebyshev polynomial. They in turn give rise to the first-order Taylor coefficients [24,37]

$$\hat{\mu}(\mathbf{q}) := -\frac{2}{T} \sum_{l=0}^{L-1} (-1)^l \hat{b}_l^{(\mathbf{q})} l^2, \quad \mathbf{q} \in \{0, 1, 2, 3\}^n \setminus \{0^n\}. \quad (\text{I10})$$

Here $\hat{\mu}(\mathbf{q})$ serves as our estimate of $\mu(\mathbf{q})$. Finally, we use these estimated values to construct our learned description:

$$\hat{H} = \sum_{\mathbf{q} \in \{0, 1, 2, 3\}^n \setminus \{0^n\}} \hat{\mu}(\mathbf{q}) \sigma_{\mathbf{q}}. \quad (\text{I11})$$

b. Performance of the protocol

The performance of our Hamiltonian learning protocol in the previous subsection is stated in the following theorem, which is a stronger version of Theorem 3 in the main text.

Theorem II. Suppose the unknown n -qubit Hamiltonian H to be learned is traceless and satisfies $\|H\| = \mathcal{O}(\text{poly}(n))$; parameters $0 < \delta, \epsilon < 1/3$; and s is the number of systems collectively measured in each experiment. With the accuracy guaranteed in Eq. (I7), the learned description $\hat{H}$ in Eq. (I11) satisfies Eq. (I4). When $s = 1$, our Hamiltonian learning protocol is ancilla-free and does not use quantum memory; its query number to H and total evolution time both scale as

$$\tilde{\mathcal{O}} \left(\left(\frac{d^2}{\epsilon^2} + \frac{d^3}{\epsilon} \right) \log(\delta^{-1}) \right). \quad (\text{I12})$$

When $s = \Theta(\sqrt{d})$, our Hamiltonian learning protocol uses quantum memory; its query number to H and total evolution time both scale as

$$\tilde{\mathcal{O}} \left(\frac{d^{2.5}}{\epsilon^2} \log(\delta^{-1}) \right). \quad (\text{I13})$$

Proof of Theorem II. We begin by proving that $\hat{H}$ given in Eq. (I11) indeed satisfies Eq. (I4). First, by the union bound, the accuracy guaranteed in Eq. (I7) implies that the following condition holds with probability at least $1 - \delta$:

$$\left| 2\hat{\alpha}_{t_j}^{(\mathbf{q})} - \text{Tr}[\sigma_{\mathbf{q}'} \mathcal{U}_{t_j}(\rho_{\mathbf{q}})] \right| \leq \tilde{\epsilon} \quad \forall \mathbf{q} \in \{0, 1, 2, 3\}^n \setminus \{0^n\} \quad \text{and} \quad 1 \leq j \leq L. \quad (\text{I14})$$

Second, according to Ref. [24, Appendix D], if Eq. (I14) holds with $\tilde{\epsilon} = 3T\epsilon/[4(L-1)L(2L-1)]$, then $\hat{\mu}(\mathbf{q})$ constructed in Eq. (I10) is an ϵ -accurate estimate of $\mu(\mathbf{q})$ for

all $\mathbf{q} \neq 0^n$, i.e.,

$$|\hat{\mu}(\mathbf{q}) - \mu(\mathbf{q})| \leq \epsilon \quad \forall \mathbf{q} \in \{0, 1, 2, 3\}^n \setminus \{0^n\}. \quad (\text{I15})$$

Third, as explained in Sec. VII B, if Eq. (I15) holds with $\epsilon = \epsilon/d$, then $\hat{H}$ in Eq. (I11) satisfies

$$\frac{1}{\sqrt{d}} \|\hat{H} - H\|_2 \leq \epsilon. \quad (\text{I16})$$

In conclusion, our protocol learns the unknown Hamiltonian within error ϵ in NFN, with success probability at least $1 - \delta$.

Next, we calculate the number of queries to H and total evolution time required by our protocol. Notice that the Hamiltonian H is queried only in the first step of our protocol, in which we use the CSEU method to estimate linear properties of the time-evolution channel $\mathcal{U}_{t_j}$. Let $1 \leq s \leq d$ be the number of systems collectively measured by our CSEU protocol in each experiment. According to Theorem G1 in Appendix G, for each $1 \leq j \leq L$, in order to guarantee the relation Eq. (I7), our CSEU protocol needs to query $\mathcal{U}_{t_j}$ (i.e., the Hamiltonian H) the following number of times:

$$\begin{aligned} N_1 &= \mathcal{O} \left[\left(\frac{s}{\tilde{\epsilon}^2} + \frac{d^2}{s\tilde{\epsilon}} \right) \log \left(\frac{4^n - 1}{\delta/L} \right) \right] \\ &= \tilde{\mathcal{O}} \left[\left(\frac{sd^2}{\epsilon^2} + \frac{d^3}{s\epsilon} \right) \log(\delta^{-1}) \right], \end{aligned} \quad (\text{I17})$$

where we have used the relations $\text{Tr}(\rho_{\mathbf{q}}^2) = \mathcal{O}(1/d)$ and $\text{Tr}(\sigma_{\mathbf{q}'}^2) = d$, and the assumption $\|H\| = \mathcal{O}(\text{poly}(n))$. In the following, we consider two cases depending on the value of s .

