

Performance analysis of multi-shot shadow estimation

You Zhou  and Qing Liu 

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

Shadow estimation is an efficient method for predicting many observables of a quantum state with a statistical guarantee. In the multi-shot scenario, one performs projective measurement on the sequentially prepared state for K times under the same unitary evolution, and repeats this procedure for M rounds of random sampled unitary, which results in MK times measurements in total. Here we analyze the performance of shadow estimation in this multi-shot scenario, which is characterized by the variance of estimating the expectation value of some observable O . We find that in addition to the shadow-norm $\|O\|_{\text{shadow}}$ introduced in [Huang et.al. Nat. Phys. 2020[1]], the variance is also related to another norm, and we denote it as the cross-shadow-norm $\|O\|_{\text{Xshadow}}$. For both random Pauli and Clifford measurements, we analyze and show the upper bounds of $\|O\|_{\text{Xshadow}}$. In particular, we figure out the exact variance formula of Pauli observables for random Pauli measurements. Our work gives theoretical guidance for the application of multi-shot shadow estimation.

1 Introduction

Learning the properties of quantum systems is of fundamental and practical interest, which can uncover quantum physics and enable quantum technologies. Traditional quantum state tomography [2–4] is not efficient with the increasing of qubit number in various quantum simulating and computing platforms [5, 6]. Recently, randomized measurements [7], especially the shadow estimation method [1, 8] are proposed for learning the quantum systems efficiently with a statistical guarantee.

In shadow estimation, a quantum state ρ is prepared sequentially, evolved by a randomly sampled unitary $\rho \rightarrow U\rho U^\dagger$, and finally measured in the computational basis. With the measurement results and the information of U , one can construct the shadow snapshot $\hat{\rho}$, which is an unbiased estimator of ρ . Using a few independent snapshots, many properties described by the observable O , for instance, the local Pauli operators and fidelities to some entangled states [1], can be predicted efficiently. Shadow estimation has been applied to many quantum information tasks, from quantum correlation detection [9–11], quantum chaos diagnosis [12, 13], to quantum error mitigation [14, 15], quantum machine learning [16, 17], and near-term quantum algorithms [18]. There are also a few works aiming to enhance the performance of shadow estimation from various perspectives, for instance, derandomization [19] and locally biased shadow for measuring Pauli observables [20, 21], hybrid shadow for polynomial functions [22], single-setting shadow [23],

You Zhou : you_zhou@fudan.edu.cn

Qing Liu : liuqingppk@gmail.com

generalized-measurement shadow [24], shadow estimation with a shallow circuit [25–27] or restricted unitary evolution [28–31], and practical issues considering measurement noises [32, 33].

In the current framework, shadow estimation in general needs to sample say M times random unitary to generate M shadow snapshots. However, the execution of different unitary is still resource-consuming in real experiments, which generally means changing control pulses or even physical settings. One direct idea is to add more measurement shots under the same unitary evolution [7, 14], say K -shot per unitary, to compensate for the realization of random unitaries. This strategy, denoted as multi-shot shadow estimation, however, still lacks a rigorous performance analysis compared to the original one.

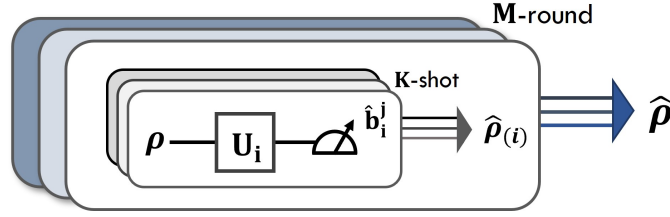


Figure 1: Multi-shot shadow estimation.

In this work, we fill this gap by giving the statistical guarantee for multi-shot shadow estimation. We estimate the variance of measuring some observable O and relate it to the introduced cross-shadow-norm $\|O\|_{\text{Xshadow}}$, which supplements the original shadow-norm $\|O\|_{\text{shadow}}$ [1]. The final variance is practically bounded by an interpolation of $\|O\|_{\text{Xshadow}}$ and $\|O\|_{\text{shadow}}$ controlled by the shot-number K . For both Pauli measurements and Clifford measurements, where the random unitary is sampled from single-qubit and global Clifford circuits respectively, we analyze $\|O\|_{\text{Xshadow}}$ and the variance. In the Pauli measurement, we find that multi-shot shadow estimation shows an advantage when estimating Pauli observable, conditioning on the expectation value of that observable. However, in the Clifford measurement, we find no clear benefit of increasing the shot-number K for a given number of random evolution M , especially for estimating the fidelity for many-qubit states. Our work could advance further applications of multi-shot shadow estimation.

2 Multi-shot shadow estimation

In this section, we introduce the background of shadow estimation and its multi-shot extension. Here, we focus on the n -qubit quantum system, that is, the state $\rho \in \mathcal{H}_D$ with $D = 2^n$, and denote the computational basis of $\mathcal{H}_D$ as $\{|\mathbf{b}\rangle\} = \{|b_1 b_2 \cdots b_n\rangle\}$ with $b_i = 0/1$. In shadow estimation [1], one prepares an unknown quantum state ρ in the experiment sequentially for M rounds. In the i -th round, one evolves the quantum system with a random unitary U sampled from some ensemble $\mathcal{E}$ to get $U\rho U^\dagger$, and measures it in the computational basis to get the result $|\mathbf{b}^{(i)}\rangle$. We call M the setting-number hereafter, since the sampled unitary U in each round determines the measurement setting. The shadow snapshot can be constructed as follows.

$$\hat{\rho}_{(i)} := \mathcal{M}^{-1} \left(U^\dagger |\hat{\mathbf{b}}^{(i)}\rangle \langle \hat{\mathbf{b}}^{(i)}| U \right), \quad (1)$$

with $\hat{\mathbf{b}}^{(i)}$ a random variable. $\hat{\rho}_{(i)}$ is an unbiased estimator of ρ , such that $\mathbb{E}_{\{U, \mathbf{b}^{(i)}\}} (\hat{\rho}_{(i)}) = \rho$. And the inverse (classical) post-processing $\mathcal{M}^{-1}$ is determined by the chosen random

unitary ensemble [1, 25, 34, 35]. For the n -qubit random Clifford circuit ensemble $\mathcal{E}_{\text{Cl}}$ and the tensor-product of random single-qubit Clifford gate ensemble $\mathcal{E}_{\text{Pauli}}$, one has $\mathcal{M}_C^{-1} = \mathcal{M}_n^{-1}$ and $\mathcal{M}_P^{-1} = \otimes_{i=1}^n \mathcal{M}_1^{-1}$ respectively, with $\mathcal{M}_n^{-1}(A) = (2^n + 1)A - \mathbb{I}_{2^n} \text{tr}(A)$ [1]. The random measurements from $\mathcal{E}_{\text{Cl}}$ and $\mathcal{E}_{\text{Pauli}}$ are denoted as Clifford and Pauli measurement primitives respectively.

Algorithm 1 Multi-shot shadow estimation

Input: $M \times K$ sequentially prepared ρ

Output: The shadow set $\{\hat{\rho}_{(i)}\}_{i=1}^M$.

```

1: for  $i = 1$  to  $M$  do
2:   Randomly choose  $U \in \mathcal{E}$  and record it.
3:   for  $j = 1$  to  $K$  do
4:     Evolve the state  $\rho$  using  $U$  to get  $U\rho U^\dagger$ .
5:     Measure the state in the computational basis  $\{|b\rangle\}$ .
6:     Construct the unbiased estimator  $\hat{\rho}_{(i)}^{(j)}$  with the result  $\mathbf{b}^{(i,j)}$  by Eq. (1), where  $i$ 
       and  $j$  denoting the  $j$ -th shot under the  $i$ -th unitary.
7:   end for
8:   Average  $K$  results under the same unitary to get  $\hat{\rho}_{(i)} = \frac{1}{K} \sum_j \hat{\rho}_{(i)}^{(j)}$ .
9: end for
10: Get the shadow set  $\{\hat{\rho}_{(1)}, \hat{\rho}_{(2)}, \dots, \hat{\rho}_{(M)}\}$ , which contains  $M$  independent estimators of
     $\rho$ .

```

In the multi-shot shadow estimation illustrated in Fig. 1, one conducts K shots by applying the same unitary U sampled in each round, that is, the measurement settings of these K shots are the same. So the total preparation-and-measurement number is MK . Denote the measurement result in i -th round and j -th shot as $\mathbf{b}^{(i,j)}$. One can construct the estimator of ρ in this shot as $\hat{\rho}_{(i)}^{(j)}$ following Eq. (1), and then average on K shots and M rounds to get

$$\hat{\rho}_{(i)} = K^{-1} \sum_{j \in [K]} \hat{\rho}_{(i)}^{(j)}, \quad \hat{\rho} = M^{-1} \sum_{i \in [M]} \hat{\rho}_{(i)}. \quad (2)$$

This procedure is listed in Algorithm 1, and we remark that as $K = 1$, it reduces to the original shadow estimation.

To estimate the expectation value of an observable O , one can construct the unbiased estimator as $\hat{O} = \text{tr}(O\hat{\rho})$. The performance of shadow estimation is mainly characterized by the variance of $\hat{O}$, which determines the experiment time to control the estimation error,

$$\text{Var}(\hat{O}) = \mathbb{E} \text{tr}(O\hat{\rho})^2 - \text{tr}(O\rho)^2. \quad (3)$$

Here, the expectation value is taken on the random unitary U in M rounds and the measurement result $\mathbf{b}^{(i,j)}$ in MK shots. In the following sections, we give a general expression for this key quantity, and then analyze the variance for Pauli and Clifford measurements respectively. We remark that even though we focus on estimating one observable O here, the variance result can be applied to simultaneously estimating a set of observables $\{O^i\}$ by using the median-of-mean technique [1], which is discussed in Appendix B.

3 The general variance expression and cross-shadow-norm

In this section, we give the general variance expression for the multi-shot shadow estimation and relate the variance to the cross-shadow-norm. First, we define two functions of a

quantum state σ and an observable O as follows,

$$\begin{aligned}\Gamma_1(\sigma, O) &:= \mathbb{E}_U \sum_{\mathbf{b}} \langle \mathbf{b} | U \sigma U^\dagger | \mathbf{b} \rangle \langle \mathbf{b} | U \mathcal{M}^{-1}(O) U^\dagger | \mathbf{b} \rangle^2, \\ \Gamma_2(\sigma, O) &:= \mathbb{E}_U \sum_{\mathbf{b}, \mathbf{b}'} \langle \mathbf{b} | U \sigma U^\dagger | \mathbf{b} \rangle \langle \mathbf{b} | U \mathcal{M}^{-1}(O) U^\dagger | \mathbf{b} \rangle \langle \mathbf{b}' | U \sigma U^\dagger | \mathbf{b}' \rangle \langle \mathbf{b}' | U \mathcal{M}^{-1}(O) U^\dagger | \mathbf{b}' \rangle,\end{aligned}\tag{4}$$

where $\mathbf{b}, \mathbf{b}'$ are computational basis for n -bit. The cross-shadow-norm is defined as follows.

Definition 1. *The cross-shadow-norm (short as XSnorm) of some observable O is defined as*

$$\|O\|_{\text{Xshadow}} = \max_{\sigma \text{ state}} \Gamma_2(\sigma, O)^{\frac{1}{2}}.\tag{5}$$

The proof of it being a norm is in Appendix A.1. Note that the original shadow norm (short as Snorm hereafter) is defined as $\|O\|_{\text{shadow}} = \max_{\sigma} \Gamma_1(\sigma, O)^{\frac{1}{2}}$ [1]. By definition, one has the upper bounds $\Gamma_1(\sigma, O) \leq \|O\|_{\text{shadow}}^2$, and $\Gamma_2(\sigma, O) \leq \|O\|_{\text{Xshadow}}^2$. Then we show the first result of this work considering the statistical variance for measuring some observable using multi-shot shadow estimation.

Theorem 1. *The statistical variance of $\text{tr}(O\hat{\rho})$ shows*

$$\text{Var}[\text{tr}(O\hat{\rho})] = \frac{1}{M} \left[\frac{1}{K} \Gamma_1(\rho, O_0) + \left(1 - \frac{1}{K}\right) \Gamma_2(\rho, O_0) - \text{tr}(O_0 \rho)^2 \right],\tag{6}$$

where $O_0 = O - \text{tr}(O)\mathbb{I}_D/D$ is the traceless part of O , and the functions Γ_1, Γ_2 are defined in Eq. (4), which can be bounded by the square of the Snorm and XSnorm respectively. As a result, one has the upper bound of the variance as

$$\text{Var}[\text{tr}(O\hat{\rho})] \leq \frac{1}{M} \left[\frac{1}{K} \|O_0\|_{\text{shadow}}^2 + \left(1 - \frac{1}{K}\right) \|O_0\|_{\text{Xshadow}}^2 \right].\tag{7}$$

The proof of Theorem 1 is left in Appendix A.2. We remark that as $K = 1$, the variance in Eq. (6) and the upper bound in Eq. (7) reduce to the result of the original shadow estimation [1]. Note that the variance is an interpolation of the two functions $\Gamma_1(\rho, O_0)$ and $\Gamma_2(\rho, O_0)$ with the shot-number K . And the advantage of introducing multi-shot should come from the condition when $\Gamma_2(\rho, O_0) \ll \Gamma_1(\rho, O_0)$. We remark that Ref. [14] also investigated on this kind of multi-shot variance with a focus on quantum error mitigation.

