

## Robust Shadow Estimation


Senrui Chen<sup>1,2,3,†</sup>, Wenjun Yu<sup>1,†</sup>, Pei Zeng<sup>1,3,\*</sup> and Steven T. Flammia<sup>4</sup>

<sup>1</sup>Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China

<sup>2</sup>Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

<sup>3</sup>Pritzker School of Molecular Engineering, The University of Chicago, Illinois 60637, USA

<sup>4</sup>AWS Center for Quantum Computing, Pasadena, California 91125, USA

 (Received 25 December 2020; accepted 23 August 2021; published 22 September 2021)

Efficiently estimating properties of large and strongly coupled quantum systems is a central focus in many-body physics and quantum information theory. While quantum computers promise speedups for many of these tasks, near-term devices are prone to noise that will generally reduce the accuracy of such estimates. Here, we propose a sample-efficient and noise-resilient protocol for learning properties of quantum states building on the shadow estimation scheme [Huang *et al.*, Nature Physics 16, 1050–1057 (2020)]. By introducing an experimentally friendly calibration procedure, our protocol can efficiently characterize and mitigate noises in the shadow estimation scheme, given only minimal assumptions on the experimental conditions. When the strength of noises can be bounded, our protocol approximately retains the same order of sample efficiency as the standard shadow estimation scheme, while also possessing a provable noise resilience. We give rigorous bounds on the sample complexity of our protocol and demonstrate its performance with several numerical experiments, including estimations of quantum fidelity, correlation functions and energy expectations, etc., which highlight a wide spectrum of potential applications of our protocol on near-term devices.

DOI: [10.1103/PRXQuantum.2.030348](https://doi.org/10.1103/PRXQuantum.2.030348)

### I. INTRODUCTION

We are in the process of building large-scale and controllable quantum systems. This not only provides new insights and tool kits for fundamental research in quantum many-body systems [1] and the quantum nature of spacetime [2], but also yields fruitful applications in computing [3–6], communication [7–9], and sensing [10,11]. Learning the properties, e.g., fidelity [12,13], entanglement [14,15], and energy [16], of generated quantum states is usually a major step in many quantum benchmarking protocols and quantum algorithms. Among various figures of merit, robustness and efficiency are two key factors to assess the practicality of any property learning protocol.

In the noisy intermediate-scale quantum (NISQ) era [17], quantum circuits inevitably suffer from noise. The robustness of a property learning protocol then refers to

the ability to tolerate such noise. In a typical property estimation process, we generate several identical copies of the target quantum states, and then measure them using some devices that might be noisy and uncharacterized. To verify the property estimates, one has to introduce new benchmarking devices, which (in the NISQ era) will also be noisy. Consequently, we will be trapped into a loop of benchmarking. To get rid of this, at least two approaches have been proposed. One is to introduce extra assumptions on the noise model, in which case we might be able to mitigate the error [18–21], but such assumptions may not be verifiable. The other is to use device-independent protocols [22–24] that do not have any assumptions on the devices, but such protocols are mostly designed for some specific property learning tasks (e.g., entanglement detection), and their requirements on devices and computational and sample complexity can be too strict to produce anything informative in practice.

Thus, while property learning and testing leads to large efficiency gains in sample and computational complexity, one must in general have a well-characterized device for these methods to be applicable. Quantum tomography [25,26] is a standard method to extract complete characterization information, but it requires exponentially many samples with respect to the number of qubits. Several

\*qubitpei@gmail.com

†These authors contributed equally to this work.

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

efficient tomographic schemes have been proposed based on some properties of the prepared states, such as the low rank property [27,28], permutation symmetry [29,30], and the locality of Schmidt decomposition [31,32]. Nevertheless, such assumptions are restrictive and not applicable in many cases. Another line of research focuses on efficiently extracting partial information of a quantum state without any prior knowledge. An example is the quantum overlapping tomography [33,34] that can simultaneously estimate all  $k$ -qubit reduced density matrices of an arbitrary quantum state in a sample-efficient manner for small  $k$ . The simplest version of this idea is to uniformly measure random Pauli strings [35], which leads to a sample complexity of  $O(k3^k \log n)$  for estimating all  $k$ -body Pauli observables to fixed precision. Machine-learning-based approaches are also proposed [36] in this direction.

Recently, a new paradigm for efficient and universal quantum property estimation has been proposed named quantum shadow estimation. Shadow estimation was first put forward in Ref. [37]. Roughly speaking, this scheme can simultaneously estimate the expectation values with respect to  $N$  observables of an unknown  $d$ -dimensional quantum state with a sample size of order  $\log d \log N$ , that is usually more efficient than either conducting full tomography or measuring the  $N$  observables one by one. Later on, a more experimentally friendly shadow estimation scheme was proposed [38], which is able to estimate many useful properties of a quantum system with a small number of samples (see also Ref. [39]). This protocol is also proven to be worst-case sample optimal in the sense that any other protocol that is able to accurately estimate any collection of arbitrary observables must consume a number of samples at least comparable to this one. Although promising for a broad spectrum of applications, the shadow estimation scheme in Ref. [38] (as well as the random Pauli scheme from Ref. [35]) assumes perfect implementation of a group of unitary gates as well as an ideal projective measurement on the computational basis. It remains unclear how experimental noise can affect the performance of this scheme.

In this work, we reexamine the shadow estimation scheme and regard it as a twirling and retrieval procedure of the measurement channel. In this way, we extend shadow estimation to the case when the unitaries and measurements are noisy. With similar techniques used in the study of randomized benchmarking [40–44], we propose a modified shadow estimation strategy that is noise resilient. When the noise in the unitary operations and measurements is small, the robust shadow tomography scheme is able to faithfully estimate the required properties with a small additional cost, subject only to the assumption that one can prepare the initial ground state  $|0\rangle^{\otimes n}$  with high fidelity. The proposed scheme is both robust and efficient, and hence highly practical for property estimation of a quantum system.

## II. PRELIMINARIES

We first introduce the Pauli-transfer matrix (PTM) representation (or Liouville representation) to simplify the notation. Note that all the linear operators  $\mathcal{L}(\mathcal{H}_d)$  on the underlying  $n$ -qubit Hilbert space  $\mathcal{H}_d$  with  $d = 2^n$  can be vectorized using the  $n$ -qubit (normalized) Pauli operator basis  $\{\sigma_a := P_a/\sqrt{d}\}_a$ , where the  $P_a$  are the usual Pauli matrices. For a linear operator  $A \in \mathcal{L}(\mathcal{H}_d)$ , we define a column vector  $|A\rangle \in \mathcal{H}_{d^2}$  with the  $a$ th entry given by  $|A\rangle_a = \text{Tr}(P_a A)/\sqrt{d}$ . The inner product on the vector space  $\mathcal{H}_{d^2}$  is defined by the *Hilbert-Schmidt inner product* as  $\langle\langle A|B\rangle\rangle := \text{Tr}(A^\dagger B)$ . The normalized Pauli basis  $\{\sigma_a\}_a$  is then an orthonormal basis in  $\mathcal{H}_{d^2}$ . Superoperators on  $\mathcal{H}_d$  are linear maps taking operators to operators  $\mathcal{L}(\mathcal{H}_d) \rightarrow \mathcal{L}(\mathcal{H}_d)$ . In the vector space  $\mathcal{H}_{d^2}$ , a superoperator  $\mathcal{E}$  can be represented by a matrix in the Pauli basis, with the entries given by  $\mathcal{E}_{ab} = \langle\langle \sigma_a | \mathcal{E}(\sigma_b) \rangle\rangle = \langle\langle \sigma_a | \mathcal{E} | \sigma_b \rangle\rangle$ . The matrix form of the superoperator with the Pauli basis is sometimes called the Pauli transfer matrix. With a slight abuse of notation, we sometimes denote a superoperator and its PTM using the same notation. A detailed introduction to the PTM is given in Appendix A 3.

In this work, we focus on the task of estimating the expectation values  $\{\text{Tr}(O_i \rho)\}_i$  of a set of observables  $\{O_i\}_i$  on an underlying unknown quantum state  $\rho$ ,

$$\text{Tr}(O_i \rho) = \langle\langle O_i | \rho \rangle\rangle, \quad 1 \leq i \leq N. \quad (1)$$

When the number of observables  $N$  is large, a direct exhaustive measurement of the (generally incompatible) observables  $\{O_i\}$  on  $\rho$  is expensive. Besides, in many cases we may want to perform tomographic experiments on  $\rho$  before deciding which observables  $\{O_i\}$  should be estimated. To realize this, a natural idea is to insert an extra prepare-and-measure superoperator between  $\langle\langle O_i |$  and  $|\rho\rangle\rangle$ ,

$$\langle\langle O_i | \rho \rangle\rangle \rightarrow \sum_x \langle\langle O_i | A_x \rangle\rangle \langle\langle E_x | \rho \rangle\rangle. \quad (2)$$

In an experiment, we first apply a positive operator-valued measure (POVM) measurement  $\{E_x\}_x$  at  $\rho$ . Then, conditioned on the outcome  $x$ , we calculate  $\langle\langle O_i | A_x \rangle\rangle$  via classical postprocessing. If we repeat this procedure then the sample average over these experiments gives an estimator for  $\langle\langle O_i | \rho \rangle\rangle$ . As long as the inserted superoperator  $\sum_x |A_x\rangle\rangle \langle\langle E_x |$  equals  $\mathcal{I}$ , this estimator will be unbiased.

To construct a realization of such a superoperator, we consider the dephasing channel in the computational basis ( $Z$  basis)  $\mathcal{M}_Z := \sum_z |z\rangle\rangle \langle\langle z |$ , where  $|z\rangle\rangle$  is the vectorization of the  $Z$ -basis eigenstate  $|z\rangle \langle z|$ , with  $z \in \{0, 1\}^{\otimes n}$ . Expanding  $\mathcal{M}_Z$  in the Pauli operator basis  $\{|\sigma_0\rangle\rangle, |\sigma_x\rangle\rangle, |\sigma_y\rangle\rangle, |\sigma_z\rangle\rangle\}^{\otimes n}$ , we have

$$\begin{aligned} \mathcal{M}_Z &= (|\sigma_0\rangle\rangle \langle\langle \sigma_0 | + |\sigma_z\rangle\rangle \langle\langle \sigma_z |)^{\otimes n} \\ &= [\text{diag}(1, 0, 0, 1)]^{\otimes n}, \end{aligned} \quad (3)$$

where  $\text{diag}(a_1, a_2, \dots)$  is a diagonal matrix with the diagonal elements  $a_1, a_2, \dots$ . If  $\mathcal{M}_Z$  were invertible, we could insert the superoperator  $\mathcal{M}_Z^{-1}\mathcal{M}_Z = \sum_z |\mathcal{M}_Z^{-1}(z)\rangle\langle z| = \mathcal{I}$ . However,  $\mathcal{M}_Z$  is not invertible due to the lack of  $X, Y$ -basis information in a  $Z$ -basis measurement. To make  $\mathcal{M}_Z$  invertible, we can introduce an extra unitary twirling [38],

$$\mathcal{M} = \mathbb{E}_{U \in \mathbb{G}} \mathcal{U}^\dagger \mathcal{M}_Z \mathcal{U}. \quad (4)$$

Here,  $\mathbb{G}$  is a subset of the unitaries  $\{U\}$  in  $U(d)$  to be specified later, and  $\mathcal{U}$  is the PTM representation of  $U$ .

When  $\mathbb{G}$  forms a group, the PTMs  $\{\mathcal{U}\}$  forms a representation of  $\mathbb{G}$ . A direct application of Schur's lemma [45] (see Appendix A 1) allows us to calculate the explicit form of  $\mathcal{M}$ ,

$$\mathcal{M} = \sum_{\lambda \in R_{\mathbb{G}}} \frac{\text{Tr}[\mathcal{M}_Z \Pi_\lambda]}{\text{Tr}[\Pi_\lambda]} \Pi_\lambda, \quad (5)$$

where  $R_{\mathbb{G}}$  is the set of irreducible subrepresentations of the group  $\mathbb{G}$ , and  $\Pi_\lambda$  is the corresponding projector onto the invariant subspace. Since the projectors are complete and orthogonal to each other,  $\mathcal{M}$  is invertible if and only if all the coefficients are nonzero. Therefore, the twirling group  $\mathbb{G}$  needs to satisfy

$$\text{Tr}[\mathcal{M}_Z \Pi_\lambda] \neq 0 \quad \text{for all } \lambda \in R_{\mathbb{G}}. \quad (6)$$

Once Eq. (6) is satisfied, we can construct a shadow estimation protocol based on the equation

$$\langle\langle O_i | \rho \rangle\rangle = \mathbb{E}_{U \in \mathbb{G}} \sum_{z \in \{0,1\}^{\otimes n}} \langle\langle O_i | \mathcal{M}^{-1} \mathcal{U}^\dagger | z \rangle\rangle \langle\langle z | \mathcal{U} | \rho \rangle\rangle. \quad (7)$$

To implement shadow estimation, one can repeat the following experiment: generate a single copy of  $\rho$ , act via a randomly sampled unitary  $U$ , and then perform a  $Z$ -basis measurement to return an output bit string  $b$ . Then  $\langle\langle O_i | \mathcal{M}^{-1} \mathcal{U}^\dagger | b \rangle\rangle$  is calculated on a classical computer. Thanks to this decoupled processing of  $\rho$  with respect to  $O_i$ , the estimation of different observables can be done in parallel with a relatively small increase in sample complexity.