- (1) First, when $s = 1$, our protocol does not use quantum memory and ancillary qubits. In this case, the total number of Hamiltonian queries is given by

$$N_{\text{total}} = LN_1 = \tilde{\mathcal{O}} \left[\left(\frac{d^2}{\epsilon^2} + \frac{d^3}{\epsilon} \right) \log(\delta^{-1}) \right], \quad (\text{I18})$$

and the total evolution time under H reads

$$\begin{aligned} T_{\text{total}} &= N_1 \sum_{j=1}^L t_j \leq N_1 LT \\ &= \tilde{\mathcal{O}} \left[\left(\frac{d^2}{\epsilon^2} + \frac{d^3}{\epsilon} \right) \log(\delta^{-1}) \right]. \end{aligned} \quad (\text{I19})$$

Equations (I18) and (I19) confirm Eq. (I12) in Theorem II.

- (2) Second, when $s = \Theta(\sqrt{d})$, our protocol uses quantum memory. In this case, the total number of queries to H is

$$\begin{aligned} N_{\text{total}} &= LN_1 \\ &= \tilde{O} \left[\left(\frac{d^2 \Theta(\sqrt{d})}{\epsilon^2} + \frac{d^3}{\Theta(\sqrt{d}) \epsilon} \right) \log(\delta^{-1}) \right] \\ &= \tilde{O} \left(\frac{d^{2.5}}{\epsilon^2} \log(\delta^{-1}) \right), \end{aligned} \quad (I20)$$

and the total evolution time reads

$$T_{\text{total}} = N_1 \sum_{j=1}^L t_j \leq N_1 L T = \tilde{O} \left(\frac{d^{2.5}}{\epsilon^2} \log(\delta^{-1}) \right). \quad (I21)$$

Equations (I20) and (I21) confirm Eq. (I13) in Theorem I1.

APPENDIX J: PROOF OF EQ. (60)

1. Main proof

When ρ is the infinite temperature thermal state I/d , and W, V are unitary, traceless, and Hermitian operators, the estimator $\hat{C}(U)$ defined in Eq. (59) can be rewritten as

$$\begin{aligned} \hat{C}(U) &:= \frac{1}{m(m-1)d} \sum_{i \neq j} \hat{D}(i, j), \\ \hat{D}(i, j) &:= \text{Tr} \left[\left(\hat{X}_i \otimes \hat{X}_j \right) (W \otimes V^\top \otimes W \otimes V^\top) T_{(1,3)} \right]. \end{aligned} \quad (J1)$$

So the variance of $\hat{C}(U)$ can be expanded as

$$\text{Var}[\hat{C}(U)] = \frac{\sum_{i \neq j} \sum_{k \neq \ell} \text{Cov}[\hat{D}(i, j), \hat{D}(k, \ell)]}{m^2(m-1)^2 d^2}. \quad (J2)$$

We need to bound all of these covariance terms to bound the variance. When all indices i, j, k, ℓ are distinct, the

covariance is 0 by independence. For the other cases, the covariance can be bounded by

$$\begin{aligned} \text{Cov}[\hat{D}(i, j), \hat{D}(k, \ell)] &= \mathbb{E}[\hat{D}(i, j) \hat{D}(k, \ell)^*] - \mathbb{E}[\hat{D}(i, j)] \mathbb{E}[\hat{D}(k, \ell)^*] \\ &= \mathbb{E}[\hat{D}(i, j) \hat{D}(k, \ell)] - \left| \mathbb{E}[\hat{D}(i, j)] \right|^2 \\ &\leq \mathbb{E}[\hat{D}(i, j) \hat{D}(k, \ell)]. \end{aligned} \quad (J3)$$

Here the second equality holds because $\hat{D}(k, \ell)$ is real. To see this, note that

$$\begin{aligned} \hat{D}(k, \ell)^* &= \text{Tr} \left[T_{(1,3)}^\dagger (W \otimes V^\top \otimes W \otimes V^\top)^\dagger (\hat{X}_k \otimes \hat{X}_\ell)^\dagger \right] \\ &= \text{Tr} \left[(\hat{X}_k \otimes \hat{X}_\ell) T_{(1,3)} (W \otimes V^\top \otimes W \otimes V^\top) \right] \\ &= \hat{D}(k, \ell), \end{aligned} \quad (J4)$$

where the second equality holds because $\hat{X}_k, \hat{X}_\ell, W, V$, and $T_{(1,3)}$ are all Hermitian operators; and the last equality holds because $W \otimes V^\top \otimes W \otimes V^\top$ and $T_{(1,3)}$ commute.