The function $\Gamma_1(\sigma, O)$ and the Snorm $\|O_0\|_{\text{shadow}}$ have been extensively studied in Ref. [1]. Thus, the remaining content of this work is to analyze $\Gamma_2(\sigma, O)$ and the XSnorm $\|O_0\|_{\text{Xshadow}}$. To proceed, one can write $\Gamma_2(\sigma, O)$ in Eq. (4) in the following twirling channel form.

$$\Gamma_2(\sigma, O) = \text{tr} \left[\sigma \otimes \mathcal{M}^{-1}(O) \otimes \sigma \otimes \mathcal{M}^{-1}(O) \Phi^{(4, \mathcal{E})}(\Lambda_n) \right].\tag{8}$$

Here, we denote the 4-fold twirling channel as $\Phi^{(4, \mathcal{E})}(\cdot) = \mathbb{E}_{\{U \in \mathcal{E}\}} U^{\dagger \otimes 4}(\cdot) U^{\otimes 4}$, which is dependent on the random unitary ensemble $\mathcal{E}$, and the 4-copy n -qubit diagonal operator is defined as

$$\Lambda_n := \sum_{\mathbf{b}, \mathbf{b}'} (|\mathbf{b}\rangle \langle \mathbf{b}|)^{\otimes 2} \otimes (|\mathbf{b}'\rangle \langle \mathbf{b}'|)^{\otimes 2}.\tag{9}$$

For the global n -qubit Clifford ensemble, we denote the twirling channel as $\Phi_n^{(4, \text{Cl})}$, and the one of local Clifford is just the tensor product $\bigotimes_{i=1}^n \Phi_{1, (i)}^{(4, \text{Cl})}$.

It is known that Clifford ensemble is not a 4-design [36]. Hence, in order to calculate $\Gamma_2(\sigma, O)$ and the XSnorm, one should apply the representation theory (rep-th) of Clifford

group in the 4-copy Hilbert space [36, 37]. In the following sections, we first figure out the exact result of random Pauli measurements, say single-qubit Clifford twirling based on direct calculation, *without* using the rep-th; and then give the bound of $\Gamma_2(\sigma, O)$ for random Clifford measurements, say n -qubit Clifford twirling with results from rep-th.

4 Variance analysis for Pauli measurements

In this section, we focus on the observable O being an n -qubit Pauli operator P , that is $P = \otimes_{i=1}^n P_i$, with $P_i \in \{\mathbb{I}_2, X, Y, Z\}$. This choice is of practical interest, for example, in the measurement problem in quantum chemistry simulation [19, 20, 38]. We say a Pauli operator P is w -weight, if there are w qubits with $P_i \neq \mathbb{I}_2$, i.e., there are single-qubit Pauli operators on w qubits and identity operators on the other $n - w$ qubits. We have the following result for the Pauli observable.

Proposition 1. *Suppose $O = P$ is a Pauli observable with weight w , the function Γ_2 defined in Eq. (4) shows*

$$\Gamma_2^{\text{Pauli}}(\sigma, P) = 3^w \text{tr}(P\sigma)^2 \quad (10)$$

for the random Pauli measurements. In this case, the XSnorm is $\|P\|_{\text{Xshadow}}^{\text{Pauli}} = 3^{w/2}$ by Eq. (5).

We remark that compared to $\Gamma_1^{\text{Pauli}}(\sigma, P) = 3^w$ in the original shadow [1], $\Gamma_2^{\text{Pauli}}(\sigma, P)$ has an extra relevant term, that is the square of the expectation value $\text{tr}(P\sigma)^2$. To prove Proposition 1, we mainly apply the following Lemma 1 of the single-qubit Clifford twirling.

Lemma 1. *The 4-fold single-qubit Clifford twirling channel maps the operator*

$$\Lambda_1 := \sum_{b=0,1} \sum_{b'=0,1} |b\rangle \langle b|^{\otimes 2} \otimes |b'\rangle \langle b'|^{\otimes 2} \quad (11)$$

into

$$\Phi_1^{(4, \text{Cl})}(\Lambda_1) = \frac{1}{3} \left(\frac{1}{2} \mathbb{I}_2^{\otimes 2} \otimes \mathbb{F}^{(2)} + \frac{1}{2} \mathbb{F}^{(2)} \otimes \mathbb{I}_2^{\otimes 2} + \mathbb{F}^{(4)} \right). \quad (12)$$

Here $\mathbb{F}^{(t)} := 2^{-t/2}(\mathbb{I}_2^{\otimes t} + X^{\otimes t} + Y^{\otimes t} + Z^{\otimes t})$ on t qubits, and $\mathbb{F}^{(2)}$ is the swap operator on two qubits.

We prove Lemma 1 by giving the general t -copy twirling result of $\Phi_1^{(t, \text{Cl})}$ on any Pauli operator in Appendix C.1. Note that the diagonal operator in Eq. (9) can be decomposed to $\Lambda_n = \otimes_{i=1}^n \Lambda_{1,(i)}$, and the twirling result in Eq. (8) is thus in the tensor-product form $\otimes_{i=1}^n \Phi_{1,(i)}^{(4, \text{Cl})}(\Lambda_{1,(i)})$. Then one can prove Proposition 1 by applying the result of Lemma 1 to Eq. (8) and further calculation, which is left in Appendix C.2.

Inserting $\Gamma_1^{\text{Pauli}}(\sigma, P) = 3^w$ [1] and the result of $\Gamma_2^{\text{Pauli}}(\sigma, P)$ of Proposition 1 into Eq. (6), one has the following exact variance.

Theorem 2. *Suppose $O = P$ is a Pauli observable with weight w , the variance to measure P using random Pauli measurements is*

$$\text{Var}[\text{tr}(P\hat{\rho})] = \frac{1}{M} \left[\frac{1}{K} 3^w + \left(1 - \frac{1}{K}\right) 3^w \text{tr}(P\rho)^2 - \text{tr}(P\rho)^2 \right]. \quad (13)$$

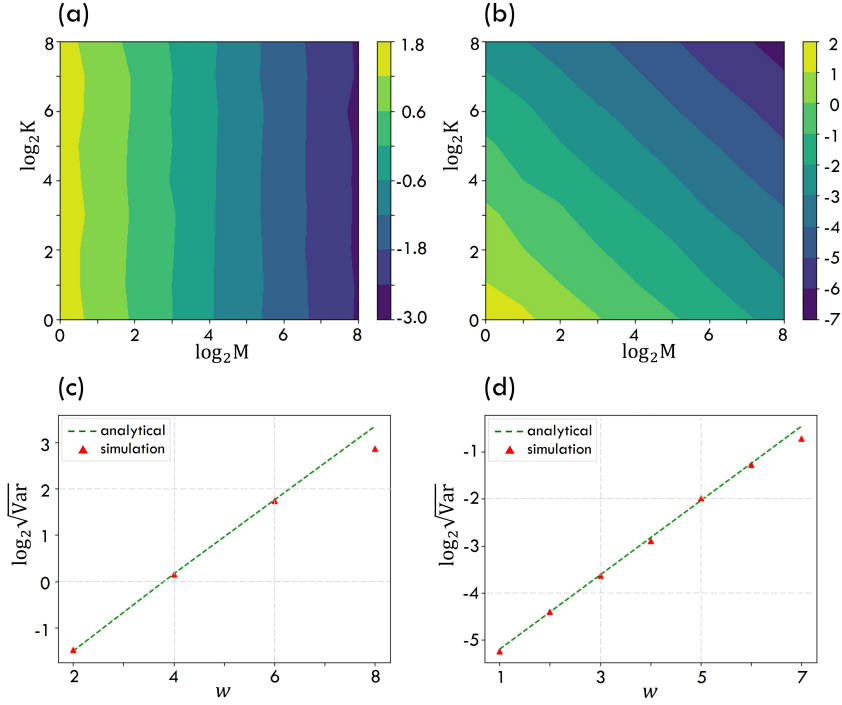


Figure 2: Scaling of the statistical variance of estimating $\text{tr}(P\hat{\rho})$ with shot-number K , setting-number M and weight w for different Pauli observables. In (a) and (b), the processed state ρ is a 5-qubit GHZ state. We investigate the dependence of $\log_2 \sqrt{\text{Var}[\text{tr}(P\hat{\rho})]}$ (colored area) on K and M with different Pauli observables $P = \sigma_z^1 \otimes \sigma_z^2 \otimes \mathbb{I}_2^{\otimes 3}$ and $P = \sigma_x^1 \otimes \sigma_x^2 \otimes \mathbb{I}_2^{\otimes 3}$ in (a) and (b) respectively. The variance in (a) is almost independent with K , and for the one in (b) the dependence on K and M is the same. Note that in (a) and (b), $\text{tr}(P\rho)^2 = 1$ and 0 respectively, which are consistent with Eq. (13) and (14). In (c) and (d), we explore the variance dependence on weight w . Here, the processed state ρ is an 8-qubit GHZ state, and we set $M = K = 64$. In (c), the Pauli observable is $P = \sigma_z^{\otimes w} \otimes \mathbb{I}_2^{\otimes(8-w)}$ with w being an even number, and in (d), $P = \sigma_x^{\otimes w} \otimes \mathbb{I}_2^{\otimes(8-w)}$, such that $\text{tr}(P\rho)^2 = 1$ and 0 respectively. The red triangles represent the simulated variance which shows consistency with the green dotted line, i.e., the analytical variance given by Eq. (13).

Theorem 2 shows that the dependence of $\text{Var}[\text{tr}(P\hat{\rho})]$ on the shot-number K is related to $\text{tr}(P\rho)^2$. For the extreme cases when $\text{tr}(P\rho)^2 = 0/1$,

$$\text{Var}[\text{tr}(P\hat{\rho})] = \begin{cases} \frac{1}{M} \frac{1}{K} 3^w & \text{tr}(P\rho)^2 = 0, \\ \frac{1}{M} (3^w - 1) & \text{tr}(P\rho)^2 = 1. \end{cases} \quad (14)$$

One can see that when $\text{tr}(P\rho)^2 = 1$, the variance is independent of the shot-number K ; for $\text{tr}(P\rho)^2 = 0$, the dependence of K is the same as the setting-number M . So it is advantageous to increase K for $\text{tr}(P\rho)^2$ being small, by considering that it is more convenient to repeat shots than change measurement settings in a real experiment. The variance result here can be extended to any w -local observable following the proof routine in Ref. [1], and this is investigated in Ref. [14]. We give more discussions on this point in Appendix B.

In Fig. 2, we show numerical results of the variance dependence on M , K and w , where we perform random Pauli measurements on the Greenberger–Horne–Zeilinger (GHZ) states. Please see Appendix F for details of the numerical simulation, and also extended discussions

about the reduction of experiment time-cost. The numerical results are consistent with the variance given by Eq. (13) and (14).

5 Variance analysis for Clifford measurements

In this section, we analyze the statistical variance of random Clifford measurements. In this case, the inverse channel maps $\mathcal{M}^{-1}(O_0) = (D+1)O_0$, and $\Gamma_2(\sigma, O_0)$ in Eq. (8) becomes

$$\Gamma_2^{\text{Cl}}(\sigma, O_0) = (D+1)^2 \text{tr} \left[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi_n^{(4, \text{Cl})}(\Lambda_n) \right]. \quad (15)$$

For the simplicity of the calculation, we decompose $\Lambda_n = \Lambda_n^0 + \Lambda_n^1$, with

$$\begin{aligned} \Lambda_n^0 &= \sum_{\mathbf{b}} |\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4}, \\ \Lambda_n^1 &= \sum_{\mathbf{b} \neq \mathbf{b}'} |\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 2} \otimes |\mathbf{b}'\rangle \langle \mathbf{b}'|^{\otimes 2}. \end{aligned} \quad (16)$$

To calculate $\Gamma_2^{\text{Cl}}(\sigma, O_0)$, one should apply the 4-fold twirling result of the Clifford group [36] on Λ_n^0 and Λ_n^1 , respectively. And we show in Appendix E that Λ_n^1 contributes to the leading term of the final result. Since the n -qubit Clifford twirling is quite sophisticated, we do not calculate it in a very exact form but give the following upper bound on $\Gamma_2^{\text{Cl}}(\sigma, O_0)$.

Proposition 2. *Suppose O_0 is a traceless observable, the function Γ_2 defined in Eq. (4) is upper bounded by*

$$\Gamma_2^{\text{Cl}}(\sigma, O_0) \leq c \|O_0\|_2^2, \quad (17)$$

for the random Clifford measurements, where $\|A\|_2 = \sqrt{\text{tr}(AA^\dagger)}$ is the Frobenius norm, and c is some constant independent of the dimension D . In this case, the XSnorm $\|O_0\|_{\text{Xshadow}}^{\text{Cl}} \leq \sqrt{c} \|O_0\|_2$ by Eq. (5).

For comparison, in the original shadow [1],

$$\Gamma_1^{\text{Cl}}(\sigma, O_0) = \frac{D+1}{D+2} \left[\|O_0\|_2^2 + 2 \text{tr}(\sigma O_0^2) \right] \leq 3 \|O_0\|_2^2, \quad (18)$$

and $\|O_0\|_{\text{shadow}} \leq \sqrt{3} \|O_0\|_2$, which is qualitatively same to the current multi-shot result. The proof of Proposition 2 is left in Appendix D and E. In Appendix D, we first analyze $\Gamma_2^{\text{Haar}}(\sigma, O_0)$ in the Haar random case, where the global unitary is sampled from Haar ensemble or any unitary 4-design ensemble. We then extend the result to the Clifford ensemble in Appendix E. For both Haar and Clifford ensembles, we can bound $\Gamma_2(\sigma, O_0)$ in the order $\mathcal{O}(1) \|O_0\|_2^2$ as in Proposition 2.

Inserting Eq. (17) in Eq. (6), one has the following variance result.

Theorem 3. *For a general observable O , the variance to measure O using the random Clifford measurements is upper bounded by*

$$\begin{aligned} \text{Var} [\text{tr}(O\hat{\rho})] &\leq \frac{1}{M} \left[\frac{3}{K} + \left(1 - \frac{1}{K}\right)c \right] \|O_0\|_2^2 \\ &\leq \frac{c'}{M} \|O_0\|_2^2, \end{aligned} \quad (19)$$

with $c' = c + 3$ some constant independent of the dimension D , and $\|A\|_2 = \sqrt{\text{tr}(AA^\dagger)}$ is the Frobenius norm.