The quantum shadow estimation procedure can be summarized as in Algorithm 1.

We refer to  $\hat{o}_i^{(r)}$  as the single-round estimator. The subroutine **MedianOfMeans** divides the  $R = NK$  single-round estimators into  $K$  groups, calculates the mean value of each group, and takes the median of these mean values

---

**Input:** Unknown  $n$ -qubit quantum state  $\rho$ , observables  $\{O_i\}_{i=1}^M$ ,  $\mathbb{G} \subseteq U(2^n)$ , quantum channel  $\mathcal{M}$ , and  $N, K \in \mathbb{N}_+$ .

**Output:** A set of estimates  $\{\hat{o}_i\}$  of  $\{\text{Tr}(\rho O_i)\}$ .

```

1:  $R := NK$ .
2: for  $r = 1$  to  $R$  do
3:   Prepare  $\rho$ , uniformly sample  $U \in \mathbb{G}$  and apply to  $\rho$ .
4:   Measure in the computational basis, outcome  $|b\rangle$ .
5:    $\hat{o}_i^{(r)} := \langle\langle O_i | \mathcal{M}^{-1} \mathcal{U}^\dagger | b \rangle\rangle, \forall i$ .
6: end for
7:  $\hat{o}_i := \text{MedianOfMeans}(\{\hat{o}_i^{(r)}\}_{r=1}^R, N, K), \forall i$ .
8: return  $\{\hat{o}_i\}$ .
```

---

Algorithm 1. Shadow estimation (**Shadow**) [38].

as the final estimator. As a formula,

$$\begin{aligned} \bar{o}_i^{(k)} &:= \frac{1}{N} \sum_{r=(k-1)N+1}^{kN} \hat{o}_i^{(r)}, \quad k = 1, 2, \dots, K, \\ \hat{o}_i &:= \text{median}\{\bar{o}_i^{(1)}, \bar{o}_i^{(2)}, \dots, \bar{o}_i^{(K)}\}. \end{aligned} \quad (8)$$

For the standard shadow estimation algorithm [38], the input quantum channel  $\mathcal{M}$  is decided by Eq. (4).

### III. ROBUST SHADOW ESTIMATION

In practice, the unitary operations and measurements used in the standard shadow estimation algorithm will be noisy. We want to mitigate the effect of this noise on the output estimate of the shadow. Our strategy to do this is simple: we first learn the noise as a simple stochastic model and then compensate for these errors using robust classical postprocessing.

In general, noise in quantum devices is not stochastic, and coherent errors must be addressed. However, thanks to the unitary twirling in shadow estimation, the stochastic nature of the noise is inherent to the protocol itself. For example, any noise map that is twirled over a Clifford group that contains the Pauli group as a subgroup will reduce the noise to a purely stochastic Pauli channel [46]. The complete characterization of such noise channels can be efficiently and accurately performed [47–49]. It is then straightforward to compensate for such errors by modifying the classical postprocessing, although a lengthy analysis is required to show the efficacy of this strategy.

In order to pursue a rigorous analysis of this strategy, we make the following two assumptions on the noise in the experimental device implementing the shadow estimation.

**Assumptions 1** (Simplifying noise assumptions).

- A1** *The noise in the circuit is gate-independent, time-stationary, Markovian noise.*
- A2** *The experimental device can generate the computational basis state  $|0\rangle \equiv |0\rangle^{\otimes n}$  with sufficiently high fidelity.*

Our first assumption is used throughout to ensure that there exists a completely positive trace-preserving (CPTP) map such that the noisy gate  $\tilde{U}$  can be decomposed into  $\Lambda\mathcal{U}$ , where  $\mathcal{U}$  is the ideal gate while  $\Lambda$  is the noise channel. The noise map  $\Lambda$  is independent of the unitary  $\mathcal{U}$  and the time  $t$ . It also implies that the noise map occurring in the measurement is fixed independent of time and hence can be absorbed into  $\Lambda$ . We remark that assumption **A1** is widely used in the analysis of randomized benchmarking protocols. The gate-independent part of the assumption is especially appropriate when the experimental unitaries are single-qubit gates, but it has been shown that the effect of weak gate dependence (a form of non-Markovianity) generally leads to weak perturbations [50,51]. We also provide numerical evidence in Sec. VIII showing that our scheme is still quite robust against realistic gate-dependent noise models in experiment.

For our second assumption, **A2**, from Sec. III to Sec. V we initially make the stronger assumption that the experimental device can prepare the  $|0\rangle$  state *exactly*. In Sec. VI we relax this to show that, when  $|0\rangle$  is not precisely prepared, but is prepared with sufficiently high fidelity, our protocol still gives a good estimation. Fortunately, the computational basis state  $|0\rangle$  is relatively easy to generate faithfully in many experimental platforms.

To see how unitary twirling helps to reduce the number of noise parameters, we calculate the noisy version of the random measurement channel  $\tilde{\mathcal{M}}$ ,

$$\begin{aligned}\tilde{\mathcal{M}} &= \mathbb{E}_{U \in \mathbb{G}} U^\dagger \mathcal{M}_Z \Lambda U \\ &= \sum_{\lambda \in R_{\mathbb{G}}} \frac{\text{Tr}[\mathcal{M}_Z \Lambda \Pi_\lambda]}{\text{Tr}[\Pi_\lambda]} \Pi_\lambda \\ &= \sum_{\lambda} f_\lambda \Pi_\lambda,\end{aligned}\quad (9)$$

where the  $\{f_\lambda\}$  are expansion coefficients of the twirled channel. Note that channel  $\Lambda$  describes both the noise in gate  $\mathcal{U}$  and in the measurement  $\mathcal{M}_Z$ , which is always possible under our assumption **A1**. The number of  $\{f_\lambda\}$  is related to the number of irreducible representations in the PTM representation of the twirling group  $\mathbb{G}$ . Later we show that the coefficients  $\{f_\lambda\}$  can be estimated in parallel, similar to the normal shadow estimation procedure (referred to as the *calibration procedure*).

Based on the observations above, we propose our robust quantum shadow estimation (**RShadow**) protocol to faithfully estimate  $\{\text{Tr}(O_i \rho)\}_i$  even with noise. The algorithm is depicted in Fig. 1 and it works as follows. We first estimate the noise channel  $\tilde{\mathcal{M}}$  of Eq. (9) with the *calibration procedure*, and then use the estimator  $\tilde{\mathcal{M}}$  as the input parameter  $\mathcal{M}$  of Algorithm 1 to predict any properties of interest (referred to as the *estimation procedure*). The procedure

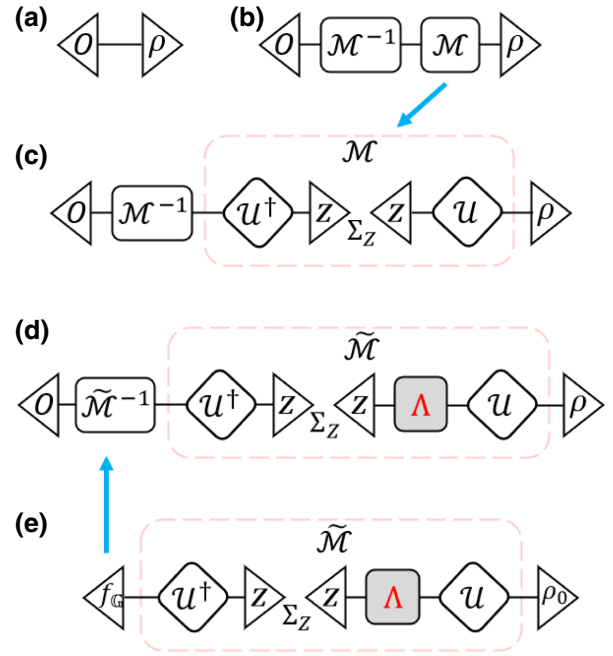


FIG. 1. Diagram of the shadow estimation protocol. (a) We want to estimate the expectation value  $\text{Tr}(O\rho) = \langle\langle O|\rho\rangle\rangle$  for a set of observables  $\{O_i\}$  and an unknown state  $\rho$ . (b) To do this, we insert a channel  $\mathcal{M}$  and its corresponding inverse map  $\mathcal{M}^{-1}$  in the middle, which will not change the expectation value. (c) The channel  $\mathcal{M}$  can be realized as a random unitary twirling  $\mathbb{E}_U U^\dagger \cdot U$  acting on the  $Z$ -basis dephasing map  $\mathcal{M}_Z = \sum_z |z\rangle\langle z|$ . (d) In practice, the implemented unitary  $\mathcal{U}$  and the measurement  $\langle\langle z|$  are noisy, causing an extra uncharacterized noise channel  $\Lambda$ . (e) In practice, the unitary  $\mathcal{U}$  and the measurement  $\langle\langle z|$  suffer from a noise channel  $\Lambda$ , causing an uncharacterized channel  $\tilde{\mathcal{M}}$  that needs to be inverted. (e) The calibration procedure of **RShadow**. By experimenting on some well-characterized state  $\rho_0$ , we can estimate the channel  $\tilde{\mathcal{M}}$  and its inverse, and hence mitigate the noise in the shadow estimation procedure. Here  $f_{\mathbb{G}}$  is the **NoiseEst $_{\mathbb{G}}$**  subroutine described in Algorithm 2.

is shown in Algorithm 2, where the subroutine **NoiseEst** is decided by  $\mathbb{G}$  and is given later.

---

**Input:** Unknown  $n$ -qubit quantum state  $\rho$ , observables  $\{O_i\}_{i=1}^M$ ,  $\mathbb{G} \subseteq U(2^n)$  and  $N_1, N_2, K_1, K_2 \in \mathbb{N}_+$ .  
**Output:** A set of estimations  $\{\hat{o}_i\}$  of  $\{\text{Tr}(\rho O_i)\}$ .

- 1:  $R := N_1 K_1$ .  $\triangleright$  Calibration
- 2: **for**  $r = 1$  **to**  $R$  **do**
- 3:   Prepare  $|0\rangle$ , sample (noisy)  $U \in \mathbb{G}$  and apply to  $|0\rangle$ .
- 4:   Measure in the computational basis, return  $|b\rangle$ .
- 5:    $\hat{f}_\lambda^{(r)} := \text{NoiseEst}_{\mathbb{G}}(\lambda, U, b)$ ,  $\forall \lambda \in R_{\mathbb{G}}$ .
- 6: **end for**
- 7:  $\hat{f}_\lambda := \text{MedianOfMeans}(\{\hat{f}_\lambda^{(r)}\}_{r=1}^R, N_1, K_1)$ ,  $\forall \lambda \in R_{\mathbb{G}}$ .
- 8:  $\hat{\mathcal{M}} := \sum_{\lambda \in R_{\mathbb{G}}} \hat{f}_\lambda \Pi_\lambda$ .
- 9:  $\{\hat{o}_i\} = \text{Shadow}(\rho, \{O_i\}, \mathbb{G}, \hat{\mathcal{M}}, N_2, K_2)$ .  $\triangleright$  Estimation
- 10: **return**  $\{\hat{o}_i\}$ .

---

Algorithm 2. Robust shadow estimation (**RShadow**).

In the following discussion, we focus on two specific groups  $\mathbb{G}$ : the  $n$ -qubit Clifford group  $\text{Cl}(2^n)$  and the  $n$ -fold tensor product of the single-qubit Clifford group  $\text{Cl}_2^{\otimes n}$ . We give a specific construction of the **NoiseEst** subroutine and show the correctness and efficiency of our **RShadow** algorithm with these two groups.

#### IV. ROBUST SHADOW ESTIMATION USING THE GLOBAL CLIFFORD GROUP

We first present a robust shadow estimation protocol using the  $n$ -qubit global Clifford group,  $\text{Cl}(2^n)$ . The  $n$ -qubit Clifford group has many useful properties, such as being a unitary 3-design [52–54], which is widely used in many tasks of quantum information and quantum computation. It is a standard result that the  $n$ -qubit Clifford group has two irreducible representations in the Liouville representation whose projectors are given by  $|\sigma_0\rangle\rangle\langle\langle\sigma_0|$  and  $I - |\sigma_0\rangle\rangle\langle\langle\sigma_0|$ . Assuming that the  $\tilde{M}$  channel defined in Eq. (9) is trace preserving, it can be written as

$$\tilde{M} = \mathbb{E}_{U \sim \text{Cl}(2^n)} U^\dagger \mathcal{M}_Z \Lambda U = |\sigma_0\rangle\rangle\langle\langle\sigma_0| + f(I - |\sigma_0\rangle\rangle\langle\langle\sigma_0|) \quad (10)$$

for some  $f \in \mathbb{R}$ , i.e., as a depolarizing channel. It is easy to obtain  $f = 1/(2^n + 1)$  for the noiseless case using Eq. (9). The noise characterization subroutine with  $\text{Cl}(2^n)$  is defined as

$$\text{NoiseEst}_{\text{Cl}(2^n)}(U, b) := \frac{2^n \langle\langle b | \mathcal{U} | \mathbf{0} \rangle\rangle - 1}{2^n - 1}, \quad (11)$$

where  $|b\rangle\rangle$  is the Liouville representation of the computational basis state  $|b\rangle$   $\langle b|$  and similarly for  $|\mathbf{0}\rangle\rangle$ .

