

Optimal Randomized Measurements for a Family of Nonlinear Quantum Properties

Zhenyu Du^{1,*}, Yifan Tang^{2,*}, Andreas Elben^{3,4}, Ingo Roth⁵, Jens Eisert², and Zhenhuan Liu^{1,†}

¹Center for Quantum Information, Institute for Interdisciplinary Information Sciences, *Tsinghua University*, Beijing 100084, China

²Dahlem Center for Complex Quantum Systems, *Freie Universität Berlin*, Berlin 14195, Germany

³PSI Center for Scientific Computing, Theory and Data, *Paul Scherrer Institute*, 5232 Villigen PSI, Switzerland

⁴ETHZ-PSI Quantum Computing Hub, *Paul Scherrer Institute*, 5232 Villigen PSI, Switzerland

⁵Quantum Research Center, *Technology Innovation Institute*, Masdar City, Abu Dhabi, United Arab Emirates



(Received 24 June 2025; accepted 12 December 2025; published 26 March 2026)

Quantum learning encounters fundamental challenges when estimating nonlinear properties, owing to the inherent linearity of quantum mechanics. Although recent advances in single-copy randomized measurement protocols have achieved optimal sample complexity for specific tasks like state purity estimation, generalizing these protocols to estimate broader classes of nonlinear properties without sacrificing optimality remains an open problem. In this work, we introduce the observable-driven randomized measurement (ORM) protocol enabling the estimation of $\text{Tr}(O\rho^2)$ for an arbitrary observable O —an essential quantity in quantum computing and many-body physics. We establish an upper bound for ORM's sample complexity and show its optimality for observables with a large trace-norm, including Pauli and local observables, closing a gap in the literature. For these observables, ORM admits an efficient implementation with Clifford circuits. Numerical experiments validate that ORM requires substantially fewer state samples to achieve the same precision compared to classical shadows. Additionally, we introduce a braiding randomized measurement protocol for multiple low-rank nonlinear observables, reducing circuit complexities in practical applications.

DOI: [10.1103/4rkx-xkwy](https://doi.org/10.1103/4rkx-xkwy)

I. INTRODUCTION

Quantum learning is pivotal for both verifying and extracting information from quantum systems, addressing the growing demands of quantum technologies while at the same time enabling transformative applications [1–5]. To date, the majority of quantum learning theory has focused on the estimation of linear properties of quantum states. However, nonlinear properties, like the state purity $\text{Tr}(\rho^2)$ and the nonlinear expectation value $\text{Tr}(O\rho^2)$, play vital roles in quantum simulation for virtual cooling [6,7], quantum error mitigation [8–10], and instances of quantum algorithms [11,12]. Specifically, estimating $\text{Tr}(O\rho^2)/\text{Tr}(\rho^2)$ yields estimates of a quantum system at a lower temperature than the physically prepared thermal

state, thereby suppressing potential incoherent noise. Further applications include inferring entropic quantities and the quantum Fisher information of an unknown quantum state [13,14] and detecting quantum phases of matter in mixed states [15–17].

Yet, the intrinsic linearity of quantum mechanics renders the estimation of nonlinear properties considerably more challenging than that of a linear expectation value $\text{Tr}(O\rho)$ of an observable O . For linear estimation, directly measuring single state copies in the observable's eigenbasis has optimal sample complexity in the Hilbert space dimension. While quantum memory enables entangled measurements across multiple state copies, such resources only offer exponential advantage in more complex tasks [11,18–20].

In contrast, when estimating a nonlinear property like $\text{Tr}(O\rho^2)$, directly applying observable-driven techniques from linear estimation unavoidably requires coherently controlled quantum memory [21–28] and renders implementations on quantum devices challenging [25–27]. Existing protocols using only single-copy quantum operations instead resort to general informationally complete measurements independent of the observable [29,30] and postprocess tensor products of so-called *classical shadows*

*Z.D. and Y.T. contributed equally to this work.

†Contact author: qubithuan@gmail.com

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.

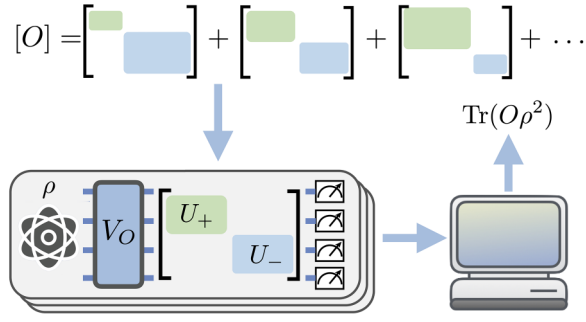


FIG. 1. The framework of observable-driven randomized measurement protocol. To estimate $\text{Tr}(O\rho^2)$ for arbitrary observables O , the observable-driven randomized measurement protocol first decomposes O into a few dichotomic observables. For each dichotomic observable, one rotates the target state into the eigenbasis of the observable, evolves it with a block-diagonal random unitary, measures in the computational basis, and postprocesses the measurement results to estimate its nonlinear expectation value.

(CS) of the state [31,32]. Current lower bounds $\Omega(\sqrt{d})$, where d is the dimension of Hilbert space, suggest that these protocols consuming $\mathcal{O}(d)$ state copies are indeed suboptimal [11]. Purity estimation can be optimally implemented using *randomized measurements* (RMs) [2,18,19,33–37]. How to optimally estimate $\text{Tr}(O\rho^2)$ given a general observable O from single state copies—a nonlinear analog of directly measuring O —is in fact an open question.

In this work, we answer this question by proposing an observable-driven protocol and establishing its optimality for a wide range of observables in the single-copy regime. Our protocol generalizes the optimal RM protocol for purity estimation to the nonlinear expectation value $\text{Tr}(O\rho^2)$ of arbitrary observables O . The RM ensemble incorporates information about the target observable O . As shown in Fig. 1, our *observable-driven randomized measurement* (ORM) protocol begins by decomposing the target observable into a few dichotomic observables,

each of which has only two distinct eigenspaces. For each dichotomic component, we design an RM protocol that employs block-diagonal random unitary evolutions to estimate its nonlinear expectation value. Our analysis shows that for general observables with bounded operator norm on a quantum system of dimension d , the ORM protocol exhibits a sample complexity scaling of $\mathcal{O}(\sqrt{d})$. Importantly, this scaling matches the fundamental lower bound for a broad class of observables, including all Pauli and local observables, across all single-copy schemes [11,38], and thereby is provably optimal. The ORM protocol also offers significant practical advantages. It is sufficient to use a constant number of measurement bases for constant additive error and failure rate, independent of the system size or the target observable. Furthermore, the protocol features exceptionally simple postprocessing, alleviating the need for precise, gate-by-gate calibration of the unitary operations.

To enhance the practicality of our ORM protocol, we construct more efficient circuit implementations for practical observables. We first simplify the protocol for Pauli observables by introducing mid-circuit measurements and subsystem random unitaries and propose a variant requiring only single-qubit gates. Built on this, we develop an alternative Clifford-circuit implementation that also achieves the optimal $\mathcal{O}(\sqrt{d})$ sample complexity. We further extend our analysis to other unitary ensembles such as approximate designs, shallow circuits, and local unitaries to establish the protocol’s robustness. Finally, via a Pauli sampling protocol, we generalize these efficient primitives to a broad class of physical observables. For local Hamiltonians and stabilizer fidelities, this approach preserves the $\mathcal{O}(\sqrt{d})$ sample complexity while significantly reducing circuit complexities.

Additionally, for constant-rank observables, we propose the *braiding randomized measurement* (BRM) protocol, which simultaneously combines the strengths of randomized measurement and classical shadows. Specifically, BRM achieves the same sample complexity scaling as both classical shadows and ORM for constant-rank observables.

TABLE I. The comparison of classical shadows and observable-driven randomized measurement in estimating $\text{Tr}(O\rho^2)$ with constant additive error and failure probability.

	Observables	Sample complexity	Measurement bases	Random unitary	Information of unitary in postprocessing?	Estimating M independent
CS	$\ O\ _\infty \leq 1$	$\mathcal{O}(\sqrt{d\text{Tr}(O^2)})$	$\mathcal{O}(\sqrt{d\text{Tr}(O^2)})$	Unitary 3-design	Yes	$\mathcal{O}(\log(M))$
ORM	$\ O\ _\infty \leq 1$	$\mathcal{O}(\sqrt{d})$	$\mathcal{O}(1)$	Block-diagonal unitary 4-design	No	$\mathcal{O}(M)$
ORM	Pauli observables, local observables, and stabilizer fidelities	$\mathcal{O}(\sqrt{d})$	$\mathcal{O}(1)$	Clifford	No	$\mathcal{O}(M)$

TABLE II. The comparison of classical shadows, observable-driven randomized measurement, and braiding randomized measurement protocols in estimating low-rank nonlinear observables with constant additive error and failure probability.

	Observables	Sample complexity	Measurement bases	Random unitary	Information of unitary in postprocessing?	Estimating M independent
CS	$\text{Tr}(O^2) = \mathcal{O}(1)$	$\mathcal{O}(\sqrt{d})$	$\mathcal{O}(\sqrt{d})$	Unitary 3-design	Yes	$\mathcal{O}(\log(M))$
ORM	$\text{Tr}(O^2) = \mathcal{O}(1)$	$\mathcal{O}(\sqrt{d})$	$\mathcal{O}(1)$	Block-diagonal unitary 4-design	No	$\mathcal{O}(M)$
ORM	Stabilizer fidelities	$\mathcal{O}(\sqrt{d})$	$\mathcal{O}(1)$	Clifford	No	$\mathcal{O}(M)$
BRM	$\text{Tr}(O^2) = \mathcal{O}(1)$	$\mathcal{O}(\sqrt{d})$	$\mathcal{O}(1)$	Unitary 6-design	Yes	$\mathcal{O}(\log(M))$

Moreover, it significantly reduces the number of required measurement bases compared to classical shadows and offers more favorable scaling when simultaneously estimating multiple observables, compared to ORM.

This work is organized as follows. Section II revisits the RM protocol for purity estimation. Section III introduces the ORM protocol and analyzes its sample complexity. Section IV further simplifies the circuit implementation of our protocols for a broad class of observables. We first present a version tailored to Pauli observables that admits a Clifford-circuit implementation and then discuss extensions using approximate unitary designs, low-depth circuits, and local measurements. We also propose a Pauli sampling scheme that enables simple implementations for estimating local observables and stabilizer fidelities. Section V demonstrates applications of ORM to error mitigation, mixed-phase detection, and the estimation of entropic quantities; in particular, we numerically apply ORM to a virtual cooling task and show its advantage over classical shadows. Section VI introduces the BRM protocol for estimating low-rank observables. Finally, Sec. VII summarizes our main results and outlines directions for future research. Our results are summarized in Tables I and II.

II. REVISIT: RANDOMIZED MEASUREMENT PROTOCOL FOR PURITY ESTIMATION

Randomness plays a crucial role in quantum information processing, such as in quantum sampling advantage [39,40], device benchmarking [41,42], noise suppression [43–45], efficient quantum state tomography [37], and quantum cryptography [46]. RM has emerged as a key tool in quantum learning [2,33,35,36], usually offering highly attractive performance compared to conventional tomography-based methods, while avoiding resource-intensive operations such as quantum memory or deep quantum circuits.

To motivate our method, we first revisit how to use an RM protocol to estimate the purity of a quantum state ρ [2,34]. Throughout this work, we assume ρ is an n -qubit

d -dimensional quantum state. The main idea is to perform repeated measurements on multiple random bases and then postprocess the measurement outcomes. Specifically, we draw N_U independent random unitaries from a suitable random unitary ensemble (often taken to be a unitary design [47]; see Appendix A 1) to rotate the measurement basis. For each choice of U , we apply it to ρ , generating $U\rho U^\dagger$, and then perform N_M computational-basis measurements on the evolved state. Each measurement produces an n -bit outcome string s_i , and we denote the outcomes of all N_M measurements as $\{s_i\}_{i=1}^{N_M}$. Subsequently, we perform classical postprocessing on these measurement outcomes to estimate the purity. When the random unitary is sampled from a unitary two-design, the postprocessing function is given by

$$X_2(s_1, s_2) = \begin{cases} d, & s_1 = s_2, \\ -1, & s_1 \neq s_2, \end{cases} \quad (1)$$

and the unbiased estimator for $\text{Tr}(\rho^2)$ is chosen as

$$\hat{M}_2^U = \binom{N_M}{2}^{-1} \sum_{1 \leq i < j \leq N_M} X_2(s_i, s_j). \quad (2)$$

We can further average the estimators with all N_U choices of unitaries $\{U_u\}_{u=1}^{N_U}$ to obtain

$$\hat{M}_2 = \frac{1}{N_U} \sum_{u=1}^{N_U} \hat{M}_2^{U_u}. \quad (3)$$

Each $\hat{M}_2^U$ is an unbiased estimator of $\text{Tr}(\rho^2)$ [2]. Moreover, if the random unitary ensemble forms a unitary 4-design, then the variance of the estimator $\hat{M}_2$ scales roughly as $d/(N_U N_M^2)$ [48,49]. Thus, by setting $N_M = \mathcal{O}(\sqrt{d})$ and $N_U = \mathcal{O}(1)$, the purity can be estimated with constant additive error, saturating the lower bound on sample complexity for purity estimation using single-copy operations [18,19]. Another feature of the RM protocol is that, while the number of measurements in a single basis N_M scales

exponentially with the system size, selecting a constant number of measurement bases N_U suffices for accurate state purity estimation [2,50]. Moreover, its postprocessing does not require knowledge of the applied unitaries, providing advantages over classical-shadow-based quantum learning protocols in terms of computational cost for postprocessing and robustness against miscalibration.

The core idea of making $\hat{M}_2^U$ an unbiased estimator of $\text{Tr}(\rho^2)$ is that the postprocessing function $X_2(s_1, s_2)$ in Eq. (1) virtually measures the SWAP observable between two copies of ρ [2]. Specifically, we define a diagonal observable,

$$X_2 = \sum_{s_i, s_j} X_2(s_i, s_j) |s_i, s_j\rangle\langle s_i, s_j| \quad (4)$$

and draw U from a unitary 2-design ensemble, we have

$$\mathbb{E}_U [U^{\dagger \otimes 2} X_2 U^{\otimes 2}] = \mathbb{S}, \quad (5)$$

where $\mathbb{S}$ is the SWAP operator. Following this line of thought, we get

$$\begin{aligned} & \mathbb{E}_{U, \{s_i\}_i} [\hat{M}_2^U] \\ &= \mathbb{E}_U \left[\sum_{s_i, s_j} X_2(s_i, s_j) \langle s_i | U \rho U^\dagger | s_i \rangle \langle s_j | U \rho U^\dagger | s_j \rangle \right] \\ &= \mathbb{E}_U \left[\sum_{s_i, s_j} X_2(s_i, s_j) \text{Tr}(|s_i, s_j\rangle\langle s_i, s_j| U^{\otimes 2} \rho^{\otimes 2} U^{\dagger \otimes 2}) \right] \\ &= \mathbb{E}_U [\text{Tr}(U^{\dagger \otimes 2} X_2 U^{\otimes 2} \rho^{\otimes 2})] = \text{Tr}(\mathbb{S} \rho^{\otimes 2}) = \text{Tr}(\rho^2). \quad (6) \end{aligned}$$

The construction of the RM protocol utilizes the property of purity that it is invariant under the action of an arbitrary unitary operation, i.e., $[\mathbb{S}, U^{\otimes 2}] = 0$. Thus, the measurement statistics on a random basis reflect the information of state purity. Following this logic, by incorporating the properties of the target nonlinear observable, we can design an RM protocol to estimate more general state functions.

III. OBSERVABLE-DRIVEN RANDOMIZED MEASUREMENT PROTOCOL

In this section, we present our main result: an observable-driven randomized measurement protocol to estimate the nonlinear quantity $\text{Tr}(O\rho^2)$ for arbitrary observables O . As the name suggests, the ORM protocol uses random unitaries drawn from an ensemble specific to the target observable O . We prove that ORM accurately estimates $\text{Tr}(O\rho^2)$ with high probability using only $\mathcal{O}(\sqrt{d})$ copies of ρ for all observables with $\|O\|_\infty \leq 1$. The result is the first to match the lower bounds derived in Refs. [11,38] and concludes an open theoretical question.

A. Estimating dichotomic observables

We first consider a dichotomic observable (i.e., a Hermitian matrix whose eigenvalues are $+1$ and -1), which is the foundation for developing an estimation protocol for general observables. A representative example is $O = Z_1 = Z \otimes I_{n-1}$, which acts as the Pauli-Z operator on the first qubit and trivially on the remaining qubits. For any single-copy protocol, estimating $\text{Tr}(Z_1 \rho^2)$ to constant precision with constant failure probability has a sample complexity lower bound of $\Omega(\sqrt{d})$ [11]. In Appendix C, we show that this lower bound can be generalized to the estimation of all Pauli observables. In the meantime, classical shadows require

$$\mathcal{O}(\sqrt{d \text{Tr}(O^2)}) = \mathcal{O}(d) \quad (7)$$

many samples [31,51,52], resulting in a quadratic gap relative to the lower bound. This gap primarily arises because the quantum experiments in classical shadows do not utilize information about the target observable. Here, we leverage the knowledge of O to design a protocol that closes this gap.

Given the dichotomic observable O with two eigenspaces $\Pi_\pm$ (eigenvalues ± 1 and projectors $P_\pm$) and corresponding dimensions $d_\pm$, the target state ρ can be written in a block form with respect to these subspaces,

$$\rho = \begin{bmatrix} \tilde{\rho}_+ & B \\ B^\dagger & \tilde{\rho}_- \end{bmatrix}, \quad (8)$$

where $\tilde{\rho}_\pm = P_\pm \rho P_\pm$. A key observation is that $\text{Tr}(O\rho^2)$ can be expressed as a difference of purities within each eigen-subspace,

$$\begin{aligned} \text{Tr}(O\rho^2) &= \text{Tr} \left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \tilde{\rho}_+ & B \\ B^\dagger & \tilde{\rho}_- \end{bmatrix}^2 \right) \\ &= \text{Tr}(\tilde{\rho}_+^2) - \text{Tr}(\tilde{\rho}_-^2). \quad (9) \end{aligned}$$

Therefore, our idea for estimating $\text{Tr}(O\rho^2)$ is to separately estimate the purities of $\tilde{\rho}_\pm$ and then compute their difference. This can be done by modifying the random unitaries and the postprocessing function. As shown in Fig. 1, given the dichotomic observable O , one first rotates the target state ρ into the eigenbasis of O using V_O (such that $V_O O V_O^\dagger$ is a diagonal matrix), which makes the density matrix the form of Eq. (8) in the computational basis. To estimate purities of $\tilde{\rho}_\pm$, we rotate the state $V_O \rho V_O^\dagger$ with the unitary $U = \tilde{U}_+ \oplus \tilde{U}_-$, where $\tilde{U}_+$ and $\tilde{U}_-$ are independent random unitaries acting on the d_+ - and d_- -dimensional subspaces, respectively. The diagonal blocks of the rotated density matrix are $\tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger$ and $\tilde{U}_- \tilde{\rho}_- \tilde{U}_-^\dagger$, which corresponds to original blocks evolved under random unitaries $\tilde{U}_+$ and $\tilde{U}_-$. Similar to the RM protocol, in each random unitary,

we perform N_M times of computational-basis measurement to collect the outcomes $\{s_i\}_{i=1}^{N_M}$. We then classify them into two groups depending on their eigenspaces. Then, one can employ the estimator of Eq. (2) for each group of data to estimate the purity of the corresponding block and evaluate the difference between them to estimate $\text{Tr}(O\rho^2)$.

To summarize, the unbiased estimator of $\text{Tr}(O\rho^2)$ can be chosen as

$$\omega_U = \binom{N_M}{2}^{-1} \sum_{1 \leq i < j \leq N_M} X_2^b(s_i, s_j). \quad (10)$$

Here, the postprocessing function is defined as

$$X_2^b(s_i, s_j) = \begin{cases} \lambda_\ell X_2^\ell(s_i, s_j), & \text{if } s_i, s_j \in \Pi_\ell \text{ for some } \ell \in \{+, -\}, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where $\lambda_\pm = \pm 1$ are the eigenvalues corresponding to $\Pi_\pm$, and $X_2^\pm(s_i, s_j)$ is given by Eq. (1) with d replaced by $d_\pm$. Note that when O is the identity operator, this estimator reduces to the estimator of the original RM protocol (using a global unitary) for purity estimation [2,33,34]. Combining the original proof of RM shown in Eq. (6) and the observation of Eq. (9), one can prove that ω_U is indeed an unbiased estimator of $\text{Tr}(O\rho^2)$ (see Appendix B for details). To further reduce the variance of the estimator, we can take the average of multiple estimators in Eq. (10) estimated with U being chosen as N_U independent random unitaries $\{U_1, \dots, U_{N_U}\}$, leading to

$$\omega = \frac{1}{N_U} \sum_{u=1}^{N_U} \omega_{U_u}. \quad (12)$$

Furthermore, by repeating the protocol $T = \mathcal{O}(\log(\delta^{-1}))$ times and applying the median-of-means method, we can suppress the failure probability to δ . We summarize the full procedure in Protocol 1.

Recall that the RM protocol described earlier can estimate the purity $\text{Tr}(\rho^2)$ of a density matrix to constant additive error using $\mathcal{O}(\sqrt{d})$ samples. The procedure of the ORM protocol for estimating a dichotomic observable is equivalent to performing the RM protocol in two blocks of the target state and then postprocessing accordingly. Consequently, the ORM protocol has a similar performance guarantee to the RM protocol [49], as stated in the following theorem (the proof is provided in Appendix B 1):

Lemma 1 (*Estimating dichotomic observables with block-diagonal unitary 4-designs*). Let O be a dichotomic observable, the dimensions of its two eigenspaces be $d_+, d_- \neq 0$, and two random unitaries $\tilde{U}_+$ and $\tilde{U}_-$ be sampled from a unitary 4-design. For any input state ρ in a

PROTOCOL 1. Estimating $\text{Tr}(O\rho^2)$ for a dichotomic observable O .

Input: T, N_U, N_M : number of repetitions, unitaries per repetition, and measurements per unitary. O : classical description of a dichotomic observable and its two eigenspaces with dimensions d_+ and d_- . $T \times N_U \times N_M$ copies of ρ .

Output: ω : an estimator of $\text{Tr}(O\rho^2)$.

```

1 Find the unitary  $V_O$  that diagonalizes  $O$ , i.e.,  $V_O O V_O^\dagger$  is a
  diagonal matrix in the computational basis.
2 for  $t = 1$  to  $T$  do
3   for  $u = 1$  to  $N_U$  do
4     Sample random unitaries  $\tilde{U}_+$  and  $\tilde{U}_-$  with
       dimensions  $d_+$  and  $d_-$ , respectively, and let
        $U = \tilde{U}_+ \oplus \tilde{U}_-$ .
5     for  $i = 1$  to  $N_M$  do
6       Rotate  $\rho$  with  $V_O$  and  $U$  to get  $UV_O \rho V_O^\dagger U^\dagger$ .
7       Perform a computational basis measurement and
         obtain an outcome  $s_{u,i}^{(t)}$ .
8       Compute  $\omega_u^{(t)} = \binom{N_M}{2}^{-1} \sum_{1 \leq i < j \leq N_M} X_2^b(s_{u,i}^{(t)}, s_{u,j}^{(t)})$ .
9     Compute  $\omega^{(t)} = \frac{1}{N_U} \sum_{u=1}^{N_U} \omega_u^{(t)}$ .
10 Return  $\omega = \text{median}\{\omega^{(1)}, \omega^{(2)}, \dots, \omega^{(T)}\}$ .
```

d -dimensional qubit system, Protocol 1 estimates the non-linear quantity $\text{Tr}(O\rho^2)$ to additive error ε with probability at least $1 - \delta$, with a sample complexity of

$$\mathcal{O} \left(\max \left\{ \frac{\sqrt{d}}{\varepsilon}, \frac{\sqrt{d/d_+}}{\varepsilon^2}, \frac{\sqrt{d/d_-}}{\varepsilon^2} \right\} \log(\delta^{-1}) \right). \quad (13)$$

Specifically, the number of repetitions is $T = \mathcal{O}(\log(\delta^{-1}))$, the number of random unitaries per repetition is $N_U = \mathcal{O}(\max\{1, 1/(d_+\varepsilon^2), 1/(d_-\varepsilon^2)\})$, and the number of measurements per unitary is $N_M = \mathcal{O}(\sqrt{N_U d}/\varepsilon)$.

Note that the sample complexity given in Lemma 1 cannot cover the sample complexity for estimating purity, which corresponds to the observable of identity $\mathbb{I}$ with $d_- = 0$. Actually, by slightly modifying our derivation in Appendix B 1, we can derive the sample complexity for purity estimation [49].

Remark 1 (*Purity estimation with unitary 4-designs*). When O is the identity matrix, $O = \mathbb{I}$, i.e., has a unique eigenvalue 1, Protocol 1 still applies by setting $d_+ = d$, $d_- = 0$. In this case, ORM is equivalent to the RM protocol measuring the purity [2]. The sample complexity is $\mathcal{O}(\max\{\sqrt{d}/\varepsilon, 1/\varepsilon^2\} \log(\delta^{-1}))$ by setting $T = \mathcal{O}(\log(\delta^{-1}))$, $N_U = \mathcal{O}(\max\{1, 1/(d\varepsilon^2)\})$, and $N_M = \mathcal{O}(\sqrt{N_U d}/\varepsilon)$.

Note that when $d_{\pm} = \Omega(d)$, Protocol 1 achieves a sample complexity upper bound as $\mathcal{O}\left(\max\left\{\sqrt{d}/\varepsilon, 1/\varepsilon^2\right\}\right)$, which is similar to the sample complexity for estimating purity. However, when $d_- = \mathcal{O}(1)$ or $d_+ = \mathcal{O}(1)$, the sample complexity upper bound can reach $\mathcal{O}(\sqrt{d}/\varepsilon^2)$, which is worse than the purity estimation. We can use an observable decomposition technique to avoid this upper bound, the proof of which is left to Appendix B 3.

Lemma 2 (Dichotomic decomposition). For any dichotomic observable O with $\min\{d_+, d_-\} \leq d/4$, one can decompose O as $O = \frac{1}{2}(\pm\mathbb{I} + O_1 + O_2 + O_3)$, where each O_i is a dichotomic observable with subspace dimensions being $\Omega(d)$.

Therefore, by estimating $\text{Tr}(\rho^2)$ and each $\text{Tr}(O_i\rho^2)$ to $\mathcal{O}(\varepsilon)$ precision and combining the results, we have:

Theorem 1 (Improved scaling in estimating dichotomic observables with block-diagonal unitary 4-designs). For any dichotomic observable O , $\text{Tr}(O\rho^2)$ can be estimated to within additive error ε with probability at least $1 - \delta$ using $\mathcal{O}\left(\max\left\{\sqrt{d}/\varepsilon, 1/\varepsilon^2\right\} \log(\delta^{-1})\right)$ sample complexity with single-copy operations.