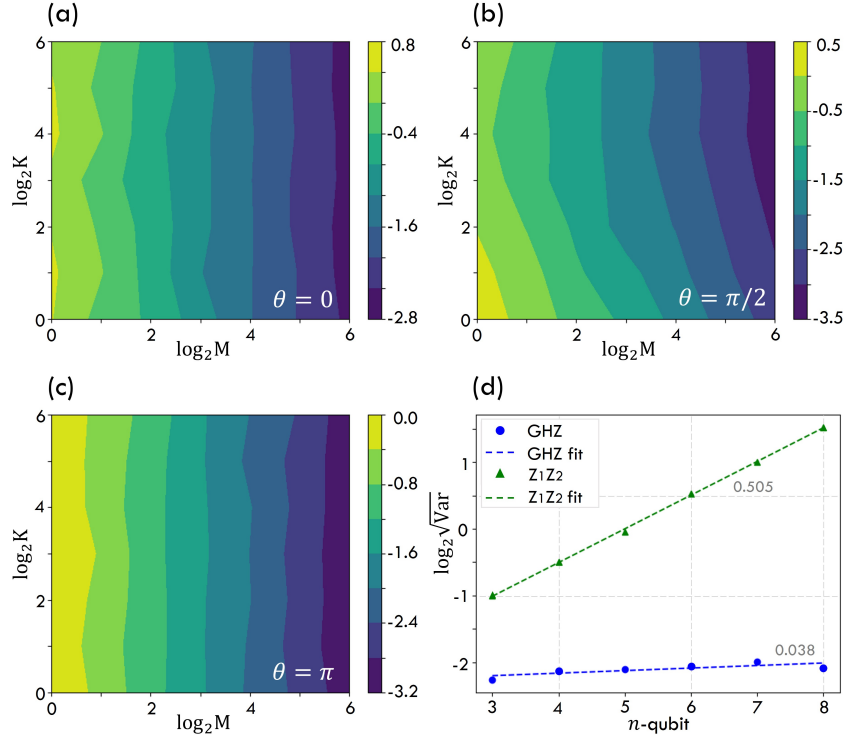


Figure 3: Scaling of the statistical variance of random Clifford measurements with K , M and qubit number n for different processed state ρ and observables. In (a), (b), and (c) the processed state ρ is a 5-qubit GHZ_θ state, with θ equals to 0, $\pi/2$ and π respectively. and the observable is $O = |GHZ_{\theta=0}\rangle\langle GHZ_{\theta=0}|$, with the expectation value $\text{tr}(O\rho) = (1 + \cos\theta)/2$ being 1, 0.5 and 0 respectively. In all three cases, the variance is almost independent of the shot-number K , in contrast to Fig. 2 (b) of Pauli measurements. In (d), we explore the variance dependence on qubit number n for different observables when $M = K = 32$. The blue dots represent the case when the observable $O = |GHZ_{\theta=0}\rangle\langle GHZ_{\theta=0}|$, and the green triangles are the case when $O = \sigma_z^{\otimes 2} \otimes \mathbb{I}_2^{\otimes (n-2)}$. Note that the corresponding Frobenius norm $\|O_0\|_2$ is about a constant and $2^{n/2}$, respectively. The dotted lines and the slopes represent the fitting curves for the corresponding cases, which is consistent with Eq. (19).

In the original shadow, the variance bound is about $\mathcal{O}(1)M^{-1}\|O_0\|_2^2$. Even though Eq. (19) is an upper bound, it already gives a hint that it is not very helpful to increase the shot-number K in the random Clifford measurements. We demonstrate this phenomenon with numerical simulation considering measuring the fidelity. In Fig. 3, we take the state $\rho = |GHZ_\theta\rangle\langle GHZ_\theta|$, with $|GHZ_\theta\rangle = 1/\sqrt{2}(|0\rangle^{\otimes n} + e^{i\theta}|1\rangle^{\otimes n})$, i.e., the GHZ state with some phase. And the observable is taken to be $O = |GHZ_{\theta=0}\rangle\langle GHZ_{\theta=0}|$, that is, measuring the fidelity of the quantum state to the GHZ state. We find that the variance is almost independent with the shot-number K for various choices of θ which gives different expectation values of $\text{tr}(O\rho)$. This is very different from the results of random Pauli measurements as shown in Fig. 2.

6 Conclusion and outlook

In this work, we systemically analyze the statistical performance of multi-shot shadow estimation by introducing a key quantity —the cross-shadow-norm. We find the advantage of this framework in Pauli measurements, however, the advantage in Clifford measurements is

not significant. There are a few interesting points that merit further investigation. First, the application of the result of Pauli measurements to quantum chemistry problems is promising, especially combined with other measurement techniques, such as derandomization and local-bias [19, 20]. Second, considering that the advantage of multi-shot for Clifford measurements is minor, it is interesting to investigate whether this hold or not for shallow circuit [25–27] or restricted unitary evolution [28–31]. Third, it is also intriguing to extend the current analysis to boson [39–41] and fermion systems [42, 43], and to non-linear functions of quantum states [9, 11, 22]. Finally, we expect the result here would also benefit the statistical analysis for other randomized measurement tasks, especially those related to high-order functions [44–47].

While completing this work, we became aware of a related work [48] by Jonas Helsen and Michael Walter which also considers shadow estimation under multi-shot. Different from Ref. [48], we analyze the performance under Pauli measurements. Our negative conclusion on Clifford measurements is consistent with theirs. Ref. [48] also demonstrates the weakness of Clifford measurements by giving the exact variance for estimating stabilizer-fidelity. Additionally, they nicely extend the Clifford unitary to doped Clifford circuits to bypass this weakness.

7 Acknowledgements

Y.Z. thanks Huangjun Zhu for the useful discussion on unitary design. This work is supported by NSFC Grant No. 12205048 and startup funding from Fudan University.

References

- [1] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020. DOI: [10.1038/s41567-020-0932-7](https://doi.org/10.1038/s41567-020-0932-7).
- [2] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017. DOI: [10.1109/TIT.2017.2719044](https://doi.org/10.1109/TIT.2017.2719044).
- [3] Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012. DOI: [10.1088/1367-2630/14/9/095022](https://doi.org/10.1088/1367-2630/14/9/095022).
- [4] Martin Kliesch and Ingo Roth. Theory of quantum system certification. *PRX Quantum*, 2(1):010201, 2021. DOI: [10.1103/PRXQuantum.2.010201](https://doi.org/10.1103/PRXQuantum.2.010201).
- [5] Ehud Altman, Kenneth R Brown, Giuseppe Carleo, Lincoln D Carr, Eugene Demler, Cheng Chin, Brian DeMarco, Sophia E Economou, Mark A Eriksson, Kai-Mei C Fu, et al. Quantum simulators: Architectures and opportunities. *PRX Quantum*, 2(1):017003, 2021. DOI: [10.1103/PRXQuantum.2.017003](https://doi.org/10.1103/PRXQuantum.2.017003).
- [6] Yuri Alexeev, Dave Bacon, Kenneth R Brown, Robert Calderbank, Lincoln D Carr, Frederic T Chong, Brian DeMarco, Dirk Englund, Edward Farhi, Bill Fefferman, et al. Quantum computer systems for scientific discovery. *PRX Quantum*, 2(1):017001, 2021. DOI: [10.1103/PRXQuantum.2.017001](https://doi.org/10.1103/PRXQuantum.2.017001).
- [7] Andreas Elben, Steven T Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5(1):9–24, 2023. DOI: [10.1038/s42254-022-00535-2](https://doi.org/10.1038/s42254-022-00535-2).
- [8] Scott Aaronson. Shadow tomography of quantum states. *SIAM Journal on Computing*, 49(5):STOC18–368, 2019. DOI: [10.1137/18M120275X](https://doi.org/10.1137/18M120275X).

- [9] Andreas Elben, Richard Kueng, Hsin-Yuan Robert Huang, Rick van Bijnen, Christian Kokail, Marcello Dalmonte, Pasquale Calabrese, Barbara Kraus, John Preskill, Peter Zoller, et al. Mixed-state entanglement from local randomized measurements. *Physical Review Letters*, 125(20):200501, 2020. DOI: [10.1103/PhysRevLett.125.200501](https://doi.org/10.1103/PhysRevLett.125.200501).
- [10] Aniket Rath, Cyril Branciard, Anna Minguzzi, and Benoît Vermersch. Quantum fisher information from randomized measurements. *Phys. Rev. Lett.*, 127:260501, Dec 2021. DOI: [10.1103/PhysRevLett.127.260501](https://doi.org/10.1103/PhysRevLett.127.260501).
- [11] Zhenhuan Liu, Yifan Tang, Hao Dai, Pengyu Liu, Shu Chen, and Xiongfeng Ma. Detecting entanglement in quantum many-body systems via permutation moments. *Physical Review Letters*, 129(26):260501, 2022. DOI: [10.1103/PhysRevLett.129.260501](https://doi.org/10.1103/PhysRevLett.129.260501).
- [12] Roy J. Garcia, You Zhou, and Arthur Jaffe. Quantum scrambling with classical shadows. *Phys. Rev. Research*, 3:033155, Aug 2021. DOI: [10.1103/PhysRevResearch.3.033155](https://doi.org/10.1103/PhysRevResearch.3.033155).
- [13] Max McGinley, Sebastian Leontica, Samuel J Garratt, Jovan Jovanovic, and Steven H Simon. Quantifying information scrambling via classical shadow tomography on programmable quantum simulators. *Physical Review A*, 106(1):012441, 2022. DOI: [10.1103/PhysRevA.106.012441](https://doi.org/10.1103/PhysRevA.106.012441).
- [14] Alireza Seif, Ze-Pei Cui, Sisi Zhou, Senrui Chen, and Liang Jiang. Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors. *PRX Quantum*, 4(1):010303, 2023. DOI: [10.1103/PRXQuantum.4.010303](https://doi.org/10.1103/PRXQuantum.4.010303).
- [15] Hong-Ye Hu, Ryan LaRose, Yi-Zhuang You, Eleanor Rieffel, and Zhihui Wang. Logical shadow tomography: Efficient estimation of error-mitigated observables. *arXiv preprint arXiv:2203.07263*, 2022. DOI: [10.48550/arXiv.2203.07263](https://doi.org/10.48550/arXiv.2203.07263).
- [16] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, et al. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, 2022. DOI: [10.1126/science.abn7293](https://doi.org/10.1126/science.abn7293).
- [17] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V Albert, and John Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613):eabk3333, 2022. DOI: [10.1126/science.abk3333](https://doi.org/10.1126/science.abk3333).
- [18] Stefan H. Sack, Raimel A. Medina, Alexios A. Michailidis, Richard Kueng, and Maksym Serbyn. Avoiding barren plateaus using classical shadows. *PRX Quantum*, 3:020365, Jun 2022. DOI: [10.1103/PRXQuantum.3.020365](https://doi.org/10.1103/PRXQuantum.3.020365).
- [19] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Efficient estimation of pauli observables by derandomization. *Physical review letters*, 127(3):030503, 2021. DOI: [10.1103/PhysRevLett.127.030503](https://doi.org/10.1103/PhysRevLett.127.030503).
- [20] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of quantum hamiltonians with locally-biased classical shadows. *Communications in Mathematical Physics*, 391(3):951–967, 2022. DOI: [10.1007/s00220-022-04343-8](https://doi.org/10.1007/s00220-022-04343-8).
- [21] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. Overlapped grouping measurement: A unified framework for measuring quantum states. *Quantum*, 7:896, 2023. DOI: [10.22331/q-2023-01-13-896](https://doi.org/10.22331/q-2023-01-13-896).
- [22] You Zhou and Zhenhuan Liu. A hybrid framework for estimating nonlinear functions of quantum states. *arXiv preprint arXiv:2208.08416*, 2022. DOI: [10.48550/arXiv.2208.08416](https://doi.org/10.48550/arXiv.2208.08416).
- [23] Roman Stricker, Michael Meth, Lukas Postler, Claire Edmunds, Chris Ferrie, Rainer Blatt, Philipp Schindler, Thomas Monz, Richard Kueng, and Martin Ringbauer. Ex-