Next, define the  $Z$ -basis average fidelity of a noise channel  $\Lambda$  as  $F_Z(\Lambda) = (1/2^n) \sum_{b \in \{0,1\}^n} \langle\langle b | \Lambda | b \rangle\rangle$ . The following theorem demonstrates the correctness and sample efficiency of our protocol. We remark that the validity of this theorem relies on Assumptions 1.

**Theorem 1** (Informal). *For **RShadow** with  $\mathbb{G} = \text{Cl}(2^n)$ , if the number of samples for the calibration procedure satisfies*

$$R = \tilde{O}(\varepsilon^{-2} F_Z^{-2}), \quad (12)$$

where  $F_Z \equiv F_Z(\Lambda)$  and we assume that  $F_Z \gg 2^{-n}$ , then the subsequent estimation procedure with high probability satisfies

$$|\mathbb{E}(\hat{\delta}^{(r)}) - \text{Tr}(O\rho)| \leq \varepsilon \|O\|_\infty \quad (13)$$

for any observable  $O$  and quantum state  $\rho$ , where  $\hat{\delta}^{(r)}$  is the single-round estimator defined as in Algorithm 1.

Here and throughout the paper, we use  $\tilde{O}$  to represent the big-O notation with polylogarithmic factors suppressed. The more rigorous version of Theorem 1 is Theorem 7 in Appendix B. We see that our protocol indeed eliminates the *systematic error* of shadow estimation in a sample-efficient manner, since without the calibration step the empirical expectation value would converge to a value that conflated the noise map  $\Lambda$  into the estimate, whereas  $\Lambda$  does not appear in Eq. (13). More specifically, if the  $Z$ -basis average fidelity of the noise channel  $\Lambda$  is lower bounded by some constant (e.g., constant-strength depolarizing noise), then the sample complexity of our calibration stage is approximately independent of the system size  $n$ .

A more realistic noise model to consider is that of local noise with fixed strength, where  $\Lambda := \bigotimes_{i=1}^n \Lambda_i$  and each single-qubit noise channel  $\Lambda_i$  satisfies  $F_Z(\Lambda_i) \geq 1 - \xi$ . In that case, we have  $F_Z(\Lambda)^{-2} \approx \exp(2n\xi)$  for small  $\xi$ , so we can efficiently deal with a system size  $n$  that is comparable to  $\xi^{-1}$ .

Next, we consider the sample complexity of the estimation procedure. Following a similar methodology of bounding the sample complexity in the noise-free standard shadow estimation scheme [38], we bound the sample complexity of our **RShadow** estimation procedure as follows.

**Theorem 2** (Informal). *For **RShadow** with  $\text{Cl}(2^n)$ , if the number of calibration samples  $R_C$  and the number of estimation samples  $R_E$  satisfy*

$$\begin{aligned} R_C &= \tilde{O}(\varepsilon_1^{-2} F_Z^{-2}), \\ R_E &= \tilde{O}(\varepsilon_2^{-2} F_Z^{-2} \log M), \end{aligned} \quad (14)$$

respectively, then the protocol can estimate  $M$  arbitrary linear functions  $\text{Tr}(O_1\rho), \dots, \text{Tr}(O_M\rho)$  such that  $\max_i \text{Tr}(O_i^2) \leq 1$ , up to accuracy  $\varepsilon_1 + \varepsilon_2$  with high success probability.

The rigorous version of Theorem 2 is Theorem 8 in Appendix B. Compared with results in Ref. [38], one can see that the **RShadow** scheme has nearly the same sample complexity order as the noise-free standard shadow estimation methods in a low-noise regime.

Finally, we comment on the computational complexity of **RShadow**. The computational complexity of our calibration procedure is favorable since the single-round fidelity estimator can be calculated efficiently with the Gottesman-Knill theorem [55,56]. However, an efficient computation using the Gottesman-Knill theorem for the estimation procedure would require the observable  $O$  to have additional structure, such as being a stabilizer state or being a Pauli operator. The standard shadow estimation scheme of Ref. [38] or the fast Pauli expectation estimation method of Ref. [35] also have such a requirement.

### V. ROBUST SHADOW ESTIMATION USING THE LOCAL CLIFFORD GROUP

Despite the useful properties that the global Clifford group possesses, it is often challenging to implement the full  $n$ -qubit Clifford group under current experimental conditions. The local Clifford group  $\text{Cl}_2^{\otimes n}$ , which is the  $n$ -fold tensor product of the single-qubit Clifford group, is an experimentally friendly alternative. We now present a robust shadow estimation protocol based on the local Clifford group that can efficiently calibrate and mitigate the error in estimating any *local property*.

It is known that the  $n$ -qubit local Clifford group has  $2^n$  irreducible representations [57]. Being twirled by the local Clifford group, the channel  $\tilde{\mathcal{M}}$  becomes a Pauli channel that is symmetric among the  $x, y, z$  indices, and the Pauli-Liouville representation is

$$\tilde{\mathcal{M}} = \mathbb{E}_{U \sim \text{Cl}_2^{\otimes n}} U^\dagger \mathcal{M}_Z \Lambda U = \sum_{z \in \{0,1\}^n} f_z \Pi_z, \quad (15)$$

where  $\Pi_z = \bigotimes_{i=1}^n \Pi_{z_i}$ ,

$$\Pi_{z_i} = \begin{cases} |\sigma_0\rangle\langle\sigma_0|, & z_i = 0, \\ I - |\sigma_0\rangle\langle\sigma_0|, & z_i = 1, \end{cases}$$

for  $f_z \in \mathbb{R}$ , which is called the Pauli fidelity. Here, for any string  $m \in \{0, 1\}^n$ , we define  $|m\rangle$  to be the Liouville representation of the computational basis state  $|m\rangle \langle m|$ , and define  $P_m := \bigotimes_{i=1}^n P_Z^{m_i}$  and the  $\sigma_m$  to be the corresponding normalized Pauli operators. In the noiseless case, one can obtain  $f_z = 3^{-|z|}$  using Eq. (9), where  $|z|$  is the number of 1s in  $z$ .

The noise characterization subroutine with  $\text{Cl}_2^{\otimes n}$  is defined as

$$\text{NoiseEst}_{\text{Cl}_2^{\otimes n}}(z, U, b) := \langle\langle b|U|P_z\rangle\rangle \quad \text{for all } z \in \{0, 1\}^n. \quad (16)$$

In the standard shadow estimation using  $\text{Cl}_2^{\otimes n}$  [38] (and in the earlier work [35]), one can only efficiently estimate observables with small Pauli weight. An  $n$ -qubit observable  $O$  is called  $k$  local if it can be written as  $O = \tilde{O}_S \otimes I_{[n] \setminus S}$  for some  $k$ -element index set  $S \subset [n]$  and a  $k$ -qubit observable  $\tilde{O}$ . Similarly, our **RShadow** protocol with  $\text{Cl}_2^{\otimes n}$  is also designed for predicting  $k$ -local observables. The correctness and efficiency is given by the following theorem.

**Theorem 3 (Informal).** *For RShadow with  $\text{Cl}_2^{\otimes n}$ , if the number of samples for the calibration procedure satisfies*

$$R = \tilde{O}(3^k \varepsilon^{-2} F_Z^{-2}) \quad (17)$$

*then the subsequent estimation procedure with high probability satisfies*

$$|\mathbb{E}(\hat{o}^{(r)}) - \text{Tr}(O\rho)| \leq \varepsilon 2^k \|O\|_\infty \quad (18)$$

*for any  $k$ -local observable  $O$  and quantum state  $\rho$ , where  $\hat{o}^{(r)}$  is the single-round estimator defined as in Algorithm 1.*

The rigorous version of Theorem 3 is Theorem 9 in Appendix C. Indeed, this protocol can calibrate the shadow estimation process for all  $k$ -local observables using a small number of samples that depends only on  $k$  (but basically not on the system size  $n$ ). Note that Theorem 3 holds for any gate-independent noise model, even for *global unitary noise*.

Now we investigate the sample complexity of the estimation procedure. We are currently unable to bound the sample complexity against the most general noise channel, but we do have a bound for a local noise model, as shown in the following theorem.

**Theorem 4 (Informal).** *For RShadow with  $\text{Cl}_2^{\otimes n}$ , suppose that the noise is local, i.e.,  $\Lambda := \bigotimes_{i=1}^n \Lambda_i$ , and satisfies  $F_Z(\Lambda_i) \geq 1 - \xi$  for all  $i$  and some  $\xi \ll \frac{1}{2}$ . If the number of calibration samples  $R_C$  and the number of estimation samples  $R_E$  satisfy*

$$\begin{aligned} R_C &= \tilde{O}(12^k e^{4k\xi} \varepsilon_1^{-2}), \\ R_E &= \tilde{O}(4^k e^{4k\xi} \varepsilon_2^{-2} \log M), \end{aligned} \quad (19)$$

*respectively, then the protocol can estimate  $M$  arbitrary linear functions  $\text{Tr}(O_1\rho), \dots, \text{Tr}(O_M\rho)$  such that every  $O_i$  is  $k$  local and  $\|O_i\|_\infty \leq 1$ , up to accuracy  $\varepsilon_1 + \varepsilon_2$  with high success probability.*

The rigorous version of Theorem 4 is Theorem 10 in Appendix C. Again, we see that **RShadow** using  $\text{Cl}_2^{\otimes n}$  has a sample complexity similar to the noiseless standard shadow estimation protocol when the noise is local and not too strong. We also remark that, although we do not have a sample complexity bound against a more general noise model, our numerical results show that **RShadow** can still perform well in that case (see Appendix E). Furthermore, in realistic experiments, one can monitor the standard deviation of estimators in real time, which means that they can still suppress statistical fluctuations to an acceptable level even without a theoretical sample complexity bound.

Regarding the computational complexity, it is obviously impractical to calibrate all  $2^n$  parameters  $f_z$ . However, since we only care about  $k$ -local observables, only  $\hat{f}_z^{(r)}$  such that  $|z| \leq k$  needs to be computed, the number of which is no greater than  $n^k$ . Furthermore, note that  $\hat{f}_z^{(r)}$  can be decomposed as  $\prod_{i=1}^n \langle\langle b_i|U_i|P_Z^{z_i}\rangle\rangle$ , so all these  $\hat{f}_z^{(r)}$  can be computed within  $O(n^k)$  time using dynamic programming.

If there is extra structure of the observables to be predicted (e.g., spatially local), the number of necessary  $\hat{f}_z^{(r)}$  can be further reduced. In practice, one may store the raw data of the calibration procedure and see what observables are to be predicted, before deciding which set of  $f_z$  needs to be calculated. An example is given below in our numerical experiments. The computational complexity for the estimation procedure is therefore low when the observables are  $k$  local for reasonably small  $k$ .

## VI. ROBUSTNESS AGAINST STATE PREPARATION NOISE

In the last two sections, we proved the performance of the **RShadow** protocol based on the assumption of perfect  $|0\rangle$  preparation. Although  $|0\rangle$  is relatively easy to prepare on most current quantum computing platforms, state preparation (SP) noise is still inevitable. In this section, we show that the **RShadow** protocol is also robust against small SP noise in the following sense: when  $|0\rangle$  can be prepared with high fidelity during the calibration procedure, the estimators for the estimation procedure will not be too biased, and the sample complexity will not increase drastically.

Formally, in a realistic calibration procedure, one prepares some  $\rho_0$  instead of  $|0\rangle$  for each round. We assume that  $\rho_0$  is time independent, which is reasonable if the experimental conditions do not change much during this process. We have the following theorems.