The ORM protocol offers several practical advantages. First, it only requires a minimal number of measurement settings. Notably, for any estimation accuracy ε that is not exponentially small, even a constant number of measurement bases, N_U , is sufficient to estimate $\text{Tr}(O\rho^2)$ accurately. This feature significantly reduces the experimental overhead, as reconfiguring the measurements is on many platforms more demanding than repeated sampling in a fixed basis [53–56].

Another crucial advantage of our protocol lies in its simplified gate calibration and postprocessing. The data processing is independent of the specific random unitary applied, requiring only comparing the sampled bitstrings [see Eq. (11)]. This implies that the unitaries do not need meticulous, gate-by-gate calibration; it is sufficient that they are drawn from the required unitary design. In contrast to methods like classical shadows, ORM avoids computationally intensive steps like classical state reconstruction, and the postprocessing function can be evaluated efficiently in practice. The runtime of computing the estimator over T samples scales as $\mathcal{O}(nT^2)$ under a brute-force enumeration of pairwise contributions and can be improved to $\mathcal{O}(nT)$ using standard data structures such as a trie. Thus, we do not expect the protocol to be limited in its scalability by the postprocessing complexity but rather by the data acquisition.

Furthermore, the ORM protocol achieves the optimal sample complexity for Pauli observables. According to Lemma 1, when ε and δ are constants, the sample complexity for estimating Pauli observables is upper bounded

by $\mathcal{O}(\sqrt{d})$, which is more efficient than other prevailing single-copy randomized measurement protocols such as classical shadows (see Appendix A 3 for an analysis of the classical shadows). In Appendix C, we generalize the proof in Ref. [11] and show that $\mathcal{O}(\sqrt{d})$ is optimal for all Pauli observables, thus answering an open question in quantum learning. In the next section, we will show that our protocol is actually optimal for a much broader class of observables (see Fact 1).

B. Estimating general observables

We now discuss how to generalize the protocol from dichotomic observables to estimating general observables O subject to $\|O\|_{\infty} \leq 1$ (otherwise rescale O according to $O/\|O\|_{\infty}$). The main observation is that an arbitrary observable can be approximated, to a desired precision ε , by a sum of $\mathcal{O}(\log \varepsilon^{-1})$ dichotomic observables. With the decomposition, we can then apply Protocol 1 to estimate the expectation value of each component and sum the resulting estimators to obtain the final estimator. We first show how to perform such a decomposition.

Lemma 3 (Dichotomic decomposition of general observables). For any observable O satisfying $\|O\|_{\infty} \leq 1$, one can decompose O as

$$O = \sum_{l=1}^k 2^{-l} O_l + O_{\Delta}, \quad (14)$$

where $k \in \mathbb{N}^+$, each O_l is a dichotomic observable with eigenvalues ± 1 , and $\|O_{\Delta}\|_{\infty} \leq 2^{-k}$.

Proof. First, suppose O is a diagonal matrix, $O = \text{diag}(a_0, a_1, \dots, a_{d-1})$, where $d = 2^n$ is the dimension of the Hilbert space. Given $k \in \mathbb{N}_+$, each diagonal element a_i can be written as

$$a_i = \sum_{l=1}^k 2^{-l} a_{i,l} + a_{i,\Delta}, \quad (15)$$

where $a_{i,l} \in \{\pm 1\}$ and $0 \leq |a_{i,\Delta}| \leq 2^{-k}$. To see this, define $a_{i,1} = \text{sgn}(a_i)$, recursively

$$a_{i,l} = \text{sgn}\left(a_i - \sum_{j=1}^{l-1} 2^{-j} a_{i,j}\right) \quad (16)$$

for $2 \leq l \leq k$, and $a_{i,\Delta} = a_i - \sum_{l=1}^k 2^{-l} a_{i,l}$, where the (slightly modified) sign function is given by

$$\text{sgn}(x) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0. \end{cases} \quad (17)$$

By construction, $|a_i - \sum_{j=1}^l 2^{-j} a_{ij}| \leq 2^{-l}$ for all $1 \leq l \leq k$.
Let

$$\begin{aligned} O_l &= \text{diag}(a_{0,l}, a_{1,l}, \dots, a_{d-1,l}), \quad 1 \leq l \leq k, \\ O_\Delta &= \text{diag}(a_{0,\Delta}, a_{1,\Delta}, \dots, a_{d-1,\Delta}). \end{aligned} \quad (18)$$

We have that

$$O = \sum_{l=1}^k 2^{-l} O_l + O_\Delta, \quad (19)$$

where each O_l is diagonal with entries in $\{\pm 1\}$, and $\|O_\Delta\|_\infty = \max_i \{|a_{i,\Delta}|\} \leq 2^{-k}$.

If O is not diagonal, let $O = V_O^\dagger D V_O$ be its diagonalization, where D is a real diagonal matrix and V_O is unitary. Decompose D into a sum of diagonal dichotomic matrices D_l as above, and define $O_l = V_O^\dagger D_l V_O$, $O_\Delta = V_O^\dagger D_\Delta V_O$. This yields the desired decomposition. ■

Lemma 3 shows that a general observable O can be decomposed into a sum of dichotomic observables $\{O_l\}_{l=0}^k$. The contribution of residual term O_Δ to $\text{Tr}(O\rho^2)$ decreases exponentially as k increases. Notice that the constructive proof presented here does not necessarily yield the optimal number of decomposed terms, but provides a general decomposing procedure of an arbitrary observable to dichotomic ones. By applying Protocol 1 to estimate each O_l and ignoring O_Δ for $k = \mathcal{O}(\log \varepsilon^{-1})$, we obtain the protocol for estimating $\text{Tr}(O\rho^2)$ for an arbitrary observable O to a desired precision ε . The full procedure is summarized in Protocol 2, and the resulting sample complexity is stated below, with the proof being presented in Appendix B 4.

Theorem 2 (Estimating general observables with block-diagonal unitary 4-designs). Let O be an observable with $\|O\|_\infty \leq 1$. For any input state ρ in a d -dimensional qubit system, Protocol 2 estimates $\text{Tr}(O\rho^2)$ to within additive error ε with probability at least $1 - \delta$, using a sample complexity of

$$\mathcal{O}\left(\max\left\{\frac{\sqrt{d}}{\varepsilon}, \frac{1}{\varepsilon^2}\right\} \log[\delta^{-1} \log(\varepsilon^{-1})]\right). \quad (20)$$

Furthermore, a matching lower bound exists for a broad class of observables, as established by the following result.

Fact 1 (Lower bound of sample complexity for estimating observables with large trace-norm [38]). Given an observable O with $\|O\|_1 = \Omega(d)$, any measurement protocol that uses single-copy operations on a d -dimensional input state ρ to estimate $\text{Tr}(O\rho^2)$ to within additive error $\varepsilon = \mathcal{O}(1)$ requires at least $\Omega(\sqrt{d})$ sample complexity.

PROTOCOL 2. Estimating $\text{Tr}(O\rho^2)$ for general observables (ORM).

Input: ε : desired precision of additive error. δ : desired upper bound of failure probability. O : classical description of an observable with $\|O\|_\infty \leq 1$.
 $\mathcal{O}\left(\max\left\{\frac{\sqrt{d}}{\varepsilon}, \frac{1}{\varepsilon^2}\right\} \log[\delta^{-1} \log(\varepsilon^{-1})]\right)$ copies of ρ .

Output: ω : an estimator of $\text{Tr}(O\rho^2)$, satisfying

$$\Pr\left[\left|\omega - \text{Tr}(O\rho^2)\right| < \varepsilon\right] \geq 1 - \delta.$$

1 Set $k = \lceil \log_2(\varepsilon^{-1}) \rceil + 1$, get the dichotomic deposition

$$O = \sum_{l=1}^k 2^{-l} O_l + O_\Delta \text{ using Lemma 3.}$$

2 **for** $l = 1$ **to** k **do**

3 Set $\varepsilon'_l = (\frac{3}{2})^l \frac{\varepsilon}{8}$, $\delta' = \delta/k$. Obtain an estimator ω_l for $\text{Tr}(O_l \rho^2)$ with precision ε' and failure probability δ' using Remark 1 and Theorem 1.

4 **Return** $\omega = \sum_{l=1}^k 2^{-l} \omega_l$.

The condition $\|O\|_1 = \Omega(d)$ is met by most observables of physical interest, including Pauli observables, local observables, and typical physical Hamiltonians, establishing the broad optimality of our protocol for these key applications. See Sec. IV for a detailed discussion and efficient implementations for estimating these observables.

While the ORM protocol offers significant advantages in terms of sample and computational complexity in postprocessing, it also presents challenges to a practical implementation. The protocol requires classically diagonalizing the target observable. Furthermore, it necessitates the implementation of random unitaries drawn from block-diagonal unitary designs, with their structure tailored to the specific observable being measured. Implementing these specialized ensembles generally has a similar circuit complexity as realizing global random unitaries. Additionally, the postmeasurement data from an experiment cannot be reused for estimating multiple observables simultaneously. A comparison between classical shadows and ORM is summarized in Table I.

IV. EFFICIENT CIRCUIT IMPLEMENTATIONS

While the ORM protocol is theoretically appealing due to its optimal sample complexity, its practical utility additionally requires scalable implementation of the measurements. In this section, we develop efficient circuit implementations of ORM for several physically relevant observables. First, we present the ORM protocol tailored for Pauli observables. Then, we provide performance guarantees for ORM when replacing the unitary 4-design with Clifford unitaries for Pauli observables and discuss its performance when using approximate unitary designs and local gates. Finally, we develop a Pauli sampling protocol that decomposes the target observables into several Pauli

strings and performs importance sampling on those strings to estimate their nonlinear expectation values. We show that our ORM achieves $\mathcal{O}(\sqrt{d})$ sample complexity for a wide range of physical observables and is sample-optimal for many of them.

A. Estimating Pauli observables

The unitary transformation that rotates an arbitrary observable into its eigenbasis can be prohibitively complex, often requiring a circuit depth that scales exponentially with the system size. This complexity presents a barrier even for the simpler task of estimating a linear expectation value. Fortunately, many observables of practical interest possess sufficient structure to be diagonalized efficiently. In this section, we focus on one of the most important classes of such observables—Pauli operators—and present an implementation of the ORM protocol that is not only efficient in terms of circuit resources but also achieves the optimal $\mathcal{O}(\sqrt{d})$ sample complexity.

Starting from the simplest case where the target observable is $O = Z_1$, i.e., the Pauli-Z operator on the first qubit and identity on the others. In this case, the eigenspace with $+1$ eigenvalue is spanned by state vectors of the form $|0\rangle \otimes |\psi_{n-1}\rangle$, where $|\psi_{n-1}\rangle$ is an arbitrary pure state defined on the second to the n th qubits. Similarly, the eigenspace with -1 eigenvalue is spanned by state vectors of the form $|1\rangle \otimes |\psi_{n-1}\rangle$. Now, the gate implementation procedure in Protocol 1 can be greatly simplified into the following.

- (1) Draw random unitaries $U_{\pm}$ from a unitary 4-design that act on the last $n - 1$ qubits.
- (2) Measure the first qubit in the computational basis to obtain $i \in \{0, 1\}$.
- (3) When the measurement result is $i = 0$, apply U_+ on the other $(n - 1)$ qubits, or apply U_- otherwise.

By judging the first bits of two measurement outcomes, it is easy to decide whether the two outcomes are from the same eigenspace or not and thus decide the value of the postprocessing function. To summarize, when the target observable is Z_1 , one does not need to perform the diagonalization unitary but instead two independent $(n - 1)$ -qubit random unitaries. At the same time, the sample complexity remains $\mathcal{O}(\sqrt{d})$ as the procedure described above is actually equivalent to that of Protocol 1. From an experimental perspective, chaotic evolution in certain analog systems is known to approximate global random unitaries [58–61], which can be beneficial for quantum learning tasks [62–64]. Recent experimental efforts have already used this chaotic behavior to benchmark quantum devices [65]. Implementing our protocol on such experimental platforms is therefore both interesting and promising.

Furthermore, the $(n - 1)$ -qubit unitary could also be replaced by the tensor product of $n - 1$ local unitaries drawn from a single-qubit unitary 2-design to further reduce the implementation difficulty. The postprocessing function needs to be modified correspondingly to

$$X_2^b(s_i, s_j) = \begin{cases} 2^{n-1}(-2)^{-\Delta(s_i, s_j)}, & \text{if } s_i[1] = s_j[1] = 0, \\ -2^{n-1}(-2)^{-\Delta(s_i, s_j)}, & \text{if } s_i[1] = s_j[1] = 1, \\ 0, & \text{otherwise,} \end{cases} \quad (21)$$

where $\Delta(s_i, s_j)$ is the Hamming distance between s_i and s_j , and $s_i[1]$ is the measurement outcome of the first qubit [34]. However, in this case, we cannot guarantee that the protocol can reach the $\mathcal{O}(\sqrt{d})$ sample complexity for estimating $\text{Tr}(Z_1 \rho^2)$. We will use numerical experiments later to assess the performance of the local-version protocol.

Notice that when O is a general Pauli observable, we could always find a Clifford circuit V_C rotating it to a single-qubit Pauli Z acting on O 's support (say qubit 1), i.e., $Z_1 = V_C O V_C^\dagger$. Experimentally, this means one can first apply the circuit V_C to the state ρ , then run the ORM protocol for the simple, fixed observable Z_1 . Crucially, when O is a local Pauli operator, the required Clifford circuit V_C is also local and can be implemented with much lower depth. As illustrated in Fig. 2, this approach provides an efficient circuit implementation for the entire class of Pauli observables.

B. Replacing the exact 4-design with Clifford unitaries, approximate design, and local gates

While our protocol for Pauli observables achieves the optimal performance with an exact 4-design, implementing such an ensemble is demanding [66]. In this section,

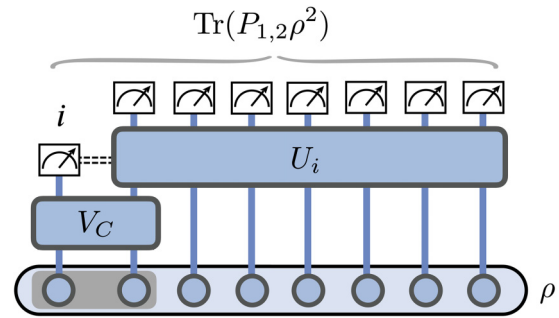


FIG. 2. Protocol for measuring any Pauli observable, with the example of a two-qubit Pauli observable defined on the first two qubits, $O = P_{1,2}$. Here, V_C is a Clifford unitary that rotates O to the Pauli-Z operator on the first qubit. Depending on the measurement outcome of the first qubit, we then apply U_i to the remaining part of the system, which can be either a global random unitary or a tensor product of single-qubit random unitaries.

we demonstrate that this requirement can be significantly relaxed by using a more practical ensemble of random Clifford unitaries [67,68], which can be implemented efficiently with $\mathcal{O}(n)$ circuit depth on 1D architectures [69,70]. This requirement parallels the global Clifford protocol used in classical shadows [31]. Specifically, we prove that replacing the exact 4-design with random Clifford unitaries preserves the optimal sample complexity scaling of $\mathcal{O}(\sqrt{d})$, albeit at the cost of a less favorable dependence on the desired precision ϵ .

Lemma 4 (Estimating traceless dichotomic observables with block-diagonal Clifford gates). Let O be a traceless dichotomic observable, i.e., the dimensions of its two eigenspaces are $d_+, d_- = d/2 = 2^{n-1}$, and two random unitaries $\tilde{U}_+$ and $\tilde{U}_-$ be sampled from the $(n-1)$ -qubit Clifford group. For any input n -qubit state ρ with $n \geq 3$, Protocol 1 estimates the nonlinear quantity $\text{Tr}(O\rho^2)$ to additive error ϵ with probability at least $1 - \delta$, with a sample complexity of

$$\mathcal{O}\left(\frac{\sqrt{d}}{\epsilon^2} \log(\delta^{-1})\right). \quad (22)$$

Specifically, the number of repetitions is $T = \mathcal{O}(\log(\delta^{-1}))$, the number of random unitaries per repetition is $N_U = \mathcal{O}(1/\epsilon^2)$, and the number of measurements per unitary is $N_M = \mathcal{O}(\sqrt{d})$.

The proof of Lemma 4 is detailed in Appendix B 2. With this result, we can now replace the unitary U_i in Fig. 2 by a random Clifford unitary and achieve the sample-optimal single-copy estimation of Pauli observables with solely Clifford circuits. Similar to the unitary 4-design case, our analysis also extends to the estimation of purity and inner product between two states using multi-qubit Clifford gates with only minor modifications, which is consistent with the results reported in Ref. [71].

Moreover, the requirement for an exact unitary 4-design can be relaxed to an approximate one. Although a rigorous proof is beyond the scope of this work, the reasoning is straightforward. If one uses an ϵ -approximate unitary design, the resulting deviations in quantities like the expectation value [e.g., Eq. (6)] and its variance will be bounded by $d'\epsilon$, where d' is a constant. By selecting an approximation error that is polynomially small in the system dimension, specifically $\epsilon = d^{-c}$ for a constant $c > c'$, the error in the final estimate becomes negligible. The performance thus remains nearly identical to that achieved with an exact design. Crucially, such approximate designs can

be implemented using a random circuit with depth of [57]

$$\log(\epsilon^{-1}) = c \log d = \mathcal{O}(n), \quad (23)$$

which is comparable to that of the random Clifford ensemble.

The unitary ensemble can be simplified even further to shallow random circuits or even local random unitaries by adjusting the estimator form [see Eq. (21)]. While these more practical ensembles do introduce some additional sample complexity overhead—for example, average sample complexities of $\mathcal{O}(\sqrt{2.5^n})$ and $\mathcal{O}(\sqrt{2.18^n})$ for local and shallow unitary ensembles [71]—the cost remains significantly lower than the $\mathcal{O}(2^n)$ complexity guarantee of classical-shadow protocols. In Sec. V, we present numerical experiments that validate this advantage, showing that our protocol achieves better sample complexity than classical shadows even when using local unitaries.

C. Estimating physical observables via Pauli sampling

Building upon our schemes for estimating Pauli observables with Clifford unitaries, we now show that the nonlinear expectation values of a wide range of observables—such as local observables and fidelity to stabilizer states—can be estimated using efficient circuit implementation schemes while maintaining the $\mathcal{O}(\sqrt{d})$ sample complexity. This is optimal for most of these observables.

Our central method is built on a powerful technique: importance sampling of Pauli operators [72,73]. Given any observable O , we can decompose it in the Pauli basis as follows:

$$O = \frac{1}{\sqrt{d}} \sum_{i=1}^K \chi_O(P_i) P_i, \quad (24)$$

where $P_i \in \{I, X, Y, Z\}^{\otimes n}$ is a Pauli operator, and $\chi_O(P) := \text{Tr}(OP)/\sqrt{d}$ represents the corresponding Pauli coefficient. Here, K denotes the number of Pauli operators needed to decompose O . The nonlinear expectation values can thus be expressed as

$$\text{Tr}(O\rho^2) = \sum_{i=1}^K \chi_O(P_i) \chi_{\rho^2}(P_i). \quad (25)$$

This expression can be further simplified to

$$\text{Tr}(O\rho^2) = \sum_{i=1}^K p_i X_i, \quad (26)$$

where the terms p_i and X_i are defined as

$$\begin{aligned} p_i &= \frac{\chi_O^2(P_i)}{\|O\|_2^2}, \\ X_i &= \|O\|_2^2 \frac{\chi_{\rho^2}(P_i)}{\chi_O(P_i)}. \end{aligned} \quad (27)$$

Note that $\{p_i\}_{i=1}^K$ forms a probability distribution over the Pauli operators that decompose M .

Our protocol proceeds as follows.

- (1) Sample $l = \mathcal{O}(K\|O\|_2^2/(d\varepsilon^2))$ Pauli strings P_i according to the probability distribution $\{p_i\}$.
- (2) For each $1 \leq i \leq l$, use

$$m_i = \mathcal{O}\left(\frac{\|O\|_2^4}{\sqrt{dl}\varepsilon^2\chi_O^2(P_i)}\right) \quad (28)$$

samples of ρ to obtain an estimator ω_i of $\text{Tr}(P_i\rho^2)$ in Eq. (12) using Protocol 1 and Lemma 4 with random Clifford gates.

- (3) Compute the averaged estimator

$$\omega = \frac{1}{l} \sum_{i=1}^l \frac{\|O\|_2^2}{\text{Tr}(OP_i)} \omega_i. \quad (29)$$

- (4) Repeat steps 1–3 for $T = \mathcal{O}(\log \delta^{-1})$ times and apply the median-of-means method to the obtained T estimators to output the final estimators.

We prove that the above Pauli sampling protocol achieves the following performance in Appendix D:

Theorem 3 (Estimating observables via Pauli sampling and block-diagonal Clifford gates). Let O be an observable that admits a Pauli decomposition, with the number of Pauli basis elements given by K [see Eq. (24)]. Then, the Pauli sampling protocol estimates the nonlinear quantity $\text{Tr}(O\rho^2)$ to additive error ε with probability at least $1 - \delta$, with an expected sample complexity of

$$\mathcal{O}\left(\frac{K\|O\|_2^2}{\sqrt{d}\varepsilon^2} \log(\delta^{-1})\right). \quad (30)$$

The actual sample complexity is unlikely to exceed its expected value by much, as guaranteed by Markov's inequality. Furthermore, we note that the sample complexity can be further improved using unitary 4-designs, which can result in a better dependence on ε^{-1} rather than ε^{-2} for the precision ε .

D. Local observables and Hamiltonians

We now apply the Pauli sampling protocol to a wide range of observables and demonstrate that it achieves $\mathcal{O}(\sqrt{d})$ complexity.

Local observables, which act nontrivially on only a constant number of qubits, are ubiquitous in quantum many-body physics (e.g., correlation functions and local energy terms). For any k -local observable $O = O_A \otimes \mathbb{I}_{[n]\setminus A}$ with $|A| = k$, we can always decompose it using $K = 4^k$ Pauli operators, each acting nontrivially only on A . Assuming the proper normalization condition $\|O\|_\infty \leq 1$, we have $\|O\|_2^2 \leq d$, leading to the following performance guarantee.

Corollary 1 (Performance guarantee for local observables and Hamiltonians). Let O be a k -local observable satisfying $\|O\|_\infty \leq 1$. Then, the Pauli sampling protocol estimates $\text{Tr}(O\rho^2)$ to additive precision ε with probability $1 - \delta$, with a sample complexity of $\mathcal{O}\left(K(\sqrt{d}/\varepsilon^2) \log(\delta^{-1})\right)$, where $K \leq 4^k$ is the number of nonzero Pauli terms in the decomposition of O .

This method extends directly to local Hamiltonians, which model most practical quantum systems and give the energy of the target states. Such Hamiltonians can be written as

$$H = \sum_{i=1}^K a_i P_i, \quad (31)$$

where $K = \text{poly}(n)$ and each P_i is a local Pauli operator. The 2-norm of H is given by $\|H\|_2^2 = \sum_i a_i^2 d$, which yields the sample complexity

$$\mathcal{O}\left(\left(\sum_i a_i^2\right) K \frac{\sqrt{d}}{\varepsilon^2} \log(\delta^{-1})\right). \quad (32)$$

For k -local observables, we have that $\|O\|_1 = 2^{n-k} \|O_A\|_1 = \Omega(d)$ for constant k and $\|O_A\|_1 = \Omega(1)$. By Fact 1, estimating the nonlinear expectation value of such observables requires at least $\Omega(\sqrt{d})$ sample complexity, so our protocol is sample-optimal across all single-copy schemes. Similarly, since the energy is nonvanishing for most eigenstates of practical Hamiltonians H , we have $\|H\|_1 = \Omega(d)$, and thus our protocol achieves the optimal sample complexity $\tilde{\mathcal{O}}(\sqrt{d})$ up to a logarithmic factor $(\sum_i a_i^2) K$.

E. Stabilizer fidelities and subspaces

Another important class of observables is the fidelities to stabilizer states $|\psi\rangle$, which play a fundamental role in quantum information theory [74]. A stabilizer state $|\psi\rangle$ can

be represented as a uniform sum over Pauli strings in its stabilizer group,

$$|\psi\rangle\langle\psi| = \frac{1}{2^n} \sum_{i=1}^{2^n} P_i, \quad P_i \in \mathcal{G}_\psi, \quad (33)$$

where $|\mathcal{G}_\psi| = 2^n$. This can be generalized to a stabilizer subspace, represented by the projector

$$\Pi = \frac{1}{2^k} \sum_{i=1}^{2^k} P_i, \quad P_i \in \mathcal{G}_\Pi, \quad (34)$$

where $\mathcal{G}_\Pi$ contains all Pauli strings that stabilize this subspace, and 2^{n-k} is the rank of the subspace. For example, when Π represents the set of all code states of a stabilizer code, Π becomes the projector onto the code space. This is important in tasks like subspace verification for quantum error-correcting codes [75]. When $k = n$, this reduces to a single stabilizer state.

For any $1 \leq k \leq n$, we have $\|\Pi\|_2^2 = 2^{n-k}$ and $K = 2^k$, which leads to the following sample complexity guarantees.

Corollary 2. Let $1 \leq k \leq n$ and Π be the projector onto a stabilizer subspace as given in Eq. (34). Then, the Pauli sampling protocol estimates $\text{Tr}(\Pi\rho^2)$ to additive precision ε with probability $1 - \delta$, with a sample complexity of $\mathcal{O}\left((\sqrt{d}/\varepsilon^2) \log(\delta^{-1})\right)$.

When k is constant, our protocol achieves optimal sample complexity by Fact 1. To the best of our knowledge, it remains open whether the sample complexity $\Omega(\sqrt{d})$ is optimal for low-rank projectors. See Sec. VI for further discussion on estimating nonlinear expectation values of low-rank observables. Moreover, for states with low non-stabilizerness, our protocol can potentially preserve the $\mathcal{O}(\sqrt{d})$ sample complexity [76].

V. APPLICATIONS

Estimating higher-order expectation values of quantum states is both a fundamental task and a key primitive in various fields. In this section, we highlight three applications: quantum virtual cooling, mixed-state phase detection, and entropy estimation. We perform numerical experiments to confirm the advantages of ORM compared with classical shadows.

A. Quantum virtual cooling

A particularly important case that has attracted much recent attention is error mitigation on noisy intermediate-scale quantum devices and slightly beyond [8]. In quantum experiments, we often aim to obtain the expectation values

of observables on a quantum state prepared by a quantum device. However, due to imperfections such as gate errors or insufficient cooling, the actual state is typically a noisy mixed state ρ rather than a pure state. Although some severe obstructions against the scalability of classes of quantum error mitigation methods have been identified [77–80], various such notions (also involving quantum processing aspects) remain a key ingredient in realizing functioning near-term quantum architectures.

The most important information is usually encoded in the principal component $|\psi\rangle$ of ρ , which can be expressed as

$$\rho = (1 - p) |\psi\rangle\langle\psi| + p \rho_\perp, \quad (35)$$

where $\rho_\perp$ has orthogonal support to $|\psi\rangle$. When the noise strength p is small, the error contributed by the noisy part $\rho_\perp$ can be substantially suppressed by considering the second-order expectation value $\text{Tr}(O\rho^2)/\text{Tr}(\rho^2)$ [9,10]. When $\rho_\perp$ has exponentially large rank, such as in the case of white noise, $\text{Tr}(O\rho^2)/\text{Tr}(\rho^2)$ can even exponentially suppress the noise with respect to the number of qubits. Therefore, estimating $\text{Tr}(O\rho^2)$ and the purity $\text{Tr}(\rho^2)$ provides a significantly improved estimation of $\langle\psi|O|\psi\rangle$ compared to directly measuring $\text{Tr}(O\rho)$.