- perimental single-setting quantum state tomography. *PRX Quantum*, 3:040310, Oct 2022. DOI: [10.1103/PRXQuantum.3.040310](https://doi.org/10.1103/PRXQuantum.3.040310).
- [24] H. Chau Nguyen, Jan Lennart Bönsel, Jonathan Steinberg, and Otfried Gühne. Optimizing shadow tomography with generalized measurements. *Phys. Rev. Lett.*, 129: 220502, Nov 2022. DOI: [10.1103/PhysRevLett.129.220502](https://doi.org/10.1103/PhysRevLett.129.220502).
 - [25] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical shadow tomography with locally scrambled quantum dynamics. *Physical Review Research*, 5(2):023027, 2023. DOI: [10.1103/PhysRevResearch.5.023027](https://doi.org/10.1103/PhysRevResearch.5.023027).
 - [26] Christian Bertonni, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, and Hakop Pashayan. Shallow shadows: Expectation estimation using low-depth random clifford circuits. *arXiv preprint [arXiv:2209.12924](https://arxiv.org/abs/2209.12924)*, 2022. DOI: [10.48550/arXiv.2209.12924](https://doi.org/10.48550/arXiv.2209.12924).
 - [27] Mirko Arienzo, Markus Heinrich, Ingo Roth, and Martin Kliesch. Closed-form analytic expressions for shadow estimation with brickwork circuits. *arXiv preprint [arXiv:2211.09835](https://arxiv.org/abs/2211.09835)*, 2022. DOI: [10.48550/arXiv.2211.09835](https://doi.org/10.48550/arXiv.2211.09835).
 - [28] Hong-Ye Hu and Yi-Zhuang You. Hamiltonian-driven shadow tomography of quantum states. *Phys. Rev. Research*, 4:013054, Jan 2022. DOI: [10.1103/PhysRevResearch.4.013054](https://doi.org/10.1103/PhysRevResearch.4.013054).
 - [29] Minh C Tran, Daniel K Mark, Wen Wei Ho, and Soonwon Choi. Measuring arbitrary physical properties in analog quantum simulation. *Physical Review X*, 13(1):011049, 2023. DOI: [10.1103/PhysRevX.13.011049](https://doi.org/10.1103/PhysRevX.13.011049).
 - [30] Max McGinley and Michele Fava. Shadow tomography from emergent state designs in analog quantum simulators. *arXiv preprint [arXiv:2212.02543](https://arxiv.org/abs/2212.02543)*, 2022. DOI: [10.48550/arXiv.2212.02543](https://doi.org/10.48550/arXiv.2212.02543).
 - [31] Katherine Van Kirk, Jordan Cotler, Hsin-Yuan Huang, and Mikhail D Lukin. Hardware-efficient learning of quantum many-body states. *arXiv preprint [arXiv:2212.06084](https://arxiv.org/abs/2212.06084)*, 2022. DOI: [10.48550/arXiv.2212.06084](https://doi.org/10.48550/arXiv.2212.06084).
 - [32] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. *PRX Quantum*, 2:030348, Sep 2021. DOI: [10.1103/PRXQuantum.2.030348](https://doi.org/10.1103/PRXQuantum.2.030348).
 - [33] Dax Enshan Koh and Sabee Grewal. Classical shadows with noise. *Quantum*, 6:776, 2022. DOI: [10.22331/q-2022-08-16-776](https://doi.org/10.22331/q-2022-08-16-776).
 - [34] Matthias Ohliger, Vincent Nesme, and Jens Eisert. Efficient and feasible state tomography of quantum many-body systems. *New Journal of Physics*, 15(1):015024, 2013. DOI: [10.1088/1367-2630/15/1/015024](https://doi.org/10.1088/1367-2630/15/1/015024).
 - [35] Ahmed A Akhtar, Hong-Ye Hu, and Yi-Zhuang You. Scalable and flexible classical shadow tomography with tensor networks. *Quantum*, 7:1026, 2023. DOI: [10.22331/q-2023-06-01-1026](https://doi.org/10.22331/q-2023-06-01-1026).
 - [36] Huangjun Zhu, Richard Kueng, Markus Grassl, and David Gross. The clifford group fails gracefully to be a unitary 4-design. *arXiv:1609.08172*, 2016. DOI: [10.48550/arXiv.1609.08172](https://doi.org/10.48550/arXiv.1609.08172).
 - [37] David Gross, Sepehr Nezami, and Michael Walter. Schur–weyl duality for the clifford group with applications: Property testing, a robust hudson theorem, and de finetti representations. *Communications in Mathematical Physics*, 385(3):1325–1393, 2021. DOI: [10.1007/s00220-021-04118-7](https://doi.org/10.1007/s00220-021-04118-7).
 - [38] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. Quantum computational chemistry. *Rev. Mod. Phys.*, 92:015003, Mar 2020. DOI: [10.1103/RevModPhys.92.015003](https://doi.org/10.1103/RevModPhys.92.015003).
 - [39] Srilekha Gandhari, Victor V Albert, Thomas Gerrits, Jacob M Taylor, and Michael J

- Gullans. Continuous-variable shadow tomography. *arXiv preprint arXiv:2211.05149*, 2022. DOI: [10.48550/arXiv.2211.05149](https://doi.org/10.48550/arXiv.2211.05149).
- [40] Simon Becker, Nilanjana Datta, Ludovico Lami, and Cambyse Rouzé. Classical shadow tomography for continuous variables quantum systems. *arXiv preprint arXiv:2211.07578*, 2022. DOI: [10.48550/arXiv.2211.07578](https://doi.org/10.48550/arXiv.2211.07578).
 - [41] Tianren Gu, Xiao Yuan, and Bujiao Wu. Efficient measurement schemes for bosonic systems. *arXiv preprint arXiv:2210.13585*, 2022. DOI: [10.48550/arXiv.2210.13585](https://doi.org/10.48550/arXiv.2210.13585).
 - [42] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. Fermionic partial tomography via classical shadows. *Phys. Rev. Lett.*, 127:110504, Sep 2021. DOI: [10.1103/PhysRevLett.127.110504](https://doi.org/10.1103/PhysRevLett.127.110504).
 - [43] Guang Hao Low. Classical shadows of fermions with particle number symmetry. *arXiv preprint arXiv:2208.08964*, 2022. DOI: [10.48550/arXiv.2208.08964](https://doi.org/10.48550/arXiv.2208.08964).
 - [44] You Zhou, Pei Zeng, and Zhenhuan Liu. Single-copies estimation of entanglement negativity. *Phys. Rev. Lett.*, 125:200502, Nov 2020. DOI: [10.1103/PhysRevLett.125.200502](https://doi.org/10.1103/PhysRevLett.125.200502).
 - [45] Andreas Ketterer, Nikolai Wyderka, and Otfried Gühne. Characterizing multipartite entanglement with moments of random correlations. *Phys. Rev. Lett.*, 122:120505, Mar 2019. DOI: [10.1103/PhysRevLett.122.120505](https://doi.org/10.1103/PhysRevLett.122.120505).
 - [46] Xiao-Dong Yu, Satoya Imai, and Otfried Gühne. Optimal entanglement certification from moments of the partial transpose. *Phys. Rev. Lett.*, 127:060504, Aug 2021. DOI: [10.1103/PhysRevLett.127.060504](https://doi.org/10.1103/PhysRevLett.127.060504).
 - [47] Zhenhuan Liu, Pei Zeng, You Zhou, and Mile Gu. Characterizing correlation within multipartite quantum systems via local randomized measurements. *Phys. Rev. A*, 105:022407, Feb 2022. DOI: [10.1103/PhysRevA.105.022407](https://doi.org/10.1103/PhysRevA.105.022407).
 - [48] Jonas Helsen and Michael Walter. Thrifty shadow estimation: re-using quantum circuits and bounding tails. *arXiv preprint arXiv:2212.06240*, 2022. DOI: [10.48550/arXiv.2212.06240](https://doi.org/10.48550/arXiv.2212.06240).
 - [49] Benoit Collins and Piotr Sniady. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006. ISSN 1432-0916. DOI: [10.1007/s00220-006-1554-3](https://doi.org/10.1007/s00220-006-1554-3).
 - [50] Daniel A. Roberts and Beni Yoshida. Chaos and complexity by design. *Journal of High Energy Physics*, 2017(4):121, 2017. ISSN 1029-8479. DOI: [10.1007/JHEP04\(2017\)121](https://doi.org/10.1007/JHEP04(2017)121).
 - [51] 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, Oct 2018. DOI: [10.1103/PhysRevLett.121.170502](https://doi.org/10.1103/PhysRevLett.121.170502).
 - [52] Lorenzo Leone, Salvatore F. E. Oliviero, You Zhou, and Alioscia Hama. Quantum Chaos is Quantum. *Quantum*, 5:453, May 2021. ISSN 2521-327X. DOI: [10.22331/q-2021-05-04-453](https://doi.org/10.22331/q-2021-05-04-453).

A Proofs of cross-shadow-norm

A.1 Proof of cross-shadow-norm as a norm

First one has $\|0\|_{\text{Xshadow}} = 0$, and then we verify the triangle inequality as follows.

$$\begin{aligned}
& \|O^1 + O^2\|_{\text{Xshadow}}^2 \\
&= \mathbb{E}_U \sum_{\mathbf{b}, \mathbf{b}'} \langle \mathbf{b} | U \sigma U^\dagger | \mathbf{b} \rangle \langle \mathbf{b} | U [\mathcal{M}^{-1}(O^1) + \mathcal{M}^{-1}(O^2)] U^\dagger | \mathbf{b} \rangle \\
&\quad \times \langle \mathbf{b}' | U \sigma U^\dagger | \mathbf{b}' \rangle \langle \mathbf{b}' | U [\mathcal{M}^{-1}(O^1) + \mathcal{M}^{-1}(O^2)] U^\dagger | \mathbf{b}' \rangle \\
&= \Gamma_2(\sigma, O^1) + \Gamma_2(\sigma, O^2) \\
&\quad + 2 \mathbb{E}_U \sum_{\mathbf{b}, \mathbf{b}'} \langle \mathbf{b} | U \sigma U^\dagger | \mathbf{b} \rangle \langle \mathbf{b} | U \mathcal{M}^{-1}(O^1) U^\dagger | \mathbf{b} \rangle \langle \mathbf{b}' | U \sigma U^\dagger | \mathbf{b}' \rangle \langle \mathbf{b}' | U \mathcal{M}^{-1}(O^2) U^\dagger | \mathbf{b}' \rangle \\
&\leq \|O^1\|_{\text{Xshadow}}^2 + \|O^2\|_{\text{Xshadow}}^2 \\
&\quad + 2 \mathbb{E}_U \sum_{\mathbf{b}} \langle \mathbf{b} | U \sigma U^\dagger | \mathbf{b} \rangle \langle \mathbf{b} | U \mathcal{M}^{-1}(O^1) U^\dagger | \mathbf{b} \rangle \sum_{\mathbf{b}'} \langle \mathbf{b}' | U \sigma U^\dagger | \mathbf{b}' \rangle \langle \mathbf{b}' | U \mathcal{M}^{-1}(O^2) U^\dagger | \mathbf{b}' \rangle \\
&\leq \|O^1\|_{\text{Xshadow}}^2 + \|O^2\|_{\text{Xshadow}}^2 + 2 \sqrt{\Gamma_2(\sigma, O^1)} \sqrt{\Gamma_2(\sigma, O^2)} \\
&\leq \|O^1\|_{\text{Xshadow}}^2 + \|O^2\|_{\text{Xshadow}}^2 + 2 \sqrt{\|O^1\|_{\text{Xshadow}} \|O^2\|_{\text{Xshadow}}} \\
&= (\|O^1\|_{\text{Xshadow}} + \|O^2\|_{\text{Xshadow}})^2.
\end{aligned} \tag{20}$$

Here the first and third inequalities are due to the definition of XSnorm in Eq. (5), and σ is assumed to be the optimal state in the maximization for $O^1 + O^2$, but may be not the optimal one for O^1 and O^2 respectively. The second inequality is by the Cauchy–Schwarz inequality for the domain of the random unitary U . That is, we can take the formula of the cross term as the inner product of two vectors indexed by U .

A.2 Proof of Theorem 1

As $\hat{\rho}_{(i)}$ are M i.i.d. random variables, we only need to focus on a specific i_0 , and get the final variance by directly dividing it by M . By definition the variance of $\text{tr}(O \hat{\rho}_{(i_0)})$ shows

$$\begin{aligned}
\text{Var} \left[\text{tr}(O \hat{\rho}_{(i_0)}) \right] &= \mathbb{E} \left[\text{tr}(O \hat{\rho}_{(i_0)})^2 \right] - \text{tr}(O \rho)^2 \\
&= \mathbb{E} \left[\text{tr}(O_0 \hat{\rho}_{(i_0)})^2 \right] - \text{tr}(O_0 \rho)^2.
\end{aligned} \tag{21}$$

Here one can shift the operator to its traceless part without changing the variance. The expectation value can be written explicitly as

$$\begin{aligned}
\mathbb{E} \left[\text{tr}(O \hat{\rho}_{(i_0)})^2 \right] &= \mathbb{E} \left[\left(\frac{1}{K} \sum_j \text{tr}(O \hat{\rho}_{(i_0)}^{(j)}) \right)^2 \right] \\
&= \frac{1}{K^2} \sum_{j, j'} \mathbb{E} \left[\text{tr}(O \hat{\rho}_{(i_0)}^{(j)}) \text{tr}(O \hat{\rho}_{(i_0)}^{(j')}) \right].
\end{aligned} \tag{22}$$

The expectation value of the terms in the summation depends on the coincidence of

the index j . For $j = j'$ with totally K terms, one has

$$\begin{aligned}
& \mathbb{E} \operatorname{tr}(O\hat{\rho}_{(i_0)}^{(j)}) \operatorname{tr}(O\hat{\rho}_{(i_0)}^{(j)}) \\
&= \mathbb{E} \langle \hat{\mathbf{b}}^{(j)} | U\mathcal{M}^{-1}(O)U^\dagger | \hat{\mathbf{b}}^{(j)} \rangle^2 \\
&= \mathbb{E}_U \sum_{\mathbf{b}^{(j)}} \Pr(\mathbf{b}^{(j)}|U) \langle \mathbf{b}^{(j)} | U\mathcal{M}^{-1}(O)U^\dagger | \mathbf{b}^{(j)} \rangle^2 \\
&= \mathbb{E}_U \sum_{\mathbf{b}^{(j)}} \langle \mathbf{b}^{(j)} | U\rho U^\dagger | \mathbf{b}^{(j)} \rangle \langle \mathbf{b}^{(j)} | U\mathcal{M}^{-1}(O)U^\dagger | \mathbf{b}^{(j)} \rangle^2 = \Gamma_1(\rho, O).
\end{aligned} \tag{23}$$

by definition in Eq. (4). Here in the first equality we insert the definition of $\hat{\rho}_{(i_0)}^{(j)}$, and use the self-adjoint property of the inverse channel $\mathcal{M}^{-1}$.