**Theorem 5.** *For **RShadow** using  $\text{Cl}(2^n)$ , if the state preparation fidelity satisfies*

$$F(|0\rangle\langle 0|, \rho_0) \geq 1 - \varepsilon_{\text{SP}} \quad (20)$$

*then, with the same number of calibration samples as in Theorem 1, the subsequent estimation procedure with high probability satisfies*

$$|\mathbb{E}(\hat{\rho}^{(r)}) - \text{Tr}(O\rho)| \leq (\varepsilon + 2\varepsilon_{\text{SP}})\|O\|_{\infty} \quad (21)$$

*up to first orders in  $\varepsilon$  and  $\varepsilon_{\text{SP}}$ .*

**Theorem 6.** *For **RShadow** using  $\text{Cl}_2^{\otimes n}$ , if the state is prepared as a product state  $\rho_0 = \bigotimes_{i=1}^n \rho_{0,i}$  and the single-qubit state preparation fidelity satisfies*

$$F(|0\rangle\langle 0|, \rho_{0,i}) \geq 1 - \xi_{\text{SP}} \quad \text{for all } i \in [n] \quad (22)$$

*then, with the same number of calibration samples as in Theorem 3, the subsequent estimation procedure with high probability satisfies*

$$|\mathbb{E}(\hat{\rho}^{(r)}) - \text{Tr}(O\rho)| \leq (\varepsilon + 2k\xi_{\text{SP}})2^k\|O\|_{\infty} \quad (23)$$

*up to first orders in  $\varepsilon$  and  $k\xi_{\text{SP}}$ . Here  $k$  is the locality of observable  $O$ .*

The proof is given in Appendix D. The above two theorems show that the effect of SP noise can indeed be bounded for **RShadow**. They also enable an experimentalist to decide a practical sample number according to how well his device can prepare  $|0\rangle\langle 0|$ .

## VII. NUMERICAL RESULTS

Here, we design several numerical experiments to demonstrate the practicality of the robust shadow estimation (**RShadow**) protocol. We first benchmark the robustness of the **RShadow** protocol under various types of noise model in the task of estimating the fidelity of the Greenberger-Horne-Zeilinger (GHZ) state. After that, we show the application of **RShadow** in estimating the two-point correlation as well as the energy of the ground state of the antiferromagnetic transverse-field Ising model (TFIM). These tasks frequently appear in the field of quantum computational chemistry [58]. In all the numerical experiments, we assume that the states to be tested are perfectly prepared while the shadow estimation circuits are noisy. We compare the performance of the **RShadow** protocol with the standard quantum shadow estimation scheme (standard **Shadow**) [38] in all the tasks. Our numerical simulation makes use of Qiskit [59], an open-source python-based quantum information toolkit.

For the plots in this section, the error bars represent the standard deviation of the estimation procedure (which means that we ran the calibration procedure of **RShadow** only once for each data point), and are calculated via the empirical bootstrapping method [60], where we randomly sample the same size of data points *with replacement* from the original data and calculate the estimator as a bootstrap sample. Repeat this  $B = 200$  times, and take the standard deviation among these bootstrap samples as an approximation to the standard deviation of our **RShadow** estimator.

In the first experiment, we numerically prepare a ten-qubit GHZ state, and use the shadow estimation protocol to estimate its fidelity with the ideal GHZ state. Each protocol uses  $R = 10^5$  ( $N = 10^4, K = 10$ ) samples for the estimation stage, while our **RShadow** uses an extra  $R = 10^5$  ( $N = 10^4, K = 10$ ) samples for its calibration stage. We simulate the three noise models depolarizing, amplitude damping, and measurement bit flip, each with several different levels of strength. The random circuits are set to be global Clifford gates. Figure 2 shows the results. One can see that, for these three noise models, when the noise level increases, the standard shadow estimation deviates from the true value, while the robust shadow estimation remains faithful.

Note that there exist some numerical results from our robust procedure exceeding the ground truth in the above figure. That is due to the nature of the shadow protocol to eliminate the effect of noisy fidelity parameters  $\{\hat{f}_\lambda\}$ .

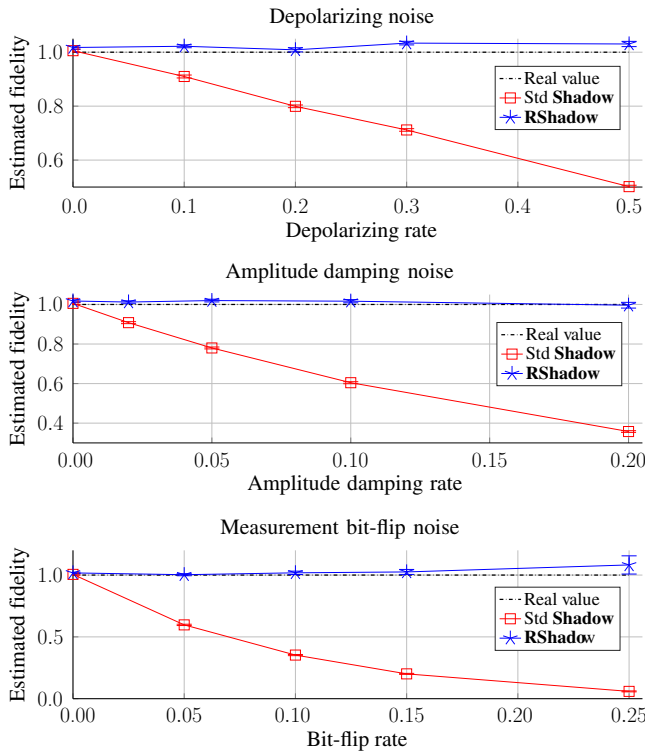


FIG. 2. Comparison of the GHZ-state fidelity estimation using standard **Shadow** and **RShadow** with respect to different noise models and noise levels. The black dashed line represents the true value. The red crosses and the blue stars represent the estimated values of standard **Shadow** and **RShadow**, respectively.

To eliminate these fidelity parameters and extract the estimation of a desired observable, the protocol will use a ratio estimator, which tends to have results with systematically biased errors. Moreover, the statistical fluctuation will affect the estimation of our procedure. Fortunately, Theorems 2 and 4 allow us to bound the size of fluctuation errors along with the systematic biases. For practical considerations, the estimation results of the observables  $\{\hat{\delta}^{(r)}\}$  are allowed to be truncated given some physical ranges from prior knowledge, and this can help to improve the accuracy and circumvent some nonphysical estimation.

On the same task of estimating the GHZ-state fidelity, we further test the performance of our **RShadow** method when the size of the system increases from 4 to 12 qubits. During the measurement procedure, we set a noise model where all the qubits undergo a local  $X$  rotation  $U_X(\theta) = e^{-i\theta X}$ . We remark that such coherent noise cannot be modeled as a classical error that occurs in the measurement results. We fix the number of trials to be  $R = 10^5$  ( $N = 2500, K = 40$ ) for both the calibration and estimation stages. Meanwhile, we choose the rotation angles to be  $\theta = \pi/25, 2\pi/25$ , and  $3\pi/25$ . In Fig. 3, we compare the fidelity estimation results of standard **Shadow** and **RShadow**. When local noises occur, the performance

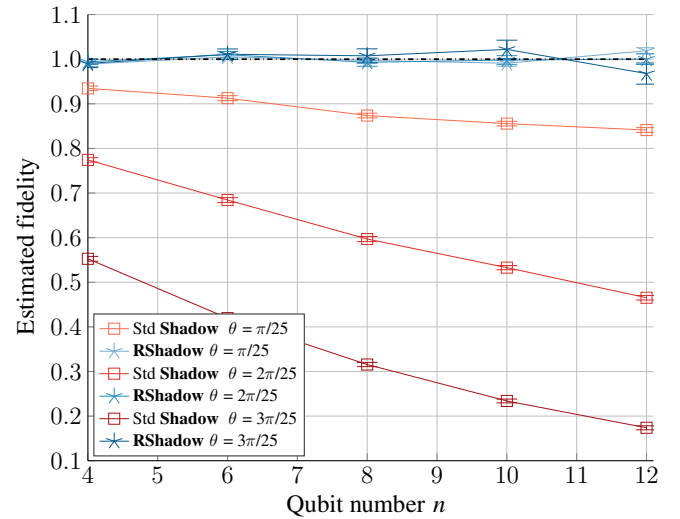


FIG. 3. Comparison of the GHZ fidelity estimation using standard **Shadow** and **RShadow** with respect to different qubit numbers  $n$ . Here, we assume that all the qubits will experience a local  $X$ -rotation error  $U_X(\theta) = e^{-i\theta X}$  with  $\theta = \pi/25, 2\pi/25$ , and  $3\pi/25$ . In the experiment, we set the number of trials  $R = 10^5$  ( $N = 2500, K = 40$ ) for both calibration and estimation stages.

of standard **Shadow** decreases when the system size increases. In contrast, the estimation of **RShadow** is still accurate. This highlights the necessity of noise suppression especially when the system size gets larger.

The next experiment is designed for shadow estimation with the local Clifford group. We estimate the two-point  $ZZ$ -correlation functions and energy expectation of the ground state of an antiferromagnetic TFIM in one dimension with open boundary, whose Hamiltonian is  $H = J \sum_i Z_i Z_{i+1} + h \sum_i X_i$ , and we focus on the case  $J = h = 1$ . The ground state is approximated using the density matrix renormalization group (DMRG) method, represented by a matrix-product state (MPS). Codes from Ref. [36] are modified here to sample random Pauli measurements on the MPS. We compare the performance of **RShadow** and the standard shadow estimation [38] scheme in the presence of measurement bit-flip noise, which means that each qubit measurement outcome has an independent probability  $p$  to be flipped. Our **RShadow** uses  $R = 500\,000$  ( $N = 20\,000, K = 25$ ) calibration samples and  $R = 500\,000$  ( $N = 10\,000, K = 50$ ) estimation samples, while standard shadow estimation uses  $R = 500\,000$  ( $N = 10\,000, K = 50$ ) samples.

We first generate a 50-spin TFIM ground state, and estimate the  $ZZ$ -correlation functions between the leftmost spin and all other spins  $\langle Z_0 Z_i \rangle$ , where the bit-flip probability is set to be 5%. Figure 4 shows the estimation values and absolute errors of both **RShadow** and standard **Shadow**. It can be seen that **RShadow** in general gives a much more precise estimation than standard **Shadow**.



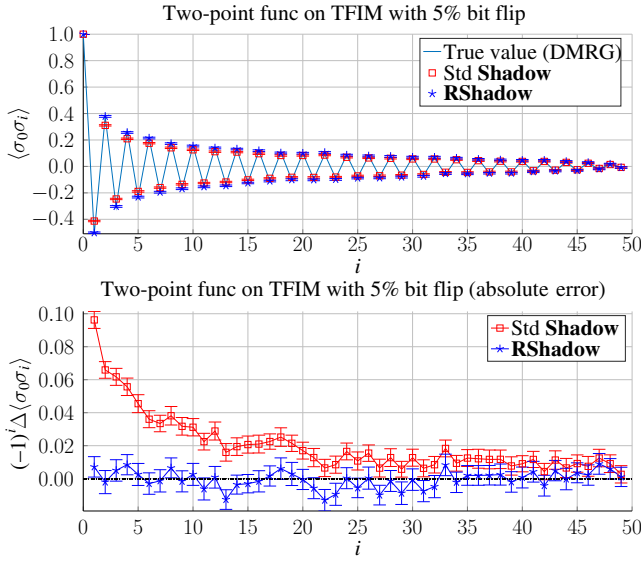


FIG. 4. Two-point correlation function estimation on a 50-spin one-dimensional (1D) TFIM ground state.

We then estimate the energy expectation. In Fig. 5 we plot the energy estimation results on a 50-spin TFIM ground state under three different noise models (similarly to the above numerical experiments of the GHZ fidelity estimation). One can see that the estimation error of standard **Shadow** increases when the noise level increases, while **RShadow** remains giving precise results. Then we fix the noise model to be a 5% measurement bit-flip error

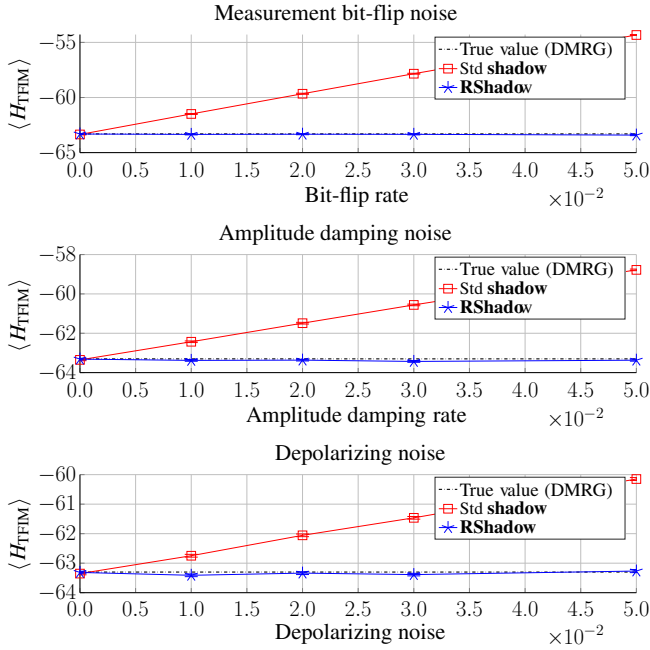


FIG. 5. Energy expectation estimation on a 50-spin 1D TFIM ground state.