When the noise is primarily introduced by thermal excitations, the target state can often be described by a Gibbs state of the form $\rho \propto \exp(-\beta H)$, where H is the Hamiltonian and $\beta > 0$ is the inverse temperature. In this setting, estimating $\text{Tr}(O\rho^2)/\text{Tr}(\rho^2)$ effectively predicts the expectation value on a lower-temperature state, as $\rho^2 \propto \exp(-2\beta H)$, providing access to physical properties at temperatures beyond the experimental cooling capability [7].

Here, we numerically simulate the quantum virtual cooling task using the global and local versions of ORM. We consider the Gibbs state of a system of L spins governed by the Heisenberg XX Hamiltonian,

$$H_{XX} = \sum_{1 \leq i < j \leq L} 2hJ_{ij}(X_i X_j + Y_i Y_j) + hB_z \sum_{i=1}^L Z_i, \quad (36)$$

with a transverse field, which is well studied and implemented in trapped-ion platforms [34,81]. The antiferromagnetic coupling strength decays according to a power law $J_{ij} = J_0/|i - j|^\alpha$ with $J_0 = 420 \text{ s}^{-1}$ and $\alpha = 1.24$. The transverse field is set to $B_z = 50J_0$ so that H_{XX} well approximates the system of hopping hard-core bosons [81]. The observable of interest is the two-local Pauli correlator $O = Z_1 Z_2$. Starting from a reference temperature $T_0 = 200J_0$ (i.e., $\beta_0 = 1/T_0$), we examine the thermal states

$$\rho(\beta) = \frac{e^{-\beta H_{XX}}}{\text{Tr}(e^{-\beta H_{XX}})} \quad (37)$$

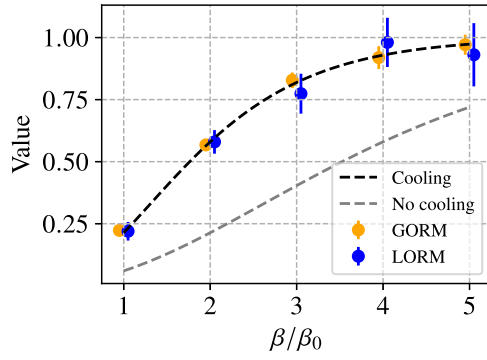


FIG. 3. Virtual cooling of 6-qubit thermal states using global-unitary observable-driven randomized measurement and local-unitary observable-driven randomized measurement protocols. The black dashed curve shows the theoretical value $\text{Tr}[Z_1 Z_2 \rho(2\beta)]$ on cooled states, while the gray dashed curve shows the value $\text{Tr}[Z_1 Z_2 \rho(\beta)]$ on original thermal states. Markers give observable-driven randomized measurement estimates of $\text{Tr}[Z_1 Z_2 \rho(\beta)^2] / \text{Tr}[\rho(\beta)^2]$ for five values of β and are slightly displaced along the horizontal axis to enhance visibility. Each point is the mean of ten independent runs with $N_U = 10$ random unitaries (drawn from the global Haar measure or the tensor product of single-qubit Haar measure) and $N_M = 1000$ measurements per unitary, with error bars denoting the standard error. Both protocols accurately reproduce the cooled-state expectation values.

for $\beta \in \{\beta_0, 2\beta_0, 3\beta_0, 4\beta_0, 5\beta_0\}$. For each β , we estimate

$$\frac{\text{Tr}[Z_1 Z_2 \rho(\beta)^2]}{\text{Tr}[\rho(\beta)^2]} = \text{Tr}[Z_1 Z_2 \rho(2\beta)] \quad (38)$$

using both *global-unitary ORM* (GORM) and *local-unitary ORM* (LORM). We numerically simulate a 6-qubit system and present the results in Fig. 3, where each data point is the mean of ten experiments with $N_U = 10$ random unitaries and $N_M = 1000$ measurements per unitary, with error bars indicating the standard error. The random unitaries are drawn from the global Haar measure or the tensor product of single-qubit Haar measure. Figure 3 also shows the exact theoretical values of $\text{Tr}[Z_1 Z_2 \rho(\beta)]$ and $\text{Tr}[Z_1 Z_2 \rho(2\beta)]$ (dashed lines) for comparison. The numerical results confirm that both GORM and LORM accurately reproduce the expectation value of $Z_1 Z_2$ at half the temperature of the original thermal state with reasonable experiment repetitions. Because local unitaries are easier to implement experimentally, LORM is often preferred despite its moderately larger estimator variance.

Furthermore, we compare the number of state copies needed to estimate the nonlinear expectation value $\text{Tr}(Z_1 Z_2 \rho^2)$ to a target precision using four protocols: global and local versions of classical shadows and ORM. The input state throughout is the Gibbs state at temperature $\beta = 8\beta_0$. For each system size and each protocol, we perform $T = 100$ experiments, obtaining estimators

$\omega_1, \dots, \omega_T$. We quantify accuracy via the *mean-squared error* (MSE),

$$\Delta^2 = \frac{1}{T} \sum_{i=1}^T [\omega_i - \text{Tr}(Z_1 Z_2 \rho^2)]^2, \quad (39)$$

and evaluate the minimum number of state copies required to achieve $\Delta^2 \leq 0.01$.

For ORM protocols, we sweep over the number N_U of random unitaries drawn from the Haar measure and, for every N_U , increase the number of measurements per unitary N_M until the MSE falls below 0.01. The sample complexity is given as $\min_{\Delta^2(N_U, N_M) \leq 0.01} (N_U N_M)$.

For classical shadows, we vary the total number of shadows N_s . In each experiment $i = 1, 2, \dots, T$, we generate $2N_s$ classical shadows $\{\hat{\rho}_j^{(i)}\}_{j=1}^{2N_s}$ using random unitaries drawn from the Haar measure, split them into two equally sized groups,

$$\hat{\sigma}^{(i,1)} = \frac{1}{N_s} \sum_{j=1}^{N_s} \hat{\rho}_j^{(i)}, \quad \hat{\sigma}^{(i,2)} = \frac{1}{N_s} \sum_{j=N_s+1}^{2N_s} \hat{\rho}_j^{(i)}, \quad (40)$$

and form the estimator

$$\omega_i = \text{Tr}[Z_1 Z_2 \frac{1}{2} (\hat{\sigma}^{(i,1)} \hat{\sigma}^{(i,2)} + \hat{\sigma}^{(i,2)} \hat{\sigma}^{(i,1)})]. \quad (41)$$

Because each ω_i uses only N_s^2 distinct shadow pairs, rather than the full $(2N_s)^2/2 = 2N_s^2$ possible pairs, we count the effective sample complexity as $\sqrt{2}N_s$ instead of $2N_s$. The reported sample complexity is the smallest $\sqrt{2}N_s$ that achieves an MSE below 0.01.

The numerical results are summarized in Fig. 4. Both global and local versions of ORM require far fewer state copies than their classical-shadow counterparts (GCS and LCS, respectively), especially when the number of qubits

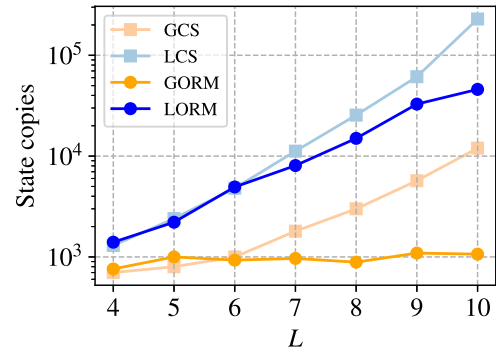


FIG. 4. Number of state copies needed to estimate $\text{Tr}(Z_1 Z_2 \rho^2)$ with $\text{MSE} \leq 0.01$ averaged over 100 experiments, as a function of system size L . Within both the global-unitary and local-unitary settings, observable-driven randomized measurement requires far fewer copies than classical shadows, offering a significant advantage in sample complexity with increasing system size.

is large. These findings confirm that classical shadows do not attain optimal sample complexity scaling. Whereas the classical-shadow curves grow exponentially with system size, the global ORM curve increases only slowly, remaining almost flat up to $L \leq 10$ qubits, underscoring its excellent sample efficiency in practice.

B. Detection of mixed-state quantum phases

The study of phases of matter and phase transitions is one of the most fundamental branches of physics, providing the conceptual foundation for understanding diverse phenomena across condensed matter physics, statistical mechanics, and even emerging connections to artificial intelligence [82]. Traditionally, investigations of quantum phase transitions have largely focused on pure quantum states, particularly at zero temperature, where the ground state or low-lying excitations dominate the system's behavior. In reality, perfectly pure states at zero temperature do not exist. Any physical system inevitably interacts with its environment, leading to decoherence and resulting in mixed states. As such, pure states should be viewed as a highly useful, yet ultimately idealized abstraction.

Motivated by this realization, recent efforts have sought to extend the study of quantum phases and phase transitions beyond pure states to encompass mixed states. These investigations have revealed novel physical phenomena with no direct counterparts in the pure-state regime [15,83,84]. One prominent example is the discovery of symmetry-protected topological phases for mixed states, where researchers have found that certain topological properties, traditionally thought to be fragile under decoherence, can in fact remain robust even in noisy, open-system conditions [16,17,85] in a precise sense.

In the course of studying mixed-state phases, researchers have realized that traditional linear quantities, such as correlation functions, are often insufficient to fully capture the phase behavior and critical phenomena of mixed states. This has motivated the use of various nonlinear quantities, such as entanglement entropy, Rényi correlators, and fidelity correlators, that are more sensitive to quantum correlations and coherence. Among these, our target quantity in this work, $\text{Tr}(O\rho^2)$ with O being a Pauli correlator, plays a particularly important role in characterizing mixed-state phases and their transitions [15,16]. However, previous studies have primarily relied on the classical shadows to estimate such quantities [17]. As demonstrated in Fig. 4, the ORM protocol proposed in this work offers a significant advantage in sample complexity compared to classical shadows, where the observable is also set as the Pauli correlator. This further highlights the value and the practical usefulness of the ORM approach as a powerful tool for probing mixed-state phases and uncovering

new physical insights beyond what is accessible through traditional techniques.

C. Estimation of entropic quantities

We here briefly mention a further application of $\text{Tr}(O\rho^2)$ not considered before, that of the estimation of certain entropic quantities. Specifically, the *Petz-Rényi relative entropy* [86] is defined for $\alpha \in (0, 1) \cup (1, \infty)$ as

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log_2 \text{Tr}(\rho^\alpha \sigma^{1-\alpha}), \quad (42)$$

a quantity with applications and a crisp interpretation in quantum information theory [87]. For $\alpha = 2$ and σ being a full-rank density matrix, this gives

$$D_2(\rho\|\sigma) := \log_2 \text{Tr}(\rho^2 \sigma^{-1}). \quad (43)$$

As σ^{-1} is Hermitian and can be seen as a quantum observable O , the ORM protocol can be used to estimate the value of $D_2(\rho\|\sigma)$.

When σ is the maximally mixed state, $\sigma = \mathbb{I}/d$, $D_2(\rho\|\sigma)$ provides an alternative metric of how close ρ is to the maximally mixed state, a task fundamental to the resource theory of purity [88] and closely related to quantum mixedness testing [89]. Let

$$c := D_2\left(\rho \left\| \frac{\mathbb{I}}{d}\right.\right) = n + \log \text{Tr}(\rho^2), \quad (44)$$

so that $\text{Tr}(\rho^2) = 2^{-(n-c)}$. To estimate c within additive error ε , we need to produce an estimator ω of $\text{Tr}(\rho^2)$ satisfying $\omega \in [2^{-(n-c+\varepsilon)}, 2^{-(n-c-\varepsilon)}]$, which translates to an additive precision requirement $\varepsilon_\omega = \Theta(\varepsilon 2^{-(n-c)})$. By Remark 1, achieving this accuracy requires

$$\mathcal{O}\left(\frac{1}{\varepsilon_\omega^2}\right) = \mathcal{O}\left(\frac{d^2}{\varepsilon^2}\right) \quad (45)$$

copies of ρ , reflecting the fundamentally exponential sample complexity of mixedness testing [89].

The ORM protocol applies to estimating $D_2(\rho\|\sigma)$ for a general state σ . Let

$$\kappa = \lambda_{\max}(\sigma)/\lambda_{\min}(\sigma) \quad (46)$$

be the condition number of σ , where $\lambda_{\max}$ and $\lambda_{\min}$ are its largest and smallest eigenvalues. We can rewrite

$$c := D_2(\rho\|\sigma) = n + \log \kappa + \log \text{Tr}(\rho^2(\kappa d\sigma)^{-1}). \quad (47)$$

To estimate c within additive error ε , it suffices to obtain an estimator ω of $\text{Tr}(\rho^2(\kappa d\sigma)^{-1})$ satisfying

$\omega \in [\kappa^{-1}2^{-(n-c+\varepsilon)}, \kappa^{-1}2^{-(n-c-\varepsilon)}]$, which requires an additive precision

$$\varepsilon_\omega = \Theta(\kappa^{-1}\varepsilon 2^{-(n-c)}). \quad (48)$$

The matrix $(\kappa d\sigma)^{-1}$ is Hermitian and obeys

$$\|(\kappa d\sigma)^{-1}\|_\infty \leq 1. \quad (49)$$

Theorem 2 guarantees us to achieve this precision with $\tilde{O}(\kappa^2 d^2 / \varepsilon^2)$ copies of ρ . When $\kappa = \mathcal{O}(1)$, the resulting sample complexity (up to logarithmic factors) matches that of the maximally mixed reference state.

VI. AN ALTERNATE APPROACH FOR LOW-RANK OBSERVABLE ESTIMATION

In the previous section, we have utilized the local structure of the target observable to simplify the implementation of the ORM protocol, replacing the block-diagonal random unitary with two independent random unitaries. In this section, we will develop an alternative RM protocol to show that, when estimating $\text{Tr}(O\rho^2)$ for low-rank observables, we can also replace the block-diagonal random unitary with a normal random unitary. This protocol is partly inspired by the thrifty shadow protocol [90,91]. Note that estimating such quantities is particularly important for error-mitigated fidelity estimation, where $O = |\psi\rangle\langle\psi|$ and $|\psi\rangle$ is the target state vector.

For low-rank observable, Protocol 2 achieves a sample complexity of $\mathcal{O}(\sqrt{d})$, as shown in Theorem 2. Interestingly, this favorable scaling does not necessarily require an observable-dependent protocol. In fact, the classical shadows can also estimate $\text{Tr}(O\rho^2)$ with sample complexity $\mathcal{O}(\sqrt{d})$ (see Appendix A 3). However, a major limitation of the classical shadows lies in its large number of required measurement settings. For example, when $\varepsilon, \delta = \mathcal{O}(1)$, our protocol only requires $N_U = \mathcal{O}(1)$ distinct measurement settings, whereas classical shadows demands $\mathcal{O}(\sqrt{d})$ random unitaries. Reducing the number of measurement settings is crucial for practical quantum platforms such as superconducting and optical qubits [92,93], where measurements under a single setting can be repeated efficiently, but changing the measurement basis is time-consuming. Furthermore, the measurement cost can be exponentially reduced on platforms that support ensemble-average measurements across many copies, such as nuclear magnetic resonance systems, where a thermodynamically large number of quantum states can be simultaneously accessed [50].

The new protocol combines the central ideas of the RM protocol introduced in Sec. II and the classical shadows and thus is named BRM. Specifically, it shares the same quantum data-obtaining process with RM and borrows the idea from classical shadows in data postprocessing. To illustrate this, we first show how the purity-estimation protocol can be generalized to estimate $\text{Tr}(\rho^3)$. Recall that the

key step in estimating $\text{Tr}(\rho^2)$ is to virtually implement a SWAP operator $\mathbb{S}$ using the data postprocessing function [see Eqs. (1) and (6)]. By a similar argument, an estimator for $\text{Tr}(\rho^3)$ should virtually implement the third-order permutation operator. Concretely, the estimator takes the form

$$\hat{M}_3^U = \binom{N_M}{3}^{-1} \sum_{1 \leq i < j < k \leq N_M} X_3(s_i, s_j, s_k), \quad (50)$$

where $X_3(s_i, s_j, s_k)$ is chosen so that

$$\mathbb{E}_U [U^{\otimes 3} X_3 U^{\dagger \otimes 3}] = \frac{1}{2} (\mathbb{P}_{(132)}^3 + \mathbb{P}_{(123)}^3) \quad (51)$$

for U being sampled from a unitary 3-design [47]. Here,

$$X_3 = \sum_{s_1, s_2, s_3} X_3(s_1, s_2, s_3) |s_1, s_2, s_3\rangle\langle s_1, s_2, s_3|, \quad (52)$$

and $\mathbb{P}_{(132)}^3, \mathbb{P}_{(123)}^3$ are third-order cyclic permutation operators satisfying $\mathbb{P}_{(132)}^3 |\phi_1, \phi_2, \phi_3\rangle = |\phi_2, \phi_3, \phi_1\rangle$ and $\mathbb{P}_{(123)}^3 |\phi_1, \phi_2, \phi_3\rangle = |\phi_3, \phi_1, \phi_2\rangle$ for arbitrary $|\phi_1\rangle, |\phi_2\rangle$ and $|\phi_3\rangle$. To fulfill Eq. (51) and make $\hat{M}_3^U$ unbiased, one can select $X_3(s_1, s_2, s_3) = X_3(|\{s_1, s_2, s_3\}\rangle)$, with

$$X_3(1) = \frac{1+d^2}{2}, \quad X_3(2) = \frac{1-d}{2}, \quad X_3(3) = 1 \quad (53)$$

for $d = 2^n$ [94]. We now make a key observation: since

$$\text{Tr}(\rho^3) = \text{Tr}(\rho^2 \rho), \quad (54)$$

replacing one instance of ρ in $\text{Tr}(\rho^3)$ with an observable O yields $\text{Tr}(O\rho^2)$. Leveraging the unbiasedness of the estimator $\hat{M}_3^U$, we replace one “physical probability” $\langle s| U \rho U^\dagger |s\rangle$ with a “virtual probability” $\langle s| U O U^\dagger |s\rangle$, which can be computed classically from O and U . This replacement still preserves unbiasedness, producing an estimator for $\text{Tr}(O\rho^2)$. Specifically, the resulting single-round estimator is

$$\omega_u = \binom{N_M}{2}^{-1} \sum_{1 \leq i < j \leq N_M} \sum_{\sigma \in \{0,1\}^n} X_3(s_i, s_j, \sigma) \times \langle \sigma | U O U^\dagger | \sigma \rangle. \quad (55)$$

In fact, by a minor modification to the estimator ω_u , we could remove its dependency on O and construct the unbiased estimator

$$\hat{\rho}^2 = \binom{N_M}{2}^{-1} \sum_{1 \leq i < j \leq N_M} \sum_{\sigma \in \{0,1\}^n} X_3(s_i, s_j, \sigma) U^\dagger |\sigma\rangle\langle\sigma| U \quad (56)$$

for ρ^2 . This could be regarded as the “shadow” of ρ^2 while estimating properties in the form of $\text{Tr}(O\rho^2)$. Similarly to

PROTOCOL 3. Estimating $\text{Tr}(O\rho^2)$ for many low-rank observables (BRM).

Input: T, N_U, N_M : Number of repetitions, unitaries per repetition, and measurements per unitary.
 $\{O_m\}_{m=1}^M$: classical description of M observables satisfying $\text{Tr}(O_i^2) = \mathcal{O}(1)$.
 $T \cdot N_U \cdot N_M$ copies of ρ .

Output: $\{\omega_m\}_{m=1}^M$: estimators of $\{\text{Tr}(O_m\rho^2)\}_{m=1}^M$, respectively.

```

1 for  $t = 1$  to  $T$  do
2   for  $u = 1$  to  $N_U$  do
3     Sample a random unitary  $U_u$ .
4     for  $i = 1$  to  $N_M$  do
5       Apply  $U_u$  to  $\rho$ . Perform
       computational-basis measurement and
       obtain an outcome  $s_{u,i}^{(t)}$ .
6   for  $m = 1$  to  $M$  do
7     for  $t = 1$  to  $T$  do
8       Compute  $\omega_m^{(t)} = \frac{1}{N_U} \sum_{u=1}^{N_U} \binom{N_M}{2}^{-1} \sum_{1 \leq i < j \leq N_M} \sum_{\sigma \in \{0,1\}^n} X_3(s_{u,i}^{(t)}, s_{u,j}^{(t)}, \sigma) \langle \sigma | U O_m U^\dagger | \sigma \rangle$ .
9     Compute  $\omega_m = \text{median}\{\omega_m^{(1)}, \omega_m^{(2)}, \dots, \omega_m^{(T)}\}$ .
10 Return  $\{\omega_m\}_{m=1}^M$ .
```

the protocols proposed in Sec. III, we can average over N_U random unitaries and then apply the median-of-means procedure to further suppress the failure probability. We summarize the full procedure in Protocol 3.

An advantage of BRM over ORM is that BRM does not require a block-diagonal unitary. Instead, it uses a global random unitary, which simplifies implementation in real experiments. Moreover, this observable-independent BRM protocol offers a notable advantage in estimating multiple low-rank observables compared to ORM. Specifically, suppose we wish to estimate M observables $O_1, O_2, \dots, O_M$, each satisfying $\text{Tr}(O_m^2) = \mathcal{O}(1)$, with a failure probability δ . By setting $\delta' = \delta/M$ and selecting the number of experiments to guarantee failure probability δ' for estimating a single observable, the union bound ensures that all M observables can be estimated simultaneously with failure probability δ . This approach introduces only an $\mathcal{O}[\log(\delta'^{-1})] = \mathcal{O}(\log M)$ overhead, rather than a multiplicative factor of M that would result from naively repeating the entire protocol M times.

We further remark that third-order twirling is required to virtually implement the permutation operators, so unitaries should be sampled from a unitary 3-design to ensure the estimator is unbiased. Furthermore, sampling from a 6-design is sufficient to guarantee the performance stated in Theorem 4 (see proof in Appendix E for the

single-observable case and the discussion above for the multiobservable case).

Theorem 4 (Estimating many low-rank observables). Let $\{O_m\}_{m=1}^M$ be a collection of observables satisfying $\text{Tr}(O_m^2) = \mathcal{O}(1)$ for all i . For an input state ρ in a d -dimension system, Protocol 3 simultaneously estimates all quantities $\text{Tr}(O_m\rho^2)$ to additive error at most $\varepsilon > 0$ with probability at least $1 - \delta$, with a sample complexity of $\mathcal{O}(\sqrt{d}\varepsilon^{-2} \log(M/\delta))$. Specifically, the number of repetitions is $T = \mathcal{O}(\log(M/\delta))$, the number of random unitaries per repetition is $N_U = \mathcal{O}(\varepsilon^{-2})$, and the number of measurements per unitary is $N_M = \mathcal{O}(\sqrt{d})$.

We use Table II to compare the three protocols (classical shadows, ORM, and BRM) in estimating low-rank nonlinear expectation value. Subject to constant additive error and failure probability, all three protocols have a sample complexity scaling with the square root of the system dimension. Our proposed protocols, both ORM and BRM, remarkably reduce the number of required measurement bases to a constant number, compared to the exponentially many with system size in classical shadows. BRM also shares the favorable scaling in simultaneously estimating multiple observables, since it is also an observable-independent protocol.

VII. SUMMARY AND OUTLOOK

In this work, we have presented observable-driven randomized measurement protocols for estimating nonlinear properties of the form $\text{Tr}(O\rho^2)$. Our approach requires only single-copy quantum experiments and achieves provably optimal sample complexity among all observables with $\|O\|_1 = \Omega(d)$. While previous single-copy protocols were known to be optimal only for purity and inner product estimation, we bridge this gap by incorporating observable information into the RM framework, achieving a sample complexity of $\mathcal{O}(\sqrt{d})$ for estimating any observable with bounded norm. Note that the observable-driven idea was also adopted for classical shadows to reduce the complexity in estimating linear properties [73,95].

For practical applications, we present an efficient Clifford-circuit implementation for estimating Pauli observables that preserves the optimal sample complexity. We then extend this implementation to a broad class of physical observables, such as local Hamiltonians and stabilizer fidelities, via a Pauli sampling protocol. We further simplify circuit requirements by employing approximate unitary designs, shallow circuits, and local unitaries. A systematic study of randomized measurement under approximate state designs—particularly the performance of shallow-depth approximate unitary designs [57]—is an interesting direction for future work.

The theoretical intuition behind our protocol is rooted in Schur-Weyl duality, which refers to the fact that the action of a group of unitaries can filter out properties invariant under their action. This is the fundamental reason why the RM protocol can estimate the purity $\text{Tr}(\rho^2)$, as the action of any unitary on the state ρ leaves the purity unchanged. Following this intuition, when estimating $\text{Tr}(O\rho^2)$, it is natural to choose a unitary ensemble consisting of all unitaries that commute with the observable O . In the case of a dichotomic observable, this insight leads naturally to Protocol 1, which employs block-diagonal random unitaries. Therefore, an intriguing direction for future work is to develop a unified framework capable of estimating functions like $\text{Tr}(O_1\rho O_2\rho)$, which plays a vital role in mixed-state quantum phase transition [84] or more general functions $\text{Tr}(O\rho^{\otimes t})$ by appropriately adjusting the random unitary ensemble. From a practical standpoint, one of the key motivations for estimating nonlinear expectation values is to amplify the information of the principal component in a noisy target state. While our work addresses the estimation of the second-order nonlinear expectation value, a natural question is how to extend these methods to estimate higher-order values, $\text{Tr}(O\rho^t)$ for $t > 2$, or even to directly obtain $\langle \psi | O | \psi \rangle$ using single-copy operations, where $|\psi\rangle$ is the principal component of the noisy mixed state ρ . A related approach that combines randomized measurements with quantum error mitigation has also been developed in Ref. [96]. The distinction lies in that Ref. [96] mitigates the measurement process itself through the estimation of an appropriate frame.

In addition to this, we show that to estimate low-rank observables, the observable-independent BRM protocol proposed in Sec. VI can achieve sample complexity and measurement bases equivalent to those of ORM. This shows once again that low-rank structure can reduce the sample complexity of quantum learning protocols [97]. The BRM protocol integrates techniques from both RM and classical shadows, offering distinct advantages over these conventional approaches: compared with RM, its capability extends beyond the estimation of state moments, and compared with classical shadows, it significantly reduces the number of required measurement bases. This result represents an intriguing step toward a unified framework combining RM and classical shadows, two of the most prevalent approaches based on measurements in random bases [35,90,91].

We further note that, to the best of our knowledge, it remains an open question whether $\Omega(\sqrt{d})$ is a fundamental lower bound for estimating $\text{Tr}(O\rho^2)$ for low-rank observables under single-copy measurement schemes. Therefore, there is still potential to further optimize the sample complexity by exploring novel ways of incorporating observable information into the randomized measurement framework. While our work has focused on single-copy measurement schemes, a promising avenue for

future research is the estimation of nonlinear properties in other settings. These include multicopy measurement scenarios [98], models with access to oracles preparing purifications of the input state [99,100], and protocols designed for a limited number of copies [38,101]. We hope that the present work stimulates further endeavors in these directions.