The following lemma summarizes the upper bounds for $\mathbb{E}[\hat{D}(i, j) \hat{D}(k, \ell)]$ in different cases.

Lemma J1. For each combination of i, j, k, ℓ with $i \neq j$ and $k \neq \ell$, the term $\mathbb{E}[\hat{D}(i, j) \hat{D}(k, \ell)]$ is

- (1) $\mathcal{O}\left((d^2/s^2 + d/s + 1)d^2\right)$ when exactly one index matches ($|\{i, j\} \cap \{k, \ell\}| = 1$).
- (2) $\mathcal{O}\left((d^2/s^2 + d/s + 1)^2 d^4\right)$ when both indices match ($|\{i, j\} \cap \{k, \ell\}| = 2$).

For all combinations of $i, j, k, \ell \in \{1, 2, \dots, m\}$ with $i \neq j$ and $k \neq \ell$, there are $\mathcal{O}(m^4)$ combinations where all four indices i, j, k, ℓ are distinct; $\mathcal{O}(m^3)$ combinations where exactly one index matches ($|\{i, j\} \cap \{k, \ell\}| = 1$); and $\mathcal{O}(m^2)$ combinations where both indices match ($|\{i, j\} \cap \{k, \ell\}| = 2$). These facts, together with Eqs. (J2), (J3) and Lemma J1, imply that

$$\text{Var}[\hat{C}(U)] \leq \frac{1}{m^2(m-1)^2 d^2} \mathcal{O} \left[m^3 \left(\frac{d^2}{s^2} + \frac{d}{s} + 1 \right) d^2 + m^2 \left(\frac{d^2}{s^2} + \frac{d}{s} + 1 \right)^2 d^4 \right] = \mathcal{O} \left[\frac{1}{m} \left(\frac{d^2}{s^2} + 1 \right) + \frac{d^2}{m^2} \left(\frac{d^4}{s^4} + 1 \right) \right], \quad (J5)$$

which confirms Eq. (60) in the main text.

2. Proof of the auxiliary Lemma J1

To complete the proof of Eq. (60), in the following we prove Lemma J1, which bounds the term $\mathbb{E}[\hat{D}(i,j)\hat{D}(k,\ell)]$ for all combinations of (i,j,k,ℓ) with $i \neq j$ and $k \neq \ell$. In particular, Appendix J2 a deals with the case in which exactly one index matches ($|\{i,j\} \cap \{k,\ell\}| = 1$); Appendix J2 b deals with the case in which both indices match ($|\{i,j\} \cap \{k,\ell\}| = 2$).

a. Proof of the first statement in Lemma J1

Suppose indices $i,j,k \in \{1, \dots, m\}$ are all distinct. Then we have

$$\begin{aligned} \mathbb{E}[\hat{D}(i,j)\hat{D}(i,k)] &= \mathbb{E}\left\{ \text{Tr}\left[(\hat{X}_i \otimes \hat{X}_j)(W \otimes V^\top \otimes W \otimes V^\top)T_{(1,3)}\right] \text{Tr}\left[(\hat{X}_i \otimes \hat{X}_k)(W \otimes V^\top \otimes W \otimes V^\top)T_{(1,3)}\right] \right\} \\ &= \mathbb{E}\left\{ \left[\text{Tr}\left((\hat{X} \otimes \Upsilon_U)(W \otimes V^\top \otimes W \otimes V^\top)T_{(1,3)}\right) \right]^2 \right\}, \end{aligned} \quad (\text{J6})$$

where the second equality holds because $\mathbb{E}[\hat{X}_j] = \mathbb{E}[\hat{X}_k] = \Upsilon_U$ [see Eq. (10) in the main text] and that $\hat{X}_i, \hat{X}_j, \hat{X}_k$ are independent snapshots. Drawing tensor-network diagrams, we have

$$\begin{aligned} &\text{Tr}[(\hat{X} \otimes \Upsilon_U)(W \otimes V^\top \otimes W \otimes V^\top)T_{(1,3)}] \\ &= \text{Diagram 1} = \text{Diagram 2} = \text{Diagram 3} = \text{Tr}[\hat{X}(WUVU^\dagger W \otimes V^\top)]. \end{aligned} \quad (\text{J7})$$