For $j \neq j'$ with totally $K^2 - K$ terms, the measurements are under the same setting i_0 but for different shots, one has

$$\begin{aligned}
& \mathbb{E} \operatorname{tr}(O\hat{\rho}_{(i_0)}^{(j)}) \operatorname{tr}(O\hat{\rho}_{(i_0)}^{(j')}) \\
&= \mathbb{E} \langle \mathbf{b}^{(j)} | U\mathcal{M}^{-1}(O)U^\dagger | \hat{\mathbf{b}}^{(j)} \rangle \langle \mathbf{b}^{(j')} | U\mathcal{M}^{-1}(O)U^\dagger | \hat{\mathbf{b}}^{(j')} \rangle \\
&= \mathbb{E}_U \sum_{\mathbf{b}^{(j)}, \mathbf{b}^{(j')}} \Pr(\mathbf{b}^{(j)}|U) \Pr(\mathbf{b}^{(j')}|U) \langle \mathbf{b}^{(j)} | U\mathcal{M}^{-1}(O)U^\dagger | \hat{\mathbf{b}}^{(j)} \rangle \langle \mathbf{b}^{(j')} | U\mathcal{M}^{-1}(O)U^\dagger | \hat{\mathbf{b}}^{(j')} \rangle \\
&= \mathbb{E}_U \sum_{\mathbf{b}^{(j)}, \mathbf{b}^{(j')}} \langle \mathbf{b}^{(j)} | U\rho U^\dagger | \mathbf{b}^{(j)} \rangle \langle \mathbf{b}^{(j')} | U\rho U^\dagger | \mathbf{b}^{(j')} \rangle \langle \mathbf{b}^{(j)} | U\mathcal{M}^{-1}(O)U^\dagger | \mathbf{b}^{(j)} \rangle \langle \mathbf{b}^{(j')} | U\mathcal{M}^{-1}(O)U^\dagger | \mathbf{b}^{(j')} \rangle \\
&= \Gamma_2(\rho, O),
\end{aligned} \tag{24}$$

by definition in Eq. (4), and here the subscript i_0 is omitted without ambiguity.

As a result, the total variance shows

$$\begin{aligned}
\operatorname{Var}[\operatorname{tr}(O\hat{\rho})] &= \frac{1}{M} \operatorname{Var}[\operatorname{tr}(O\hat{\rho}_{(i_0)})] = \frac{1}{M} \operatorname{Var}[\operatorname{tr}(O_0\hat{\rho}_{(i_0)})] \\
&= \frac{1}{M} \left[\frac{1}{K} \Gamma_1(\rho, O_0) + (1 - \frac{1}{K}) \Gamma_2(\rho, O_0) - \operatorname{tr}(O_0\rho)^2 \right],
\end{aligned} \tag{25}$$

and we finish the proof.

B Variance analysis of a collection of observables

In this section, we extend the variance results on a single observable O in the main text to a collection of observable $\{O^i\}_{i=1}^L$. In particular, we are interested in the summation of them $H = \sum_i O^i$, and each $O^i = \alpha_i P_i$ could be proportional to a Pauli operator P_i . For instance, H can be a Hamiltonian of an n -qubit system, and we aim to estimate the expectation value of H within error ϵ . We first give general formulas of the measurement budget or the variance without a specific chosen random measurements, and then show the result of estimating H under random Pauli measurements, which is of practical interest. For simplicity, we denote the estimator based on the shadow $\hat{\rho}$ as $\hat{O} = \operatorname{tr}(O\hat{\rho})$ and the expectation value as $\bar{O} = \operatorname{tr}(O\rho)$ in the following discussion.

We mainly give two solutions to this task. The first one is directly using the median-of-mean technique for all $\{O^i\}_{i=1}^L$ following Theorem S1 of Ref. [1]. Remembering that we totally collect M i.i.d. estimators $\{\hat{\rho}_{(k)}\}$, and each of them is composed of measurement data from repeated K -shots as shown in Eq. (2). For the median-of-mean post-processing, $\{\hat{\rho}_{(k)}\}$

is grouped into M_1 sets with each set containing M_2 elements (please refer to Eq. (S12) in Ref. [1] for more details). The estimator for each O^i is denoted as $\hat{O}^i(M_1, M_2)$, and $\hat{H}(M_1, M_2)$ is thus the summation of them. Different from estimating each O^i within error ϵ there, we should instead keep the whole error within ϵ . Let $\epsilon_i = \epsilon \|O^i\|_\infty / \sum_i \|O^i\|_\infty$ and $\epsilon = \sum_i \epsilon_i$, one can find the following result.

Corollary 1. *Set $M_1 = 2 \log(2L/\delta)$ and $M_2 = 34\epsilon^{-2} (\sum_i \|O^i\|_\infty)^2 \max_{j \in [L]} \text{Var}(\hat{O}^j / \|O^j\|_\infty)$, with the variance dependent on the particular random measurement primitive. A collection of $M = M_1 M_2$ shadow estimators $\{\hat{\rho}_{(k)}\}_{k=1}^M$ via the the median-of-mean prediction is sufficient to make the estimating error of $H = \sum_i O^i$, $|\hat{H}(M_1, M_2) - \bar{H}| \leq \epsilon$, under confidence level $1 - \delta$.*

In particular, for each $O^i = \alpha_i P_i$ and the random Pauli measurements, the number of independent estimators is at most

$$M = 68 \log(2L/\delta) \epsilon^{-2} \|\vec{\alpha}\|_1^2 \max_{j \in [L]} \text{Var}(\hat{P}_j), \quad (26)$$

Here $\|\vec{\alpha}\|_1 = \sum_i |\alpha_i|$ is the 1-norm of the coefficient vector, and $\text{Var}(\hat{P}_j)$ of Pauli operator P_j on a single estimator $\hat{\rho}_{(k_0)}$ given in Eq. (13) (i.e., $M = 1$ there).

Here the error of estimating H is bounded by the triangle inequality of each error ϵ_i of estimating O^i .

The second way is to calculate and bound the variance of $\hat{H}$, based on the variance of each $\hat{O}^i$. Let us first define some functions of a state σ and two observable O^1 and O^2 generalized from Eq. (4).

$$\begin{aligned} \Gamma_1(\sigma, O^1, O^2) &:= \mathbb{E}_U \sum_{\mathbf{b}} \langle \mathbf{b} | U \sigma U^\dagger | \mathbf{b} \rangle \langle \mathbf{b} | U \mathcal{M}^{-1}(O^1) U^\dagger | \mathbf{b} \rangle \langle \mathbf{b} | U \mathcal{M}^{-1}(O^2) U^\dagger | \mathbf{b} \rangle, \\ \Gamma_2(\sigma, O^1, O^2) &:= \mathbb{E}_U \sum_{\mathbf{b}, \mathbf{b}'} \langle \mathbf{b} | U \sigma U^\dagger | \mathbf{b} \rangle \langle \mathbf{b} | U \mathcal{M}^{-1}(O^1) U^\dagger | \mathbf{b} \rangle \langle \mathbf{b}' | U \sigma U^\dagger | \mathbf{b}' \rangle \langle \mathbf{b}' | U \mathcal{M}^{-1}(O^2) U^\dagger | \mathbf{b}' \rangle. \end{aligned} \quad (27)$$

One has the following variance result of estimating H .

Theorem 4. *By using multi-shot shadow estimation with M -round and K -shot, the statistical variance of $\hat{H}$ with $H = \sum_i O^i$ shows*

$$\text{Var}(\hat{H}) = \sum_i \text{Var}(\hat{O}^i) + 2 \sum_{i < j} \text{Cov}(\hat{O}^i, \hat{O}^j). \quad (28)$$

The variance of $\text{Var}(\hat{O}^i)$ is given in Eq. (6), and the covariance can be calculated via

$$\text{Cov}(\hat{O}^i, \hat{O}^j) = \frac{1}{M} \left[\frac{1}{K} \Gamma_1(\rho, O^i, O^j) + (1 - \frac{1}{K}) \Gamma_2(\rho, O^i, O^j) - \bar{O}^i \bar{O}^j \right]. \quad (29)$$

The proof follows similarly to that of Theorem 1 in Sec. A.2. In principle, one can calculate $\text{Cov}(\hat{O}^i, \hat{O}^j)$ via Eq. (29), similar as we have done in main text for $\text{Var}(\hat{O}^i)$. For the random Pauli measurements, Ref. [1] and Ref. [14] investigate $\Gamma_1(\rho, O^i, O^j)$ and $\Gamma_2(\rho, O^i, O^j)$ for two Pauli observables, respectively.

Here for a supplement, we further give some bound for the total variance $\text{Var}(\hat{H})$. Note that $\text{Cov}(\hat{O}^i, \hat{O}^j) \leq \sqrt{\text{Var}(\hat{O}^i) \text{Var}(\hat{O}^j)}$ by Cauchy-Schwarz inequality, and thus one has $\text{Var}(\hat{H}) \leq \left[\sum_i \sqrt{\text{Var}(\hat{O}^i)} \right]^2$. Applying this to random Pauli measurements, one finally gets the following result.

Corollary 2. For $H = \sum_i \alpha_i P_i$ and the random Pauli measurements, the variance of estimating H is upper bounded by

$$\text{Var}(\hat{H}) \leq \left[\sum_i \sqrt{\text{Var}(\alpha_i \hat{P}_i)} \right]^2 \leq \|\vec{\alpha}\|_1^2 \max_{j \in [L]} \text{Var}(\hat{P}_j). \quad (30)$$

with $\text{Var}(\hat{P}_j)$ given in Eq. (13).

The second inequality in Eq. (30) is by the Holder's inequality ($p = 1, q = \infty$).

By adopting the median-of-mean technique only for a single observable H , one sets $M_1 = 2 \log(2/\delta)$ and $M_2 = 34\epsilon^{-2} \text{Var}(\hat{H})$, with the variance on a single estimator $\hat{\rho}_{(k_0)}$. In this way, a collection shadow estimators $\{\hat{\rho}_{(k)}\}_{i=1}^M$ is sufficient to make the estimator error less than ϵ under confidence $1 - \delta$. By using upper bound in Eq. (30), one has M at most being

$$\begin{aligned} M &= 68 \log(2/\delta) \epsilon^{-2} \left[\sum_i \sqrt{\text{Var}(\alpha_i \hat{P}_i)} \right]^2 \\ &\leq 68 \log(2/\delta) \epsilon^{-2} \|\vec{\alpha}\|_1^2 \max_{j \in [L]} \text{Var}(\hat{P}_j). \end{aligned} \quad (31)$$

It is clear that the upper bound of M obtained here is better than the one in Eq. (26).

We remark that one can also apply Cauchy–Schwarz inequality to Eq. (30) to find another upper bound $\text{Var}(\hat{H}) \leq \|\vec{\alpha}\|_2^2 \sum_j \text{Var}(\hat{P}_j)$. Note that $\text{Var}(\hat{P}_j)$ is a function of w_i , $\bar{P}_i$ and the shot-number K . One may further utilizes the commuting relation among the operators $\{P_i\}$ of interest to minimize the measurement budget, and we leave it to some further study.

C Proof for random Pauli measurements

C.1 Proof of Lemma 1

We first decompose Λ_1 defined in Lemma 1 into the Pauli operator representation as follows.

$$\begin{aligned} \Lambda_1 &= \sum_b |b\rangle \langle b|^{\otimes 2} \sum_{b'} |b'\rangle \langle b'|^{\otimes 2} \\ &= \frac{1}{2} (\mathbb{I}_2 \otimes \mathbb{I}_2 + Z \otimes Z) \otimes \frac{1}{2} (\mathbb{I}_2 \otimes \mathbb{I}_2 + Z \otimes Z) \\ &= \frac{1}{4} (\mathbb{I}_2^{\otimes 4} + \mathbb{I}_2^{\otimes 2} Z^{\otimes 2} + Z^{\otimes 2} \mathbb{I}_2^{\otimes 2} + Z^{\otimes 4}) \end{aligned} \quad (32)$$

Here $\mathbb{I}_2$ and Z are single-qubit identity and Pauli Z operator. By inserting this decomposition inside the 4-copy single-qubit Clifford twirling channel,

$$\begin{aligned} &\Phi_1^{(4, \text{Cl})} \left[\frac{1}{4} (\mathbb{I}_2^{\otimes 4} + \mathbb{I}_2^{\otimes 2} Z^{\otimes 2} + Z^{\otimes 2} \mathbb{I}_2^{\otimes 2} + Z^{\otimes 4}) \right] \\ &= \frac{1}{4} [\mathbb{I}_2^{\otimes 4} + \mathbb{I}_2^{\otimes 2} \otimes \Phi_1^{(2, \text{cl})}(Z^{\otimes 2}) + \Phi_1^{(2, \text{cl})}(Z^{\otimes 2}) \otimes \mathbb{I}_2^{\otimes 2} + \Phi_1^{(4, \text{cl})}(Z^{\otimes 4})] \\ &= \frac{1}{4} [\mathbb{I}_2^{\otimes 4} + \mathbb{I}_2^{\otimes 2} \frac{1}{3} (2\mathbb{F}^{(2)} - \mathbb{I}^{\otimes 2}) + \frac{1}{3} (2\mathbb{F}^{(2)} - \mathbb{I}^{\otimes 2}) \mathbb{I}_2^{\otimes 2} + (4\mathbb{F}^{(4)} - \mathbb{I}_2^{\otimes 4})/3] \\ &= 1/6 \mathbb{I}_2^{\otimes 2} \mathbb{F}^{(2)} + 1/6 \mathbb{F}^{(2)} \mathbb{I}_2^{\otimes 2} + 1/3 \mathbb{F}^{(4)}. \end{aligned} \quad (33)$$

In the second line: the first term is due to $\Phi_1^{(4, \text{Cl})}$ being unital, the two terms in the middle reduce to 2-copy twirling, and the last term is still a 4-copy twirling. In the third line, we insert the result of the following Lemma.