ACKNOWLEDGMENTS

We appreciate the valuable discussion with Hong-Ye Hu, Weiyan Gong, Shu Chen, You Zhou, Pei Zeng, and Zihao Li. We thank Lorenzo Leone for his help in proving Lemma B4. Z.D. and Z.L. acknowledge the support from the National Natural Science Foundation of China (Grant No. 12174216) and the Quantum Science and Technology-National Science and Technology Major Project (Grants No. 2021ZD0300804 and No. 2021ZD0300702). Y.T. is supported by the Quantum Flagship MILLENION. J.E. is supported by the BMFTR (DAQC, MuniQC-Atoms, Hybrid++, QuSol), the Munich Quantum Valley, Berlin Quantum, the Quantum Flagship programmes MILLENION and PASQUANS2, the DFG (CRC 183), and the European Research Council (DebuQC). Part of the numerical results in this work is obtained using the Python package QISKIT [102].

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: PRELIMINARIES

1. Random unitaries

In this Appendix, we will briefly introduce the foundations of random unitary integral, especially the twirling function. One could refer to the review article Ref. [103] for more information on random unitaries. A unitary distribution is called Haar measure if and only if the following holds:

$$\int_{\text{Haar}} f(U) dU = \int_{\text{Haar}} f(UV) dU = \int_{\text{Haar}} f(VU) dU \quad (\text{A1})$$

for arbitrary continuous function $f(\cdot)$ and unitary V . According to the Schur-Weyl duality, the t -th order integral over Haar-random unitaries is related to the t -th order permutation group. For $X \in \mathcal{L}((\mathbb{C}^d)^{\otimes t})$, we define its t -th order twirling and write it as a linear combination of permutation operators,

$$\begin{aligned} \Phi_t(X) &:= \int_{\text{Haar}} U^{\otimes t} X U^{\dagger \otimes t} dU \\ &= \sum_{\pi, \sigma \in S_t} \text{Wg}(\pi^{-1}\sigma, d) \text{Tr}(\mathbb{P}_{\sigma^{-1}}^t X) \mathbb{P}_{\pi}^t, \end{aligned} \quad (\text{A2})$$

where $\mathbb{P}_\pi^t \in \mathcal{L}((\mathbb{C}^d)^{\otimes t})$ acts as permutation $\pi \in S_t$ on t states in $\mathbb{C}^d$ according to

$$\mathbb{P}_\pi^t = \sum_{s_1, \dots, s_t} |s_{\pi^{-1}(1)}, \dots, s_{\pi^{-1}(t)}\rangle \langle s_1, \dots, s_t|. \quad (\text{A3})$$

The Weingarten coefficient $\text{Wg}(\pi^{-1}\sigma, d)$ [104–106] is defined as the element in the pseudoinverse of the Gram matrix,

$$\begin{aligned} \text{Wg}(\pi^{-1}\sigma, d) &:= (G^+)_{\pi, \sigma}, \\ \text{where } G_{\pi, \sigma} &= \text{Tr}(\mathbb{P}_{\pi^{-1}}^t \mathbb{P}_\sigma^t). \end{aligned} \quad (\text{A4})$$

The design of postprocessing functions in the RM protocol aims to construct a suitable operator X_t , such that $\Phi_t(X_t)$ is a combination of desired permutation operators. For example, when estimating purities, we want to realize $\Phi_2(X_2) = \mathbb{S}$, which is then reduced to solving the set

$$\begin{aligned} \text{Tr}(X_2) &= d, \\ \text{Tr}(\mathbb{S}X_2) &= d^2 \end{aligned} \quad (\text{A5})$$

of linear equations [2]. A direct construction is to make X_2 diagonal as $X_2 = \sum_{s_1, s_2} X(s_1, s_2) |s_1, s_2\rangle \langle s_1, s_2|$ and determine a function $X_2(s_1, s_2)$ satisfying

$$\begin{aligned} \sum_{s_1, s_2} X_2(s_1, s_2) &= d, \\ \sum_{s_1} X_2(s_1, s_1) &= d^2. \end{aligned} \quad (\text{A6})$$

The function in Eq. (1) provides such a solution. Furthermore, the function X_3 defined in Eq. (53) constructs an operator whose third-order twirling evaluates as a linear combination of cyclic permutation operators. Similarly, one can generalize this theory to design functions for constructing high-order permutation operators. It is worth mentioning that the inherent symmetry of the symmetric group may forbid a solution to equations corresponding to an arbitrary combination of permutation operators in Eq. (A2). One example is that when X_3 is subjected to a diagonal operator, the two cyclic permutation operators $\mathbb{P}_{(123)}^3, \mathbb{P}_{(132)}^3$ always share the same coefficients [94].

2. Median-of-means method

Here, we briefly introduce how the median-of-means method exponentially suppresses the failure probability. The median-of-means estimator is summarized in the following lemma:

Lemma A1 (Median-of-means estimator). For given parameters $0 < \delta < 1$ and $0 < c < \frac{1}{2}$, let

$$T = \frac{1}{2(\frac{1}{2} - c)^2} \log(\delta^{-1}). \quad (\text{A7})$$

Suppose $\{\omega^{(t)}\}_{t=1}^T$ are independent random variables satisfying $\mathbb{E}[\omega^{(t)}] = \bar{\omega}$ and $\text{Var}[\omega^{(t)}] \leq c\varepsilon^2$ for any t . Define the median-of-means estimator

$$\omega = \text{median}\{\omega^{(1)}, \omega^{(2)}, \dots, \omega^{(T)}\}. \quad (\text{A8})$$

Then

$$\Pr[|\omega - \bar{\omega}| > \varepsilon] \leq \delta. \quad (\text{A9})$$

Proof. The event $\{|\omega - \bar{\omega}| > \varepsilon\}$ is contained by the event

$$\frac{1}{T} \sum_{t=1}^T \mathbb{1}(|\omega^{(t)} - \bar{\omega}| > \varepsilon) \geq \frac{1}{2}, \quad (\text{A10})$$

where $\mathbb{1}(\cdot)$ is the indicator function. Define

$$X_t = \frac{1}{T} \mathbb{1}(|\omega^{(t)} - \bar{\omega}| > \varepsilon). \quad (\text{A11})$$

Then X_t are independent random variables taking values in $\{0, 1/T\}$. By Markov's inequality,

$$\Pr[|\omega^{(t)} - \bar{\omega}| > \varepsilon] \leq \frac{\text{Var}[\omega^{(t)}]}{\varepsilon^2} \leq c. \quad (\text{A12})$$

Hence, $\mathbb{E}[X_t] \leq c/T$ and $\mathbb{E}\left[\sum_{t=1}^T X_t\right] \leq c$. Applying Hoeffding's inequality, we have

$$\begin{aligned} \Pr\left[\sum_{t=1}^T X_t \geq \frac{1}{2}\right] &\leq \Pr\left[\sum_{t=1}^T X_t - \mathbb{E}\left[\sum_{t=1}^T X_t\right] \geq \frac{1}{2} - c\right] \\ &\leq \exp\left(-2\left(\frac{1}{2} - c\right)^2 T\right). \end{aligned} \quad (\text{A13})$$

Substituting

$$T = \frac{1}{2(\frac{1}{2} - c)^2} \log(\delta^{-1}) \quad (\text{A14})$$

completes the proof. $\blacksquare$

Now suppose we have a protocol that produces an estimator $\omega^{(t)}$ with expectation $\bar{\omega}$ and variance

$$\text{Var}[\omega^{(t)}] \leq c\varepsilon^2. \quad (\text{A15})$$

We run this protocol $T = \mathcal{O}(\log(\delta^{-1}))$ times [given by Eq. (A7)], thereby obtaining estimator ω to be the median

of these T estimators,

$$\omega = \text{median} \{ \omega^{(1)}, \omega^{(2)}, \dots, \omega^{(T)} \}. \quad (\text{A16})$$

By Lemma A1, ω satisfies

$$\Pr[|\omega - \bar{\omega}| \geq \varepsilon] \leq \delta. \quad (\text{A17})$$

Thus, the median-of-means method exponentially suppresses the failure probability compared to the average estimator $(1/T) \sum_{t=1}^T \omega^{(t)}$.

3. Classical shadows

Here, we briefly review the classical shadows using global random unitaries introduced in Ref. [31] and discuss its relevant performance guarantees. In this scheme, one performs N_s single-copy experiments on the input state ρ . In each round, one draws a global random unitary U from a unitary 3-design and applies it to the state: $\rho \rightarrow U\rho U^\dagger$. A computational-basis measurement is then performed on $U\rho U^\dagger$. Suppose the measurement outcome for the i th unitary U_i is $|b_i\rangle$, a “classical shadow” of ρ is constructed as

$$\hat{\rho}_i = (2^n + 1) U_i^\dagger |b_i\rangle \langle b_i| U_i - \mathbb{I}, \quad (\text{A18})$$

which satisfies $\mathbb{E}[\hat{\rho}_i] = \rho$. To estimate $\text{Tr}(O\rho^2)$, one defines the estimator [31]

$$\omega_{\text{CS}} = \binom{N_s}{2}^{-1} \sum_{1 \leq i < j \leq N_s} \text{Tr}(O_{\text{sym}} \hat{\rho}_i \otimes \hat{\rho}_j), \quad (\text{A19})$$

where

$$O_{\text{sym}} = \frac{1}{2} \mathbb{S}(O \otimes \mathbb{I} + \mathbb{I} \otimes O) \quad (\text{A20})$$

is the symmetrized version of O , and $\mathbb{S}$ denotes the swap operator acting on $(\mathbb{C}^{2^n})^{\otimes 2}$. It follows that

$$\begin{aligned} \mathbb{E}[\omega_{\text{CS}}] &= \binom{N_s}{2}^{-1} \sum_{1 \leq i < j \leq N_s} \text{Tr}(O_{\text{sym}} \mathbb{E}[\hat{\rho}_i] \otimes \mathbb{E}[\hat{\rho}_j]) \\ &= \binom{N_s}{2}^{-1} \sum_{1 \leq i < j \leq N_s} \text{Tr}(O_{\text{sym}} \rho \otimes \rho) = \text{Tr}(O\rho^2). \end{aligned} \quad (\text{A21})$$

Hence, ω_{CS} is an unbiased estimator of $\text{Tr}(O\rho^2)$. Next, we show that the associated variance obeys the following bound.

Lemma A2. The variance of the estimator ω_{CS} satisfies

$$\text{Var}[\omega_{\text{CS}}] = \frac{1}{N_s^2} \mathcal{O}(\text{Tr}(O)^2 + d\text{Tr}(O^2)) + \frac{1}{N_s} \mathcal{O}(\|O\|_\infty^2). \quad (\text{A22})$$

We postpone the proof of Lemma A2 and first discuss its implications. When $O = \mathbb{I}$, the classical-shadow estimator provides an estimation of state purity $\text{Tr}(\rho^2)$, yielding $\text{Var}[\omega_{\text{CS}}] = \mathcal{O}(d^2/N_s^2 + 1/N_s)$. To estimate $\text{Tr}(\rho^2)$ to constant additive error with high probability, the variance should be suppressed to a constant. This implies a sample complexity of $N_s = \mathcal{O}(d)$, which is worse than the $\mathcal{O}(\sqrt{d})$ scaling of the RM protocol in Sec. II. For a general traceless observable O_0 satisfying $\text{Tr}(O_0) = 0$, achieving a constant-precision estimate of $\text{Tr}(O\rho_0^2)$ requires a sample complexity

$$N_s = \mathcal{O}\left(\sqrt{d\text{Tr}(O_0^2)}\right). \quad (\text{A23})$$

Before proving Lemma A2, we first establish the following auxiliary lemma.

Lemma A3 (Auxiliary lemma). For any state ρ and Hermitian observable O act on ρ , $\text{Tr}(\rho O \rho O) \leq \|O^2\|_\infty$.

Proof. Let $\rho = \sum_i p_i |v_i\rangle \langle v_i|$ be the spectral decomposition of ρ , then

$$\begin{aligned} \text{Tr}(\rho O \rho O) &\leq \|O \rho O\|_\infty = \left\| \sum_i p_i O |v_i\rangle \langle v_i| O \right\|_\infty \\ &\leq \sum_i p_i \|O |v_i\rangle \langle v_i| O\|_\infty \leq \sum_i p_i \|O^2\|_\infty \\ &= \|O^2\|_\infty. \end{aligned} \quad (\text{A24})$$

■

Proof of Lemma A2. We use the following properties of the classical-shadow estimator.

Fact A1 (Lemma S1 of Ref. [31]). Fix observable O and set $\hat{o} = \text{Tr}(O\hat{\rho})$, where ρ is a classical shadow in Eq. (A18). Then the variance of $\hat{o}$ satisfies

$$\text{Var}[\hat{o}] \leq 3\text{Tr}(O^2). \quad (\text{A25})$$

Fact A2 (Lemma S5 of Ref. [31]). The variance of ω_{CS} satisfies

$$\begin{aligned} \text{Var}[\omega_{\text{CS}}] &\leq \frac{4}{N_s^2} \text{Var}[\text{Tr}(O_{\text{sym}} \hat{\rho}_1 \otimes \hat{\rho}_2)] \\ &\quad + \frac{4}{N_s} \text{Var}[\text{Tr}(O_{\text{sym}} \hat{\rho}_1 \otimes \rho)], \end{aligned} \quad (\text{A26})$$

where $\hat{\rho}_1$ and $\hat{\rho}_2$ are two shadow estimators defined in Eq. (A18) constructed in independent experiments.

We bound these two terms in Fact A2 separately. For the first term, we use the following fact:

Fact A3 (Lemma S6 of Ref. [31]). For the symmetrized observable O_{sym} ,

$$\text{Var}[\text{Tr}(O_{\text{sym}}\hat{\rho}_1 \otimes \hat{\rho}_2)] = \mathcal{O}\left(\text{Tr}(O_{\text{sym}}^2)\right). \quad (\text{A27})$$

Incorporating Eq. (A20) into the above, we find

$$\begin{aligned} \frac{4}{N_s^2} \text{Var}[\text{Tr}(O_{\text{sym}}\hat{\rho}_1 \otimes \hat{\rho}_2)] &= \frac{1}{N_s^2} \mathcal{O}(\text{Tr}(O_{\text{sym}}^2)) \\ &= \frac{1}{N_s^2} \mathcal{O}(\text{Tr}(O)^2 + d\text{Tr}(O^2)). \end{aligned} \quad (\text{A28})$$

For the second term, we have

$$\begin{aligned} \frac{4}{N_s} \text{Var}[\text{Tr}(O_{\text{sym}}\hat{\rho}_1 \otimes \rho)] &= \frac{2}{N_s} \text{Var}[\text{Tr}((\rho O + O\rho)\hat{\rho}_1)] \\ &= \frac{1}{N_s} \mathcal{O}(\text{Tr}(\rho O \rho O + O^2 \rho^2)) \\ &= \frac{1}{N_s} \mathcal{O}(\|O\|_\infty^2), \end{aligned} \quad (\text{A29})$$

where in the second step we have used Fact A1 and in the last step we apply Lemma A3. Combining Eqs. (A28) and (A29), we obtain the stated variance bound for ω_{CS} . ■

APPENDIX B: ADDITIONAL ANALYSIS ON ESTIMATING DICHOTOMIC OBSERVABLES

Here, we provide the analysis of the protocols introduced in Sec. III A. First, we present the proof of Lemma 1. Then, we give a decomposition of the dichotomic observable that leads to the sample complexity stated in Theorem 1.

1. Proof of Lemma 1

First, we analyze the expectation value of the estimator ω defined in Eq. (12).

Lemma B1 (Expectation value of estimator ω). Given N_M and a dichotomic observable O , the estimator ω_U defined in Eq. (12) is unbiased, satisfying

$$\mathbb{E}[\omega_U] = \text{Tr}(O\rho^2). \quad (\text{B1})$$

Proof. First consider a general observable O in the Hilbert space of an n -qubit system $\mathcal{H}_n = (\mathbb{C}^2)^{\otimes n}$. Let

$$O = \sum_{\ell=1}^L \lambda_\ell P_\ell \quad (\text{B2})$$

be its spectral decomposition, where P_ℓ is the projector onto the eigenspace Π_ℓ corresponding to the eigenvalue λ_ℓ .

Our protocol proceeds as follows. For each eigenspace Π_ℓ , we draw a Haar-random unitary $\tilde{U}_\ell$, and then apply their direct sum

$$U = \bigoplus_{\ell=1}^L \tilde{U}_\ell \quad (\text{B3})$$

to the state ρ . Next, we measure the resulting state on the eigenbasis of O and record the measurement outcomes $\{s_1, \dots, s_{N_M}\}$. Suppose we generate these random unitaries $\{\tilde{U}_\ell\}_{\ell=1}^L$ for N_U times, and in each time we perform N_M repeated measurements. For each unitary U , we construct the estimator

$$\omega_U = \binom{N_M}{2}^{-1} \sum_{\{i,j\} \in \binom{[N_M]}{2}} X_2^b(s_i, s_j), \quad (\text{B4})$$

where

$$X_2^b(s_i, s_j) = \begin{cases} \lambda_\ell X_2^\ell(s_i, s_j), & s_i, s_j \in \Pi_\ell \text{ for some } \ell \in [L], \\ 0, & \text{otherwise.} \end{cases} \quad (\text{B5})$$

Here, $\{i, j\} \in \binom{[N_M]}{2}$ means $1 \leq i < j \leq N_M, i \neq j$, and $X_2^\ell(s_i, s_j)$ is defined in Eq. (1), with d replaced by d_ℓ , the dimension of Π_ℓ . The expectation value of ω_U is

$$\begin{aligned} \mathbb{E}[\omega_U] &= \mathbb{E}_U \left[\sum_{s_1, s_2} \langle s_1 | U \rho U^\dagger | s_1 \rangle \langle s_2 | U \rho U^\dagger | s_2 \rangle X_2^b(s_1, s_2) \right] \\ &= \mathbb{E}_U \left[\sum_{\ell=1}^L \sum_{s_1, s_2} \langle s_1 | P_\ell U \rho U^\dagger P_\ell | s_1 \rangle \right. \\ &\quad \left. \times \langle s_2 | P_\ell U \rho U^\dagger P_\ell | s_2 \rangle \lambda_\ell X_2^\ell(s_1, s_2) \right]. \end{aligned} \quad (\text{B6})$$

Note that U , constructed as in Eq. (B3), commutes with each projector P_ℓ (and thus commutes with O). Define $U_\ell = \tilde{U}_\ell \oplus \mathbf{0}$ to be the unitary acting on the whole space $\mathcal{H}_n$ but keeping only terms in subspace Π_ℓ , and denote

$$\tilde{\rho}_\ell = (P_\ell \rho P_\ell)|_{\Pi_\ell}, \quad |\tilde{s}\rangle = |s\rangle|_{\Pi_\ell} \quad \text{for } s \in \Pi_\ell. \quad (\text{B7})$$

Then it follows that

$$\begin{aligned} U_\ell &= P_\ell U P_\ell = P_\ell U = U P_\ell, \\ U_\ell^\dagger &= P_\ell U^\dagger P_\ell = P_\ell U^\dagger = U^\dagger P_\ell. \end{aligned} \quad (\text{B8})$$

Using these definitions, we can simplify the terms in Eq. (B6),

$$\begin{aligned}
\mathbb{E}[\omega_U] &= \mathbb{E}_{\{U_\ell\}} \left[\sum_{\ell=1}^L \sum_{s_1, s_2} \langle s_1 | U_\ell \rho U_\ell^\dagger | s_1 \rangle \langle s_2 | U_\ell \rho U_\ell^\dagger | s_2 \rangle \lambda_\ell X_2^\ell(s_1, s_2) \right] \\
&= \mathbb{E}_{\{\tilde{U}_\ell\}} \left[\sum_{\ell=1}^L \sum_{\tilde{s}_1, \tilde{s}_2} \langle \tilde{s}_1 | \tilde{U}_\ell \tilde{\rho}_\ell \tilde{U}_\ell^\dagger | \tilde{s}_1 \rangle \langle \tilde{s}_2 | \tilde{U}_\ell \tilde{\rho}_\ell \tilde{U}_\ell^\dagger | \tilde{s}_2 \rangle \lambda_\ell X_2^\ell(\tilde{s}_1, \tilde{s}_2) \right] \\
&= \sum_{\ell=1}^L \lambda_\ell \text{Tr}(\tilde{\rho}_\ell^2).
\end{aligned} \tag{B9}$$

In the last step, we use Eq. (6). When O is a dichotomic observable with eigenvalues ± 1 , Eq. (B9) simplifies to

$$\mathbb{E}[\omega_U] = \text{Tr}(\tilde{\rho}_+^2) - \text{Tr}(\tilde{\rho}_-^2) = \text{Tr}(O\rho^2), \tag{B10}$$

where the last equality uses Eq. (9). ■

Furthermore, we could sample U for N_U times independently and construct ω_u according to Eq. (12) for each $u \in [N_U]$. Then their average ω satisfies

$$\mathbb{E}[\omega] = \frac{1}{N_U} \sum_{u=1}^{N_U} \mathbb{E}[\omega_u] = \text{Tr}(O\rho^2), \tag{B11}$$

thus is also an unbiased estimator. We establish the variance of the estimator ω .

Lemma B2 (Variance of estimator ω with block-diagonal unitary 4-design). Given N_U, N_M , and a dichotomic observable O , the estimator ω defined in Eq. (12), obtained by following Protocol 1 with $\tilde{U}_\pm$ sampled from a unitary 4-design, has variance

$$\text{Var}(\omega) = \mathcal{O} \left(\frac{1}{N_U} \left(\frac{d}{N_M^2} + \frac{1}{N_M} + \frac{1}{d_+} + \frac{1}{d_-} \right) \right), \tag{B12}$$

where d_+ and d_- denote the dimensions of the eigenspaces Π_+ and Π_- of O , respectively, and $d = d_+ + d_-$.

Proof. We calculate the variance of the estimator ω_u [see Eq. (B4)] to get

$$\begin{aligned}
\text{Var}(\omega_u) &= \mathbb{E}[\omega_u^2] - \mathbb{E}[\omega_u]^2 \\
&= \binom{N_M}{2}^{-2} \sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2}}} \mathbb{E}[X_2^b(s_i, s_j) X_2^b(s_{i'}, s_{j'})] - \mathbb{E}[\omega_u]^2.
\end{aligned} \tag{B13}$$

Define $\Delta(i, j; i', j') = |\{i, j\} \cup \{i', j'\}| \in \{2, 3, 4\}$ and split the above summation into three terms,

$$\sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2}}} = \sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2} \\ \Delta(i,j;i',j')=2}} + \sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2} \\ \Delta(i,j;i',j')=3}} + \sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2} \\ \Delta(i,j;i',j')=4}}. \tag{B14}$$

(1) The first term is

$$\begin{aligned}
& \sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2} \\ \Delta(i,j;i',j')=2}} \mathbb{E} [X_2^b(s_i, s_j) X_2^b(s_{i'}, s_{j'})] \\
&= \binom{N_M}{2} \mathbb{E}_{U_+} \left[\sum_{s_1, s_2} \langle s_1 | U_+ \rho U_+^\dagger | s_1 \rangle \langle s_2 | U_+ \rho U_+^\dagger | s_2 \rangle X_2^+(s_1, s_2)^2 \right] + (+ \iff -) \\
&= \binom{N_M}{2} \mathbb{E}_{\tilde{U}_+} \left[\sum_{\tilde{s}_1, \tilde{s}_2} \langle \tilde{s}_1 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_1 \rangle \langle \tilde{s}_2 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_2 \rangle X_2^+(\tilde{s}_1, \tilde{s}_2)^2 \right] + (+ \iff -) \\
&= \binom{N_M}{2} \mathbb{E}_{\tilde{U}_+} \left[\text{Tr} \left(\tilde{U}_+^{\dagger \otimes 2} \left(\sum_{\tilde{s}_1, \tilde{s}_2} X_2^+(\tilde{s}_1, \tilde{s}_2)^2 |\tilde{s}_1 \tilde{s}_2\rangle \langle \tilde{s}_1 \tilde{s}_2| \right) \tilde{U}_+^{\otimes 2} \tilde{\rho}_+^{\otimes 2} \right) \right] + (+ \iff -), \quad (\text{B15})
\end{aligned}$$

where $(+ \iff -)$ is obtained by replacing all the label “+” with label “−” (and vice versa) in the previous terms. We define

$$X_+^{(1,2)2} = \sum_{\tilde{s}_1, \tilde{s}_2} X_2^+(\tilde{s}_1, \tilde{s}_2)^2 |\tilde{s}_1 \tilde{s}_2\rangle \langle \tilde{s}_1 \tilde{s}_2| \quad (\text{B16})$$

and calculate its second-order twirling [94],

$$\Phi_2(X_+^{(1,2)2}) = d_+ \mathbb{I} + (d_+ - 1) \mathbb{S}, \quad (\text{B17})$$

where $\mathbb{I}$ and $\mathbb{S}$ are respectively the identity and SWAP operator acting on $\Pi_+^{\otimes 2}$. In this way, the contribution of this term becomes

$$A_2 = \binom{N_M}{2} (d_+ \text{Tr}(\tilde{\rho}_+)^2 + (d_+ - 1) \text{Tr}(\tilde{\rho}_+^2)) + (+ \iff -). \quad (\text{B18})$$

(2) The second term is

$$\begin{aligned}
& \sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2} \\ \Delta(i,j;i',j')=3}} \mathbb{E} [X_2^b(s_i, s_j) X_2^b(s_{i'}, s_{j'})] \\
&= 6 \binom{N_M}{3} \mathbb{E}_{U_+} \left[\sum_{s_1, s_2, s_3} \langle s_1 | U_+ \rho U_+^\dagger | s_1 \rangle \langle s_2 | U_+ \rho U_+^\dagger | s_2 \rangle \langle s_3 | U_+ \rho U_+^\dagger | s_3 \rangle X_2^+(s_1, s_2) X_2^+(s_1, s_3) \right] + (+ \iff -) \\
&= 6 \binom{N_M}{3} \mathbb{E}_{\tilde{U}_+} \left[\sum_{\tilde{s}_1, \tilde{s}_2, \tilde{s}_3} \langle \tilde{s}_1 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_1 \rangle \langle \tilde{s}_2 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_2 \rangle \langle \tilde{s}_3 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_3 \rangle X_2^+(\tilde{s}_1, \tilde{s}_2) X_2^+(\tilde{s}_1, \tilde{s}_3) \right] + (+ \iff -) \\
&= 6 \binom{N_M}{3} \mathbb{E}_{\tilde{U}_+} \left[\text{Tr} \left(\tilde{U}_+^{\dagger \otimes 3} \left(\sum_{\tilde{s}_1, \tilde{s}_2, \tilde{s}_3} X_2^+(\tilde{s}_1, \tilde{s}_2) X_2^+(\tilde{s}_1, \tilde{s}_3) |\tilde{s}_1 \tilde{s}_2 \tilde{s}_3\rangle \langle \tilde{s}_1 \tilde{s}_2 \tilde{s}_3| \right) \tilde{U}_+^{\otimes 3} \tilde{\rho}_+^{\otimes 3} \right) \right] + (+ \iff -). \quad (\text{B19})
\end{aligned}$$