It follows that

$$\mathbb{E}[\hat{D}(i,j)\hat{D}(i,k)] = \mathbb{E}\left\{ \left(\text{Tr}[\hat{X}(WUVU^\dagger W \otimes V^\top)] \right)^2 \right\} = \text{Tr}\left(\mathbb{E}[\hat{X} \otimes \hat{X}] (WUVU^\dagger W \otimes V^\top \otimes WUVU^\dagger W \otimes V^\top)\right). \quad (\text{J8})$$

The expression of $\mathbb{E}[\hat{X} \otimes \hat{X}]$ in Lemma F1 has 7 terms. Plugging them into Eq. (J8), we get

$$\text{Tr}\left[\tilde{\Delta}_\ell (WUVU^\dagger W \otimes V^\top \otimes WUVU^\dagger W \otimes V^\top)\right] = \begin{cases} \mathcal{O}(d^2) & \ell = 1, 2, 3, 4, \\ 0 & \ell = 5, 6, 7, \end{cases} \quad (\text{J9})$$

which can be derived using tensor-network diagrams. For example,

$$\begin{aligned}
& \text{Tr}[\tilde{\Delta}_2(WUVU^\dagger W \otimes V^\top \otimes WUVU^\dagger W \otimes V^\top)] \\
&= \text{Diagram 1} = \frac{1}{3! \cdot 2!} \sum_{\pi_1 \in S_3} \sum_{\pi_2 \in S_2} \text{Diagram 2} \\
&\stackrel{(a)}{=} \frac{1}{12} (\text{Tr}(V^2) \text{Tr}[(WUVU^\dagger W)^2] + 2 \text{Tr}[(WUVU^\dagger W)^2 U V^2 U^\dagger]) \\
&\stackrel{(b)}{=} \frac{1}{12} [(\text{Tr}(I))^2 + 2 \text{Tr}(I)] = \frac{d^2 + 2d}{12} = \mathcal{O}(d^2), \tag{J10}
\end{aligned}$$

Diagram 1: A tensor network diagram with four input wires at the top. The first and second wires pass through a box labeled $WUVU^\dagger W$. The third and fourth wires pass through a box labeled $WUVU^\dagger W$ and then a box labeled $V^\top$. These four wires then enter a red dashed box labeled $\tilde{\Delta}_2$. Inside this box, the first and second wires are connected to a box labeled U , which is connected to a box labeled $\Pi_{\text{sym}}^{(3)}$. The third and fourth wires are connected to a box labeled $U^\dagger$, which is connected to a box labeled $\Pi_{\text{sym}}^{(2)}$. The output wires from $\Pi_{\text{sym}}^{(3)}$ and $\Pi_{\text{sym}}^{(2)}$ are connected to each other.

Diagram 2: A tensor network diagram with four input wires at the top. The first and second wires pass through a box labeled $WUVU^\dagger W$. The third and fourth wires pass through a box labeled $WUVU^\dagger W$ and then a box labeled V . These four wires then enter a box labeled T_{π_1} . The output wires from T_{π_1} are connected to a box labeled $U^\dagger$, which is connected to a box labeled T_{π_2} . The output wires from T_{π_2} are connected to each other.

where (a) holds because $\text{Tr}(V) = \text{Tr}(WUVU^\dagger W) = 0$, and (b) holds because $(WUVU^\dagger W)^2 = V^2 = I$.

Equations (J8), (J9), and Lemma F1 together imply that

$$\mathbb{E}[\hat{D}(i,j)\hat{D}(i,k)] = \frac{2d^2(d+1)^2(d+s)}{s^2(d+s+1)} \left(\frac{1}{\kappa_2} + \frac{s}{\kappa_3} + \frac{s}{\kappa_3} + \frac{s(s-1)}{2\kappa_4} \right) \mathcal{O}(d^2) = \mathcal{O}\left[\left(\frac{d^2}{s^2} + \frac{d}{s} + 1\right)d^2\right]. \tag{J11}$$

In addition, we have

$$\mathbb{E}[\hat{D}(i,j)\hat{D}(k,i)] = \mathbb{E}[\hat{D}(i,j)\hat{D}(i,k)] = \mathcal{O}\left[\left(\frac{d^2}{s^2} + \frac{d}{s} + 1\right)d^2\right], \tag{J12}$$

where the first equality holds because $\hat{D}(k,i) = \hat{D}(i,k)$ by symmetry. Equations (J11) and (J12) together confirm the first statement in Lemma J1.