Lemma 2. *The t -fold single-qubit Clifford twirling channel maps any Pauli operator $P_1^{\otimes t}$ with $P_1 \in \{X, Y, Z\}$ into*

$$\Phi_1^{(t, \text{Cl})}(P_1^{\otimes t}) = \begin{cases} \frac{1}{3}(X^{\otimes t} + Y^{\otimes t} + Z^{\otimes t}) = (2^{t/2}\mathbb{F}^{(t)} - \mathbb{I}_2^{\otimes t})/3, & t \text{ even} \\ 0, & t \text{ odd} \end{cases} \quad (34)$$

where in the first line we define $\mathbb{F}^{(t)} := 2^{-t/2}(\mathbb{I}_2^{\otimes t} + X^{\otimes t} + Y^{\otimes t} + Z^{\otimes t})$, and $\mathbb{F}^{(2)}$ is the swap operator on two qubits.

Proof. By definition, random single-qubit Clifford gate takes $P_1 \in \{X, Y, Z\}$ uniformly to all six directions in the Bloch sphere, i.e., $\{\pm X, \pm Y, \pm Z\}$. Consequently, the final result is the average of six terms, i.e., $1/6 \sum_{P_1 \in \{\pm X, \pm Y, \pm Z\}} P_1^{\otimes t}$. In the even t case, we thus have the equal weight summation of $X^{\otimes t}, Y^{\otimes t}, Z^{\otimes t}$. In the odd k case, $X^{\otimes t}$ and $(-X)^{\otimes t}$ cancel with each other, same for Y, Z terms, thus it returns zero. $\square$

C.2 Proof of Proposition 1

Proof. Without loss of generality, suppose $P = P_1 \otimes \cdots \otimes P_w \otimes \mathbb{I}_{[n-w]}$ with the first w -qubit owning Pauli operator, and of course traceless. The inverse channel maps $\mathcal{M}^{-1}(O) = \bigotimes_{i=1}^n \mathcal{M}_1^{-1}(O) = \bigotimes_{i=1}^w 3P_i \otimes \mathbb{I}_{[n-w]}$. By definition of Eq. (4), one has

$$\begin{aligned} & \mathbb{E}_U \sum_{\mathbf{b}, \mathbf{b}'} \langle \mathbf{b} | U \sigma U^\dagger | \mathbf{b} \rangle \langle \mathbf{b}' | U \sigma U^\dagger | \mathbf{b}' \rangle \langle \mathbf{b} | U \mathcal{M}^{-1}(O) U^\dagger | \mathbf{b} \rangle \langle \mathbf{b}' | U \mathcal{M}^{-1}(O) U^\dagger | \mathbf{b}' \rangle \\ &= \mathbb{E}_{U = \bigotimes_{i=1}^n u_i} \sum_{b_1, \dots, b_n, b'_1, \dots, b'_n} \text{tr} \left[\sigma^{\otimes 2} \bigotimes_{i=1}^n (u_i^\dagger | b_i \rangle \langle b_i | u_i) \bigotimes_{i=1}^n (u_i^\dagger | b'_i \rangle \langle b'_i | u_i) \right] \\ & \quad \times \prod_{i=1}^w \langle b_i | u_i 3P_i u_i^\dagger | b_i \rangle \prod_{i=1}^w \langle b'_i | u_i 3P_i u_i^\dagger | b'_i \rangle \\ &= \mathbb{E}_{\bigotimes_{i=1}^w u_i} \sum_{b_1, \dots, b_w, b'_1, \dots, b'_w} \text{tr} \left[\sigma^{\otimes 2} \bigotimes_{i=1}^w (u_i^\dagger | b_i \rangle \langle b_i | u_i) \otimes \mathbb{I}_{[n-w]} \bigotimes_{i=1}^w (u_i^\dagger | b'_i \rangle \langle b'_i | u_i) \otimes \mathbb{I}_{[n-w]} \right] \\ & \quad \times \prod_{i=1}^w \langle b_i | u_i 3P_i u_i^\dagger | b_i \rangle \prod_{i=1}^w \langle b'_i | u_i 3P_i u_i^\dagger | b'_i \rangle \\ &= \mathbb{E}_{\bigotimes_{i=1}^w u_i} \sum_{b_1, \dots, b_w, b'_1, \dots, b'_w} \text{tr} \left[\sigma_{[m]}^{\otimes 2} \bigotimes_{i=1}^w (u_i^\dagger | b_i \rangle \langle b_i | u_i) \bigotimes_{i=1}^w (u_i^\dagger | b'_i \rangle \langle b'_i | u_i) \right] \\ & \quad \times \prod_{i=1}^w \langle b_i | u_i 3P_i u_i^\dagger | b_i \rangle \prod_{i=1}^w \langle b'_i | u_i 3P_i u_i^\dagger | b'_i \rangle \\ &= 3^{2w} \text{tr} \left[(\sigma_{[w]} \bigotimes_{i=1}^w P_i)^{\otimes 2} \mathbb{E}_{\bigotimes_{i=1}^w u_i} \sum_{b_1, \dots, b_w, b'_1, \dots, b'_w} \bigotimes_{i=1}^w [(u_i^\dagger | b_i \rangle \langle b_i | u_i)^{\otimes 2} \otimes (u_i^\dagger | b'_i \rangle \langle b'_i | u_i)^{\otimes 2}] \right] \\ &= 3^{2w} \text{tr} \left[(\sigma_{[w]} \bigotimes_{i=1}^w P_i)^{\otimes 2} \bigotimes_{i=1}^w \Phi_{1, (i)}^{(4, \text{Cl})} \left(\sum_{b_i} |b_i \rangle \langle b_i|^{\otimes 2} \sum_{b'_i} |b'_i \rangle \langle b'_i|^{\otimes 2} \right) \right] \end{aligned} \quad (35)$$

Here in the first equality, we write the random unitary $U = \bigotimes_{i=1}^n u_i$ and computational basis summation $\mathbf{b} = \{b_1 b_2 \cdots b_n\}$, $\mathbf{b}' = \{b'_1 b'_2 \cdots b'_n\}$ on the single-qubit level; in the second

equality we sum the b_i, b'_i for $i > w$ in the trace, and it gives $\mathbb{I}_{[n-w]}$ on the last $n-w$ qubits no matter what u_i is chosen. And the problem is reduced on the first w -qubit in the third equality, with $\sigma_{[m]}$ the reduced state of the first m -qubit from σ . In the last two lines, we arrange these operators and relate the result to the single-qubit Clifford twirling.

By inserting the Eq. (12) of Lemma. 1, one finally has

$$\begin{aligned}\Gamma_2(\sigma, P) &= 3^w \text{tr} \left[(\sigma_{[w]} \bigotimes_{i=1}^w P_i)^{\otimes 2} \bigotimes_{i=1}^w \left(\frac{1}{2} \mathbb{I}_i^{\otimes 2} \otimes \mathbb{F}_i^{(2)} + \frac{1}{2} \mathbb{F}_i^{(2)} \otimes \mathbb{I}_i^{\otimes 2} + \mathbb{F}_i^{(4)} \right) \right] \\ &= 3^w \text{tr} \left[(\sigma_{[w]} \bigotimes_{i=1}^w P_i)^{\otimes 2} \bigotimes_{i=1}^w \left(\mathbb{F}_i^{(4)} \right) \right] \\ &= 3^w \text{tr} \left(\sigma_{[w]} \bigotimes_{i=1}^w P_i \right)^2 = 3^w \text{tr}(\sigma P)^2.\end{aligned}\tag{36}$$

Here the second line is by the following fact. Suppose there is an appearance of $\mathbb{I}_i^{\otimes 2} \otimes \mathbb{F}_i^{(2)}$ or $\mathbb{F}_i^{(2)} \otimes \mathbb{I}_i^{\otimes 2}$ on i -th qubit, then the identity operator on the first or last two-copy would result in $\text{tr}(P_i) = 0$, such that gives no contribution to the final result. Consequently, we are left only $\mathbb{F}_i^{(4)}$. In the final line, $\bigotimes_{i=1}^w \left(\mathbb{F}_i^{(4)} \right)$ contains all possible w -qubit Pauli operators, and only the one $\bigotimes_{i=1}^w P_i$ contributes to the final result. $\square$

D Statistical analysis for Haar random measurements

In this section, we aim to give the statistical analysis for the global Haar random measurement, that is, the unitary ensemble $\mathcal{E}$ is Haar random on $\mathcal{H}_D$ (or any unitary 4-design ensemble). The central result shows as follows, as an analog of Proposition 2 in main text.

Proposition 3. *Suppose O_0 is a traceless observable, the function Γ_2 defined in Eq. (4) is upper bounded by*

$$\Gamma_2^{\text{Haar}}(\sigma, O_0) \leq c_1 \|O_0\|_2^2, \tag{37}$$

for the random Haar measurements, where $\|A\|_2 = \sqrt{\text{tr}(AA^\dagger)}$ is the Frobenius norm, and c_1 is some constant independent of the dimension D . In this case, the XSnorm $\|O_0\|_{\text{Xshadow}}^{\text{Haar}} \leq \sqrt{c_1} \|O_0\|_2$ by Eq. (5).

The main task of this section is to prove this proposition, which is helpful for the discussion of random Clifford measurements shown in the next section.

Recall the essential quantity in Eq. (15) we would like to calculate shows

$$\begin{aligned}\Gamma_2^{\text{Haar}} &= (D+1)^2 \text{tr} \left[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi_n^{(4, \text{Haar})}(\Lambda_n) \right] \\ &= (D+1)^2 \text{tr} \left[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi_n^{(4, \text{Haar})}(\Lambda_n^0 + \Lambda_n^1) \right]\end{aligned}\tag{38}$$

with the only difference being that we use the Haar twirling $\Phi^{(4, \text{Haar})}$ here, and Λ_n is decomposed into two parts as in main text,

$$\begin{aligned}\Lambda_n^0 &= \sum_{\mathbf{b}} |\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4}, \\ \Lambda_n^1 &= \sum_{\mathbf{b} \neq \mathbf{b}'} |\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 2} \otimes |\mathbf{b}'\rangle \langle \mathbf{b}'|^{\otimes 2}.\end{aligned}$$

Before we calculate their contributions separately in the next subsections, we give a brief review of the result of the t -copy Haar twirling, and more details can be found in, for example Ref. [49–51]. Denote the permutation elements of the t -th order symmetric group as $\pi \in S_t$, and there is a unitary representation of π on the t -copy Hilbert space $\mathcal{H}_D^{\otimes t}$ as

$$T_\pi |\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_t\rangle = |\mathbf{b}_{\pi(1)}, \mathbf{b}_{\pi(2)}, \dots, \mathbf{b}_{\pi(t)}\rangle, \quad (39)$$

with $|\mathbf{b}\rangle$ the basis state for one copy. By the celebrated Schur–Weyl duality, the twirling result is related to the irreducible representation (ir-rep) of S_t as follows.

Lemma 3. *The t -fold Haar twirling channel maps $A \in \mathcal{H}_D^{\otimes t}$ into*

$$\Phi^{(t, \text{Haar})}(A) = \frac{1}{t!} \sum_{\pi \in S_t} \text{tr}(AT_\pi) T_{\pi^{-1}} \sum_{\lambda} \frac{d_\lambda}{D_\lambda} P_\lambda, \quad (40)$$

where λ denotes the irreducible representation of S_t , and P_λ is the corresponding projector showing

$$P_\lambda = \frac{d_\lambda}{t!} \sum_{\pi \in S_t} \chi^\lambda(\pi) T_\pi, \quad (41)$$

with $\chi^\lambda(\pi)$ being the character of π .

D.1 The contribution of Λ_n^0 in Haar case

Inserting the term Λ_n^0 in the 4-copy Haar twirling channel in Eq. (40) with $t = 4$ one has

$$\begin{aligned} \Phi^{(4, \text{Haar})}(\Lambda_n^0) &= \frac{1}{4!} \sum_{\pi \in S_4} \sum_{\mathbf{b}} \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4} T_{\pi^{-1}}) T_\pi \sum_{\lambda} \frac{d_\lambda}{D_\lambda} P_\lambda \\ &= \frac{D}{4!} \sum_{\pi \in S_4} T_\pi \sum_{\lambda} \frac{d_\lambda}{D_\lambda} P_\lambda \\ &= DP_{\text{sym}} \sum_{\lambda} \frac{d_\lambda}{D_\lambda} P_\lambda \\ &= \frac{D}{D_{\text{sym}}} P_{\text{sym}} = \frac{D}{4! D_{\text{sym}}} \sum_{\pi \in S_k} \pi \end{aligned} \quad (42)$$

Here the second line is by the fact $\text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4} T_{\pi^{-1}}) = 1$ for any $\mathbf{b}$ and π ; the third line is by the definition of the symmetric subspace as $P_{\text{sym}} = \frac{1}{4!} \sum_{\pi \in S_4} T_\pi$; the final line is by the fact that symmetric subspace is one of the ir-rep subspace and orthogonal to others. Inserting this twirling result and the dimension $D_{\text{sym}} = (D+3)(D+2)(D+1)D/4!$, one has

$$\begin{aligned} &(D+1)^2 \text{tr}[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi^{(4, \text{Haar})}(\Lambda_n^0)] \\ &= \frac{(D+1)^2 D}{(D+3)(D+2)(D+1)D} \sum_{\pi \in S_k} \text{tr}(\sigma \otimes O_0 \otimes \sigma \otimes O_0 T_\pi) \\ &= \mathcal{O}(D^{-1}) \sum_{\pi \in S_4} \text{tr}(\sigma \otimes O_0 \otimes \sigma \otimes O_0 T_\pi) = \mathcal{O}(D^{-1}) \text{tr}(O_0^2) \end{aligned} \quad (43)$$

The final line shows that the result is in the order $\mathcal{O}(D^{-1}) \text{tr}(O_0^2)$, by the following lemma.