We further define the operator

$$X_+^{(1,2),(1,3)} = \sum_{\tilde{s}_1, \tilde{s}_2, \tilde{s}_3} X_2^+(\tilde{s}_1, \tilde{s}_2) X_2^+(\tilde{s}_1, \tilde{s}_3) |\tilde{s}_1 \tilde{s}_2 \tilde{s}_3\rangle \langle \tilde{s}_1 \tilde{s}_2 \tilde{s}_3| \quad (\text{B20})$$

and calculate its third-order twirling [94]

$$\Phi_3(X_+^{(1,2),(1,3)}) = -\frac{1}{d_+ + 2} (\mathbb{P}_{\mathbb{I}}^3 + \mathbb{P}_{(12)}^3 + \mathbb{P}_{(13)}^3) + \frac{d_+ + 1}{d_+ + 2} (\mathbb{P}_{(23)}^3 + \mathbb{P}_{(123)}^3 + \mathbb{P}_{(132)}^3). \quad (\text{B21})$$

Here $\mathbb{P}_\pi^3$ is the permutation operator corresponding to $\pi \in S_3$ acting on $\Pi_+^{\otimes 3}$ and we write π in its cycle decomposition. Notice that each l -cycle contributes a factor $\text{Tr}(\tilde{\rho}_+^l)$, so the summation simplifies to

$$A_3 = 6 \binom{N_M}{3} \left(\frac{d_+ - 1}{d_+ + 2} \text{Tr}(\tilde{\rho}_+^2) \text{Tr}(\tilde{\rho}_+) + \frac{2(d_+ + 1)}{d_+ + 2} \text{Tr}(\tilde{\rho}_+^3) - \frac{1}{d_+ + 2} \text{Tr}(\tilde{\rho}_+^3) \right) + (+ \iff -). \quad (\text{B22})$$

(3) The last term is given by

$$\begin{aligned} & \sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2} \\ \Delta(i,j;i',j')=4}} \mathbb{E} [X_2^b(s_i, s_j) X_2^b(s_{i'}, s_{j'})] \\ &= 6 \binom{N_M}{4} \mathbb{E}_{U_+} \left[\sum_{s_1, s_2, s_3, s_4} \langle s_1 | U_+ \rho U_+^\dagger | s_1 \rangle \langle s_2 | U_+ \rho U_+^\dagger | s_2 \rangle \langle s_3 | U_+ \rho U_+^\dagger | s_3 \rangle \langle s_4 | U_+ \rho U_+^\dagger | s_4 \rangle X_2^+(s_1, s_2) X_2^+(s_3, s_4) \right] \\ & \quad - 6 \binom{N_M}{4} \mathbb{E}_{U_+, U_-} \left[\sum_{s_1, s_2, s_3, s_4} \langle s_1 | U_+ \rho U_+^\dagger | s_1 \rangle \langle s_2 | U_+ \rho U_+^\dagger | s_2 \rangle \langle s_3 | U_- \rho U_-^\dagger | s_3 \rangle \langle s_4 | U_- \rho U_-^\dagger | s_4 \rangle X_2^+(s_1, s_2) X_2^-(s_3, s_4) \right] \\ & \quad + (+ \iff -) \\ &= 6 \binom{N_M}{4} \mathbb{E}_{\tilde{U}_+} \left[\sum_{\tilde{s}_1, \tilde{s}_2, \tilde{s}_3, \tilde{s}_4} \langle \tilde{s}_1 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_1 \rangle \langle \tilde{s}_2 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_2 \rangle \langle \tilde{s}_3 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_3 \rangle \langle \tilde{s}_4 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_4 \rangle X_2^+(\tilde{s}_1, \tilde{s}_2) X_2^+(\tilde{s}_3, \tilde{s}_4) \right] \\ & \quad - 6 \binom{N_M}{4} \mathbb{E}_{\tilde{U}_+, \tilde{U}_-} \left[\sum_{\tilde{s}_1, \tilde{s}_2, \tilde{s}_3, \tilde{s}_4} \langle \tilde{s}_1 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_1 \rangle \langle \tilde{s}_2 | \tilde{U}_+ \tilde{\rho}_+ \tilde{U}_+^\dagger | \tilde{s}_2 \rangle \langle \tilde{s}_3 | \tilde{U}_- \tilde{\rho}_- \tilde{U}_-^\dagger | \tilde{s}_3 \rangle \right. \\ & \quad \times \left. \langle \tilde{s}_4 | \tilde{U}_- \tilde{\rho}_- \tilde{U}_-^\dagger | \tilde{s}_4 \rangle X_2^+(\tilde{s}_1, \tilde{s}_2) X_2^-(\tilde{s}_3, \tilde{s}_4) \right] \\ & \quad + (+ \iff -) \\ &= 6 \binom{N_M}{4} \mathbb{E}_{\tilde{U}_+} \left[\text{Tr} \left(\tilde{U}_+^{\dagger \otimes 4} \left(\sum_{\tilde{s}_1, \tilde{s}_2, \tilde{s}_3, \tilde{s}_4} X_2^+(\tilde{s}_1, \tilde{s}_2) X_2^+(\tilde{s}_3, \tilde{s}_4) |\tilde{s}_1 \tilde{s}_2 \tilde{s}_3 \tilde{s}_4\rangle \langle \tilde{s}_1 \tilde{s}_2 \tilde{s}_3 \tilde{s}_4| \right) \tilde{U}_+^{\otimes 4} \tilde{\rho}_+^{\otimes 4} \right) \right] \\ & \quad - 6 \binom{N_M}{4} \mathbb{E}_{\tilde{U}_+, \tilde{U}_-} \left[\text{Tr} \left(\tilde{U}_+^{\dagger \otimes 2} \left(\sum_{\tilde{s}_1, \tilde{s}_2} X_2^+(\tilde{s}_1, \tilde{s}_2) |\tilde{s}_1 \tilde{s}_2\rangle \langle \tilde{s}_1 \tilde{s}_2| \right) \tilde{U}_+^{\otimes 2} \tilde{\rho}_+^{\otimes 2} \right) \right. \\ & \quad \times \left. \text{Tr} \left(\tilde{U}_-^{\dagger \otimes 2} \left(\sum_{\tilde{s}_3, \tilde{s}_4} X_2^+(\tilde{s}_3, \tilde{s}_4) |\tilde{s}_3 \tilde{s}_4\rangle \langle \tilde{s}_3 \tilde{s}_4| \right) \tilde{U}_-^{\otimes 2} \tilde{\rho}_-^{\otimes 2} \right) \right] \\ & \quad + (+ \iff -). \end{aligned} \quad (\text{B23})$$

We define

$$X_+^{(1,2) \otimes 2} = \sum_{\tilde{s}_1, \tilde{s}_2, \tilde{s}_3, \tilde{s}_4} X_2^+(\tilde{s}_1, \tilde{s}_2) X_2^+(\tilde{s}_3, \tilde{s}_4) |\tilde{s}_1 \tilde{s}_2 \tilde{s}_3 \tilde{s}_4\rangle \langle \tilde{s}_1 \tilde{s}_2 \tilde{s}_3 \tilde{s}_4|, \quad (\text{B24})$$

and its fourth-order twirling is [107]

$$\begin{aligned}\Phi_4(X_+^{(1,2)\otimes 2}) &= \left(1 + \frac{2}{d_+(d_+ + 2)(d_+ + 3)}\right) \mathbb{P}_{(12)(34)}^4 + \frac{2}{d_+(d_+ + 2)(d_+ + 3)} (\mathbb{P}_{\mathbb{I}}^4 + \mathbb{P}_{(12)}^4 + \mathbb{P}_{(34)}^4) \\ &+ \frac{d_+ + 1}{d_+(d_+ + 3)} (\mathbb{P}_{(13)(24)}^4 + \mathbb{P}_{(14)(23)}^4 + \mathbb{P}_{(1324)}^4 + \mathbb{P}_{(1423)}^4) \\ &- \frac{d_+ + 1}{d_+(d_+ + 2)(d_+ + 3)} (\text{the other elements in } S_4),\end{aligned}\quad (\text{B25})$$

where “the other elements in S_4 ” contain all the elements not listed previously. Represented by the lengths of cycles in cycle decomposition, they consist of four of type $1 + 1 + 2$, eight of type $1 + 3$, and four of type 4 . Consequently, the contribution of this summation is

$$\begin{aligned}A_4 &= 6 \binom{N_M}{4} \left[\frac{2}{d_+(d_+ + 2)(d_+ + 3)} \text{Tr}(\tilde{\rho}_+)^4 - \frac{4}{(d_+ + 2)(d_+ + 3)} \text{Tr}(\tilde{\rho}_+)^2 \text{Tr}(\tilde{\rho}_+^2) \right. \\ &- \frac{8(d_+ + 1)}{d_+(d_+ + 2)(d_+ + 3)} \text{Tr}(\tilde{\rho}_+) \text{Tr}(\tilde{\rho}_+^3) + \left(1 + \frac{2d_+^2 + 6d_+ + 6}{d_+(d_+ + 2)(d_+ + 3)}\right) \text{Tr}(\tilde{\rho}_+^2)^2 \\ &\left. + \frac{2(d_+ + 1)}{(d_+ + 2)(d_+ + 3)} \text{Tr}(\tilde{\rho}_+^4) - \text{Tr}(\tilde{\rho}_+^2) \text{Tr}(\tilde{\rho}_-^2) \right] + (+ \iff -).\end{aligned}\quad (\text{B26})$$

Inserting Eqs. (B18), (B22), (B26), and (B10) to Eq. (B13), we have

$$\begin{aligned}\text{Var}(\omega_u) &= \binom{N_M}{2}^{-2} (A_2 + A_3 + A_4) - (\text{Tr}(\tilde{\rho}_+^2) - \text{Tr}(\tilde{\rho}_-^2))^2 \\ &= \frac{2d_+}{N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+)^2 + \frac{2d_-}{N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-)^2 \\ &+ \frac{2(d_+ - 1)}{N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+^2) + \frac{2(d_- - 1)}{N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-^2) \\ &+ \frac{4(d_+ - 1)(N_M - 2)}{(d_+ + 2)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+^2) \text{Tr}(\tilde{\rho}_+) + \frac{4(d_- - 1)(N_M - 2)}{(d_- + 2)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-^2) \text{Tr}(\tilde{\rho}_-) \\ &+ \frac{8(d_+ + 1)(N_M - 2)}{(d_+ + 2)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+^3) + \frac{8(d_- + 1)(N_M - 2)}{(d_- + 2)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-^3) \\ &- \frac{4(N_M - 2)}{(d_+ + 2)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+)^3 - \frac{4(N_M - 2)}{(d_- + 2)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-)^3 \\ &+ \frac{2(N_M - 2)(N_M - 3)}{d_+(d_+ + 2)(d_+ + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+)^4 + \frac{2(N_M - 2)(N_M - 3)}{d_-(d_- + 2)(d_- + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-)^4 \\ &- \frac{4(N_M - 2)(N_M - 3)}{(d_+ + 2)(d_+ + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+)^2 \text{Tr}(\tilde{\rho}_+^2) - \frac{4(N_M - 2)(N_M - 3)}{(d_- + 2)(d_- + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-)^2 \text{Tr}(\tilde{\rho}_-^2) \\ &- \frac{8(d_+ + 1)(N_M - 2)(N_M - 3)}{d_+(d_+ + 2)(d_+ + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+) \text{Tr}(\tilde{\rho}_+^3) - \frac{8(d_- + 1)(N_M - 2)(N_M - 3)}{d_-(d_- + 2)(d_- + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-) \text{Tr}(\tilde{\rho}_-^3) \\ &+ \frac{(2d_+^2 + 6d_+ + 6)(N_M - 2)(N_M - 3)}{d_+(d_+ + 2)(d_+ + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+^2)^2 + \frac{(2d_-^2 + 6d_- + 6)(N_M - 2)(N_M - 3)}{d_-(d_- + 2)(d_- + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-^2)^2 \\ &+ \frac{2(d_+ + 1)(N_M - 2)(N_M - 3)}{(d_+ + 2)(d_+ + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_+^4) + \frac{2(d_- + 1)(N_M - 2)(N_M - 3)}{(d_- + 2)(d_- + 3)N_M(N_M - 1)} \text{Tr}(\tilde{\rho}_-^4) \\ &- \frac{4N_M - 6}{N_M(N_M - 1)} [\text{Tr}(\tilde{\rho}_+^2) - \text{Tr}(\tilde{\rho}_-^2)]^2.\end{aligned}\quad (\text{B27})$$

Notice that $\tilde{\rho}_{\pm}$ are principal submatrices of ρ , so that they are positive semidefinite and $\text{Tr}(\tilde{\rho}_{\pm}^l) \leq 1$ for $l \in \mathbb{N}_+$. We obtain the upper bound for $\text{Var}(\omega_u)$,

$$\begin{aligned} \text{Var}(\omega_u) &\leq \frac{4}{N_M(N_M - 1)}(d_+ + d_-) + \frac{4(N_M - 2)}{N_M(N_M - 1)} \left(\frac{3d_+ + 1}{d_+ + 2} + \frac{3d_- + 1}{d_- + 2} \right) \\ &\quad + \frac{4(N_M - 2)(N_M - 3)}{N_M(N_M - 1)} \left(\frac{d_+^2 + 2d_+ + 2}{d_+(d_+ + 2)(d_+ + 3)} + \frac{d_-^2 + 2d_- + 2}{d_-(d_- + 2)(d_- + 3)} \right) \\ &= \mathcal{O} \left(\frac{d}{N_M^2} + \frac{1}{N_M} + \frac{1}{d_+} + \frac{1}{d_-} \right). \end{aligned} \quad (\text{B28})$$

Furthermore, suppose we draw N_U independent random unitaries and construct an estimator by averaging the individual estimators obtained for each choice of random unitary, i.e.,

$$\omega = \frac{1}{N_U} \sum_{u=1}^{N_U} \omega_u. \quad (\text{B29})$$

Then we can estimate $\text{Tr}(O\rho^2)$ unbiasedly, with the variance given by

$$\text{Var}(\omega) = \frac{1}{N_U} \text{Var}(\omega_u) = \frac{1}{N_U} \mathcal{O} \left(\frac{d}{N_M^2} + \frac{1}{N_M} + \frac{1}{d_+} + \frac{1}{d_-} \right). \quad (\text{B30})$$

■

Based on Lemmas A1, B1, and B2, we can now prove Lemma 1 in the main text.

Proof of Lemma 1. To apply Lemma A1, it is sufficient to suppress $\text{Var}(\omega) \leq \varepsilon^2/3 < \varepsilon^2/2$. Define $\eta = \max\{1/d_+, 1/d_-\} = \Omega(1/d)$. When $\varepsilon \geq \Omega(\eta^{1/2})$, we can set $N_U = \mathcal{O}(1)$ and $N_M = \mathcal{O}(d^{1/2}/\varepsilon)$ and the total sample complexity to be $N_U \times N_M = \mathcal{O}(d^{1/2}/\varepsilon)$. When $\varepsilon \leq \mathcal{O}(\eta^{1/2})$, we should set $N_U = \mathcal{O}(\eta/\varepsilon^2)$ and $N_M = \mathcal{O}(d^{1/2}/\eta^{1/2})$, which results in the total sample complexity $N_U \times N_M = \mathcal{O}(d^{1/2}\eta^{1/2}/\varepsilon^2)$. In conclusion, the sample complexity is upper bounded by $\mathcal{O}(\max\{\sqrt{d}/\varepsilon, \sqrt{d\eta}/\varepsilon^2\})$. ■

2. Proof of Lemma 4

In this Appendix, we prove Lemma 4. Notice that it only differs from Lemma 1 by drawing $\tilde{U}_{\pm}$ randomly from the group of multi-qubit Clifford gates (which is a unitary 3-design [67] but not a 4-design [68]), the proof is similar to that in Appendix B 1. The unbiasedness of ω follows directly from Lemma B1. For the variance, we obtain the following upper bound (cf. Lemma B2):

Lemma B3 (Variance of estimator ω with block-diagonal Clifford gates). Given N_U, N_M , and a traceless dichotomic observable O , the estimator ω defined in Eq. (12), obtained by following Protocol 1 with $\tilde{U}_{\pm}$ sampled from the multi-qubit Clifford group, has variance

$$\text{Var}^{\text{Cl}}(\omega) = \mathcal{O} \left(\frac{1}{N_U} \left(\frac{d}{N_M^2} + \frac{1}{N_M} + 1 \right) \right), \quad (\text{B31})$$

where d denotes the dimension of the qubit system.

a. Notations

We first introduce some notations for a given state ρ that will be used in the proof. We define the state moments $R_k = \text{Tr}(\rho^k)$ for $k \in \mathbb{N}_+$. It is direct to see $R_k \leq 1$ for any k . We denote the n -qubit Pauli group as $\mathcal{P}_n = \{I, X, Y, Z\}^{\otimes n}$, the projector $Q = (1/d^2) \sum_{P \in \mathcal{P}_n} P^{\otimes 4}$, and its complementary projector $Q^{\perp} = I - Q$. For a 4-order permutation operator $\sigma \in S_4$, define $Q_{\sigma} = \text{Tr}(\mathbb{P}_{\sigma}^4 Q \rho^{\otimes 4})$. Since Q_{σ} is determined by the cycle type λ of σ , we also denote $Q_{\lambda} = Q_{\sigma}$,

$$\begin{aligned} Q_{1,1,1,1} &= \frac{1}{d^2} \sum_{P \in \mathcal{P}_n} \text{Tr}(P\rho)^4, \\ Q_{2,1,1} &= \frac{1}{d^2} \sum_{P \in \mathcal{P}_n} \text{Tr}((P\rho)^2) \text{Tr}(P\rho)^2, \\ Q_{2,2} &= \frac{1}{d^2} \sum_{P \in \mathcal{P}_n} \text{Tr}((P\rho)^2)^2, \\ Q_{3,1} &= \frac{1}{d^2} \sum_{P \in \mathcal{P}_n} \text{Tr}((P\rho)^3) \text{Tr}(P\rho), \\ Q_4 &= \frac{1}{d^2} \sum_{P \in \mathcal{P}_n} \text{Tr}((P\rho)^4). \end{aligned} \quad (\text{B32})$$

We obtain a tight upper bound for these quantities as follows.

Lemma B4 (Upper bound of Q_{σ}). For n -qubit state ρ and $\sigma \in S_4$, we have $Q_{\sigma} = \text{Tr}(\mathbb{P}_{\sigma}^4 Q \rho^{\otimes 4}) \leq 1/d$, where

$d = 2^n$. The bound is saturated when ρ is a pure stabilizer state.

Proof of Lemma B4. Let the spectral decomposition be $\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|$. Then we have

$$Q_\sigma = \sum_{i,j,k,l} p_i p_j p_k p_l \text{Tr}(\mathbb{P}_\sigma^4 Q \Psi_{ijkl}), \quad (\text{B33})$$

where $\Psi_{ijkl} = \bigotimes_{\alpha=i,j,k,l} |\psi_\alpha\rangle \langle \psi_\alpha|$. Notice that Q and Ψ_{ijkl} are projectors and $[\mathbb{P}_\sigma^4, Q] = 0, \forall \sigma \in S_4$, we then have

$$\begin{aligned} \text{Tr}(\mathbb{P}_\sigma^4 Q \Psi_{ijkl}) &= \text{Tr}(\mathbb{P}_\sigma^4 Q Q \Psi_{ijkl} \Psi_{ijkl}) \\ &= \text{Tr}(\Psi_{ijkl} Q \mathbb{P}_\sigma^4 Q \Psi_{ijkl}). \end{aligned} \quad (\text{B34})$$

Using the Cauchy-Schwarz inequality of the Hilbert-Schmidt norm, we have

$$\begin{aligned} Q_\sigma &= \sum_{i,j,k,l} p_i p_j p_k p_l \text{Tr}(\Psi_{ijkl} Q \mathbb{P}_\sigma^4 Q \Psi_{ijkl}) \\ &\leq \sum_{i,j,k,l} p_i p_j p_k p_l |\text{Tr}(\Psi_{ijkl} Q \mathbb{P}_\sigma^4 Q \Psi_{ijkl})| \\ &\leq \sum_{i,j,k,l} p_i p_j p_k p_l \sqrt{\text{Tr}(Q \Psi_{ijkl} \Psi_{ijkl} Q) \text{Tr}(Q \Psi_{ijkl} \mathbb{P}_\sigma^{4\dagger} \mathbb{P}_\sigma^4 \Psi_{ijkl} Q)} \\ &= \sum_{i,j,k,l} p_i p_j p_k p_l \text{Tr}(Q \Psi_{ijkl}) \\ &= \text{Tr}(Q \rho^{\otimes 4}) = \frac{1}{d^2} \sum_{P \in \mathcal{P}_n} \text{Tr}(P \rho)^4 \leq \frac{1}{d^2} \sum_{P \in \mathcal{P}_n} \text{Tr}(P \rho)^2 \\ &= \frac{1}{d} R_2 \leq \frac{1}{d}, \end{aligned} \quad (\text{B35})$$

where we use $\text{Tr}(P \rho) \in [-1, 1]$ for $P \in \mathcal{P}_n$. When $\rho = |\psi\rangle \langle \psi|$ and $|\psi\rangle$ is a stabilizer states, there are exactly $2^n = d$ Pauli operators in $\mathcal{P}_n$ such that $\text{Tr}(P \rho) = 1$ and for others $\text{Tr}(P \rho) = 0$, we have $Q_\sigma = 1/d$. ■

In the following, we use $R_k^\pm$ and $Q_\lambda^\pm$ to denote the corresponding quantities defined for $\tilde{\rho}_\pm$.

b. Formula of 4-order Clifford twirling

The multi-qubit Clifford forms a unitary 3-design but fails gracefully to become a unitary 4-design [67,68,108]. Based on the proof of Lemma B2, we only need to reanalyze the 4-order term [cf. A_4 in Eq. (B26)], by replacing the 4-order unitary twirling formula in Eq. (B25) with the corresponding Clifford twirling formula. Denote the n -qubit Clifford group as $\text{Cl}(d)$ with $d = 2^n$, its 4-order twirling acting on an operator $X \in \mathcal{L}(\mathbb{C}^d)$ is stated as Theorem S.5

TABLE III. Character table of the symmetric group S_4 . Rows correspond to irreducible representations (partitions $\lambda \vdash 4$), and columns to conjugacy classes (cycle types of $\sigma \in S_4$).

$\chi^\lambda(\sigma)$	(1, 1, 1, 1)	(2, 1, 1)	(2, 2)	(3, 1)	(4)
(4)	1	1	1	1	1
(3, 1)	3	1	-1	0	-1
(2, 2)	2	0	2	-1	0
(2, 1, 1)	3	-1	-1	0	1
(1, 1, 1, 1)	1	-1	1	1	-1

in Ref. [109],

$$\begin{aligned} E_{\Delta_{\text{Cl}(d)}}^4(X) &:= \int_{\text{Cl}(d)} U^{\otimes 4} X U^{\dagger \otimes 4} dU \\ &= \frac{1}{24} \sum_{\lambda \vdash 4, l(\lambda) \leq d} d_\lambda \sum_{\sigma \in S_4} \left[\frac{1}{D_\lambda^+} \text{Tr}(X Q \mathbb{P}_{\sigma^{-1}}^4) Q \right. \\ &\quad \left. + \frac{1}{D_\lambda^-} \text{Tr}(X Q^\perp \mathbb{P}_{\sigma^{-1}}^4) Q^\perp \right] \mathbb{P}_\sigma^4 P_\lambda, \end{aligned} \quad (\text{B36})$$

where $\lambda \vdash 4$ is a partition of 4 and $l(\lambda)$ is its length. The dimensions of the Specht modules d_λ and Weyl modules $D_\lambda^\pm$ are provided in Table 1 of Ref. [67]. The orthogonal projectors to the submodules are [cf. Eq. (S.8) in Ref. [109]]

$$P_\lambda = \frac{d_\lambda}{24} \sum_{\sigma \in S_4} \chi^\lambda(\sigma) \mathbb{P}_\sigma^4, \quad (\text{B37})$$

where the character values $\chi^\lambda(\sigma)$ can be calculated using the Murnaghan-Nakayama rule and are listed in Table III.