b. Proof of the second statement in Lemma J1

Suppose indices $i, j \in \{1, \dots, m\}$ are distinct. Then we have

$$\begin{aligned}
\mathbb{E}[\hat{D}(i,j)\hat{D}(i,j)] &= \mathbb{E}\left\{\left[\text{Tr}\left((\hat{X}_i \otimes \hat{X}_j)(W \otimes V^\top \otimes W \otimes V^\top)T_{(1,3)}\right)\right]^2\right\} \\
&= \text{Tr}\left[\left(\mathbb{E}[\hat{X} \otimes \hat{X}] \otimes \mathbb{E}[\hat{X} \otimes \hat{X}]\right)(W \otimes V^\top \otimes W \otimes V^\top \otimes W \otimes V^\top \otimes W \otimes V^\top)T_{(1,5)(3,7)}\right]. \tag{J13}
\end{aligned}$$

Define

$$\gamma(\mu, \nu) := \text{Tr}\left[\left(\tilde{\Delta}_\mu \otimes \tilde{\Delta}_\nu\right)(W \otimes V^\top \otimes W \otimes V^\top \otimes W \otimes V^\top \otimes W \otimes V^\top)T_{(1,5)(3,7)}\right], \tag{J14}$$

which can be bounded via tensor-network diagrams:

$$\gamma(\mu, \nu) = \begin{cases} \mathcal{O}(d^4) & \mu, \nu \in \{1, 2, 3, 4\}, \\ 0 & \text{otherwise.} \end{cases} \tag{J15}$$

The proof of Eq. (J15) is a simple analog of Eq. (J9) and is omitted.

According to Lemma F1, we can then expand the right-hand side of Eq. (J13) as

Right-hand side of Eq. (J13)

$$\begin{aligned}
&= \mathcal{O}\left(\frac{d^4}{s^4}\gamma(1,1) + \frac{d^3}{s^3}[\gamma(1,2) + \gamma(1,3) + \gamma(2,1) + \gamma(3,1)] + \frac{d}{s}[\gamma(2,4) + \gamma(3,4) + \gamma(4,2) + \gamma(4,3)] \right. \\
&\quad + \frac{d^2}{s^2}[\gamma(2,2) + \gamma(2,3) + \gamma(1,4) + \gamma(3,2) + \gamma(3,3) + \gamma(4,1)] + \gamma(4,4) + \frac{(d+s)^4}{s^4}\gamma(7,7) \\
&\quad + \frac{d(d+s)}{s^2}[\gamma(3,5) + \gamma(3,6) + \gamma(2,5) + \gamma(2,6) + \gamma(5,2) + \gamma(5,3) + \gamma(6,2) + \gamma(6,3)] \\
&\quad + \frac{d^2(d+s)}{s^3}[\gamma(1,5) + \gamma(1,6) + \gamma(5,1) + \gamma(6,1)] + \frac{(d+s)}{s}[\gamma(4,5) + \gamma(4,6) + \gamma(5,4) + \gamma(6,4)] \\
&\quad + \frac{(d+s)^3}{s^3}[\gamma(5,7) + \gamma(6,7) + \gamma(7,5) + \gamma(7,6)] + \frac{d(d+s)^2}{s^3}[\gamma(2,7) + \gamma(3,7) + \gamma(7,2) + \gamma(7,3)] \\
&\quad \left. + \frac{(d+s)^2}{s^2}[\gamma(5,5) + \gamma(5,6) + \gamma(6,5) + \gamma(6,6) + \gamma(4,7) + \gamma(7,4)] + \frac{d^2(d+s)^2}{s^4}[\gamma(1,7) + \gamma(7,1)]\right) \\
&= \mathcal{O}\left[\left(\frac{d^2}{s^2} + \frac{d}{s} + 1\right)^2 d^4\right], \tag{J16}
\end{aligned}$$

where the last equality follows from Eq. (J15). Therefore, we have

$$\mathbb{E}[\hat{D}(i,j)\hat{D}(j,i)] = \mathbb{E}[\hat{D}(i,j)\hat{D}(i,j)] = \mathcal{O}\left[\left(\frac{d^2}{s^2} + \frac{d}{s} + 1\right)^2 d^4\right], \tag{J17}$$

where the first equality holds because $\hat{D}(j,i) = \hat{D}(i,j)$. This equation confirms the second statement in Lemma J1.