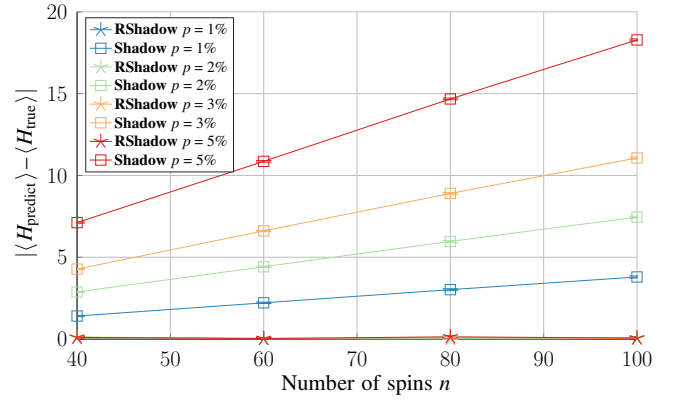


FIG. 6. Energy expectation estimation on a 1D TFIM ground state for different numbers of spins and different bit-flip probabilities.

and conduct estimation on different system sizes. In Fig. 6 we plot the absolute estimation error. This error increases when the system size grows for standard **Shadow**, but it remains close to zero for the **RShadow** scheme. This provides a strong reason why the **RShadow** scheme should be applied as the size of the quantum system becomes increasingly large.

As a remark regarding the computational complexity, we do not calibrate all  $f_z$  such that  $|z| \leq 2$ , the number of which scales as  $\mathcal{O}(n^2)$ . Instead, we only calibrate the nearest-neighbor terms of  $f_z$  for the energy expectation estimation, and the  $f_z$  terms that act on the first qubit and any other qubit for the correlation function estimation. In both cases, there are only  $\mathcal{O}(n)$  parameters to be calibrated. Therefore, when the system size gets large, the **RShadow** protocol remains efficient.

To demonstrate the noise-resilience of the **RShadow** scheme against two-qubit correlated noise, we present more numerical results in Appendix E in the task of estimating the two-point correlation function of the  $n$ -qubit GHZ state. These numerical experiments justify that the **RShadow** scheme can indeed mitigate the experimental errors and reproduce faithful estimation with a small number of benchmarking trials.

## VIII. GATE-DEPENDENT NOISE

Perhaps the strongest assumption we made is the gate independence of the noise channel  $\Lambda$  with respect to the unitary gate  $U$  being sampled. In this section, we present numerical evidence showing that even with an experimentally realistic gate-dependent noise model, **RShadow** can still greatly reduce noise bias. Throughout this section, we focus on **RShadow** with the local Clifford group, which is experimentally implementable on most near-term platforms. The task we consider here is the *electronic structure problem*: decide the ground-state energy of a molecule

with an unknown electronic structure. This is an important problem in quantum chemistry, and is viewed as one of the most promising applications of near-term quantum algorithms; see, e.g., Ref. [58]. Several recent works have already applied shadow estimation related methods to study this problem [61,62].

Specifically, we choose a benchmark molecule and use a certain encoding scheme to map the molecular Hamiltonian into a qubit Hamiltonian. Then, given the ground state of this Hamiltonian, we numerically run the (standard and robust) shadow estimation protocols to estimate its energy, and compare the estimation with the classically computed true value, in the presence of noise. In our setting, we choose  $H_2$  and apply the Bravyi-Kitaev encoding [63] to map it to a four-qubit Hamiltonian.

To come up with a realistic gate-dependent noise model, we first need to decide how the local Clifford group is implemented on real experimental platforms. One common approach is to decompose all unitary gates into a small set of generators. Here, we consider the generating set consisting of the three single-qubit generators

$$\left\{ R_P\left(\frac{\pi}{2}\right) = \exp\left(-i\frac{\pi}{4}P\right), P = X, Y, Z \right\}, \quad (24)$$

which can be understood as  $\pi/2$  rotations along the  $X, Y, Z$  axes, respectively. Every single-qubit Clifford gate can be decomposed into two subsequent rotations along two out of these three axes. For example, the Hadamard gate can be implemented by first applying a  $(\pi/2)$ -rotation pulse along the  $Y$  axis, and then a  $\pi$ -rotation pulse along the  $X$  axis, which is in turn implemented by concatenating two  $\pi/2$   $X$  pulses. (See Appendix E 2 for more details.) This generating set is widely used in real experiments.

Our numerical simulations will deal with the following two kinds of errors that naturally appear in experiments.

1. *Pulse miscalibration.* The  $\pi/2$  pulses have some fixed error due to, e.g., an uncharacterized constant magnetic field. These noisy generators would look like

$$\tilde{R}_P = \exp\left[-i\frac{1}{2}\left(\frac{\pi}{2}P + \Delta_0\right)\right]$$

for  $P = X, Y, Z$  and some traceless Hermitian operator  $\Delta_0$  representing the uncalibrated Hamiltonian. Although  $\Delta_0$  is the same for all three generators, the commutator  $[P, \Delta_0]$  is in general different for different  $P$ . Thus, one can verify that this is a gate-dependent noise model by expanding  $\tilde{R}_P$  using the Baker-Campbell-Hausdorff formula.

2. *Random over-rotation.* Because of imperfect pulse control, the actual rotation angle for each generator could be modeled as  $\pi/2 + \delta$  for some zero-mean Gaussian random variable  $\delta$ . The noisy generators

look like

$$\tilde{R}_P = \exp\left[-i\frac{1}{2}\left(\frac{\pi}{2} + \delta\right)P\right].$$

We assume that the value of  $\delta$  is resampled every time a generator is applied. One can verify that this noise model is equivalent to a dephasing noise on the eigenbasis of Pauli  $P$  following the noiseless generator  $R_P$ ; thus, it is a gate-dependent noise model. (See Appendix E 2.)

Our numerical results are presented in Fig. 7, where we plot the energy estimation outcome of both standard **Shadow** and **RShadow** in the presence of different levels of noise strength. The noise model is *pulse miscalibration* for the upper figure and *random over-rotation* for the lower one. For both noise models, we use  $R = 30\,000$  ( $N = 3000, K = 10$ ) calibration samples and  $R = 10\,000$

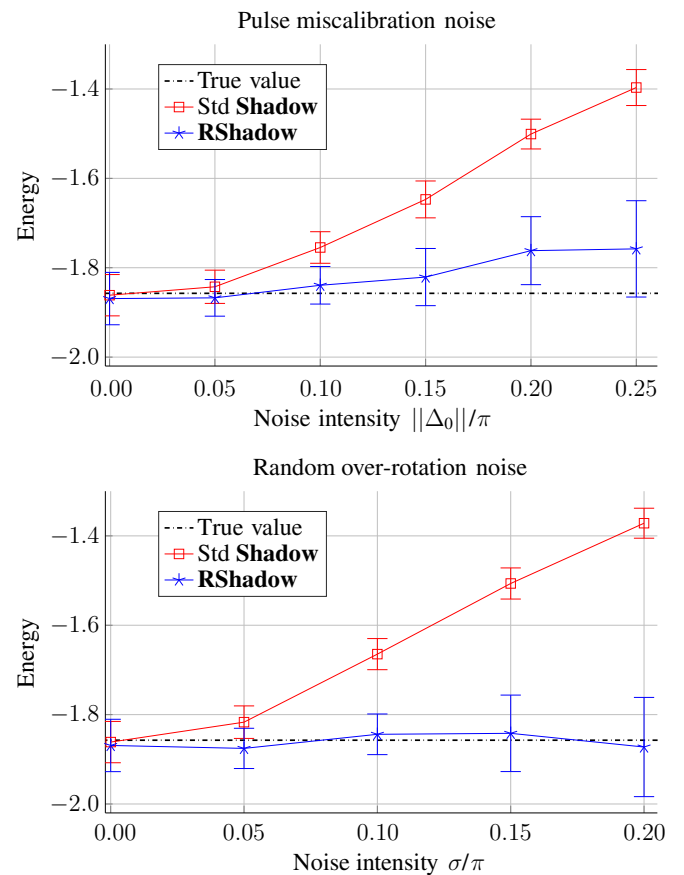


FIG. 7. Upper: ground-state energy estimation of  $H_2$  with pulse miscalibration noise. We choose the uncalibrated Hamiltonian  $\Delta_0 := \|\Delta_0\| * (aX + bY + cZ)$ , where  $[a, b, c] = [-0.5500, 0.2878, 0.7840]$  is a fixed randomly generated unit vector. See Appendix E 2 for evidence that this choice of vector is not special. Lower: ground-state energy estimation of  $H_2$  with random over-rotation noise. Here  $\sigma$  is the standard deviation of the random over-rotation angle  $\delta$ .

( $N = 1000, K = 10$ ) estimation samples for **RShadow**, and  $R = 10\,000$  ( $N = 1000, K = 10$ ) samples for standard **Shadow** [64]. The data points and the error bars are the average values and the standard deviations over 30 independent runs [65].

These numerical results provide evidence for the advantage of **RShadow** over standard shadow estimation even with realistic gate-dependent noise. Specifically, for *pulse miscalibration* noise, it seems that **RShadow** cannot eliminate all biases when the noise strength becomes very large. (Note that noise intensity  $\|\Delta_0\| = 0.1\pi$  is already fairly high in practice.) Yet, even in this regime **RShadow** still significantly outperforms standard shadow estimation and greatly suppresses the bias in the estimated ground-state energy. For *random over-rotation* noise, **RShadow** completely eliminates all bias even at large noise strengths.

So why does **RShadow** work for these gate-dependent noise models? One possible explanation is as follows. The key subroutine of **RShadow** is to learn a Pauli channel. Even though the noise has strong gate dependence, if the  $\widetilde{\mathcal{M}}$  channel defined in Eq. (15) is approximately a Pauli channel, and the random unitary gates are “good enough” to twirl the input probe state  $|0\rangle\rangle$  into an approximate complex projective 2-design, then we expect **RShadow** to still perform well. On the other hand, we note that there are contrived noise models that **RShadow** is not able to detect and mitigate. As an example pointed out by one of the Referees, if the only error in the circuit is a  $Z$ -basis dephasing error right before a single perfect unitary gate, then our calibration procedure is not able to see any noise. However, this is arguably not a practically relevant noise model for real-world experimental platforms. A more rigorous and comprehensive analysis of **RShadow**’s noise resilience against general gate-dependent noise is left for future work.

## IX. CONCLUDING REMARKS

We have analyzed the shadow estimation protocol proposed in Ref. [38] by considering the gate and measurement errors that occur during the process, and have proposed a modified protocol that is robust against such noise. We have proven that, in both the global and the local random Clifford group versions of the robust shadow protocol, we can efficiently benchmark and suppress the effects caused by the noise. On account of the broad application prospects of the shadow estimation protocol in predicting various important properties of quantum states, e.g., entanglement witness, fidelity estimation, correlation functions, etc., we expect that our robust protocol is practical and feasible for current experiments.

While we only focus on estimating linear properties in this work, **RShadow** can also be used to calibrate

the estimation of higher-order properties such as the subsystem Rényi-2 entropy with similar methods shown in Ref. [38]. An exploration into the corresponding sample complexity bound is left for future studies. It is also interesting to explore how **RShadow** can be incorporated with other variants of shadow estimation, such as the locally biased classical shadow [61] and derandomized classical shadow [62]. These two methods can greatly improve the sample efficiency of shadow estimation when one has prior knowledge about which properties are to be predicted, like in the *electronic structure problem*. We believe that these techniques can also be applied to **RShadow** and have been actively developing these methods.

The idea of using additional calibration processes and classical postprocessing to eliminate noise effects also appears in the field of error mitigation [18–21]. Among them, it is particularly interesting to compare our work with Refs. [20,21], which mitigate the measurement readout error on multiqubit devices using a measurement calibration (or detector tomography) process. The spirit of these works is quite similar to ours, but their assumptions on the noise model are much stronger, and their calibration algorithm is more like a heuristic one without an explicit bound on the sample complexity. One reason why we are able to obtain a useful sample complexity bound against a more general noise model is that the random twirling in **RShadow** greatly simplifies our analysis of the noise estimation. For future research, it is interesting to explore the relationship between **RShadow** and other error mitigation schemes, and see if any general results for error mitigation [66] can be applied to our scenario. Very recently, error mitigation has been shown to be helpful even for fault-tolerant quantum computing [67]. We expect **RShadow** to be a useful protocol in the fault-tolerant regime as well.

For our performance guarantee of the robust shadow estimation protocol, the noise in the random gates is allowed to be coherent and highly correlated, but cannot depend on the unitary gate to be implemented. This assumption is reasonable in many cases, especially in the protocol with local Clifford gates, where the noise is mainly caused by amplitude damping and decoherence of the system to the environment [68]. Nevertheless, it is important to analyze how gate dependency and non-Markovianity of the noise can affect the performance of **RShadow**. We have provided some numerical evidence for **RShadow**’s resilience against gate-dependent noise in Sec. VIII, and left more rigorous analysis for future research.