We will use the 4-order Clifford twirling of the operator [cf. Eqs. (1) and (B24), ignoring script +],

$$X^{(1,2)\otimes 2} = \sum_{\tilde{s}_1, \tilde{s}_2, \tilde{s}_3, \tilde{s}_4} X_2(\tilde{s}_1, \tilde{s}_2) X_2(\tilde{s}_3, \tilde{s}_4) |\tilde{s}_1 \tilde{s}_2 \tilde{s}_3 \tilde{s}_4\rangle \langle \tilde{s}_1 \tilde{s}_2 \tilde{s}_3 \tilde{s}_4|. \quad (\text{B38})$$

We consider $d \geq 4$ to include all partitions of 4, consistent with implied the assumption $d \geq 8$ in Lemma 4. Here, we list all the terms in

$$\text{Tr}(E_{\Delta_{\text{Cl}(d)}}^4(X^{(1,2)\otimes 2}) \rho^{\otimes 4}) = \frac{1}{24} \sum_{\lambda \vdash 4, l(\lambda) \leq d} \sum_{\sigma \in S_4} T(\lambda, \sigma), \quad (\text{B39})$$

where we define

$$\begin{aligned}
 T(\lambda, \sigma) &:= \text{Tr} \left(d_\lambda \times \left[\frac{1}{D_\lambda^+} \text{Tr} (X^{(1,2)\otimes 2} Q \mathbb{P}_{\sigma^{-1}}^4) Q + \frac{1}{D_\lambda^-} \text{Tr} (X^{(1,2)\otimes 2} Q^\perp \mathbb{P}_{\sigma^{-1}}^4) Q^\perp \right] \mathbb{P}_\sigma^4 P_\lambda \rho^{\otimes 4} \right) \\
 &= d_\lambda \left(\frac{1}{D_\lambda^+} \text{Tr} (X^{(1,2)\otimes 2} Q \mathbb{P}_{\sigma^{-1}}^4) - \frac{1}{D_\lambda^-} \text{Tr} (X^{(1,2)\otimes 2} \mathbb{P}_{\sigma^{-1}}^4) + \frac{1}{D_\lambda^-} \text{Tr} (X^{(1,2)\otimes 2} Q \mathbb{P}_{\sigma^{-1}}^4) \right) \text{Tr} (Q \mathbb{P}_\sigma^4 P_\lambda \rho^{\otimes 4}) \\
 &\quad + d_\lambda \left(\frac{1}{D_\lambda^-} \text{Tr} (X^{(1,2)\otimes 2} \mathbb{P}_{\sigma^{-1}}^4) - \frac{1}{D_\lambda^-} \text{Tr} (X^{(1,2)\otimes 2} Q \mathbb{P}_{\sigma^{-1}}^4) \right) \text{Tr} (\mathbb{P}_\sigma^4 P_\lambda \rho^{\otimes 4}). \tag{B40}
 \end{aligned}$$

$$(1) \ \lambda = [1, 1, 1, 1]: d_\lambda = 1, D_\lambda^+ = (d-1)(d-2)/6, D_\lambda^- = (d-4)(d-2)(d-1)(d+1)/24.$$

$$(a) \ \sigma \sim [1, 1, 1, 1]$$

$$(i) \ \sigma = \mathbb{P}_{(1)}^4$$

$$\begin{aligned}
 T(\lambda, \sigma) &= \frac{d^2(d^2 - 3d - 1)}{4(d-4)(d-2)(d-1)} (1 \times Q_{1,1,1,1} + (-6) \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + (-6) \times Q_4) \\
 &\quad - \frac{d}{(d-4)(d-2)} (1 \times R_1 + (-6) \times R_2 + 3 \times R_2^2 + 8 \times R_3 + (-6) \times R_4), \tag{B41}
 \end{aligned}$$

$$(b) \ \sigma \sim [2, 1, 1]$$

$$(i) \ \sigma = \mathbb{P}_{(12)}^4, \mathbb{P}_{(34)}^4$$

$$\begin{aligned}
 T(\lambda, \sigma) &= \frac{d^3}{4(d-2)(d-1)} ((-1) \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + (-3) \times Q_{2,2} + (-8) \times Q_{3,1} + 6 \times Q_4), \tag{B42}
 \end{aligned}$$

$$(ii) \ \sigma = \mathbb{P}_{(13)}^4, \mathbb{P}_{(14)}^4, \mathbb{P}_{(23)}^4, \mathbb{P}_{(24)}^4$$

$$\begin{aligned}
 T(\lambda, \sigma) &= \frac{d(d^2 - 3d - 1)}{4(d-4)(d-2)(d-1)} ((-1) \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + (-3) \times Q_{2,2} + (-8) \times Q_{3,1} + 6 \times Q_4) \\
 &\quad - \frac{1}{(d-4)(d-2)} ((-1) \times R_1 + 6 \times R_2 + (-3) \times R_2^2 + (-8) \times R_3 + 6 \times R_4), \tag{B43}
 \end{aligned}$$

$$(c) \ \sigma \sim [2, 2]$$

$$(i) \ \sigma = \mathbb{P}_{(12)(34)}^4$$

$$\begin{aligned}
 T(\lambda, \sigma) &= \frac{d^2(d^2 - 7d + 3)}{4(d-4)(d-2)(d-1)} (1 \times Q_{1,1,1,1} + (-6) \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + (-6) \times Q_4) \\
 &\quad + \frac{d(d-1)}{(d-4)(d-2)} (1 \times R_1 + (-6) \times R_2 + 3 \times R_2^2 + 8 \times R_3 + (-6) \times R_4), \tag{B44}
 \end{aligned}$$

$$(ii) \ \sigma = \mathbb{P}_{(13)(24)}^4, \mathbb{P}_{(14)(23)}^4$$

$$\begin{aligned}
 T(\lambda, \sigma) &= \frac{d^3 + d^2 - d}{4(d-2)(d-1)} (1 \times Q_{1,1,1,1} + (-6) \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + (-6) \times Q_4), \tag{B45}
 \end{aligned}$$

(d) $\sigma \sim [3, 1]$

$$(i) \sigma = \mathbb{P}_{(123)}^4, \mathbb{P}_{(132)}^4, \mathbb{P}_{(124)}^4, \mathbb{P}_{(142)}^4, \mathbb{P}_{(134)}^4, \mathbb{P}_{(143)}^4, \mathbb{P}_{(234)}^4, \mathbb{P}_{(243)}^4$$

$$T(\lambda, \sigma) = \frac{d^2}{4(d-2)(d-1)} (1 \times Q_{1,1,1,1} + (-6) \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + (-6) \times Q_4), \quad (B46)$$

(e) $\sigma \sim [4]$

$$(i) \sigma = \mathbb{P}_{(1234)}^4, \mathbb{P}_{(1243)}^4, \mathbb{P}_{(1342)}^4, \mathbb{P}_{(1432)}^4$$

$$T(\lambda, \sigma) = \frac{d(d^2 - 7d + 3)}{4(d-4)(d-2)(d-1)} ((-1) \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + (-3) \times Q_{2,2} + (-8) \times Q_{3,1} + 6 \times Q_4) \\ + \frac{d-1}{(d-4)(d-2)} ((-1) \times R_1 + 6 \times R_2 + (-3) \times R_2^2 + (-8) \times R_3 + 6 \times R_4), \quad (B47)$$

$$(ii) \sigma = \mathbb{P}_{(1324)}^4, \mathbb{P}_{(1423)}^4$$

$$T(\lambda, \sigma) = \frac{d^3}{4(d-2)(d-1)} ((-1) \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + (-3) \times Q_{2,2} + (-8) \times Q_{3,1} + 6 \times Q_4), \quad (B48)$$

(2) $\lambda = [2, 1, 1]: d_\lambda = 3, D_\lambda^+ = 0, D_\lambda^- = d(d-2)(d-1)(d+1)/8.$ (a) $\sigma \sim [1, 1, 1, 1]$

$$(i) \sigma = \mathbb{P}_{(1)}^4$$

$$T(\lambda, \sigma) = \frac{3}{d-2} (3 \times Q_{1,1,1,1} + (-6) \times Q_{2,1,1} + (-3) \times Q_{2,2} + 0 \times Q_{3,1} + 6 \times Q_4) \\ - \frac{3}{d-2} (3 \times R_1 + (-6) \times R_2 + (-3) \times R_2^2 + 0 \times R_3 + 6 \times R_4), \quad (B49)$$

(b) $\sigma \sim [2, 1, 1]$

$$(i) \sigma = \mathbb{P}_{(12)}^4, \mathbb{P}_{(34)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B50)$$

$$(ii) \sigma = \mathbb{P}_{(13)}^4, \mathbb{P}_{(14)}^4, \mathbb{P}_{(23)}^4, \mathbb{P}_{(24)}^4$$

$$T(\lambda, \sigma) = \frac{3}{d(d-2)} ((-1) \times Q_{1,1,1,1} + 2 \times Q_{2,1,1} + 1 \times Q_{2,2} + 0 \times Q_{3,1} + (-2) \times Q_4) \\ - \frac{3}{d(d-2)} ((-1) \times R_1 + 2 \times R_2 + 1 \times R_2^2 + 0 \times R_3 + (-2) \times R_4), \quad (B51)$$

(c) $\sigma \sim [2, 2]$

$$(i) \sigma = \mathbb{P}_{(12)(34)}^4$$

$$T(\lambda, \sigma) = \frac{3(1-d)}{d-2} ((-1) \times Q_{1,1,1,1} + 2 \times Q_{2,1,1} + 1 \times Q_{2,2} + 0 \times Q_{3,1} + (-2) \times Q_4) \\ + \frac{3(d-1)}{d-2} ((-1) \times R_1 + 2 \times R_2 + 1 \times R_2^2 + 0 \times R_3 + (-2) \times R_4), \quad (B52)$$

$$(ii) \sigma = \mathbb{P}_{(13)(24)}^4, \mathbb{P}_{(14)(23)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B53)$$

$$(d) \sigma \sim [3, 1]$$

$$(i) \sigma = \mathbb{P}_{(123)}^4, \mathbb{P}_{(132)}^4, \mathbb{P}_{(124)}^4, \mathbb{P}_{(142)}^4, \mathbb{P}_{(134)}^4, \mathbb{P}_{(143)}^4, \mathbb{P}_{(234)}^4, \mathbb{P}_{(243)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B54)$$

$$(e) \sigma \sim [4]$$

$$(i) \sigma = \mathbb{P}_{(1234)}^4, \mathbb{P}_{(1243)}^4, \mathbb{P}_{(1342)}^4, \mathbb{P}_{(1432)}^4$$

$$\begin{aligned} T(\lambda, \sigma) = & \frac{3(1-d)}{d(d-2)} (1 \times Q_{1,1,1,1} + (-2) \times Q_{2,1,1} + (-1) \times Q_{2,2} + 0 \times Q_{3,1} + 2 \times Q_4) \\ & + \frac{3(d-1)}{d(d-2)} (1 \times R_1 + (-2) \times R_2 + (-1) \times R_2^2 + 0 \times R_3 + 2 \times R_4), \end{aligned} \quad (B55)$$

$$(ii) \sigma = \mathbb{P}_{(1324)}^4, \mathbb{P}_{(1423)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B56)$$

$$(3) \lambda = [2, 2]: d_\lambda = 2, D_\lambda^+ = (d-1)(d+1)/3, D_\lambda^- = (d-2)(d-1)(d+1)(d+2)/12.$$

$$(a) \sigma \sim [1, 1, 1, 1]$$

$$(i) \sigma = \mathbb{P}_{(1)}^4$$

$$\begin{aligned} T(\lambda, \sigma) = & \frac{d^2}{2(d-2)(d-1)(d+1)(d+2)} (2 \times Q_{1,1,1,1} + 0 \times Q_{2,1,1} + 6 \times Q_{2,2} + (-8) \times Q_{3,1} + 0 \times Q_4) \\ & - \frac{2d}{(d-2)(d+2)} (2 \times R_1 + 0 \times R_2 + 6 \times R_2^2 + (-8) \times R_3 + 0 \times R_4), \end{aligned} \quad (B57)$$

$$(b) \sigma \sim [2, 1, 1]$$

$$(i) \sigma = \mathbb{P}_{(12)}^4, \mathbb{P}_{(34)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B58)$$

$$(ii) \sigma = \mathbb{P}_{(13)}^4, \mathbb{P}_{(14)}^4, \mathbb{P}_{(23)}^4, \mathbb{P}_{(24)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B59)$$

$$(c) \sigma \sim [2, 2]$$

$$(i) \sigma = \mathbb{P}_{(12)(34)}^4$$

$$\begin{aligned} T(\lambda, \sigma) = & \frac{d^3(d^2 - 3d - 1)}{2(d-2)(d-1)(d+1)(d+2)} (2 \times Q_{1,1,1,1} + 0 \times Q_{2,1,1} + 6 \times Q_{2,2} + (-8) \times Q_{3,1} + 0 \times Q_4) \\ & + \frac{2d(d-1)}{(d-2)(d+2)} (2 \times R_1 + 0 \times R_2 + 6 \times R_2^2 + (-8) \times R_3 + 0 \times R_4), \end{aligned} \quad (B60)$$

$$(ii) \sigma = \mathbb{P}_{(13)(24)}^4, \mathbb{P}_{(14)(23)}^4$$

$$T(\lambda, \sigma) = \frac{d^3 + d^2 - d}{2(d-1)(d+1)} (2 \times Q_{1,1,1,1} + 0 \times Q_{2,1,1} + 6 \times Q_{2,2} + (-8) \times Q_{3,1} + 0 \times Q_4), \quad (B61)$$

$$(d) \sigma \sim [3, 1]$$

$$(i) \sigma = \mathbb{P}_{(123)}^4, \mathbb{P}_{(132)}^4, \mathbb{P}_{(124)}^4, \mathbb{P}_{(142)}^4, \mathbb{P}_{(134)}^4, \mathbb{P}_{(143)}^4, \mathbb{P}_{(234)}^4, \mathbb{P}_{(243)}^4$$

$$T(\lambda, \sigma) = \frac{d^2}{2(d-1)(d+1)} ((-1) \times Q_{1,1,1,1} + 0 \times Q_{2,1,1} + (-3) \times Q_{2,2} + 4 \times Q_{3,1} + 0 \times Q_4), \quad (B62)$$

$$(e) \sigma \sim [4]$$

$$(i) \sigma = \mathbb{P}_{(1234)}^4, \mathbb{P}_{(1243)}^4, \mathbb{P}_{(1342)}^4, \mathbb{P}_{(1432)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B63)$$

$$(ii) \sigma = \mathbb{P}_{(1324)}^4, \mathbb{P}_{(1423)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B64)$$

$$(4) \lambda = [3, 1]: d_\lambda = 3, D_\lambda^+ = 0, D_\lambda^- = d(d-1)(d+1)(d+2)/8.$$

$$(a) \sigma \sim [1, 1, 1, 1]$$

$$(i) \sigma = \mathbb{P}_{(1)}^4$$

$$T(\lambda, \sigma) = \frac{3}{d+2} (3 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + (-3) \times Q_{2,2} + 0 \times Q_{3,1} + (-6) \times Q_4) \\ - \frac{3}{d+2} (3 \times R_1 + 6 \times R_2 + (-3) \times R_2^2 + 0 \times R_3 + (-6) \times R_4), \quad (B65)$$

$$(b) \sigma \sim [2, 1, 1]$$

$$(i) \sigma = \mathbb{P}_{(12)}^4, \mathbb{P}_{(34)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B66)$$

$$(ii) \sigma = \mathbb{P}_{(13)}^4, \mathbb{P}_{(14)}^4, \mathbb{P}_{(23)}^4, \mathbb{P}_{(24)}^4$$

$$T(\lambda, \sigma) = \frac{3}{d(d+2)} (1 \times Q_{1,1,1,1} + 2 \times Q_{2,1,1} + (-1) \times Q_{2,2} + 0 \times Q_{3,1} + (-2) \times Q_4) \\ - \frac{3}{d(d+2)} (1 \times R_1 + 2 \times R_2 + (-1) \times R_2^2 + 0 \times R_3 + (-2) \times R_4), \quad (B67)$$

$$(c) \sigma \sim [2, 2]$$

$$(i) \sigma = \mathbb{P}_{(12)(34)}^4$$

$$T(\lambda, \sigma) = \frac{3(1-d)}{d+2} ((-1) \times Q_{1,1,1,1} + (-2) \times Q_{2,1,1} + 1 \times Q_{2,2} + 0 \times Q_{3,1} + 2 \times Q_4) \\ + \frac{3(d-1)}{d+2} ((-1) \times R_1 + (-2) \times R_2 + 1 \times R_2^2 + 0 \times R_3 + 2 \times R_4), \quad (B68)$$

$$(ii) \sigma = \mathbb{P}_{(13)(24)}^4, \mathbb{P}_{(14)(23)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B69)$$

$$(d) \sigma \sim [3, 1]$$

$$(i) \sigma = \mathbb{P}_{(123)}^4, \mathbb{P}_{(132)}^4, \mathbb{P}_{(124)}^4, \mathbb{P}_{(142)}^4, \mathbb{P}_{(134)}^4, \mathbb{P}_{(143)}^4, \mathbb{P}_{(234)}^4, \mathbb{P}_{(243)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B70)$$

$$(e) \sigma \sim [4]$$

$$(i) \sigma = \mathbb{P}_{(1234)}^4, \mathbb{P}_{(1243)}^4, \mathbb{P}_{(1342)}^4, \mathbb{P}_{(1432)}^4$$

$$\begin{aligned} T(\lambda, \sigma) = & \frac{3(1-d)}{d(d+2)} \left((-1) \times Q_{1,1,1,1} + (-2) \times Q_{2,1,1} + 1 \times Q_{2,2} + 0 \times Q_{3,1} + 2 \times Q_4 \right) \\ & + \frac{3(d-1)}{d(d+2)} \left((-1) \times R_1 + (-2) \times R_2 + 1 \times R_2^2 + 0 \times R_3 + 2 \times R_4 \right), \end{aligned} \quad (B71)$$

$$(ii) \sigma = \mathbb{P}_{(1324)}^4, \mathbb{P}_{(1423)}^4$$

$$T(\lambda, \sigma) = 0, \quad (B72)$$

$$(5) \lambda = [4]: d_\lambda = 1, D_\lambda^+ = (d+1)(d+2)/6, D_\lambda^- = (d-1)(d+1)(d+2)(d+4)/24.$$

$$(a) \sigma \sim [1, 1, 1, 1]$$

$$(i) \sigma = \mathbb{P}_{(1)}^4$$

$$\begin{aligned} T(\lambda, \sigma) = & \frac{d^2(d^2 + 5d + 7)}{4(d+1)(d+2)(d+4)} \left(1 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + 6 \times Q_4 \right) \\ & - \frac{d}{(d+2)(d+4)} \left(1 \times R_1 + 6 \times R_2 + 3 \times R_2^2 + 8 \times R_3 + 6 \times R_4 \right), \end{aligned} \quad (B73)$$

$$(b) \sigma \sim [2, 1, 1]$$

$$(i) \sigma = \mathbb{P}_{(12)}^4, \mathbb{P}_{(34)}^4$$

$$T(\lambda, \sigma) = \frac{d^3}{4(d+1)(d+2)} \left(1 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + 6 \times Q_4 \right), \quad (B74)$$

$$(ii) \sigma = \mathbb{P}_{(13)}^4, \mathbb{P}_{(14)}^4, \mathbb{P}_{(23)}^4, \mathbb{P}_{(24)}^4$$

$$\begin{aligned} T(\lambda, \sigma) = & \frac{d(d^2 + 5d + 7)}{4(d+1)(d+2)(d+4)} \left(1 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + 6 \times Q_4 \right) \\ & - \frac{1}{(d+2)(d+4)} \left(1 \times R_1 + 6 \times R_2 + 3 \times R_2^2 + 8 \times R_3 + 6 \times R_4 \right), \end{aligned} \quad (B75)$$

$$(c) \sigma \sim [2, 2]$$

$$(i) \sigma = \mathbb{P}_{(12)(34)}^4$$

$$\begin{aligned} T(\lambda, \sigma) = & \frac{d^2(d^2 + d + 3)}{4(d+1)(d+2)(d+4)} \left(1 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + 6 \times Q_4 \right) \\ & + \frac{d(d-1)}{(d+2)(d+4)} \left(1 \times R_1 + 6 \times R_2 + 3 \times R_2^2 + 8 \times R_3 + 6 \times R_4 \right), \end{aligned} \quad (B76)$$

$$(ii) \sigma = \mathbb{P}_{(13)(24)}^4, \mathbb{P}_{(14)(23)}^4$$

$$T(\lambda, \sigma) = \frac{d^3 + d^2 - d}{4(d+1)(d+2)} (1 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + 6 \times Q_4), \quad (B77)$$

$$(d) \sigma \sim [3, 1]$$

$$(i) \sigma = \mathbb{P}_{(123)}^4, \mathbb{P}_{(132)}^4, \mathbb{P}_{(124)}^4, \mathbb{P}_{(142)}^4, \mathbb{P}_{(134)}^4, \mathbb{P}_{(143)}^4, \mathbb{P}_{(234)}^4, \mathbb{P}_{(243)}^4$$

$$T(\lambda, \sigma) = \frac{d^2}{4(d+1)(d+2)} (1 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + 6 \times Q_4), \quad (B78)$$

$$(e) \sigma \sim [4]$$

$$(i) \sigma = \mathbb{P}_{(1234)}^4, \mathbb{P}_{(1243)}^4, \mathbb{P}_{(1342)}^4, \mathbb{P}_{(1432)}^4$$

$$T(\lambda, \sigma) = \frac{d(d^2 + d + 3)}{4(d+1)(d+2)(d+4)} (1 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + 6 \times Q_4) \\ + \frac{d-1}{(d+2)(d+4)} (1 \times R_1 + 6 \times R_2 + 3 \times R_2^2 + 8 \times R_3 + 6 \times R_4), \quad (B79)$$

$$(ii) \sigma = \mathbb{P}_{(1324)}^4, \mathbb{P}_{(1423)}^4$$

$$T(\lambda, \sigma) = \frac{d^3}{4(d+1)(d+2)} (1 \times Q_{1,1,1,1} + 6 \times Q_{2,1,1} + 3 \times Q_{2,2} + 8 \times Q_{3,1} + 6 \times Q_4). \quad (B80)$$

Summarizing all the terms, we obtain

$$\text{Tr} \left(E_{\Delta_{\text{Cl}(d)}}^4 (X^{(1,2)\otimes 2}) \rho^{\otimes 4} \right) = \frac{5d^5 - d^4 - 5d^3 - 43d^2 + 48}{24(d-2)(d-1)(d+1)(d+2)} Q_{1,1,1,1} + \frac{d^2 + d + 2}{2(d+2)} Q_{2,1,1} \\ + \frac{5d^5 - 9d^4 - 21d^3 + 29d^2 + 16d - 16}{8(d-2)(d-1)(d+1)(d+2)} Q_{2,2} + \frac{d^2(d^3 + d^2 - d - 5)}{6(d-2)(d-1)(d+1)(d+2)} Q_{3,1} \\ + \frac{d-1}{2} Q_4 - \frac{1}{d+2} R_1 - \frac{2}{d+2} R_2 + \frac{d+1}{d+2} R_2^2. \quad (B81)$$

c. Proof of Lemma B3

Proof. The proof is similar to that of Lemma B2. The first two terms A_2^{Cl} and A_3^{Cl} are the same as Eqs. (B18) and (B22) since the multi-qubit Clifford group is a unitary 3-design [67]. Inserting $d_+ = d_- = d/2$, we have

$$A_2^{\text{Cl}} = \binom{N_M}{2} \left(d_+ R_1^{+2} + (d_+ - 1) R_2^+ \right) + (+ \iff -) \\ = \binom{N_M}{2} \left[\frac{d}{2} \left(R_1^{+2} + R_1^{-2} \right) + \left(\frac{d}{2} - 1 \right) \left(R_2^+ + R_2^- \right) \right] \quad (B82)$$

and

$$A_3^{\text{Cl}} = 6 \binom{N_M}{3} \left(\frac{d_+ - 1}{d_+ + 2} R_2^{+2} R_1^+ + \frac{2(d_+ + 1)}{d_+ + 2} R_3^+ - \frac{1}{d_+ + 2} R_1^{+3} \right) + (+ \iff -) \\ = 6 \binom{N_M}{3} \left(\frac{d-2}{d+4} \left(R_2^{+2} R_1^+ + R_2^{-2} R_1^- \right) + \frac{2(d+2)}{d+4} \left(R_3^+ + R_3^- \right) - \frac{1}{d+4} \left(R_1^{+3} + R_1^{-3} \right) \right). \quad (B83)$$

Instead of A_4 in Eq. (B26), we obtain

$$\begin{aligned}
A_4^{\text{Cl}} &= 6 \binom{N_M}{4} \text{Tr} \left(E_{\Delta_{\text{Cl}(d_+)}} \left(X_+^{(1,2) \otimes 2} \right) \tilde{\rho}_+^{\otimes 4} \right) - 6 \binom{N_M}{4} \text{Tr} \left(\Phi_2 \left(X_+^{(1,2)} \right) \tilde{\rho}_+^{\otimes 2} \right) \text{Tr} \left(\Phi_2 \left(X_-^{(1,2)} \right) \tilde{\rho}_-^{\otimes 2} \right) + (+ \iff -) \\
&= 6 \binom{N_M}{4} \left[\frac{5d_+^5 - d_+^4 - 5d_+^3 - 43d_+^2 + 48}{24(d_+ - 2)(d_+ - 1)(d_+ + 1)(d_+ + 2)} Q_{1,1,1,1}^+ + \frac{d_+^2 + d_+ + 2}{2(d_+ + 2)} Q_{2,1,1}^+ \right. \\
&\quad + \frac{5d_+^5 - 9d_+^4 - 21d_+^3 + 29d_+^2 + 16d_+ - 16}{8(d_+ - 2)(d_+ - 1)(d_+ + 1)(d_+ + 2)} Q_{2,2}^+ + \frac{d_+^2(d_+^3 + d_+^2 - d_+ - 5)}{6(d_+ - 2)(d_+ - 1)(d_+ + 1)(d_+ + 2)} Q_{3,1}^+ \\
&\quad + \left. \frac{d_+ - 1}{2} Q_4^+ - \frac{1}{d_+ + 2} R_1^+ - \frac{2}{d_+ + 2} R_2^+ + \frac{d_+ + 1}{d_+ + 2} R_2^{+2} - R_2^+ R_2^- \right] + (+ \iff -) \\
&= 6 \binom{N_M}{4} \left[\frac{5d^5 - 2d^4 - 20d^3 - 344d^2 + 1536}{48(d - 4)(d - 2)(d + 2)(d + 4)} (Q_{1,1,1,1}^+ + Q_{1,1,1,1}^-) + \frac{d^2 + 2d + 8}{4(d + 4)} (Q_{2,1,1}^+ + Q_{2,1,1}^-) \right. \\
&\quad + \frac{5d^5 - 18d^4 - 84d^3 + 232d^2 + 256d - 512}{16(d - 4)(d - 2)(d + 2)(d + 4)} (Q_{2,2}^+ + Q_{2,2}^-) + \frac{d^2(d^3 + 2d^2 - 4d - 40)}{12(d - 4)(d - 2)(d + 2)(d + 4)} (Q_{3,1}^+ + Q_{3,1}^-) \\
&\quad + \left. \frac{d - 2}{4} (Q_4^+ + Q_4^-) - \frac{2}{d + 4} (R_1^+ + R_1^-) - \frac{4}{d + 4} (R_2^+ + R_2^-) + \frac{d + 2}{d + 4} (R_2^{+2} + R_2^{-2}) - 2R_2^+ R_2^- \right]. \quad (\text{B84})
\end{aligned}$$

We calculate the variance of the estimator ω_u [see Eq. (B4), cf. Eq. (B27)],

$$\begin{aligned}
\text{Var}^{\text{Cl}}(\omega_u) &= \mathbb{E}[\omega_u^2] - \mathbb{E}[\omega_u]^2 = \binom{N_M}{2}^{-2} \sum_{\substack{\{i,j\} \in \binom{[N_M]}{2} \\ \{i',j'\} \in \binom{[N_M]}{2}}} \mathbb{E}[X_2(s_i, s_j) X_2(s_{i'}, s_{j'})] - \mathbb{E}[\omega_u]^2 \\
&= \binom{N_M}{2}^{-2} (A_2^{\text{Cl}} + A_3^{\text{Cl}} + A_4^{\text{Cl}}) - (R_2^+ - R_2^-)^2. \quad (\text{B85})
\end{aligned}$$

Inserting Eqs. (B82), (B83), and (B84), we have

$$\begin{aligned}
\text{Var}^{\text{Cl}}(\omega_u) &= \binom{N_M}{2}^{-2} (A_2^{\text{Cl}} + A_3^{\text{Cl}} + A_4^{\text{Cl}}) - (R_2^+ - R_2^-)^2 \\
&= -\frac{2(N_M - 2)(N_M - 3)}{N_M(N_M - 1)(d + 4)} + \frac{d}{N_M(N_M - 1)} (R_1^{+2} + R_1^{-2}) - \frac{4(N_M - 2)}{N_M(N_M - 1)(d + 4)} (R_1^{+3} + R_1^{-3}) \\
&\quad + \frac{4(N_M - 2)(d - 2)}{N_M(N_M - 1)(d + 4)} (R_1^+ R_2^{+2} + R_1^- R_2^{-2}) + \left(\frac{d - 2}{N_M(N_M - 1)} - \frac{4(N_M - 2)(N_M - 3)}{N_M(N_M - 1)(d + 4)} \right) (R_2^+ + R_2^-) \\
&\quad - \frac{4N_M - 6}{N_M(N_M - 1)} (R_2^+ - R_2^-)^2 - \frac{2(N_M - 2)(N_M - 3)}{N_M(N_M - 1)(d + 4)} (R_2^{+2} + R_2^{-2}) + \frac{8(N_M - 2)(d + 2)}{N_M(N_M - 1)(d + 4)} (R_3^+ + R_3^-) \\
&\quad + \frac{(N_M - 2)(N_M - 3)(5d^5 - 2d^4 - 20d^3 - 344d^2 + 1536)}{48N_M(N_M - 1)(d - 4)(d - 2)(d + 2)(d + 4)} (Q_{1,1,1,1}^+ + Q_{1,1,1,1}^-) \\
&\quad + \frac{(N_M - 2)(N_M - 3)(d^2 + 2d + 8)}{4N_M(N_M - 1)(d + 4)} (Q_{2,1,1}^+ + Q_{2,1,1}^-) \\
&\quad + \frac{(N_M - 2)(N_M - 3)(5d^5 - 18d^4 - 84d^3 + 232d^2 + 256d - 512)}{16N_M(N_M - 1)(d - 4)(d - 2)(d + 2)(d + 4)} (Q_{2,2}^+ + Q_{2,2}^-) \\
&\quad + \frac{(N_M - 2)(N_M - 3)d^2(d^3 + 2d^2 - 4d - 40)}{12N_M(N_M - 1)(d - 4)(d - 2)(d + 2)(d + 4)} (Q_{3,1}^+ + Q_{3,1}^-) + \frac{(N_M - 2)(N_M - 3)(d - 2)}{4N_M(N_M - 1)} (Q_4^+ + Q_4^-). \quad (\text{B86})
\end{aligned}$$

We have $R_k^\pm \leq 1$ and $Q_\sigma^\pm \leq 2/d$ from Lemma B4, so that we obtain an upper bound for $\text{Var}^{\text{Cl}}(\omega_u)$,

$$\begin{aligned} \text{Var}^{\text{Cl}}(\omega_u) &\leq \frac{4d}{N_M(N_M - 1)} + \frac{24(N_M - 2)}{N_M(N_M - 1)} \\ &\quad + \frac{4(N_M - 2)(N_M - 3)}{N_M(N_M - 1)} \\ &= \mathcal{O}\left(\frac{d}{N_M^2} + \frac{1}{N_M} + 1\right). \end{aligned} \quad (\text{B87})$$

Furthermore, suppose we draw N_U independent random unitaries and construct an estimator by averaging the individual estimators obtained for each choice of random unitary, i.e.,

$$\omega = \frac{1}{N_U} \sum_{u=1}^{N_U} \omega_u. \quad (\text{B88})$$

Then we can estimate $\text{Tr}(O\rho^2)$ unbiasedly, with the variance given by

$$\text{Var}^{\text{Cl}}(\omega) = \frac{1}{N_U} \text{Var}^{\text{Cl}}(\omega_u) = \frac{1}{N_U} \mathcal{O}\left(\frac{d}{N_M^2} + \frac{1}{N_M} + 1\right). \quad (\text{B89})$$

■

Based on Lemmas A1, B1, and B3, we can now prove Lemma 4 in the main text.