-
- [1] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, *Nat. Rev. Phys.* **5**, 141 (2023).
 - [2] J. Carrasco, A. Elben, C. Kokail, B. Kraus, and P. Zoller, Theoretical and experimental perspectives of quantum verification, *PRX Quantum* **2**, 010102 (2021).
 - [3] M. Kliesch and I. Roth, Theory of quantum system certification, *PRX Quantum* **2**, 010201 (2021).
 - [4] M. Mohseni and D. A. Lidar, Direct characterization of quantum dynamics, *Phys. Rev. Lett.* **97**, 170501 (2006).
 - [5] M. C. Caro, H.-Y. Huang, N. Ezzell, J. Gibbs, A. T. Sornborger, L. Cincio, P. J. Coles, and Z. Holmes, Out-of-distribution generalization for learning quantum dynamics, *Nat. Commun.* **14**, 3751 (2023).
 - [6] J. Haah, R. Kothari, R. O'Donnell, and E. Tang, Query-optimal estimation of unitary channels in diamond distance, in *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Santa Cruz, CA, USA, 2023), pp. 363–390.
 - [7] S. Aaronson, Shadow tomography of quantum states, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, Los Angeles, CA, USA, 2018), pp. 325–338.

- [8] C. Bădescu and R. O'Donnell, Improved Quantum data analysis, in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, Virtual, Italy, 2021), pp. 1398–1411.
- [9] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [10] C. Bertoni, J. Haferkamp, M. Hinsche, M. Ioannou, J. Eisert, and H. Pashayan, Shallow shadows: Expectation estimation using low-depth random Clifford circuits, *Phys. Rev. Lett.* **133**, 020602 (2024).
- [11] D. Grier, H. Pashayan, and L. Schaeffer, Sample-optimal classical shadows for pure states, *Quantum* **8**, 1373 (2024).
- [12] H.-Y. Hu, S. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, *Phys. Rev. Res.* **5**, 023027 (2023).
- [13] B. Vermersch, A. Rath, B. Sundar, C. Branciard, J. Preskill, and A. Elben, Enhanced estimation of quantum properties with common randomized measurements, *PRX Quantum* **5**, 010352 (2024).
- [14] J. Helsen and M. Walter, Thrifty shadow estimation: Reusing quantum circuits and bounding tails, *Phys. Rev. Lett.* **131**, 240602 (2023).
- [15] Y. Zhou and Q. Liu, Performance analysis of multi-shot shadow estimation, *Quantum* **7**, 1044 (2023).