Lemma 4. *For any $\pi \in S_4$, the following inequality holds for a quantum state σ and a traceless observable O_0 ,*

$$\text{tr}(\sigma \otimes O_0 \otimes \sigma \otimes O_0 T_\pi) \leq \text{tr}(O_0^2). \quad (44)$$

Proof. For the permutation operator T_π , the formula could take the following values: $\text{tr}(\sigma^2) \text{tr}(O_0^2)$, $\text{tr}(O_0^2)$, $\text{tr}(O_0^2 \sigma)$, $\text{tr}(O_0^2 \sigma^2)$, $\text{tr}(O_0 \sigma O_0 \sigma)$. All these can be bounded by $\text{tr}(O_0^2)$. We bound the last one with Cauchy–Schwarz inequality for operator as follows.

$$\text{tr}(O_0 \sigma O_0 \sigma) \leq \|O_0 \sigma\|_2^2 = \sqrt{\text{tr}(O_0 \sigma \sigma O_0)}^2 = \text{tr}(O_0^2 \sigma^2) \leq \|O_0^2\|_\infty \leq \text{tr}(O_0^2). \quad (45)$$

□

D.2 The contribution of Λ_n^1 in Haar case

For the second term of Λ_n^1 , by inserting the Haar twirling channel in Eq. (40) with $t = 4$, one has

$$\begin{aligned} \Phi^{(4, \text{Haar})}(\Lambda_n^1) &= \frac{1}{4!} \sum_{\pi \in S_4} \sum_{\mathbf{b} \neq \mathbf{b}'} \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 2} \otimes |\mathbf{b}'\rangle \langle \mathbf{b}'|^{\otimes 2} T_{\pi^{-1}}) T_\pi \sum_\lambda \frac{d_\lambda}{D_\lambda} P_\lambda \\ &= \frac{1}{4!} (D^2 - D) \sum_{\pi \in S_r} \pi \sum_\lambda \frac{d_\lambda}{D_\lambda} P_\lambda \\ &= \frac{1}{6} (D^2 - D) \sum_\lambda \frac{d_\lambda}{D_\lambda} P_r P_\lambda. \end{aligned} \quad (46)$$

Here the second line is by the fact that the trace formula gives zero unless $\pi \in S_r$, with the set $S_r = \{(), (12), (34), (12)(34)\}$ being a subgroup of S_4 . And we define the corresponding projector as $P_r = \frac{1}{4} \sum_{\pi \in S_r} T_\pi$. One thus further gets

$$(D+1)^2 \text{tr}[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi^{(4, \text{Haar})}(\Lambda_n^1)] = \frac{1}{6} (D^2 - D) (D+1)^2 \sum_\lambda \frac{d_\lambda}{D_\lambda} \text{tr}(\sigma \otimes O \otimes \sigma \otimes O P_r P_\lambda) \quad (47)$$

Recall the definition of P_λ in Eq. (41), and for each ir-rep λ , one has

$$\begin{aligned} \frac{d_\lambda}{D_\lambda} \text{tr}(\sigma \otimes O_0 \otimes \sigma \otimes O_0 P_r P_\lambda) &= \frac{d_\lambda^2}{D_\lambda} \frac{1}{4 * 4!} \sum_{\pi' \in S_r, \pi \in S_4} \chi^\lambda(\pi) \text{tr}(\sigma \otimes O_0 \otimes \sigma \otimes O_0 T_{\pi'} T_\pi) \\ &= \frac{d_\lambda^2}{D_\lambda} \frac{1}{4 * 4!} \sum_{\pi' \in S_r, \pi \in S_4} \chi^\lambda(\pi) \text{tr}(\sigma \otimes O_0 \otimes \sigma \otimes O_0 T_{\pi' \pi}) \\ &\leq \frac{d_\lambda^2}{D_\lambda} |\chi^\lambda(\pi)|_{\max \pi \in S_4} \text{tr}(O_0^2) \end{aligned} \quad (48)$$

where the last line we use Lemma 4 for each $T_{\pi' \pi}$. Here $|\chi^\lambda(\pi)|_{\max \pi \in S_4} \leq 3$, $\frac{d_\lambda^2}{D_\lambda} = \mathcal{O}(D^{-4})$ for any ir-rep λ , and there are totally 5 ir-reps [36]. As a result one has

$$\begin{aligned} (D+1)^2 \text{tr}[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi^{(4, \text{Haar})}(\Lambda_n^1)] &< \frac{1}{6} (D^2 - D) (D+1)^2 \left[\sum_\lambda \frac{d_\lambda^2}{D_\lambda} |\chi^\lambda(\pi)|_{\max \pi \in S_4} \right] \text{tr}(O_0^2) \\ &= \mathcal{O}(1) \text{tr}(O_0^2) \end{aligned} \quad (49)$$

Inserting Eq. (43) and Eq. (49) into Eq. (38), we finish the proof of Proposition 3.

E Statistical analysis for random Clifford measurements

Similar as the Haar case, here the we should calculate the quantity in Eq. (15) shown as follows,

$$\Gamma_2^{\text{Cl}}(\sigma, O_0) = (D+1)^2 \text{tr}[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi_n^{(4, \text{Cl})}(\Lambda_n^0 + \Lambda_n^1)]. \quad (50)$$

In the following subsections, as in the Haar case, we evaluate the contributions from Λ_n^0 and Λ_n^1 separately. As Clifford circuit is not a 4-design, the 4-copy twirling result is a little different compared to Eq. (40), which is shown as follows. One can refer to Ref. [36, 51, 52] for more details.

Lemma 5. *The 4-fold n -qubit Clifford twirling channel maps $A \in \mathcal{H}_D^{\otimes 4}$ into*

$$\Phi_n^{(4, \text{Cl})}(A) = \frac{1}{4!} \sum_{\lambda} d_{\lambda} \sum_{\pi \in S_4} \left[\frac{1}{D_{\lambda}^+} \text{tr}(A Q T_{\pi}) T_{\pi^{-1}} Q + \frac{1}{D_{\lambda}^-} \text{tr}(A Q^{\perp} T_{\pi}) T_{\pi^{-1}} Q^{\perp} \right] P_{\lambda} \quad (51)$$

where λ denotes the irreducible representation of S_t , and P_{λ} is the corresponding projector shown in Eq. (41) with $t = 4$. The operator Q is a stabilizer code subspace projector, commuting with any T_{π} and P_{λ} ,

$$Q = \frac{1}{D^2} \sum_k W_k^{\otimes 4} \quad (52)$$

with W_k running on all D^2 n -qubit Pauli operators including the identity; and $Q^{\perp} = \mathbb{I}_D^{\otimes 4} - Q$.

E.1 The contribution of Λ_n^0 in Clifford case

For the first term Λ_n^0 , by using the twirling formula in Eq. (51) one has

$$\begin{aligned} \Phi_n^{(4, \text{Cl})}(\Lambda_n^0) &= \frac{1}{4!} \sum_{\lambda} d_{\lambda} \sum_{\pi \in S_4} \sum_{\mathbf{b}} \left[\frac{1}{D_{\lambda}^+} \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4} Q T_{\pi^{-1}}) T_{\pi} Q + \frac{1}{D_{\lambda}^-} \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4} Q^{\perp} T_{\pi^{-1}}) T_{\pi} Q^{\perp} \right] P_{\lambda} \\ &= \frac{1}{4!} \sum_{\lambda} d_{\lambda} \sum_{\pi \in S_4} \sum_{\mathbf{b}} \left[\frac{1}{D_{\lambda}^+} \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4} Q) T_{\pi} Q + \frac{1}{D_{\lambda}^-} \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4} Q^{\perp}) T_{\pi} Q^{\perp} \right] P_{\lambda} \\ &= \frac{1}{4!} \sum_{\lambda} d_{\lambda} \sum_{\pi \in S_4} D \left[\frac{1}{D_{\lambda}^+ D} Q + \frac{1}{D_{\lambda}^-} (1 - 1/D) Q^{\perp} \right] T_{\pi} P_{\lambda} \\ &= \sum_{\lambda} d_{\lambda} \left[\frac{1}{D_{\lambda}^+} Q + \frac{D-1}{D_{\lambda}^-} Q^{\perp} \right] \sum_{\pi \in S_4} \frac{1}{4!} T_{\pi} P_{\lambda} \\ &= \left[\frac{1}{D_{\text{sym}}^+} Q + \frac{D-1}{D_{\text{sym}}^-} Q^{\perp} \right] P_{\text{sym}}. \end{aligned} \quad (53)$$

Here in the second line, we use the fact that $T_{\pi} |\mathbf{b}\rangle^{\otimes 4} = |\mathbf{b}\rangle^{\otimes 4}$ for any π . In the third line, $\text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4} Q) = 1/D$, since W_k which only contains $\{\mathbb{I}_2, Z\}$ on each qubit contributes to it and there are totally $2^n = D$ such terms. Therefore, $\text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 4} Q^{\perp}) = 1 - 1/D$. In the final line, we only have P_{sym} due to the orthogonality of each P_{λ} . Inserting this twirling result into the trace formula in Eq. (50) one gets

$$\begin{aligned}
& (D+1)^2 \text{tr} \left[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi_n^{(4, \text{Cl})} \left(\Lambda_n^0 \right) \right] \\
&= (D+1)^2 \left(\frac{1}{D_{\text{sym}}^+} - \frac{D-1}{D_{\text{sym}}^-} \right) \text{tr} [\sigma \otimes O_0 \otimes \sigma \otimes O_0 Q P_{\text{sym}}] + (D+1)^2 \frac{D-1}{D_{\text{sym}}^-} \text{tr} [\sigma \otimes O_0 \otimes \sigma \otimes O_0 P_{\text{sym}}] \\
&= \mathcal{O}(1) \text{tr} [\sigma \otimes O_0 \otimes \sigma \otimes O_0 Q P_{\text{sym}}] + \mathcal{O}(D^{-1}) \text{tr} [\sigma \otimes O_0 \otimes \sigma \otimes O_0 P_{\text{sym}}] \\
&= \mathcal{O}(D^{-1}) \text{tr} (O_0^2)
\end{aligned} \tag{54}$$

where in the third line we use the fact $D_{\text{sym}}^+ = \mathcal{O}(D^{-2})$ and $D_{\text{sym}}^- = \mathcal{O}(D^{-4})$ [36]; in the last line, the second term is directly from the result in Eq. (43) of the Haar case, and the first term is on account of the following Lemma 6, by taking $P_0 = P_{\text{sym}}$ and $P_1 = \mathbb{I}$ there. In total, we thus show that the contribution from Λ_n^0 is still in the order $\mathcal{O}(D^{-1}) \text{tr}(O^2)$ as in the Haar case.

Lemma 6. *For any two projectors P_0 and P_1 , which may not commute with each other but commute with Q , the following inequality holds for a quantum state σ and an observable O ,*

$$\text{tr}(\sigma \otimes O \otimes \sigma \otimes O Q P_0 P_1) \leq D^{-1} \text{tr}(O^2). \tag{55}$$

Proof. Suppose the spectrum decomposition of the observable is $O = \sum_j a_j |\Psi_j\rangle \langle \Psi_j|$, and we define the operator $O' = \sum_j i^{\delta(a_j < 0)} |a_j|^{\frac{1}{2}} |\Psi_j\rangle \langle \Psi_j|$, such that $O'^2 = O$ and $O'O'^\dagger = \sqrt{O^2}$.

$$\begin{aligned}
\text{tr}(\sigma \otimes O \otimes \sigma \otimes O Q P_0 P_1) &= \text{tr} \left(\sigma^{\frac{1}{2}} \otimes O' \otimes \sigma^{\frac{1}{2}} \otimes O' Q P_0 P_1 Q \sigma^{\frac{1}{2}} \otimes O' \otimes \sigma^{\frac{1}{2}} \otimes O' \right) \\
&\leq \|\sigma^{\frac{1}{2}} \otimes O' \otimes \sigma^{\frac{1}{2}} \otimes O' Q\|_2 \|P_0 P_1 Q \sigma^{\frac{1}{2}} \otimes O' \otimes \sigma^{\frac{1}{2}} \otimes O'\|_2 \\
&\leq \|\sigma^{\frac{1}{2}} \otimes O' \otimes \sigma^{\frac{1}{2}} \otimes O' Q\|_2^2 = \text{tr}(\sigma \otimes O' O'^\dagger \otimes \sigma \otimes O' O'^\dagger Q)
\end{aligned} \tag{56}$$

the first inequality is by using the Cauchy–Schwarz inequality, and the second inequality is by

$$\begin{aligned}
\|P_0 P_1 Q \sigma^{\frac{1}{2}} \otimes O' \otimes \sigma^{\frac{1}{2}} \otimes O'\|_2 &= \sqrt{\text{tr}(\sigma \otimes O' O'^\dagger \otimes \sigma \otimes O' O'^\dagger Q P_1 P_0 P_1 Q)} \\
&\leq \sqrt{\text{tr}(\sigma \otimes O' O'^\dagger \otimes \sigma \otimes O' O'^\dagger Q)} = \|\sigma^{\frac{1}{2}} \otimes O' \otimes \sigma^{\frac{1}{2}} \otimes O' Q\|_2,
\end{aligned} \tag{57}$$

since $P_1 P_0 P_1 \leq P_1 \leq \mathbb{I}$.