In the PTM representation, the picture of quantum state shadow estimation can be easily extended to the shadow estimation of quantum measurements and quantum channels. For example, in order to estimate  $\langle\langle O_i | \mathcal{E} | \rho_j \rangle\rangle$  for some unknown  $n$ -qubit quantum channel  $\mathcal{E}$  and a set of given observables  $\{O_i\}$  and states  $\{\rho_j\}$ , one may insert two

random measurement channels into the expression,

$$\begin{aligned} \langle\langle O_i | \mathcal{E} | \rho_j \rangle\rangle &= \langle\langle O_i | \mathcal{M}^{-1} \mathcal{M} \mathcal{E} \mathcal{M} \mathcal{M}^{-1} | \rho_j \rangle\rangle \\ &= \mathbb{E}_{U, V \in \mathbb{G}} \sum_{x, y} \langle\langle O_i | \mathcal{M}^{-1} \mathcal{U}^\dagger | x \rangle\rangle \langle\langle x | \mathcal{U} \mathcal{E} \mathcal{V} | y \rangle\rangle \\ &\quad \times \langle\langle y | \mathcal{V}^\dagger \mathcal{M}^{-1} | \rho_j \rangle\rangle. \end{aligned} \quad (25)$$

In the experiment, one can randomly prepare a computational basis state  $|y\rangle$ , apply a random unitary  $V$ , and send to the channel  $\mathcal{E}$ , then apply another random unitary  $U$  and measure in the computational basis, getting outcome  $|x\rangle$ . Then  $2^{-n} \langle\langle O_i | \mathcal{M}^{-1} \mathcal{U}^\dagger | x \rangle\rangle \langle\langle y | \mathcal{V}^\dagger \mathcal{M}^{-1} | \rho_j \rangle\rangle$  is an unbiased estimator of  $\langle\langle O_i | \mathcal{E} | \rho_j \rangle\rangle$ . This is only the most straightforward way to extend robust shadow estimation to quantum channels; there may exist other schemes that have even better performance. We believe a complete analysis of the channel version of shadow estimation will be an interesting direction for further study.

Finally, one can also consider applying (standard or robust) shadow estimation to qudit systems, boson or fermion systems, and other continuous-variable systems using the techniques developed in this work.

## ACKNOWLEDGMENTS

We thank You Zhou for discussions on random unitary schemes, Yihong Zhang for suggestions about gate-dependent noise, Hsin-Yuan Huang and Jinguo Liu for helpful suggestions on the numerical simulations, and John Preskill for discussions on the noise assumptions. We also thank Xiongfeng Ma and Richard Kueng for many helpful comments. C.S., W.Y., and P.Z. are supported by the National Natural Science Foundation of China under Grants No. 11875173 and No. 11674193, the National Key Research and Development Program of China under Grants No. 2019QY0702 and No. 2017YFA0303903, and the Zhongguancun Haihua Institute for Frontier Information Technology.

*Note added.*—After posting this paper to arXiv, two related but independent works subsequently appeared. The independent work by Koh and Grewal [69] also studies how to mitigate noise in the shadow estimation protocol. In their work, the noise channel is assumed to be completely precharacterized. The main results, Theorems 1.1 and 1.2, are similar to our Theorems 2 and 4 if our noise calibration procedure is assumed to be done perfectly. The other independent work by Berg *et al.* [70] also contains some similar ideas as presented in our work. We thank the authors for communicating their work with us.

## APPENDIX A: PRELIMINARIES

In this work, we focus on the  $n$ -qubit quantum systems with Hilbert space dimension  $d = 2^n$ . Define  $\mathcal{H}_d$  to be a finite-dimensional Hilbert space with dimension  $d$ . Define

$\mathcal{L}(\mathcal{H}_d) : \mathcal{H}_d \rightarrow \mathcal{H}_d$  to be the space of linear operators on  $\mathcal{H}_d$ . Define  $\text{Herm}(\mathcal{H}_d)$  to be the space of Hermitian operators on  $\mathcal{H}_d$ , define  $\mathcal{P}(\mathcal{H}_d)$  to be the set of positive operators on  $\mathcal{H}_d$ , and define  $\mathcal{D}(\mathcal{H}_d) \subset \mathcal{P}(\mathcal{H}_d)$  to be the set of quantum states on  $\mathcal{L}(\mathcal{H}_d)$  that are the positive operators with trace equal to 1. Sometimes we also write  $\mathcal{D}(\mathcal{H}_d)$  as  $\mathcal{D}(d)$  for notational simplicity.

## 1. Groups and representations

The group representation theory plays an important role in the shadow estimation protocol. Denote a generic group as  $\mathbb{G} = \{g_i\}_i$ , where  $g_i$  is one of the group elements. Denote a unitary representation of  $\mathbb{G}$  to be a map

$$\phi : \mathbb{G} \rightarrow \mathcal{L}(\mathcal{H}_d) : \mathbb{G} \mapsto \phi(\mathbb{G}) \quad (\text{A1})$$

with the homomorphism

$$\phi(g)\phi(h) = \phi(gh) \quad \text{for all } g, h \in \mathbb{G}. \quad (\text{A2})$$

Moreover, we denote all the irreducible representations (irreps) of the group  $\mathbb{G}$  as  $R_{\mathbb{G}} = \{\phi_\lambda(\mathbb{G})\}_\lambda$ . The Maschke lemma ensures that every representation of a group can be written as a direct sum of irreps,

$$\phi(g) \simeq \bigoplus_{\lambda \in R_{\mathbb{G}}} \phi_\lambda(g)^{\otimes m_\lambda} \quad \text{for all } g \in \mathbb{G}, \quad (\text{A3})$$

where  $m_\lambda$  is an integer implying the multiplicity of the irrep  $\phi_\lambda$ .

In the discussion below, we frequently come across the *twirling* of a linear operator  $O$  on Hilbert space  $\mathcal{H}$  with respect to a group representation  $\phi(\mathbb{G})$ ,

$$\mathcal{T}_\phi(O) := \frac{1}{|\mathbb{G}|} \sum_{g \in \mathbb{G}} \phi(g) O \phi(g)^\dagger. \quad (\text{A4})$$

As a result of the group structure  $\mathbb{G}$ , the twirling result  $\mathcal{T}_\phi(O)$  owns a simple structure, which is related to the irreps in  $\phi(\mathbb{G})$ . The following lemma is a corollary of Schur's lemma.

**Lemma 1** (Lemma 1.7 and Proposition 1.8 of Ref. [45], rephrased by Helsen *et al.* [44]). *For a finite group  $\mathbb{G}$  and a representation  $\phi$  of  $\mathbb{G}$  on a complex vector space  $\mathcal{H}$  with decomposition*

$$\phi(g) \simeq \bigoplus_{\lambda \in R_{\mathbb{G}}} \phi_\lambda(g)^{\otimes m_\lambda} \quad \text{for all } g \in \mathbb{G}, \quad (\text{A5})$$

where  $\{\phi_\lambda\}$  are the irreps of  $\phi(\mathbb{G})$  and  $m_\lambda$  is the multiplicity of  $\phi_\lambda$ . Then, for any linear map  $O \in GL(\mathcal{H})$ , the twirling

of  $O$  with respect to  $\phi$  can be written as

$$\mathcal{T}_\phi(O) = \frac{1}{|\mathbb{G}|} \sum_{g \in \mathbb{G}} \phi(g) O \phi(g)^\dagger = \sum_{\lambda \in R_{\mathbb{G}}} \sum_{j_\lambda, j'_\lambda=1}^{m_\lambda} \frac{\text{Tr}(O \Pi_{j'_\lambda}^{j_\lambda})}{\text{Tr}(\Pi_{j'_\lambda}^{j_\lambda})} \Pi_{j'_\lambda}^{j_\lambda}, \quad (\text{A6})$$

where  $\Pi_{j'_\lambda}^{j_\lambda}$  is a linear map from the support of the  $j'_\lambda$ th copy of  $\phi_\lambda$  to the support of the  $j_\lambda$ th copy of  $\phi_\lambda$ .

In this work, we focus on the group representation  $\phi$  with no multiplicities,

$$\phi(g) \simeq \bigoplus_{\lambda \in R_{\mathbb{G}}} \phi_\lambda(g) \quad \text{for all } g \in \mathbb{G}. \quad (\text{A7})$$

In this case, Eq. (A6) can be simplified as

$$\mathcal{T}_\phi(O) = \frac{1}{|\mathbb{G}|} \sum_{g \in \mathbb{G}} \phi(g) O \phi(g)^\dagger = \sum_{\lambda \in R_{\mathbb{G}}} \frac{\text{Tr}(O \Pi_\lambda)}{\text{Tr}(\Pi_\lambda)} \Pi_\lambda, \quad (\text{A8})$$

where  $\Pi_\lambda$  is the projector onto the support of  $\phi_\lambda$ .

Here, we introduce some common groups that will be frequently used. Note that all the linear operators in  $\mathcal{L}(\mathcal{H}_d)$  form a Lie group  $\text{GL}(d, \mathbb{C})$ . The unitaries in  $\mathcal{L}(\mathcal{H}_d)$  also form a Lie group  $U(d)$ .

Let  $\mathbb{Z}_2 = \{0, 1\}$  be the two-element cyclic group. Then  $\mathbb{Z}_2^n := (\mathbb{Z}_2)^{\otimes n}$  is the  $n$ -copy tensor of the  $\mathbb{Z}_2$  group. Let  $\mathbb{A} = \langle \{a_i\} \rangle$  with  $\{a_i\}$  the generators of the group. In the discussion below, we also slightly abuse the  $\mathbb{Z}_2^n$  notation by allowing it to also denote the set of  $n$ -bit binary strings.

For an  $n$ -qubit quantum system, the Pauli group is

$$\mathbb{P}^n = \langle \{i\} \otimes \{I, X, Y, Z\} \rangle^{\otimes n} \quad (\text{A9})$$

with  $I, X, Y, Z$  the qubit Pauli matrices. Denote the quotient of  $\mathbb{P}^n$  by  $\mathbf{P}^n = \mathbb{P}^n / \langle i \rangle$ , which is an Abelian group and isomorphic to  $\mathbb{Z}_2^{2n}$ . Therefore, we use a  $2n$ -bit string to denote the elements in  $\mathbf{P}^n$  and choose the elements to be

$$P_a = P_{(a_x, a_z)} = i^{a_x \cdot a_z} X^{\otimes a_x} Z^{\otimes a_z}. \quad (\text{A10})$$

The multiplication and commutation of elements in  $\mathbf{P}^n$  are given by

$$\begin{aligned} P_a P_b &= (-i)^{(a,b)} P_{a+b}, \\ P_a P_b &= (-1)^{(a,b)} P_b P_a, \end{aligned} \quad (\text{A11})$$

with

$$\langle a, b \rangle := a_x \cdot b_z - a_z \cdot b_x \pmod{4}, \quad (\text{A12})$$

a binary symplectic product. This symplectic product has the properties

$$\langle a, b \rangle = -\langle b, a \rangle, \quad (\text{A13a})$$

$$(-i)^{\langle a, b \rangle} = i^{-\langle a, b \rangle}, \quad (\text{A13b})$$

$$(-1)^{\langle a, b \rangle} = (-1)^{\langle b, a \rangle}. \quad (\text{A13c})$$

The  $n$ -qubit Clifford group  $\text{Cl}(2^n)$  is defined to be

$$\text{Cl}(2^n) = \{g | g P_a g^{-1} \in \mathbb{P}^n \text{ for all } P_a \in \mathbf{P}^n\} / U(1), \quad (\text{A14})$$

where the  $U(1)$  represents the global phase. Obviously,  $\mathbf{P}^n$  is a subgroup of  $\mathbb{C}^n$ . The single-qubit Clifford group is then  $\text{Cl}_2 := \text{Cl}(2)$ . Later we also come across the tensored  $n$ -fold single-qubit Clifford group  $\text{Cl}_2^{\otimes n}$ .

## 2. Random unitaries and $t$ designs

The shadow estimation is a direct application of twirling in random unitaries. The ideal ‘‘uniformly distributed’’ randomized unitaries over the Lie group  $\text{GL}(d, \mathbb{C})$  is characterized by the *Haar measure*  $\mu(\mathcal{H}_d)$  [71]. The Haar measure is defined to be the unique countably additive, nontrivial measure of the group  $U$  such that

$$\begin{aligned} \int_{\mu(\mathcal{H}_d)} dU &= 1, \\ \int_{\mu(\mathcal{H}_d)} dU f(U) &= \int_{\mu(\mathcal{H}_d)} dU f(UV) = \int_{\mu(\mathcal{H}_d)} dU f(VU), \end{aligned} \quad (\text{A15})$$

where  $f(U)$  is any matrix function of  $U$ .

In practice, to sample unitaries with respect to the Haar measure is challenging due to its continuity. Alternatively, one may choose to sample from a finite subset  $\mathcal{K} = \{U_k\}_{k=1}^{|\mathcal{K}|}$  over the unitaries in  $\text{GL}(d, \mathbb{C})$ .