- [16] M. Arienzo, M. Heinrich, I. Roth, and M. Kliesch, Closed-form analytic expressions for shadow estimation with brickwork circuits, *Quantum Inf. Comput.* **23**, 961 (2023).
- [17] Y. Zhou and Z. Liu, A hybrid framework for estimating nonlinear functions of quantum states, *npj Quantum Inf.* **10**, 62 (2024).
- [18] Q. Zhang, Q. Liu, and Y. Zhou, Minimal-Clifford shadow estimation by mutually unbiased bases, *Phys. Rev. Appl.* **21**, 064001 (2024).
- [19] Q. Liu, Z. Li, X. Yuan, H. Zhu, and Y. Zhou, Auxiliary-free replica shadow estimation, *arXiv:2407.20865*.
- [20] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, *Nat. Rev. Phys.* **5**, 9 (2023).
- [21] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, Shadow process tomography of quantum channels, *Phys. Rev. A* **107**, 042403 (2023).
- [22] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, *Phys. Rev. Res.* **6**, 013029 (2024).
- [23] H.-Y. Huang, S. Chen, and J. Preskill, Learning to predict arbitrary quantum processes, *PRX Quantum* **4**, 040337 (2023).
- [24] M. C. Caro, Learning quantum processes and Hamiltonians via the Pauli transfer matrix, *ACM Trans. Quantum Comput.* **5**, 1 (2024).
- [25] H. Zhao, L. Lewis, I. Kannan, Y. Quek, H.-Y. Huang, and M. C. Caro, Learning quantum states and unitaries of bounded gate complexity, *PRX Quantum* **5**, 040306 (2024).
- [26] H.-Y. Huang, Y. Liu, M. Broughton, I. Kim, A. Anshu, Z. Landau, and J. R. McClean, Learning shallow quantum circuits, in *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, Vancouver, BC, Canada, 2024), pp. 1343–1351.
- [27] H.-Y. Huang, R. Kueng, and J. Preskill, Information-theoretic bounds on quantum advantage in machine learning, *Phys. Rev. Lett.* **126**, 190505 (2021).
- [28] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, Generalization in quantum machine learning from few training data, *Nat. Commun.* **13**, 4919 (2022).
- [29] 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).
- [30] T. Surawy-Stepney, J. Kahn, R. Kueng, and M. Guta, Projected least-squares quantum process tomography, *Quantum* **6**, 844 (2022).
- [31] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195 (2017).
- [32] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, *Phys. Rev. Lett.* **122**, 040504 (2019).
- [33] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [34] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos *et al.*, Self-verifying variational quantum simulation of lattice models, *Nature* **569**, 355 (2019).
- [35] J. Gibbs, Z. Holmes, M. C. Caro, N. Ezzell, H.-Y. Huang, L. Cincio, A. T. Sornborger, and P. J. Coles, Dynamical simulation via quantum machine learning with provable generalization, *Phys. Rev. Res.* **6**, 013241 (2024).
- [36] D. Stilck França, L. A. Markovich, V. Dobrovitski, A. H. Werner, and J. Borregaard, Efficient and robust estimation of many-qubit Hamiltonians, *Nat. Commun.* **15**, 311 (2024).
- [37] A. Gu, L. Cincio, and P. J. Coles, Practical Hamiltonian learning with unitary dynamics and Gibbs states, *Nat. Commun.* **15**, 312 (2024).
- [38] H. Zhu, Multiqubit Clifford groups are unitary 3-designs, *Phys. Rev. A* **96**, 062336 (2017).
- [39] Z. Webb, The Clifford group forms a unitary 3-design, *Quantum Inf. Comput.* **16**, 1379 (2016).
- [40] 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, Denver, CO, USA, 2022), pp. 574–585.
- [41] Some references alternatively call Hamiltonian learning protocols without (with) quantum memory as incoherent (coherent) protocols [65], or protocols without (with) quantum control [68].
- [42] S. Massar and S. Popescu, Optimal extraction of information from finite quantum ensembles, *Phys. Rev. Lett.* **74**, 1259 (1995).
- [43] M. Hayashi, Asymptotic estimation theory for a finite-dimensional pure state model, *J. Phys. A: Math. Gen.* **31**, 4633 (1998).
- [44] D. Bruß and C. Macchiavello, Optimal state estimation for d -dimensional quantum systems, *Phys. Lett. A* **253**, 249 (1999).
- [45] H. Zhu and M. Hayashi, Universally Fisher-symmetric informationally complete measurements, *Phys. Rev. Lett.* **120**, 030404 (2018).
- [46] A. W. Harrow, The church of the symmetric subspace, *arXiv:1308.6595*.
- [47] 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* (Association for Computing Machinery, Rome, Italy, 2022), pp. 44–51.
- [48] Z. Hou, J.-F. Tang, J. Shang, H. Zhu, J. Li, Y. Yuan, K.-D. Wu, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, Deterministic realization of collective measurements via photonic quantum walks, *Nat. Commun.* **9**, 1414 (2018).
- [49] L. O. Conlon, T. Vogl, C. D. Marciniak, I. Pogorelov, S. K. Yung, F. Eilenberger, D. W. Berry, F. S. Santana, R. Blatt, T. Monz *et al.*, Approaching optimal entangling collective measurements on quantum computing platforms, *Nat. Phys.* **19**, 351 (2023).
- [50] K. Zhou, C. Yi, W.-Z. Yan, Z. Hou, H. Zhu, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, Experimental realization of genuine three-copy collective measurements for optimal information extraction, *Phys. Rev. Lett.* **134**, 210201 (2025).
- [51] K.-D. Wu, E. Bäumer, J.-F. Tang, K. V. Hovhannisyan, M. Perarnau-Llobet, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, Minimizing backaction through entangled measurements, *Phys. Rev. Lett.* **125**, 210401 (2020).