Proof of Lemma 4. To apply Lemma A1, it is sufficient to suppress $\text{Var}^{\text{Cl}}(\omega) \leq \varepsilon^2/3 < \varepsilon^2/2$. If we choose $N_U = \mathcal{O}(1/\varepsilon^2)$ and $N_M = \mathcal{O}(\sqrt{d})$, then $N_U \times N_M = \mathcal{O}(\sqrt{d}/\varepsilon^2)$ copies of states are sufficient to estimate $\text{Tr}(O\rho^2)$ for traceless dichotomic O to ε -precision within constant failure probability. ■

3. Decomposition of dichotomic observables

Now we provide the analysis of the sample complexity stated in Theorem 1. We assume that $\min\{d_-, d_+\} < d/4$;

otherwise, Lemma 1 already yields the desired sample complexity since $d_- = d_+ = \Omega(d)$. We then obtain the following decomposition, which is illustrated in Fig. 5.

Lemma B5 (Formal version of Lemma 2). Suppose O is a dichotomic observable satisfying $\min\{d_+, d_-\} < d/4$. Then, $O = \frac{1}{2}(\pm\mathbb{I} + O_1 + O_2 + O_3)$, where each O_i is a dichotomic observable with eigen-subspace dimensions $d_-^{(i)}$ and $d_+^{(i)}$ satisfying $\min\{d_-^{(i)}, d_+^{(i)}\} \geq d/3$.

Proof. Let P_+ and P_- denote the projectors onto the eigenspaces of O corresponding to the eigenvalues $+1$ and -1 , respectively, so that $O = P_+ - P_-$ and $\mathbb{I} = P_+ + P_-$. Without loss of generality, we first assume $d_- < d_+$ and $d_- < d/4$. Next, decompose the $+1$ eigenspace Π_+ into three orthogonal subspaces, denoted by $\Pi_+^{(1)}$, $\Pi_+^{(2)}$, and $\Pi_+^{(3)}$ (projectors correspondingly denoted as $P_+^{(j)}$), each having dimension at least $\lfloor d/3 \rfloor$. Define

$$O_i = \left(\sum_{j=1}^3 (-1)^{\delta_{ij}} P_+^{(j)} \right) - P_-, \quad (\text{B90})$$

where the exponents δ_{ij} are the delta function, $\delta_{ij} = 1$ if and only if $i = j$. A direct calculation then shows that $O = \frac{1}{2}(\mathbb{I} + O_1 + O_2 + O_3)$. Moreover, for each O_i , one can verify that the dimensions of its positive and negative eigenspaces satisfy

$$d_+^{(i)} \geq d_+ - \left\lfloor \frac{d_+}{3} \right\rfloor > \frac{2}{3}d_+ - 1 > \frac{d}{2} - 1 \geq \frac{d}{3}, \quad (\text{B91})$$

where the last inequality holds because $0 < d_- < d/4$ implies that $d \geq 8$. Also,

$$d_-^{(i)} = d_- + \left\lfloor \frac{d_+}{3} \right\rfloor \geq 1 + \left\lfloor \frac{d-1}{3} \right\rfloor \geq \frac{d}{3}. \quad (\text{B92})$$

This completes the proof of the cases $d_+ > d_-$.

For the case where $d_+ < d_-$, we apply the above procedure to the observable $-O = \frac{1}{2}(\mathbb{I} + O_1 + O_2 + O_3)$. This yields the decomposition $O = \frac{1}{2}(-\mathbb{I} + O'_1 + O'_2 + O'_3)$, where $O'_i = -O_i$ for $1 \leq i \leq 3$. Since negating an

$$\begin{bmatrix} \text{green block with } - \\ \text{blue block with } + \\ \text{blue block with } + \\ \text{blue block with } + \end{bmatrix} = \frac{1}{2} \left(\begin{bmatrix} \text{green block with } + \\ \text{blue block with } + \\ \text{blue block with } + \\ \text{blue block with } + \end{bmatrix} + \begin{bmatrix} \text{green block with } - \\ \text{blue block with } + \\ \text{blue block with } + \\ \text{blue block with } + \end{bmatrix} + \begin{bmatrix} \text{green block with } + \\ \text{blue block with } - \\ \text{blue block with } - \\ \text{blue block with } + \end{bmatrix} + \begin{bmatrix} \text{green block with } - \\ \text{blue block with } - \\ \text{blue block with } - \\ \text{blue block with } - \end{bmatrix} \right)$$

FIG. 5. Decomposition of a dichotomic observable. On the left, an observable O with $d_- < d/4$ is depicted, where the green and blue blocks represent the eigensubspaces corresponding to Π_- and the three partitions of Π_+ , denoted by $\Pi_+^{(1)}$, $\Pi_+^{(2)}$, and $\Pi_+^{(3)}$, respectively. The $\pm$ signs within each block denote the corresponding diagonal entries of ± 1 . On the right, the decomposition $O = \frac{1}{2}(\mathbb{I} + O_1 + O_2 + O_3)$ is shown: the first matrix represents the identity $\mathbb{I}$, while the remaining matrices correspond to O_1 , O_2 , and O_3 , each satisfying $d_\pm^{(i)} \geq d/3$.

observable swaps its eigenspaces but preserves their dimensions, the components O_l^i also satisfy the condition $\min\{d_+^{(i)}, d_-^{(i)}\} \geq d/3$. This completes the proof. ■

4. Proof of Theorem 2

Proof. Let $k = \lceil \log_2(\varepsilon^{-1}) \rceil + 1$, and use Lemma 3 to decompose O as

$$\text{Tr}(O\rho^2) = \sum_{l=1}^k 2^{-l} \text{Tr}(O_l \rho^2) + \text{Tr}(O_\Delta \rho^2). \quad (\text{B93})$$

We set $\delta' = \delta/k$ and $\varepsilon'_l = (\frac{3}{2})^l \frac{\varepsilon}{8}$. For each $1 \leq l \leq k$, we obtain an estimator ω_l of $\text{Tr}(O_l \rho^2)$ such that

$$|\omega_l - \text{Tr}(O_l \rho^2)| \leq \varepsilon'_l, \quad (\text{B94})$$

with probability at least $1 - \delta'$. By a union bound, with probability at least $1 - \delta$, the above holds for all $1 \leq l \leq k$. We then set

$$\omega = \sum_{l=1}^k 2^{-l} \omega_l \quad (\text{B95})$$

as the estimator for $\text{Tr}(O\rho^2)$. With probability at least $1 - \delta$,

$$\begin{aligned} |\omega - \text{Tr}(O\rho^2)| &= \left| \sum_{l=1}^k 2^{-l} (\omega_l - \text{Tr}(O_l \rho^2)) + \text{Tr}(O_\Delta \rho^2) \right| \\ &\leq \sum_{l=1}^k 2^{-l} |\omega_l - \text{Tr}(O_l \rho^2)| + |\text{Tr}(O_\Delta \rho^2)| \\ &\leq \sum_{l=1}^k 2^{-l} \varepsilon'_l + \|O_\Delta\|_\infty \\ &\leq \sum_{l=1}^k \left(\frac{3}{4}\right)^l \frac{\varepsilon}{8} + \frac{\varepsilon}{2} < \varepsilon. \end{aligned} \quad (\text{B96})$$

Finally, we account for the sample complexity. We obtain k estimators $\{\omega_l\}_{l=1}^k$. For each estimator ω_l , according to Theorem 1, the sample complexity is $\mathcal{O}\left(\max\left\{\sqrt{d}/\varepsilon'_l, 1/(\varepsilon'_l)^2\right\} \log(1/\delta')\right) = \mathcal{O}\left[\left(\frac{2}{3}\right)^l (\sqrt{d}/\varepsilon) + \left(\frac{2}{3}\right)^{2l} (1/\varepsilon^2) \log(1/\delta')\right]$. Summing over $l = 1, 2, \dots, k$

gives the total sample complexity

$$\begin{aligned} &\sum_{l=0}^k \left[\left(\frac{2}{3}\right)^l \left(\frac{\sqrt{d}}{\varepsilon}\right) + \left(\frac{2}{3}\right)^{2l} \left(\frac{1}{\varepsilon^2}\right) \right] \mathcal{O}(\log(1/\delta')) \\ &= \mathcal{O}\left(\left(\frac{\sqrt{d}}{\varepsilon} + \frac{1}{\varepsilon^2}\right) \log[\delta^{-1} \log(\varepsilon^{-1})]\right) \\ &= \mathcal{O}\left(\max\left\{\frac{\sqrt{d}}{\varepsilon}, \frac{1}{\varepsilon^2}\right\} \log[\delta^{-1} \log(\varepsilon^{-1})]\right). \end{aligned} \quad (\text{B97})$$

■

APPENDIX C: LOWER-BOUND ANALYSIS FOR SAMPLE COMPLEXITY

In this Appendix, we will give a brief proof of Fact 1, a direct corollary of Theorem 9 in the Supplemental Material of Ref. [11]. In that work, the authors proved the sample complexity lower bound for estimating $\text{Tr}(Z_1 \rho^2)$, where Z_1 represents the Pauli-Z observable on the first qubit, by constructing a state discrimination task. We consider specifically two state ensembles:

(1) $\mathcal{E}_1$ containing states of

$$\rho = \frac{1}{2} |0\rangle\langle 0| \otimes |\psi\rangle\langle \psi| + \frac{1}{2} |1\rangle\langle 1| \otimes \frac{\mathbb{I}}{2^{n-1}}. \quad (\text{C1})$$

(2) $\mathcal{E}_2$ containing states of

$$\rho = \frac{1}{2} |1\rangle\langle 1| \otimes |\psi\rangle\langle \psi| + \frac{1}{2} |0\rangle\langle 0| \otimes \frac{\mathbb{I}}{2^{n-1}}. \quad (\text{C2})$$

Here, $\mathbb{I}/2^{n-1}$ and $|\psi\rangle$ represent the $(n-1)$ -qubit maximally mixed state and Haar-random state vector. A Haar-random pure state can be sampled by $U|\psi_0\rangle$, where U is a Haar-random unitary and $|\psi_0\rangle$ is some fixed state vector. The task is that, given multiple copies of a state ρ sampled from either $\mathcal{E}_1$ or $\mathcal{E}_2$ with equal probabilities, how many copies of ρ should be provided to decide which ensemble it comes from. Note that, when n is large enough, for states from ensemble $\mathcal{E}_1$ we have $\text{Tr}(Z_1 \rho^2) \approx \frac{1}{4}$; while $\text{Tr}(Z_1 \rho^2) \approx -\frac{1}{4}$ for states from ensemble $\mathcal{E}_2$. Therefore, if one can accurately estimate the value of $\text{Tr}(Z_1 \rho^2)$, one can distinguish these two state ensembles with a high probability. Similarly, if the discrimination of these two state ensembles has a fundamental limitation, it also poses difficulties for the estimation of $\text{Tr}(Z_1 \rho^2)$. In Ref. [11], the authors proved that, for protocols utilizing arbitrary adaptive single-copy operations and measurements, $\Omega(\sqrt{d})$ copies of ρ should be provided to decide the state ensemble with success probability $\frac{1}{2} + \Omega(1)$, where

d represents the dimension of states in ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$.

It is direct to generalize their conclusion to other observables by defining two new state ensembles and the corresponding state discrimination task:

- (1) $\mathcal{E}_1^U$ containing states of

$$\rho = U \left(\frac{1}{2} |0\rangle\langle 0| \otimes |\psi\rangle\langle\psi| + \frac{1}{2} |1\rangle\langle 1| \otimes \frac{\mathbb{I}}{2^{n-1}} \right) U^\dagger, \quad (\text{C3})$$

- (2) $\mathcal{E}_2^U$ containing states of

$$\rho = U \left(\frac{1}{2} |1\rangle\langle 1| \otimes |\psi\rangle\langle\psi| + \frac{1}{2} |0\rangle\langle 0| \otimes \frac{\mathbb{I}}{2^{n-1}} \right) U^\dagger, \quad (\text{C4})$$

where U is an arbitrary d -dimensional unitary. Note that the new state ensemble discrimination task is no more difficult than the original discrimination task between $\mathcal{E}_1$ and $\mathcal{E}_2$. This is because the two original state ensembles can be transformed into $\mathcal{E}_1^U$ and $\mathcal{E}_2^U$ using the unitary transformation U , and the protocols we consider can implement arbitrary single-copy operations and measurements. So we can conclude that distinguishing between $\mathcal{E}_1^U$ and $\mathcal{E}_2^U$ also has a sample complexity lower bound of $\Omega(\sqrt{d})$. To prove the sample complexity lower bound for estimating $\text{Tr}(P\rho^2)$ with P being a Pauli observable, we can choose U to be the unitary that satisfies $U^\dagger P U = Z_1$. Then, it can be similarly verified that the accurate estimation of $\text{Tr}(P\rho^2)$ helps to distinguish $\mathcal{E}_1^U$ and $\mathcal{E}_2^U$ with a high probability. Therefore, we can conclude that the accurate estimation of $\text{Tr}(P\rho^2)$ with arbitrary Pauli observable P and target d -dimensional state ρ has a sample complexity lower bound of $\Omega(\sqrt{d})$.

APPENDIX D: ADDITIONAL ANALYSIS ON THE PAULI SAMPLING PROTOCOL

In this Appendix, we analyze the performance of the Pauli sampling scheme presented in Sec. IV C and prove Theorem 3. We begin by establishing the following lemma.

Lemma D1. The estimator ω defined in Eq. (29) satisfies

$$\Pr[|\omega - \text{Tr}(O\rho^2)| \geq \varepsilon] \leq \frac{1}{3}, \quad (\text{D1})$$

and the expected sample complexity is given by

$$\mathbb{E} \left[\sum_{i=1}^l m_i \right] = \mathcal{O} \left(\frac{K \|O\|_2^2}{\sqrt{d} \varepsilon^2} \right). \quad (\text{D2})$$

Proof. Set $l = 24K \|O\|_2^2 / d \varepsilon^2$. For a Pauli operator P_i , we set $N_U = \mathcal{O}(\|O\|_2^4 / d \varepsilon^2 \chi_O^2(P_i))$ and $N_M = \mathcal{O}(\sqrt{d})$, with sample complexity given by $m_i = N_U N_M = \mathcal{O}(\|O\|_2^4 / \sqrt{d} \varepsilon^2 \chi_O^2(P_i))$ for this Pauli operator. According to Lemma B3, we have the following:

$$\begin{aligned} \mathbb{E}[\omega_i | P_i] &= \text{Tr}(P_i \rho^2), \\ \text{Var}[\omega_i | P_i] &= \frac{d \varepsilon^2 \chi_O^2(P_i)}{24 \|O\|_2^4}. \end{aligned} \quad (\text{D3})$$

Now, consider the random variable

$$\tilde{X}_i = \frac{\|O\|_2^2}{\text{Tr}(O P_i)} \omega_i = \frac{\|O\|_2^2}{\chi_O(P_i) \sqrt{d}} \omega_i, \quad (\text{D4})$$

for which we have the following properties:

$$\begin{aligned} \mathbb{E}[\tilde{X}_i | P_i] &= \|O\|_2^2 \frac{\chi_{\rho^2}(P_i)}{\chi_O(P_i)} = X_i, \\ \text{Var}[\tilde{X}_i | P_i] &= \frac{l \varepsilon^2}{24}. \end{aligned} \quad (\text{D5})$$

Therefore, for the random variable $\omega = (1/l) \sum_{i=1}^l \tilde{X}_i$, the variance is given by

$$\text{Var}[\omega | \{P_i\}] = \frac{\varepsilon^2}{24}. \quad (\text{D6})$$

By Chebyshev's inequality, when fixing $\{P_i\}$, we have

$$\Pr\left[|\omega - \mathbb{E}[\omega | \{P_i\}]| \geq \frac{\varepsilon}{2}\right] \leq \frac{1}{6}, \quad (\text{D7})$$

where

$$\mathbb{E}[\omega | \{P_i\}] = \frac{1}{l} \sum_{i=1}^l \mathbb{E}[\tilde{X}_i | P_i] = \frac{1}{l} \sum_{i=1}^l X_i. \quad (\text{D8})$$

For each X_i , we have

$$\begin{aligned} \text{Var}_{P_i}[X_i] &= \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 \\ &\leq \mathbb{E}[X_i^2] \\ &= \sum_{i=1}^K p_i X_i^2 \\ &= \sum_{i=1}^K \frac{\chi_O^2(P_i)}{\|O\|_2^2} \left(\|O\|_2^2 \frac{\chi_{\rho^2}(P_i)}{\chi_O(P_i)} \right)^2 \\ &= \sum_{i=1}^K \|O\|_2^2 \frac{\text{Tr}(P_i \rho^2)^2}{d} \\ &\leq \frac{K \|O\|_2^2}{d}. \end{aligned} \quad (\text{D9})$$

Therefore,

$$\text{Var}_{\{P_i\}}(\mathbb{E}[\omega|\{P_i\}]) = \frac{1}{l} \text{Var}_{P_i}[X_i] \leq \frac{\varepsilon^2}{24}. \quad (\text{D10})$$

By Chebyshev's inequality, we then have

$$\Pr_{\{P_i\}}[|\mathbb{E}[\omega|\{P_i\}] - \mathbb{E}[\omega]| \geq \frac{\varepsilon}{2}] \leq \frac{1}{6}. \quad (\text{D11})$$

Finally, combining Eqs. (D7) and (D11) via a union bound, we obtain

$$\Pr[|\omega - \mathbb{E}[\omega]| \geq \varepsilon] \leq \frac{1}{3}. \quad (\text{D12})$$

Substituting $\mathbb{E}[\omega] = \mathbb{E}[X_i] = \sum_i p_i X_i = \text{Tr}(O\rho^2)$ proves the first inequality in Lemma D1.

Moreover, the expected sample complexity is given by

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^l m_i\right] &= l \sum_{i=1}^K p_i m_i \\ &= l \sum_{i=1}^K \frac{\chi_O^2(P_i)}{\|O\|_2^2} \mathcal{O}\left(\frac{\|O\|_2^4}{\sqrt{d}l\varepsilon^2 \chi_O^2(P_i)}\right) \\ &= l \sum_{i=1}^K \mathcal{O}\left(\frac{\|O\|_2^2}{\sqrt{d}l\varepsilon^2}\right) \\ &= \mathcal{O}\left(\frac{K\|O\|_2^2}{\sqrt{d}\varepsilon^2}\right). \end{aligned} \quad (\text{D13})$$

Proof of Theorem 3. The claimed performance guarantee is obtained by setting $T = \mathcal{O}(\log \delta^{-1})$ and applying the median-of-means estimator from Lemma A1. The expected sample complexity is given by

$$T \cdot \mathbb{E}\left[\sum_{i=1}^l m_i\right] = \mathcal{O}\left(\frac{K\|O\|_2^2 \log \delta^{-1}}{\sqrt{d}\varepsilon^2}\right). \quad (\text{D14})$$

APPENDIX E: ADDITIONAL ANALYSIS ON ESTIMATING LOW-RANK OBSERVABLES

In this Appendix, we analyze the performance of the BRM protocol proposed in Sec. VI for estimating nonlinear properties $\text{Tr}(O\rho^2)$. We begin by showing it is unbiased for estimating $\text{Tr}(O\rho^2)$ and then provide a performance analysis of the proposed protocol. Consider the estimator defined in Eq. (55),

$$\begin{aligned} \omega_u &= \binom{N_M}{2}^{-1} \sum_{1 \leq i < j \leq N_M} \sum_{\sigma \in \{0,1\}^n} X_3(s_i, s_j, \sigma) \\ &\quad \times \langle \sigma | U O U^\dagger | \sigma \rangle, \end{aligned} \quad (\text{E1})$$

where s_i and s_j are the i th and j th measurement outcomes from a set of N_M measurements in the basis defined by U_u . The coefficient function $X_3(s_1, s_2, s_3)$ should be chosen such that the corresponding coefficient matrix $X_3 = \sum_{s_1, s_2, s_3} X_3(s_1, s_2, s_3) |s_1, s_2, s_3\rangle \langle s_1, s_2, s_3|$ satisfies Eq. (51), so that

$$\mathbb{E}[U^{\otimes 3} X_3 U^{\dagger \otimes 3}] = \frac{1}{2} (\mathbb{P}_{(132)}^3 + \mathbb{P}_{(123)}^3), \quad (\text{E2})$$

i.e., its third-order twirling yields a superposition of the two third-order cyclic operators. Let $d = 2^n$ be the dimension of the Hilbert space, one possible choice is shown as Eq. (18) in the Supplemental Material of Ref. [94], which reads as

$$\begin{aligned} X_3(s_1, s_2, s_3) &= X_3(|\{s_1, s_2, s_3\}|), \quad \text{with} \\ X_3(1) &= \frac{1+d^2}{2}, \quad X_3(2) = \frac{1-d}{2}, \quad X_3(3) = 1. \end{aligned} \quad (\text{E3})$$

Then we have

$$\begin{aligned} \mathbb{E}[\omega_u] &= \mathbb{E}_U \left[\sum_{s_1, s_2, \sigma} \langle s_1 | U \rho U^\dagger | s_1 \rangle \langle s_2 | U \rho U^\dagger | s_2 \rangle \langle \sigma | U O U^\dagger | \sigma \rangle X_3(s_1, s_2, \sigma) \right] \\ &= \mathbb{E}_U \left[\text{Tr} \left(U^{\dagger \otimes 3} \left(\sum_{s_1, s_2, \sigma} X_3(s_1, s_2, \sigma) |s_1 s_2 \sigma\rangle \langle s_1 s_2 \sigma| \right) U^{\otimes 3} (\rho \otimes \rho \otimes O) \right) \right] \\ &= \frac{1}{2} (\text{Tr}(\rho O \rho) + \text{Tr}(\rho^2 O)) \\ &= \text{Tr}(O\rho^2). \end{aligned} \quad (\text{E4})$$

Lemma E1 ([Variance of ω_u in Eq. (55)]). Given an observable O , the variance of the estimator ω_u defined in Eq. (55) satisfies

$$\text{Var}[\omega_u] = \mathcal{O}\left(\left(1 + \frac{d}{N_M^2}\right) \text{Tr}(O^2)\right). \quad (\text{E5})$$

We postpone the proof of Lemma E1 and first analyze the performance of the full protocol. We further draw N_U independent random unitaries and construct the estimator,

$$\omega = \frac{1}{N_U} \sum_{u=1}^{N_U} \omega_u. \quad (\text{E6})$$

By standard variance reduction for averaging independent, identically distributed estimators and applying Lemma E1, we obtain the following corollary.

Corollary E1 (Variance of ω). Given an observable O , the variance of the estimator ω defined in Eq. (E6) satisfies

$$\text{Var}[\omega] = \mathcal{O}\left(\left(\frac{1}{N_U} + \frac{d}{N_U N_M^2}\right) \text{Tr}(O^2)\right). \quad (\text{E7})$$

This allows us to prove Theorem 4 in the main text.

Proof of Theorem 4. By Corollary E1, for any constant $0 < c < \frac{1}{2}$, we can choose $N_U = \mathcal{O}(\varepsilon^{-2})$ and $N_M = \mathcal{O}(\sqrt{d})$ such that

$$\text{Var}[\omega] \leq c\varepsilon^2. \quad (\text{E8})$$

Set the failure probability per observable to $\delta' = \delta/M$ and use the median-of-means estimator (Lemma A1) to estimate $\text{Tr}(O_m \rho^2)$ within additive error ε with success probability at least $1 - \delta'$. This requires $T = \mathcal{O}(\log(1/\delta'))$ repetitions. Applying the union bound over all M observables, we find that with probability at least $1 - \delta'M = 1 - \delta$, the estimation error satisfies

$$|\omega - \text{Tr}(O_m \rho^2)| \leq \varepsilon \quad (\text{E9})$$

for all $1 \leq m \leq M$. This completes the proof. ■

We now prove the validity of Lemma E1.