To further bound Eq. (56), we insert the formula of Q in Eq. (52) to get

$$\begin{aligned}
\text{tr}(\sigma \otimes O \otimes \sigma \otimes O Q P_0 P_1) &\leq \text{tr}(\sigma \otimes O' O'^\dagger \otimes \sigma \otimes O' O'^\dagger Q) \\
&= D^{-2} \sum_k \text{tr}(\sigma W_k)^2 \text{tr}(\sqrt{O^2} W_k)^2 \\
&\leq D^{-2} \sum_k \text{tr}(\sqrt{O^2} W_k)^2 \\
&= D^{-1} \text{tr}(O^2).
\end{aligned} \tag{58}$$

Here the second inequality is by the fact $\text{tr}(\sigma W_k)^2 \leq 1$, and the last line is by the decomposition of the n -qubit swap operator in the Pauli basis as $\mathbb{S} = D^{-1} \sum_k W_k \otimes W_k$. $\square$

E.2 The contribution of Λ_n^1 in Clifford case

For the second term Λ_n^1 , by using the twirling formula in Eq. (51) one has

$$\begin{aligned} \Phi_n^{(4,\text{Cl})}(\Lambda_n^1) &= \frac{1}{4!} \sum_{\lambda} d_{\lambda} \sum_{\pi \in S_4} \sum_{\mathbf{b} \neq \mathbf{b}'} \left[\frac{1}{D_{\lambda}^+} \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 2} \otimes |\mathbf{b}'\rangle \langle \mathbf{b}'|^{\otimes 2} Q T_{\pi}) T_{\pi^{-1}} Q \right. \\ &\quad \left. + \frac{1}{D_{\lambda}^-} \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 2} \otimes |\mathbf{b}'\rangle \langle \mathbf{b}'|^{\otimes 2} Q^{\perp} T_{\pi}) T_{\pi^{-1}} Q^{\perp} \right] P_{\lambda} \end{aligned} \quad (59)$$

We define the first trace formula as

$$\begin{aligned} Q_0(\pi) &:= \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 2} \otimes |\mathbf{b}'\rangle \langle \mathbf{b}'|^{\otimes 2} Q T_{\pi}) \\ &= D^{-2} \sum_k \langle \mathbf{b}| W_k \otimes \langle \mathbf{b}| W_k \otimes \langle \mathbf{b}'| W_k \otimes \langle \mathbf{b}'| W_k T_{\pi} |\mathbf{b}\rangle \otimes |\mathbf{b}\rangle \otimes |\mathbf{b}'\rangle \otimes |\mathbf{b}'\rangle. \end{aligned} \quad (60)$$

There are two cases which give nonzero value depending on π . In the first case, via the permutation T_{π} , $|\mathbf{b}\rangle$ is connected to $\langle \mathbf{b}|$, and this also holds for $\mathbf{b}'$. That is, $\pi \in S_r$ with the set $S_r = \{(), (12), (34), (12)(34)\}$ and one has $|\langle \mathbf{b}| W_k |\mathbf{b}\rangle|^2 |\langle \mathbf{b}'| W_k |\mathbf{b}'\rangle|^2$. As in the Haar case, one can only choose W_k with $\{\mathbb{I}_2, Z\}$ for each qubit, so it gives $D/D^2 = D^{-1}$. In the second case, all $|\mathbf{b}\rangle$ are connected to $\langle \mathbf{b}'|$, that is $\pi \in S_{r'}$, with $S_{r'} = (13)(24)S_r = S_r(13)(24)$. And one has $|\langle \mathbf{b}| W_k |\mathbf{b}'\rangle|^4$. For $\mathbf{b} \neq \mathbf{b}'$, one needs to choose $\{\mathbb{I}_2, Z\}$ for the qubit with the same bit value of $\mathbf{b}$ and $\mathbf{b}'$, and $\{X, Y\}$ for the qubit of different bit values. Thus the result is also D^{-1} . Besides these two cases, the term shows $|\langle \mathbf{b}| W_k |\mathbf{b}'\rangle|^2 |\langle \mathbf{b}| W_k |\mathbf{b}\rangle| |\langle \mathbf{b}'| W_k |\mathbf{b}'\rangle|$. It is not hard to see that it is 0 no matter what W_k is.

Define $Q_1(\pi) := \text{tr}(|\mathbf{b}\rangle \langle \mathbf{b}|^{\otimes 2} \otimes |\mathbf{b}'\rangle \langle \mathbf{b}'|^{\otimes 2} Q^{\perp} T_{\pi})$. Note that $Q_0(\pi) + Q_1(\pi) = 1$ when $\pi \in S_r$, otherwise it is 0, as shown in the Haar case. We thus have

$$Q_0(\pi) = \begin{cases} D^{-1}, & \pi \in S_r \cup S_{r'} \\ 0, & \pi \in S_4 \setminus (S_r \cup S_{r'}) \end{cases} \quad Q_1(\pi) = \begin{cases} 1 - D^{-1}, & \pi \in S_r \\ -D^{-1}, & \pi \in S_{r'} \\ 0, & \pi \in S_4 \setminus (S_r \cup S_{r'}) \end{cases} \quad (61)$$

Inserting these into Eq. (59), and define the projector $P_{r'} = \frac{1}{4} \sum_{\pi \in S_{r'}} T_{\pi}$ similar as P_r in the Haar case, one has

$$\begin{aligned} &\Phi_n^{(4,\text{Cl})}(\Lambda_n^1) \\ &= \frac{1}{4!} \sum_{\lambda} d_{\lambda} (D^2 - D) \left\{ \frac{1}{D_{\lambda}^+} D^{-1} \sum_{\pi \in S_r \cup S_{r'}} T_{\pi} Q + \frac{1}{D_{\lambda}^-} \left[(1 - D^{-1}) \sum_{\pi \in S_r} T_{\pi} - D^{-1} \sum_{\pi' \in S_{r'}} T_{\pi'} \right] Q^{\perp} \right\} P_{\lambda} \\ &= \sum_{\lambda} \Theta_1(\lambda) Q P_r P_{\lambda} + \Theta_2(\lambda) Q P_{r'} P_{\lambda} + \Theta_3(\lambda) P_r P_{\lambda} + \Theta_4(\lambda) P_{r'} P_{\lambda} \end{aligned} \quad (62)$$

with

$$\begin{aligned} \Theta_1(\lambda) &= \frac{1}{3!} d_{\lambda} (D^2 - D) \left(\frac{1}{D_{\lambda}^+ D} - \frac{1 - D^{-1}}{D_{\lambda}^-} \right) = \mathcal{O}(D^{-1}), \\ \Theta_2(\lambda) &= \frac{1}{3!} d_{\lambda} (D^2 - D) \left(\frac{1}{D_{\lambda}^+ D} + \frac{1}{D_{\lambda}^- D} \right) = \mathcal{O}(D^{-1}), \\ \Theta_3(\lambda) &= \frac{1}{3!} d_{\lambda} (D^2 - D) \frac{1 - D^{-1}}{D_{\lambda}^-} = \mathcal{O}(D^{-2}), \\ \Theta_4(\lambda) &= -\frac{1}{3!} d_{\lambda} (D^2 - D) \frac{1}{D_{\lambda}^- D} = \mathcal{O}(D^{-3}), \end{aligned} \quad (63)$$

by the fact $D_\lambda^+ = \mathcal{O}(D^2)$ (otherwise it is zero) and $D_\lambda^- = \mathcal{O}(D^4)$ [36]. For the ir-rep where $D_\lambda^+ = \text{tr}(QP_\lambda) = 0$, that is, $QP_\lambda = 0$, one has $\Theta_1(\lambda) = \Theta_2(\lambda) = 0$. Recall Eq. (53), one finally has the contribution of Λ_n^1 as

$$\begin{aligned}
& (D+1)^2 \text{tr} \left[\sigma \otimes O_0 \otimes \sigma \otimes O_0 \Phi_n^{(4,\text{Cl})} \left(\Lambda_n^1 \right) \right] \\
&= (D+1)^2 \text{tr} \left\{ \sigma \otimes O_0 \otimes \sigma \otimes O_0 \left[\sum_\lambda \Theta_1(\lambda) QP_\lambda P_\lambda + \Theta_2(\lambda) QP_{r'} P_\lambda + \Theta_3(\lambda) P_r P_\lambda + \Theta_4(\lambda) P_{r'} P_\lambda \right] \right\} \\
&\leq (D+1)^2 \left[\sum_\lambda \Theta_1(\lambda) + \Theta_2(\lambda) \right] D^{-1} \text{tr}(O_0^2) + (D+1)^2 \left[\sum_\lambda \Theta_3(\lambda) + |\Theta_4(\lambda)| \right] \mathcal{O}(1) \text{tr}(O_0^2) \\
&= \mathcal{O}(1) \text{tr}(O_0^2).
\end{aligned} \tag{64}$$

Here in the last but one line, we apply Lemma 6 for the first two terms, and the last two terms can be bounded by following Eq. (48) in the Haar case. Combining Eq. (54) and (64), we finally prove Proposition 2 in main text.

F Numerical simulations and practical time-cost analysis

F.1 Details of numerical simulations

In this section, we clarify the technique and details of the numerical simulations for the scaling of statistical variance with shot-number K , setting-number M , weight w of Pauli observables, and qubit number n in Fig. 2 and 3.

We first generate the processed state ρ , which is either a GHZ state or a more general GHZ_θ state in quantum simulators. Then we perform randomized Pauli or Clifford measurements on ρ for M different random bases, each one repeating K shots, and record the measurement results and the corresponding bases. Using the numerical results, we can construct the classical shadow $\hat{\rho}$ through Eq. (1) and Eq. (2), and calculate the value of $\hat{o} = \text{tr}(O\hat{\rho})$. Here, the observables O are set as illustrated in the captions of Fig. 2 and 3. Finally, for each pair of (M, K) which generates a total estimator $\hat{\rho}$ in Eq. (2), we repeat the whole randomized measurement procedure for $T = 2000$ times to obtain the estimation of the statistical variance, with $\text{Var}[\text{tr}(O\hat{\rho})] \sim \frac{1}{T} \sum_{t=1}^T [\text{tr}(O\hat{\rho}_t) - \text{tr}(O\rho)]^2$.

The numerical simulation results of random Pauli and Clifford measurements in Fig. 2 and 3 show consistency with the analytical results given by Eq. (13), (14), and (19).

F.2 Practical time-cost and accuracy analysis for random Pauli measurements

Here, we illustrate the advantage of the introduced multi-shot protocol, by investigating the variance and time cost enhancements against the original shadow protocol for fixed measurement budgets and accuracy requirements.

Generally, the total experimental time cost $\mathcal{C}_{tot}$ depends on both the time to change measurement settings $\mathcal{C}_M$ and the time to perform single-shot measurements $\mathcal{C}_K$. We assume, similar as in Ref. [14], $\mathcal{C}_M = 1000$ and $\mathcal{C}_K = 1$, which indicates that changing the measurement settings requires more time than performing measurements in a fixed basis. Thus, the measurement budget is approximately $\mathcal{C}_{tot} = 1000M + MK$. As indicated in Eq. (13), the variance to measure a Pauli operator P using random Pauli measurements depends on $\{M, K, w, \text{tr}(P\rho)\}$. Here, we directly set $w = 2$, $\text{tr}(P\rho) = 0.2$ and explore the scaling of the variance and total time cost with the shot-number K and setting-number M , as shown in Fig. 4.

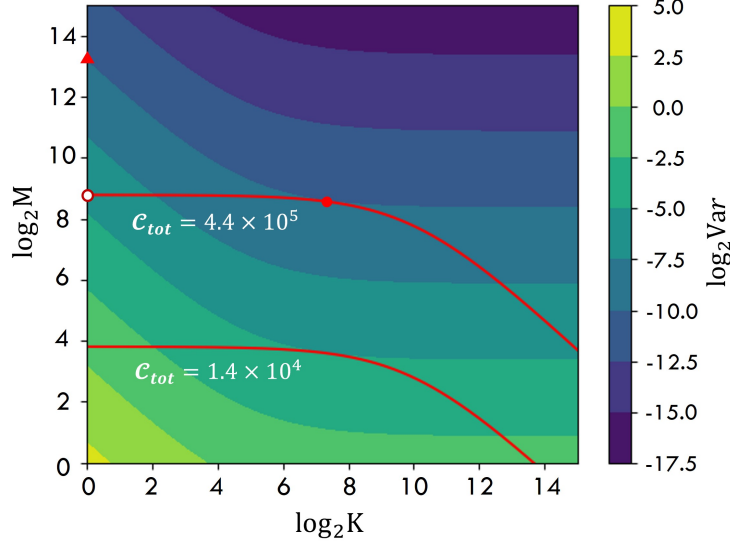


Figure 4: Scaling of the analytical variance of random Pauli measurements and total time cost with M and K . The variance (colored area) is given by Eq. (13) with $w = 2$ and $\text{tr}(P\rho) = 0.2$. The red lines show the contours of fixed $\mathcal{C}_{tot} = 1000M + MK$ with different measurement budgets $\mathcal{C}_{tot}$. The red dot represents the optimal choice of (M, K) for multi-shot protocol when $\mathcal{C}_{tot} = 4.4 \times 10^5$. The red ring on the vertical axis (i.e., $K = 1$) is the choice of M for the original shadow protocol using the same measurement budget, as that of the red dot. The red triangle on the vertical axis labels the corresponding choice of M with the same variance as the red dot.

Consider a fixed measurement budget $\mathcal{C}_{tot} = 4.4 \times 10^5$, we find the optimal choice of (M, K) using our multi-shot protocol. It provides the lowest variance for estimating $\text{tr}(P\rho)$, that is, $\text{Var}_{ms} = 0.001$, with the optimal choice approximately $M = 380$ and $K = 160$. Under the same budget, as shown in Fig. 4, the variance using the original shadow protocol Var_{os} is much larger than using the multi-shot method Var_{ms} , i.e., $\text{Var}_{os} = 0.011 > \text{Var}_{ms} = 0.001$. In the meantime, if one fixes the variance requirement as $\text{Var} = 0.001$, the total experimental time cost of the original shadow is $\mathcal{C}'_{tot} = 9.15 \times 10^6$ marked by the point of red triangle in Fig. 4, which is much larger than the one using our multi-shot technique. Therefore, the results of the example above show advantages of our multi-shot shadow protocol against the original shadow from both the measurement budgets and the accuracy aspects.