- [52] M.-D. Choi, Completely positive linear maps on complex matrices, *Linear Algebra Appl.* **10**, 285 (1975).
- [53] R. Kothari, An optimal quantum algorithm for the oracle identification problem, in *31st International Symposium on Theoretical Aspects of Computer Science* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2014), pp. 482–493.
- [54] S. Arunachalam and R. de Wolf, Guest column: A survey of quantum learning theory, *ACM SIGACT News* **48**, 41 (2017).
- [55] B. Collins and I. Nechita, Random matrix techniques in quantum information theory, *J. Math. Phys.* **57**, 015215 (2016).
- [56] I. Nechita, Asymptotics of random density matrices, *Ann. Henri Poincaré* **8**, 1521 (2007).
- [57] K. Życzkowski, K. A. Penson, I. Nechita, and B. Collins, Generating random density matrices, *J. Math. Phys.* **52**, 062201 (2011).
- [58] E. Lubkin, Entropy of an n -system from its correlation with a k -reservoir, *J. Math. Phys.* **19**, 1028 (1978).
- [59] J. L. O’Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, Quantum process tomography of a controlled-NOT gate, *Phys. Rev. Lett.* **93**, 080502 (2004).
- [60] Y. Yang, R. Renner, and G. Chiribella, Optimal universal programming of unitary gates, *Phys. Rev. Lett.* **125**, 210501 (2020).
- [61] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, *Nat. Photonics* **5**, 222 (2011).
- [62] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, *Rev. Mod. Phys.* **89**, 035002 (2017).
- [63] M. D. Shulman, S. P. Harvey, J. M. Nichol, S. D. Bartlett, A. C. Doherty, V. Umansky, and A. Yacoby, Suppressing qubit dephasing using real-time Hamiltonian estimation, *Nat. Commun.* **5**, 5156 (2014).
- [64] L. Innocenti, L. Banchi, A. Ferraro, S. Bose, and M. Paternostro, Supervised learning of time-independent Hamiltonians for gate design, *New J. Phys.* **22**, 065001 (2020).
- [65] A. Bluhm, M. C. Caro, and A. Oufkir, Hamiltonian property testing, [arXiv:2403.02968](https://arxiv.org/abs/2403.02968).
- [66] M. Ma, S. T. Flammia, J. Preskill, and Y. Tong, Learning k -body Hamiltonians via compressed sensing, [arXiv:2410.18928](https://arxiv.org/abs/2410.18928).
- [67] Q. Zhao, Y. Zhou, A. F. Shaw, T. Li, and A. M. Childs, Hamiltonian simulation with random inputs, *Phys. Rev. Lett.* **129**, 270502 (2022).
- [68] A. Dutkiewicz, T. E. O’Brien, and T. Schuster, The advantage of quantum control in many-body Hamiltonian learning, *Quantum* **8**, 1537 (2024).
- [69] E. Bairey, I. Arad, and N. H. Lindner, Learning a local Hamiltonian from local measurements, *Phys. Rev. Lett.* **122**, 020504 (2019).
- [70] W. Yu, J. Sun, Z. Han, and X. Yuan, Robust and efficient Hamiltonian learning, *Quantum* **7**, 1045 (2023).
- [71] H.-Y. Hu, M. Ma, W. Gong, Q. Ye, Y. Tong, S. T. Flammia, and S. F. Yelin, Ansatz-free Hamiltonian learning with Heisenberg-limited scaling, accepted by *PRX Quantum*.
- [72] A. Zhao, Learning the structure of any Hamiltonian from minimal assumptions, in *Proceedings of the 57th Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, Prague, Czechia, 2025), pp. 1201–1211.
- [73] H.-Y. Huang, Y. Tong, D. Fang, and Y. Su, Learning many-body Hamiltonians with Heisenberg-limited scaling, *Phys. Rev. Lett.* **130**, 200403 (2023).
- [74] J. Haah, R. Kothari, and E. Tang, Learning quantum Hamiltonians from high-temperature Gibbs states and real-time evolutions, *Nat. Phys.* **20**, 1027 (2024).
- [75] J. Castaneda and N. Wiebe, Hamiltonian learning via shadow tomography of pseudo-Choi states, *Quantum* **9**, 1700 (2025).
- [76] A. Bakshi, A. Liu, A. Moitra, and E. Tang, Structure learning of Hamiltonians from real-time evolution, in *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Chicago, IL, USA, 2024), pp. 1037–1050.
- [77] S. Xu and B. Swingle, Scrambling dynamics and out-of-time-ordered correlators in quantum many-body systems, *PRX Quantum* **5**, 010201 (2024).
- [78] C.-F. Chen, A. Lucas, and C. Yin, Speed limits and locality in many-body quantum dynamics, *Rep. Prog. Phys.* **86**, 116001 (2023).
- [79] H. Shen, P. Zhang, Y.-Z. You, and H. Zhai, Information scrambling in quantum neural networks, *Phys. Rev. Lett.* **124**, 200504 (2020).
- [80] C. W. von Keyserlingk, T. Rakovszky, F. Pollmann, and S. L. Sondhi, Operator hydrodynamics, OTOCs, and entanglement growth in systems without conservation laws, *Phys. Rev. X* **8**, 021013 (2018).
- [81] B. Vermersch, A. Elben, L. M. Sieberer, N. Y. Yao, and P. Zoller, Probing scrambling using statistical correlations between randomized measurements, *Phys. Rev. X* **9**, 021061 (2019).
- [82] R. J. Garcia, Y. Zhou, and A. Jaffe, Quantum scrambling with classical shadows, *Phys. Rev. Res.* **3**, 033155 (2021).
- [83] L. Leone, S. F. E. Oliviero, Y. Zhou, and A. Hama, Quantum chaos is quantum, *Quantum* **5**, 453 (2021).
- [84] A. A. Mele, Introduction to Haar measure tools in quantum information: A beginner’s tutorial, *Quantum* **8**, 1340 (2024).