Proof of Lemma E1. We begin by applying the total variance law to arrive at

$$\begin{aligned} \text{Var}[\omega_u] &= \mathbb{E}_U[\text{Var}(\omega_u|U)] + \text{Var}_U[\mathbb{E}_s(\omega_u|U)] \\ &= \mathbb{E}_U[\mathbb{E}_s(\omega_u^2|U) - \mathbb{E}_s(\omega_u|U)^2] + \mathbb{E}_U[\mathbb{E}_s(\omega_u|U)^2] - [\mathbb{E}_U \mathbb{E}_s(\omega_u|U)]^2 \\ &= \mathbb{E}_U \mathbb{E}_s(\omega_u^2|U) - [\mathbb{E}_U \mathbb{E}_s(\omega_u|U)]^2 \\ &\leq \mathbb{E}_U \mathbb{E}_s(\omega_u^2|U), \end{aligned} \quad (\text{E10})$$

where $\mathbf{s} = (s_1, \dots, s_{N_M}) \in \mathbb{Z}_d^{N_M}$ denotes the sequence of measurement outcomes. In the analysis below, we use the notation $A \lesssim B$ to denote $A = \mathcal{O}(B)$. Define $\Delta(i, j; i', j') = |\{i, j\} \cup \{i', j'\}| \in \{2, 3, 4\}$ as the total number of distinct indices in index pairs (i, j) and (i', j') . We now expand ω_u and analyze each term to get

$$\begin{aligned} &\mathbb{E}_U \mathbb{E}_s(\omega_u^2 | U) \\ &\lesssim \frac{1}{N_M^4} \mathbb{E}_U \mathbb{E}_s \sum_{i < j; i' < j'} \sum_{a_3, a_4 \in \mathbb{Z}_d} X_3(s_i, s_j, a_3) X_3(s'_i, s'_j, a_4) \times \langle a_3 | U O U^\dagger | a_3 \rangle \langle a_4 | U O U^\dagger | a_4 \rangle \\ &= \frac{1}{N_M^4} \left(\sum_{\Delta(i, j; i', j')=4} + \sum_{\Delta(i, j; i', j')=3} + \sum_{\Delta(i, j; i', j')=2} \right) \mathbb{E}_U \mathbb{E}_s [\text{Tr}[(|s_i\rangle\langle s_i| \otimes |s_j\rangle\langle s_j| \otimes U O U^\dagger) X_3] \\ &\quad \cdot \text{Tr}[(|s'_i\rangle\langle s'_i| \otimes |s'_j\rangle\langle s'_j| \otimes U O U^\dagger) X_3]] \\ &\lesssim \frac{1}{N_M^4} \left\{ \binom{N_M}{4} \mathbb{E}_U [\text{Tr}[(\rho_U \otimes \rho_U \otimes O_U) X_3]^2] + \binom{N_M}{3} \mathbb{E}_{U s_1} [\text{Tr}[(|s_1\rangle\langle s_1| \otimes \rho_U \otimes O_U) X_3]^2] \right. \\ &\quad \left. + \binom{N_M}{2} \mathbb{E}_{U s_1, s_2} [\text{Tr}[(|s_1\rangle\langle s_1| \otimes |s_2\rangle\langle s_2| \otimes O_U) X_3]^2] \right\} \end{aligned}$$

$$\begin{aligned}
&\lesssim \text{Tr}[\Phi_6(X_3 \otimes X_3)(\rho_1 \otimes \rho_2 \otimes O_3 \otimes \rho_4 \otimes \rho_5 \otimes O_6)] + \frac{1}{N_M} \text{Tr}\left[\Phi_5(X_3^{(1,2,3),(1,4,5)})(\rho_1 \otimes \rho_2 \otimes O_3 \otimes \rho_4 \otimes O_5)\right] \\
&\quad + \frac{1}{N_M^2} \text{Tr}\left[\Phi_4(X_3^{(1,2,3),(1,2,4)})(\rho_1 \otimes \rho_2 \otimes O_3 \otimes O_4)\right] \\
&=: \Gamma_6 + \frac{1}{N_M} \Gamma_5 + \frac{1}{N_M^2} \Gamma_4.
\end{aligned} \tag{E11}$$

Here $\rho_U = U\rho U^\dagger$, $O_U = UOU^\dagger$, and the coefficient matrices are defined as

$$\begin{aligned}
X_3^{(1,2,3),(1,4,5)} &= \sum_{\mathbf{a} \in \mathbb{Z}_d^5} X_3(a_1, a_2, a_3) X_3(a_1, a_4, a_5) |\mathbf{a}\rangle \langle \mathbf{a}|, \\
X_3^{(1,2,3),(1,2,4)} &= \sum_{\mathbf{a} \in \mathbb{Z}_d^4} X_3(a_1, a_2, a_3) X_3(a_1, a_2, a_4) |\mathbf{a}\rangle \langle \mathbf{a}|.
\end{aligned} \tag{E12}$$

To analyze the terms Γ_i , we use the Weingarten calculus and retain only the leading-order contributions in the large- d limit ($d \gg 1$). The Weingarten coefficient $\text{Wg}(\alpha, d)$ in Eq. (A4) admits the asymptotic expansion

$$\begin{aligned}
\text{Wg}(\alpha, d) &= d^{k(\alpha)-2t} \prod_{i=1}^{k(\alpha)} (-1)^{\xi_i-1} C_{\xi_i-1} + \mathcal{O}(d^{k(\alpha)-2t-2}) \\
&= \mathcal{O}(d^{k(\alpha)-2t}),
\end{aligned} \tag{E13}$$

where

- (1) $k(\alpha)$ is the number of cycles in permutation α ,
- (2) $(\xi_1, \dots, \xi_{k(\alpha)})$ is the cycle type of α ,
- (3) $C_k = (2k)!/k!(k+1)!$ is the k th Catalan number,

as given in Eq. (141) of Ref. [94]. Since $k(\alpha) \leq t$ for all $\alpha \in S_t$, it follows that

$$\text{Wg}(\alpha, d) = \mathcal{O}(d^{-t}). \tag{E14}$$

For $\pi \in S_t$ and input $Q = \rho^{\otimes t} O^{\otimes 2}$, where $2 \leq t \leq 4$, we claim that

$$\text{Tr}[\mathbb{P}_\pi^t Q] \leq \text{Tr}(O^2). \tag{E15}$$

This holds because any contraction involving ρ and O (e.g., terms like $\text{Tr}(\rho^a O \rho^b O)$ for $a, b \in \mathbb{N}_+$) can be bounded by $\|O^2\|_\infty \leq \text{Tr}(O^2)$ using similar analysis as in the proof of Lemma A3. Combining these observations and using Eq. (A2), we have that for constant t and an arbitrary

matrix X ,

$$\begin{aligned}
\text{Tr}[\Phi_t(X)Q] &= \sum_{\pi, \sigma \in S_t} \text{Wg}(\pi^{-1}\sigma, d) \text{Tr}[\mathbb{P}_{\sigma^{-1}}^t X] \text{Tr}[\mathbb{P}_\pi^t Q] \\
&= \mathcal{O}(d^{-t} \text{Tr}(O^2)) \sum_{\pi \in S_t} \text{Tr}[\mathbb{P}_\pi^t X].
\end{aligned} \tag{E16}$$

From Eq. (150) in [94], for any permutation $\pi \in S_t$, the following scaling bounds hold:

$$\begin{aligned}
\text{Tr}[\mathbb{P}_\pi^t X_3^{\otimes 2}] &= \mathcal{O}(d^6), \\
\text{Tr}[\mathbb{P}_\pi^t X_3^{(1,2,3),(1,4,5)}] &= \mathcal{O}(d^5), \\
\text{Tr}[\mathbb{P}_\pi^t X_3^{(1,2,3),(1,2,4)}] &= \mathcal{O}(d^5).
\end{aligned} \tag{E17}$$

Substituting these into Eq. (E16), we obtain the following bounds for the Γ terms in Eq. (E11):

$$\begin{aligned}
\Gamma_6 &= \mathcal{O}(d^{-6} \text{Tr}(O^2)) \cdot \mathcal{O}(d^6) = \mathcal{O}(\text{Tr}(O^2)), \\
\Gamma_5 &= \mathcal{O}(d^{-5} \text{Tr}(O^2)) \cdot \mathcal{O}(d^5) = \mathcal{O}(\text{Tr}(O^2)), \\
\Gamma_4 &= \mathcal{O}(d^{-4} \text{Tr}(O^2)) \cdot \mathcal{O}(d^5) = \mathcal{O}(d \text{Tr}(O^2)).
\end{aligned} \tag{E18}$$

Thus, we conclude that

$$\begin{aligned}
\text{Var}[\omega_u] &\lesssim \Gamma_6 + \frac{1}{N_M} \Gamma_5 + \frac{1}{N_M^2} \Gamma_4 \\
&\lesssim \left(1 + \frac{1}{N_M} + \frac{d}{N_M^2}\right) \text{Tr}(O^2) \\
&\lesssim \left(1 + \frac{d}{N_M^2}\right) \text{Tr}(O^2),
\end{aligned} \tag{E19}$$

where the last line uses the fact that $1/N_M \leq 1$. This completes the proof of Lemma E1. $\blacksquare$

-
- [1] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nat. Rev. Phys.* **2**, 382 (2020).
 - [2] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states, *Phys. Rev. A* **99**, 052323 (2019).

- [3] M. Kliesch and I. Roth, Theory of quantum system certification, *PRX Quantum* **2**, 010201 (2021).
- [4] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Challenges and opportunities in quantum machine learning, *Nat. Comput. Sci.* **2**, 567 (2022).
- [5] J. Eisert and J. Preskill, Mind the gaps: The fraught road to quantum advantage, [arXiv:2510.19928](#).
- [6] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Quantum thermalization through entanglement in an isolated many-body system, *Science* **353**, 794 (2016).
- [7] J. Cotler, S. Choi, A. Lukin, H. Gharibyan, T. Grover, M. E. Tai, M. Rispoli, R. Schittko, P. M. Preiss, A. M. Kaufman, M. Greiner, H. Pichler, and P. Hayden, Quantum virtual cooling, *Phys. Rev. X* **9**, 031013 (2019).
- [8] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).
- [9] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, *Phys. Rev. X* **11**, 041036 (2021).
- [10] B. Koczor, Exponential error suppression for near-term quantum devices, *Phys. Rev. X* **11**, 031057 (2021).
- [11] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, and J. R. McClean, Quantum advantage in learning from experiments, *Science* **376**, 1182 (2022).
- [12] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, *Nat. Phys.* **10**, 631 (2014).
- [13] A. Rath, C. Branciard, A. Minguzzi, and B. Vermersch, Quantum fisher information from randomized measurements, *Phys. Rev. Lett.* **127**, 260501 (2021).
- [14] M. Yu, D. Li, J. Wang, Y. Chu, P. Yang, M. Gong, N. Goldman, and J. Cai, Experimental estimation of the quantum fisher information from randomized measurements, *Phys. Rev. Res.* **3**, 043122 (2021).
- [15] J. Y. Lee, C.-M. Jian, and C. Xu, Quantum criticality under decoherence or weak measurement, *PRX Quantum* **4**, 030317 (2023).
- [16] J. Y. Lee, Y.-Z. You, and C. Xu, Symmetry protected topological phases under decoherence, *Quantum* **9**, 1607 (2025).
- [17] Y. Zhang, T. H. Hsieh, Y. B. Kim, and Y. Zou, Probing mixed-state phases on a quantum computer via Renyi correlators and variational decoding, [arXiv:2505.02900](#).
- [18] D. Aharonov, J. Cotler, and X.-L. Qi, Quantum algorithmic measurement, *Nat. Commun.* **13**, 887 (2022).
- [19] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, in *Proceedings of the 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)* (IEEE Computer Society, Denver, CO, USA, 2022), pp. 574–585. <https://ieeexplore.ieee.org/abstract/document/9719827/>.
- [20] W. Gong, J. Haferkamp, Q. Ye, and Z. Zhang, Sample complexity of purity and inner product estimation, *Phys. Rev. Res.* (to be published).
- [21] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Jozsa, and C. Macchiavello, Stabilization of quantum computations by symmetrization, *SIAM J. Comput.* **26**, 1541 (1997).
- [22] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, Quantum fingerprinting, *Phys. Rev. Lett.* **87**, 167902 (2001).
- [23] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Direct estimations of linear and nonlinear functionals of a quantum state, *Phys. Rev. Lett.* **88**, 217901 (2002).
- [24] R. Islam, R. Ma, P. M. Preiss, M. Eric Tai, A. Lukin, M. Rispoli, and M. Greiner, Measuring entanglement entropy in a quantum many-body system, *Nature* **528**, 77 (2015).
- [25] Y. Zhou and Z. Liu, A hybrid framework for estimating nonlinear functions of quantum states, *npj Quantum Inf.* **10**, 62 (2024).
- [26] Q. Liu, Z. Li, X. Yuan, H. Zhu, and Y. Zhou, Auxiliary-free replica shadow estimation, [arXiv:2407.20865](#).
- [27] D. Miller *et al.*, Experimental measurement and a physical interpretation of quantum shadow enumerators, [arXiv:2408.16914](#).
- [28] H. Yao, Y. Liu, T. Lin, and X. Wang, Sample-efficient estimation of nonlinear quantum state functions, [arXiv:2412.01696](#).
- [29] A. Elben, R. Kueng, H.-Y. R. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermersch, Mixed-state entanglement from local randomized measurements, *Phys. Rev. Lett.* **125**, 200501 (2020).
- [30] V. Vitale, A. Rath, P. Jurcevic, A. Elben, C. Branciard, and B. Vermersch, Robust estimation of the quantum fisher information on a quantum processor, *PRX Quantum* **5**, 030338 (2024).
- [31] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [32] C. Berton, 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).
- [33] S. J. van Enk and C. W. J. Beenakker, Measuring $\text{Tr } \rho^n$ on single copies of ρ using random measurements, *Phys. Rev. Lett.* **108**, 110503 (2012).
- [34] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, *Science* **364**, 260 (2019).
- [35] 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).
- [36] P. Cieřliński, S. Imai, J. Dziewior, O. Gühne, L. Knips, W. Laskowski, J. Meinecke, T. Paterek, and T. Vértesi, Analysing quantum systems with randomised measurements, *Phys. Rep.* **1095**, 1 (2024).
- [37] M. Ohliger, V. Nesme, and J. Eisert, Efficient and feasible state tomography of quantum many-body systems, *New J. Phys.* **15**, 015024 (2013).
- [38] Q. Ye, Z. Liu, and D.-L. Deng, Exponential advantage from one more replica in estimating nonlinear properties of quantum states, [arXiv:2509.24000](#).
- [39] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [40] D. Hangleiter and J. Eisert, Computational advantage of quantum random sampling, *Rev. Mod. Phys.* **95**, 035001 (2023).

- [41] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, *Phys. Rev. A* **77**, 012307 (2008).
- [42] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, General framework for randomized benchmarking, *PRX Quantum* **3**, 020357 (2022).
- [43] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, *Phys. Rev. A* **94**, 052325 (2016).
- [44] Z. Cai and S. C. Benjamin, Constructing smaller Pauli twirling sets for arbitrary error channels, *Sci. Rep.* **9**, 1 (2019).
- [45] K. Tsubouchi, Y. Mitsuhashi, K. Sharma, and N. Yoshioka, Symmetric Clifford twirling for cost-optimal quantum error mitigation in early FTQC regime, *npj Quantum Inf.* **11**, 104 (2025).
- [46] Z. Ji, Y.-K. Liu, and F. Song, in *Advances in Cryptology—CRYPTO 2018*, edited by H. Shacham and A. Boldyreva (Springer International Publishing, Cham, 2018), pp. 126–152. https://link.springer.com/chapter/10.1007/978-3-319-96878-0_5.
- [47] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, *J. Math. Phys.* **48**, 052104 (2007).
- [48] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller, Rényi entropies from random quenches in atomic Hubbard and spin models, *Phys. Rev. Lett.* **120**, 050406 (2018).
- [49] A. Anshu, Z. Landau, and Y. Liu, in *Proceedings of the 54th Annual ACM Symposium on Theory of Computing, STOC 2022* (Association for Computing Machinery, New York, NY, USA, 2022), pp. 44–51. <https://doi.org/10.1145/3519935.3519974>.
- [50] H. Liu, Z. Liu, S. Chen, X. Nie, X. Liu, and D. Lu, Certifying quantum temporal correlation via randomized measurements: Theory and experiment, *Phys. Rev. Lett.* **134**, 040201 (2025).
- [51] A. Seif, Z.-P. Cui, S. Zhou, S. Chen, and L. Jiang, Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors, *PRX Quantum* **4**, 010303 (2023).
- [52] H.-Y. Hu, R. LaRose, Y.-Z. You, E. Rieffel, and Z. Wang, Logical shadow tomography: Efficient estimation of error-mitigated observables, [arXiv:2203.07263](https://arxiv.org/abs/2203.07263).
- [53] A. H. Myerson, D. J. Szwer, S. C. Webster, D. T. C. Allcock, M. J. Curtis, G. Imreh, J. A. Sherman, D. N. Stacey, A. M. Steane, and D. M. Lucas, High-fidelity readout of trapped-ion qubits, *Phys. Rev. Lett.* **100**, 200502 (2008).
- [54] F. Lenzini, J. Janousek, O. Thearle, M. Villa, B. Haylock, S. Kasture, L. Cui, H.-P. Phan, D. V. Dao, H. Yonezawa, P. K. Lam, E. H. Huntington, and M. Lobino, Integrated photonic platform for quantum information with continuous variables, *Sci. Adv.* **4**, eaat9331 (2018).
- [55] R. Dassonneville, T. Ramos, V. Milchakov, L. Planat, E. Dumur, F. Foroughi, J. Puertas, S. Leger, K. Bharadwaj, J. Delaforce, C. Naud, W. Hasch-Guichard, J. J. Garcia-Ripoll, N. Roch, and O. Buisson, Fast high-fidelity quantum nondemolition qubit readout via a nonperturbative cross-Kerr coupling, *Phys. Rev. X* **10**, 011045 (2020).
- [56] A. Opremcak, C. H. Liu, C. Wilen, K. Okubo, B. G. Christensen, D. Sank, T. C. White, A. Vainsencher, M. Giustina, A. Megrant, B. Burkett, B. L. T. Plourde, and R. McDermott, High-fidelity measurement of a superconducting qubit using an on-chip microwave photon counter, *Phys. Rev. X* **11**, 011027 (2021).
- [57] T. Schuster, J. Haferkamp, and H.-Y. Huang, Random unitaries in extremely low depth, *Science* **389**, 92 (2025).
- [58] E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, and J. Eisert, Mixing properties of stochastic quantum Hamiltonians, *Commun. Math. Phys.* **355**, 905 (2017).
- [59] Y. Nakata, C. Hirche, M. Koashi, and A. Winter, Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics, *Phys. Rev. X* **7**, 021006 (2017).
- [60] W. W. Ho and S. Choi, Exact emergent quantum state designs from quantum chaotic dynamics, *Phys. Rev. Lett.* **128**, 060601 (2022).
- [61] J. S. Cotler, D. K. Mark, H.-Y. Huang, F. Hernández, J. Choi, A. L. Shaw, M. Endres, and S. Choi, Emergent quantum state designs from individual many-body wave functions, *PRX Quantum* **4**, 010311 (2023).
- [62] M. C. Tran, D. K. Mark, W. W. Ho, and S. Choi, Measuring arbitrary physical properties in analog quantum simulation, *Phys. Rev. X* **13**, 011049 (2023).
- [63] Z. Liu, Z. Hao, and H.-Y. Hu, Predicting arbitrary state properties from single Hamiltonian quench dynamics, *Phys. Rev. Res.* **6**, 043118 (2024).
- [64] D. K. Mark, H.-Y. Hu, J. Kwan, C. Kokail, S. Choi, and S. F. Yelin, Efficiently measuring d -wave pairing and beyond in quantum gas microscopes, *Phys. Rev. Lett.* **135**, 123402 (2025).
- [65] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, R. Finkelstein, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale, H. Pichler, F. G. S. L. Brandão, S. Choi, and M. Endres, Preparing random states and benchmarking with many-body quantum chaos, *Nature* **613**, 468 (2023).
- [66] Y. Nakata, D. Zhao, T. Okuda, E. Bannai, Y. Suzuki, S. Tamiya, K. Heya, Z. Yan, K. Zuo, S. Tamate, Y. Tabuchi, and Y. Nakamura, Quantum circuits for exact unitary t -designs and applications to higher-order randomized benchmarking, *PRX Quantum* **2**, 030339 (2021).
- [67] H. Zhu, Multiqubit Clifford groups are unitary 3-designs, *Phys. Rev. A* **96**, 062336 (2017).
- [68] H. Zhu, R. Kueng, M. Grassl, and D. Gross, The Clifford group fails gracefully to be a unitary 4-design, [arXiv:1609.08172](https://arxiv.org/abs/1609.08172).
- [69] D. Maslov and M. Roetteler, Shorter stabilizer circuits via Bruhat decomposition and quantum circuit transformations, *IEEE Trans. Inf. Theory* **64**, 4729 (2018).
- [70] S. Bravyi and D. Maslov, Hadamard-free circuits expose the structure of the Clifford group, *IEEE Trans. Inf. Theory* **67**, 4546 (2021).
- [71] C. Zheng, K. Wang, X. Yu, P. Xu, and Z. Zhang, Distributed quantum inner product estimation with low-depth circuits, [arXiv:2506.19574](https://arxiv.org/abs/2506.19574).
- [72] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, *Phys. Rev. Lett.* **106**, 230501 (2011).

- [73] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, *Phys. Rev. Lett.* **127**, 030503 (2021).
- [74] D. Gottesman, Stabilizer codes and quantum error correction, [arXiv:quant-ph/9705052](https://arxiv.org/abs/quant-ph/9705052).
- [75] J. Chen, P. Zeng, Q. Zhao, X. Ma, and Y. Zhou, Quantum subspace verification for error correction codes, *PRX Quantum* **6**, 040353 (2025).
- [76] M. Hinsche, M. Ioannou, S. Jerbi, L. Leone, J. Eisert, and J. Carrasco, Efficient distributed inner-product estimation via Pauli sampling, *PRX Quantum* **6**, 030354 (2025).
- [77] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, *npj Quantum Inf.* **8**, 114 (2022).
- [78] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, *Nat. Phys.* **20**, 1648 (2024).
- [79] R. Takagi, H. Tajima, and M. Gu, Universal sampling lower bounds for quantum error mitigation, *Phys. Rev. Lett.* **131**, 210602 (2023).
- [80] K. Tsubouchi, T. Sagawa, and N. Yoshioka, Universal cost bound of quantum error mitigation based on quantum estimation theory, *Phys. Rev. Lett.* **131**, 210601 (2023).
- [81] P. Jurcevic, B. P. Lanyon, P. Hauke, C. Hempel, P. Zoller, R. Blatt, and C. F. Roos, Quasiparticle engineering and entanglement propagation in a quantum many-body system, *Nature* **511**, 202 (2014).
- [82] S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, United Kingdom, 2011), 2nd ed.
- [83] G.-Y. Zhu, N. Tantivasadakarn, A. Vishwanath, S. Trebst, and R. Verresen, Nishimori's cat: Stable long-range entanglement from finite-depth unitaries and weak measurements, *Phys. Rev. Lett.* **131**, 200201 (2023).
- [84] L. A. Lessa, R. Ma, J.-H. Zhang, Z. Bi, M. Cheng, and C. Wang, Strong-to-weak spontaneous symmetry breaking in mixed quantum states, *PRX Quantum* **6**, 010344 (2025).
- [85] C. de Groot, A. Turzillo, and N. Schuch, Symmetry protected topological order in open quantum systems, *Quantum* **6**, 856 (2022).
- [86] D. Petz, Quasi-entropies for finite quantum systems, *Rep. Math. Phys.* **23**, 57 (1986).
- [87] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, United Kingdom, 2013).
- [88] A. Streltsov, H. Kampermann, S. Wölk, M. Gessner, and D. Bruß, Maximal coherence and the resource theory of purity, *New J. Phys.* **20**, 053058 (2018).
- [89] S. Chen, J. Li, B. Huang, and A. Liu, in *Proceedings of the 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Computer Society, Denver, CO, USA, 2022), pp. 1205–1213. <https://ieeexplore.ieee.org/abstract/document/9996689/>.
- [90] J. Helsen and M. Walter, Thrifty shadow estimation: Reusing quantum circuits and bounding tails, *Phys. Rev. Lett.* **131**, 240602 (2023).
- [91] Y. Zhou and Q. Liu, Performance analysis of multi-shot shadow estimation, *Quantum* **7**, 1044 (2023).
- [92] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, Circuit quantum electrodynamics, *Rev. Mod. Phys.* **93**, 025005 (2021).
- [93] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski, Multiphoton entanglement and interferometry, *Rev. Mod. Phys.* **84**, 777 (2012).
- [94] Y. Zhou, P. Zeng, and Z. Liu, Single-copies estimation of entanglement negativity, *Phys. Rev. Lett.* **125**, 200502 (2020).
- [95] K. Van Kirk, C. Kokail, J. Kunjummen, H.-Y. Hu, Y. Teng, M. Cain, J. Taylor, S. F. Yelin, H. Pichler, and M. Lukin, Derandomized shallow shadows: Efficient Pauli learning with bounded-depth circuits, [arXiv:2412.18973](https://arxiv.org/abs/2412.18973).
- [96] E. Onorati, J. Kitzinger, J. Helsen, M. Ioannou, A. Werner, I. Roth, and J. Eisert, Noise-mitigated randomized measurements and self-calibrating shadow estimation, [arXiv:2403.04751](https://arxiv.org/abs/2403.04751).
- [97] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum state tomography via compressed sensing, *Phys. Rev. Lett.* **105**, 150401 (2010).
- [98] K. Chen, Q. Wang, Z. Yu, and Z. Zhang, Simultaneous estimation of nonlinear functionals of a quantum state, [arXiv:2505.16715](https://arxiv.org/abs/2505.16715).
- [99] Y. Zhang, Y. Wu, Y. Zhou, and X. Yuan, Measuring less to learn more: Quadratic speedup in learning nonlinear properties of quantum density matrices, [arXiv:2509.01571](https://arxiv.org/abs/2509.01571).
- [100] E. Tang, J. Wright, and M. Zhandry, Conjugate queries can help, [arXiv:2510.07622](https://arxiv.org/abs/2510.07622).
- [101] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, A hierarchy for replica quantum advantage, [arXiv:2111.05874](https://arxiv.org/abs/2111.05874).
- [102] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit, [arXiv:2405.08810](https://arxiv.org/abs/2405.08810).
- [103] A. A. Mele, Introduction to Haar measure tools in quantum information: A beginner's tutorial, *Quantum* **8**, 1340 (2024).
- [104] D. Weingarten, Asymptotic behavior of group integrals in the limit of infinite rank, *J. Math. Phys.* **19**, 999 (1978).
- [105] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability, *Int. Math. Res. Not.* **2003**, 953 (2003).
- [106] G. Köstenberger, Weingarten calculus, [arXiv:2101.00921](https://arxiv.org/abs/2101.00921).
- [107] Z. Liu, Y. Tang, H. Dai, P. Liu, S. Chen, and X. Ma, Detecting entanglement in quantum many-body systems via permutation moments, *Phys. Rev. Lett.* **129**, 260501 (2022).
- [108] L. Bittel, J. Eisert, L. Leone, A. A. Mele, and S. F. E. Oliviero, A complete theory of the Clifford commutant, [arXiv:2504.12263](https://arxiv.org/abs/2504.12263).
- [109] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch, Recovering quantum gates from few average gate fidelities, *Phys. Rev. Lett.* **121**, 170502 (2018).