**Definition 1.** A finite subset  $\mathcal{K} = \{U_k\}_{k=1}^{|\mathcal{K}|} \subset U(d)$  is a unitary  $t$  design if

$$\frac{1}{|\mathcal{K}|} \sum_{k=1}^{|\mathcal{K}|} f_{(t,t)}(U_k) = \int_{\mu(\mathcal{H}_d)} dU f_{(t,t)}(U) \quad (\text{A16})$$

for all the polynomial  $f_{(t,t)}(U)$  of degree at most  $t$  in the matrix elements of  $U$  and at most  $t$  in the matrix elements of  $U^*$ .

It has been proven that, the Clifford gate set  $\text{Cl}(d) \subset U(\mathcal{H})$  is a unitary 3-design [52,53], while fails to be a unitary four design [72].

### 3. Quantum channel and the representations

Quantum channels are the linear maps  $\mathcal{E} : \mathcal{L}(\mathcal{H}_d) \rightarrow \mathcal{L}(\mathcal{H}_d)$  that are CPTP.

**Definition 2.** Let  $\mathcal{E} : \mathcal{L}(\mathcal{H}_d) \rightarrow \mathcal{L}(\mathcal{H}_d)$  be a linear map. We say that

1.  $\mathcal{E}$  is positive if  $\mathcal{E}(\rho) \in \mathcal{D}(\mathcal{H}_d)$  for any  $\rho \in \mathcal{D}(\mathcal{H}_d)$ ;
2.  $\mathcal{E}$  is completely positive (CP) if  $\mathcal{I}_{d'} \otimes \mathcal{E}$  is positive for all dimensions  $d'$ ;
3.  $\mathcal{E}$  is trace preserving (TP) if  $\text{Tr}[\mathcal{E}(\rho)] = 1$  for any  $\text{Tr}[\rho] = 1$ ;
4.  $\mathcal{E}$  is a quantum channel if it is both CP and TP.

In this work, we will come across two representations of the quantum channels: Kraus representation and Liouville representation. For a quantum channel  $\mathcal{E} : \mathcal{L}(\mathcal{H}_d) \rightarrow \mathcal{L}(\mathcal{H}_d)$ , its action on a linear operator  $O \in \mathcal{L}(\mathcal{H}_d)$  can be expressed as

$$\mathcal{E}(O) = \sum_{t=1}^k K_t O K_t^\dagger, \quad (\text{A17})$$

where  $\{K_t\}_{t=1}^k$  are the Kraus operators satisfying  $\sum_{t=1}^k K_t^\dagger K_t = I$ .

To represent the effect of quantum channels in a convenient way, we first introduce the Pauli basis  $P^n$  on  $\mathcal{L}(\mathcal{H}_d)$  to vectorize the linear operators in  $\mathcal{L}(\mathcal{H}_d)$ . Define the inner product between two operators to be the Hilbert-Schmidt product

$$\langle Q, W \rangle := \text{Tr}(QW^\dagger) \quad \text{for all } Q, W \in \text{GL}(\mathcal{H}_d). \quad (\text{A18})$$

In this case, the operators in  $P^n$  form an orthogonal basis. We introduce the operators

$$\sigma_a = P_a / \sqrt{d} \quad (\text{A19})$$

as the orthonormal basis. To vectorize the linear space spanned by  $\{\sigma_a\}$ , we introduce the notation  $\{|\sigma_a\rangle\}$ . For the single-qubit case, we also use the notation

$$\begin{aligned} \sigma_I &= \sigma_0 = \sigma_{(0,0)}, & \sigma_X &= \sigma_{(1,0)}, \\ \sigma_Z &= \sigma_1 = \sigma_{(0,1)}, & \sigma_Y &= \sigma_{(1,1)}. \end{aligned} \quad (\text{A20})$$

Then the operators on  $\mathcal{L}(\mathcal{H}_d)$  can be vectorized as

$$|Q\rangle\rangle = \sum_{a \in \mathbb{Z}_2^{2n}} \langle\langle Q | \sigma_a \rangle\rangle |\sigma_a\rangle\rangle. \quad (\text{A21})$$

The quantum channel  $\mathcal{E}$  can then be represented as

$$\mathcal{E} = \sum_{a,b \in \mathbb{Z}_2^{2n}} \langle\langle \sigma_a | \mathcal{E} | \sigma_b \rangle\rangle |\sigma_a\rangle\rangle \langle\langle \sigma_b | \quad (\text{A22})$$

with

$$\langle\langle \sigma_a | \mathcal{E} | \sigma_b \rangle\rangle := \langle \sigma_a, \mathcal{E}(\sigma_b) \rangle. \quad (\text{A23})$$

The matrix  $\mathcal{E}$  is the PTM or Pauli-Liouville representation. In this work, we slightly abuse the notation of a superoperator  $\mathcal{E}$  to represent its PTM. For a unitary matrix  $U$ , we use the calligraphic  $\mathcal{U}$  to represent its PTM.

For a quantum channel  $\mathcal{E}$  with state  $\rho$  input, and POVM measurement  $M = \{M_b\}$  with  $\sum_b M_b = I$ , the probability of getting the measurement result  $b$  is

$$p_b = \langle\langle M_b | \mathcal{E} | \rho \rangle\rangle. \quad (\text{A24})$$

Under the PTM representation, the composition and tensor product of channels  $\mathcal{E}_1$  and  $\mathcal{E}_2$  can be naturally expressed as

$$\begin{aligned} |\mathcal{E}_1 \circ \mathcal{E}_2(\rho)\rangle\rangle &= \mathcal{E}_1 \mathcal{E}_2 |\rho\rangle\rangle, \\ |\mathcal{E}_1 \otimes \mathcal{E}_2(\rho^{\otimes 2})\rangle\rangle &= \mathcal{E}_1 \otimes \mathcal{E}_2 |\rho^{\otimes 2}\rangle\rangle. \end{aligned} \quad (\text{A25})$$

The PTM of the unitaries in  $U(d)$  forms a natural group representation of  $U(d)$ . Denote the PTM of a given unitary  $U$  as  $\phi^P(U) := \mathcal{U}$ . Then we have

$$\phi^P(U)\phi^P(V) = \phi^P(UV). \quad (\text{A26})$$

The PTM representation  $\phi^P[U(d)]$  can be decomposed into two irreps:

$$\phi^P(U) \simeq \phi_I^P(U) \oplus \phi_\sigma^P(U) \quad \text{for all } U \in U(d). \quad (\text{A27})$$

Here,

$$\begin{aligned} \phi_I^P(U) &= \Pi_I \phi^P(U) \Pi_I, \\ \phi_\sigma^P(U) &= \Pi_\sigma \phi^P(U) \Pi_\sigma, \end{aligned} \quad (\text{A28})$$

with the projectors  $\Pi_I$  and  $\Pi_\sigma$  given by

$$\begin{aligned} \Pi_I &= |\sigma_0^{\otimes n}\rangle\rangle \langle\langle \sigma_0^{\otimes n}|, \\ \Pi_\sigma &= I - \Pi_I = \sum_{a \in \mathbb{Z}_2^{2n}, a \neq (0,0)^{\otimes n}} |\sigma_a\rangle\rangle \langle\langle \sigma_a|. \end{aligned} \quad (\text{A29})$$

The  $n$ -qubit Clifford group  $\text{Cl}(2^n)$ , as the subset of the  $n$ -qubit unitary group, can also be represented by PTM matrices. The PTM representation  $\phi^P[\text{Cl}(2^n)]$  can be decomposed similarly as

$$\phi^P(U) \simeq \phi_I^P(U) \oplus \phi_\sigma^P(U) \quad \text{for all } U \in \text{Cl}(2^n), \quad (\text{A30})$$

where  $\phi_I^P$  and  $\phi_\sigma^P$  are two irreps on the support  $\Pi_I$  and  $\Pi_\sigma$ , respectively.

#### 4. Weingarten function

In this part, we introduce the Weingarten function as a tool to calculate general Haar integrals [73–75]. The following presentation owes a lot to Section 2 of Ref. [76].

For an operator  $A$  acting on  $\mathcal{H}_d^{\otimes k}$ , define the  $k$ -fold Haar twirling of  $A$  as

$$\Phi_{\text{Haar}}^{(k)}(A) := \int_{\mu(\mathcal{H}_d)} dU (U^{\otimes k})^\dagger A U^{\otimes k}. \quad (\text{A31})$$

Using Schur-Weyl duality, one can show that

$$\Phi_{\text{Haar}}^{(k)}(A) = \sum_{\pi, \sigma \in S_k} c_{\pi, \sigma} W_\pi \text{Tr}(W_\sigma A). \quad (\text{A32})$$

Here,  $S_k$  is the  $k$ -element permutation group,  $W_\pi$  is the permutation operator defined as

$$W_\pi |a_1, \dots, a_k\rangle = |a_{\pi(1)}, \dots, a_{\pi(k)}\rangle$$

for all  $|a_1, \dots, a_k\rangle \in \mathcal{H}_d^{\otimes k}$ ,  $\pi \in S_k$ ,

(A33)

and the coefficients  $c_{\pi, \sigma}$  form the Weingarten matrix [74]

that can be calculated as

$$c_{\pi, \sigma} = (Q^+)_{\pi, \sigma}, \quad Q_{\pi, \sigma} := d^{\#\text{cycles}(\pi\sigma)}, \quad (\text{A34})$$

where  $Q$  is called the Gram matrix. Here  $Q^+$  stands for the Moore-Penrose pseudoinverse of  $Q$ , which is  $Q^{-1}$  when  $Q$  is invertible. (Note that, when  $Q$  is not invertible,  $c$  is not uniquely determined. It is only a conventional choice to take  $c = Q^+$  [75, 77].)

In following sections, we are interested in the case  $k = 3$ . We sort the elements of  $S_3$  in the order

$$\vec{W} := [W_(), W_{(1,2)}, W_{(1,3)}, W_{(2,3)}, W_{(1,2,3)}, W_{(1,3,2)}]. \quad (\text{A35})$$

In this basis, the Gram matrix becomes

$$Q = \begin{bmatrix} d^3 & d^2 & d^2 & d^2 & d & d \\ d^2 & d^3 & d & d & d^2 & d^2 \\ d^2 & d & d^3 & d & d^2 & d^2 \\ d^2 & d & d & d^3 & d^2 & d^2 \\ d & d^2 & d^2 & d^2 & d & d^3 \\ d & d^2 & d^2 & d^2 & d^3 & d \end{bmatrix}. \quad (\text{A36})$$

For  $d \geq 3$ , one can show that the Weingarten matrix becomes

$$c = \frac{1}{d(d^2 - 1)(d^2 - 4)} \begin{bmatrix} d^2 - 2 & -d & -d & -d & 2 & 2 \\ -d & d^2 - 2 & 2 & 2 & -d & -d \\ -d & 2 & d^2 - 2 & 2 & -d & -d \\ -d & 2 & 2 & d^2 - 2 & -d & -d \\ 2 & -d & -d & -d & 2 & d^2 - 2 \\ 2 & -d & -d & -d & d^2 - 2 & 2 \end{bmatrix}, \quad (\text{A37})$$

while, for  $d = 2$ ,  $Q$  is singular, so we take its pseudoinverse as

$$c = \frac{1}{144} \begin{bmatrix} 17 & 1 & 1 & 1 & -7 & -7 \\ 1 & 17 & -7 & -7 & 1 & 1 \\ 1 & -7 & 17 & -7 & 1 & 1 \\ 1 & -7 & -7 & 17 & 1 & 1 \\ -7 & 1 & 1 & 1 & -7 & 17 \\ -7 & 1 & 1 & 1 & 17 & -7 \end{bmatrix}. \quad (\text{A38})$$

#### APPENDIX B: SAMPLE COMPLEXITY OF RSHADOW WITH THE GLOBAL CLIFFORD GROUP

In this section, we study our robust shadow estimation protocol with  $\mathbb{G}$  chosen to be the  $n$ -qubit Clifford group  $\text{Cl}(2^n)$ .

##### 1. Calibration procedure: global

Recall that channel  $\tilde{\mathcal{M}}$  can be written in the Pauli basis as

$$\tilde{\mathcal{M}} = \mathbb{E}_{U \sim \text{Cl}(2^n)} U^\dagger \mathcal{M}_z \Lambda U = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & f & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f \end{bmatrix} \quad (\text{B1})$$

for some  $f \in \mathbb{R}$  depending on  $\Lambda$ . Note that  $f = (d + 1)^{-1}$  when the noise channel is trivial, i.e.,  $\Lambda = \text{id}$ . We rewrite the **RShadow** protocol from the main text as follows.

##### Protocol 1 (RShadow with $\text{Cl}(2^n)$ ).

1. Prepare  $|\mathbf{0}\rangle \equiv |0\rangle^{\otimes n}$ . Sample  $U$  uniformly from  $\text{Cl}(2^n)$  and apply it to  $|\mathbf{0}\rangle$ .

2. Measure the above state in the computational basis. Denote the outcome state vector as  $|b\rangle$ .
3. Calculate the single-round estimator of  $f$  as  $\hat{f}^{(r)} := (d\hat{F}^{(r)} - 1)/(d - 1)$ , where  $\hat{F}^{(r)} := |\langle b|U|\mathbf{0}\rangle|^2$ .
4. Repeat steps 1–3  $R = NK$  rounds. Then the final estimation of  $f$  is given by a median of mean estimator  $\hat{f}$  constructed from the single-round estimators  $\{\hat{f}^{(r)}\}_{r=1}^R$  with parameters  $N, K$  [see Eq. (B21)].
5. After the above steps, apply the standard classical shadow protocol of Ref. [38] on  $\rho$  with the inverse channel  $\mathcal{M}^{-1}$  replaced by

$$\widetilde{\mathcal{M}}^{-1} := \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \hat{f}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{f}^{-1} \end{bmatrix}$$

in the Liouville representation.

In Protocol 1, the unitary operations and the measurement are assumed to contain gate-independent noise, and the preparation of  $|\mathbf{0}\rangle$  is assumed to be perfect. The next theorem shows that  $\hat{f}^{(r)}$  is an unbiased estimator of  $f$  and its variance can be bounded.

**Proposition 1.** *The single-round fidelity estimator  $\hat{F}^{(r)}$  given in Protocol 1 satisfies*

$$\mathbb{E}(\hat{F}^{(r)}) = F_{\text{avg}}(\widetilde{\mathcal{M}}) = \frac{F_Z(\Lambda) + 1}{d + 1}, \quad \text{Var}(\hat{F}^{(r)}) \leq \frac{2}{d^2}, \quad (\text{B2})$$

where  $F_{\text{avg}}(\widetilde{\mathcal{M}}) = \int_{\psi \in \text{Haar}} d\psi \langle \langle \psi | \widetilde{\mathcal{M}} | \psi \rangle \rangle$  is the average fidelity of  $\widetilde{\mathcal{M}}$ , and  $F_Z(\Lambda) = (1/2^n) \sum_{b \in \{0,1\}^n} \langle \langle b | \Lambda | b \rangle \rangle$  is the Z-basis average fidelity of  $\Lambda$ .

Moreover, the single-round estimator  $\hat{f}$  satisfies

$$\mathbb{E}(\hat{f}^{(r)}) = f = \frac{dF_Z(\Lambda) - 1}{d^2 - 1}, \quad \text{Var}(\hat{f}^{(r)}) \leq \frac{2}{(d - 1)^2}. \quad (\text{B3})$$

Before we provide the proof of Proposition 1, we first introduce two lemmas.

**Lemma 2** (See, e.g., Proposition 4 of Ref. [72]). *If a group  $\mathbb{G} \subseteq U(d)$  forms a unitary  $t$  design then*

$$\mathbb{E}_{U \sim \mathbb{G}} (U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger)^{\otimes t} = P_{\text{sym}^t} / \binom{d+t-1}{t}, \quad (\text{B4})$$

where  $P_{\text{sym}^t}$  is the projector onto the  $t$ -fold symmetric space, or, equivalently,  $P_{\text{sym}^t} = (1/|S_t|) \sum_{\pi \in S_t} W_\pi$ , where  $W_\pi$  is the permutation operator defined in Eq. (A33).

**Lemma 3.** *For two operators  $A, B$  acting on  $\mathcal{H}(d)$ ,*

$$\text{Tr}(P_{\text{sym}^2} A \otimes B) = \frac{1}{2}[\text{Tr} A \text{Tr} B + \text{Tr}(AB)], \quad (\text{B5})$$

$$\begin{aligned} \text{Tr}(P_{\text{sym}^3} A \otimes B \otimes B) &= \frac{1}{6}[\text{Tr} A (\text{Tr} B)^2 + \text{Tr} A \text{Tr}(B^2) \\ &\quad + 2 \text{Tr}(AB) \text{Tr} B + 2 \text{Tr}(AB^2)]. \end{aligned} \quad (\text{B6})$$

*Proof.* For the first equation,

$$\begin{aligned} \text{Tr}(P_{\text{sym}^2} A \otimes B) &= \frac{1}{2}[\text{Tr}[I(A \otimes B)] + \text{Tr}[S(A \otimes B)]] \\ &= \frac{1}{2}[\text{Tr} A \text{Tr} B + \text{Tr}(AB)], \end{aligned} \quad (\text{B7})$$

where  $S$  is the swap operator.

For the second equation, using the language of tensor network (see, e.g., Section 3.1 of Ref. [78]), we can derive

$$\begin{aligned} \text{Tr}[\vec{W}(A \otimes B \otimes B)] &= [\text{Tr} A (\text{Tr} B)^2, \text{Tr}(AB) \text{Tr} B, \\ &\quad \text{Tr}(AB) \text{Tr} B, \text{Tr} A \text{Tr}(B^2), \text{Tr}(AB^2), \text{Tr}(AB^2)], \end{aligned} \quad (\text{B8})$$

where  $\vec{W}$  is a vectorization of  $S_3$  defined in Eq. (A35). Averaging this up gives the second equation. ■

*Proof of Proposition 1.* First, from Eq. (B1) we immediately have

$$f = \frac{\text{Tr}(\widetilde{\mathcal{M}}) - 1}{d^2 - 1}. \quad (\text{B9})$$

We also have the following relation between the average fidelity of a channel  $\mathcal{M}$  and the trace of its Pauli transformer matrix (see, e.g., Ref. [44]):

$$F_{\text{avg}}(\widetilde{\mathcal{M}}) = \frac{d^{-1} \text{Tr}(\widetilde{\mathcal{M}}) + 1}{d + 1}. \quad (\text{B10})$$

Combining the above two equations, we get

$$f = \frac{dF_{\text{avg}}(\widetilde{\mathcal{M}}) - 1}{d - 1}; \quad (\text{B11})$$

hence, Eq. (B3) follows directly from Eq. (B2). We only need to calculate the expectation and variance of  $\hat{F}^{(r)}$ .

Denote the Kraus operators of the noise channel  $\Lambda$  as  $\{K_t\}$ . The average fidelity of  $\widetilde{\mathcal{M}}$  can be explicitly written as



$$\begin{aligned}
 F_{\text{avg}}(\tilde{\mathcal{M}}) &= \int_{d\psi} \mathbb{E}_{U \sim \text{Cl}} \langle \psi | U^\dagger \circ M_z \circ \Lambda \circ U(|\psi\rangle\langle\psi|) | \psi \rangle \\
 &= \int_{d\psi} \mathbb{E}_{U \sim \text{Cl}} \sum_{b,t} \langle \psi | U^\dagger | b \rangle \langle b | K_t U | \psi \rangle \langle \psi | U^\dagger K_t^\dagger | b \rangle \\
 &\quad \langle b | U | \psi \rangle \\
 &= \int_{d\psi} \mathbb{E}_{U \sim \text{Cl}} \sum_{b,t} | \langle b | K_t U | \psi \rangle |^2 | \langle b | U | \psi \rangle |^2.
 \end{aligned} \tag{B12}$$

$$\begin{aligned}
 &= \frac{2}{(d+1)d} \sum_{b,t} \frac{1}{2} (\langle b | K_t K_t^\dagger | b \rangle + | \langle b | K_t | b \rangle |^2) \\
 &= \frac{1}{(d+1)d} \left( d + \sum_{b,t} | \langle b | K_t | b \rangle |^2 \right) \\
 &= \frac{1 + F_Z}{d+1},
 \end{aligned} \tag{B15}$$

On the other hand, the expectation of  $\hat{F}^{(r)}$  can be expressed as

$$\begin{aligned}
 \mathbb{E}(\hat{F}^{(r)}) &= \mathbb{E}_{U \sim \text{Cl}} \sum_{b,t} | \langle b | K_t U | \mathbf{0} \rangle |^2 | \langle b | U | \mathbf{0} \rangle |^2 \\
 &= \mathbb{E}_{V \sim \text{Cl}} \mathbb{E}_{W \sim \text{Cl}} \sum_{b,t} | \langle b | K_t V W | \mathbf{0} \rangle |^2 | \langle b | V W | \mathbf{0} \rangle |^2 \\
 &= \int_{d\psi} \mathbb{E}_{V \sim \text{Cl}} \sum_{b,t} | \langle b | K_t V | \psi \rangle |^2 | \langle b | V | \psi \rangle |^2,
 \end{aligned} \tag{B13}$$

where the first equality follows from the definition of the expectation, the second equality follows from the fact that sampling an element  $U$  from a group is equivalent to independently sampling two elements  $V, W$  from the group and taking  $U = V \circ W$ , and the last equality uses the fact that  $\text{Cl}(2^n)$  is a unitary 2-design. As a result, we have shown that

$$\mathbb{E}(\hat{F}^{(r)}) = F_{\text{avg}}(\tilde{\mathcal{M}}). \tag{B14}$$

Next, in order to get  $\text{Var}(\hat{F}^{(r)})$ , we calculate the values of  $\mathbb{E}(\hat{F}^{(r)})$  and  $\mathbb{E}(\hat{F}^{(r)^2})$  explicitly. Based on Lemmas 2 and 3, and recalling the fact that  $\text{Cl}(2^n)$  is a unitary 3-design [52–54], we are able to carry out the calculations

$$\begin{aligned}
 \mathbb{E}(\hat{F}^{(r)}) &= \mathbb{E}_{U \sim \text{Cl}} \sum_{b,t} | \langle b | K_t U | \mathbf{0} \rangle |^2 | \langle b | U | \mathbf{0} \rangle |^2 \\
 &= \sum_{b,t} \text{Tr} \left[ \mathbb{E}_{U \sim \text{Cl}} (U | \mathbf{0} \rangle \langle \mathbf{0} | U^\dagger)^{\otimes 2} (K_t^\dagger | b \rangle \langle b | K_t \right. \\
 &\quad \left. \otimes | b \rangle \langle b |) \right] \\
 &= \frac{2}{(d+1)d} \sum_{b,t} \text{Tr} [P_{\text{sym}^2}(K_t^\dagger | b \rangle \langle b | K_t \otimes | b \rangle \langle b |)]
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{E}(\hat{F}^{(r)^2}) &= \mathbb{E}_{U \sim \text{Cl}} \sum_{b,t} | \langle b | K_t U | \mathbf{0} \rangle |^2 | \langle b | U | \mathbf{0} \rangle |^4 \\
 &= \sum_{b,t} \text{Tr} \left[ \mathbb{E}_{U \sim \text{Cl}} (U | \mathbf{0} \rangle \langle \mathbf{0} | U^\dagger)^{\otimes 3} (K_t^\dagger | b \rangle \langle b | K_t \right. \\
 &\quad \left. \otimes | b \rangle \langle b | \otimes | b \rangle \langle b |) \right] \\
 &= \frac{6}{(d+2)(d+1)d} \sum_{b,t} \text{Tr} [P_{\text{sym}^3}(K_t^\dagger | b \rangle \langle b | K_t \\
 &\quad \otimes | b \rangle \langle b | \otimes | b \rangle \langle b |)] \\
 &= \frac{6}{(d+2)(d+1)d} \sum_{b,t} \frac{1}{3} \\
 &\quad (\langle b | K_t K_t^\dagger | b \rangle + 2 | \langle b | K_t | b \rangle |^2) \\
 &= \frac{2(1 + 2F_Z)}{(d+2)(d+1)},
 \end{aligned} \tag{B16}$$

where we write  $F_Z \equiv F_Z(\Lambda)$  as the  $Z$ -basis average fidelity of  $\Lambda$ .

Now we can bound the variance of  $\hat{F}$  as

$$\begin{aligned}
 \text{Var}(\hat{F}^{(r)}) &= \mathbb{E}(\hat{F}^{(r)^2}) - [\mathbb{E}(\hat{F}^{(r)})]^2 \\
 &= \frac{-(d+2)F_Z^2 + 2dF_Z + d}{(d+2)(d+1)^2} \\
 &\leq \frac{2}{d^2},
 \end{aligned} \tag{B17}$$

where we have used the fact that  $F_Z \leq 1$ . This completes the proof.  $\blacksquare$

Now we analyse the sample complexity of Protocol 1 in order to guarantee that the protocol succeeds within a given level of precision. Specifically, we consider using the protocol to estimate a linear function of  $\rho$ , i.e.,  $\langle O | \rho \rangle$ . Given that one makes sufficiently many samples in the estimation procedure, the estimation of this function will be close to

$\langle\langle O | \widehat{\mathcal{M}}^{-1} \widetilde{\mathcal{M}} | \rho \rangle\rangle$ . Hence, we are concerned about the error

$$\begin{aligned} & \left| \langle\langle O | \widehat{\mathcal{M}}^{-1} \widetilde{\mathcal{M}} | \rho \rangle\rangle - \langle\langle O | \rho \rangle\rangle \right| \\ &= \left| \langle\langle O | \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \hat{f}^{-1}f - 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{f}^{-1}f - 1 \end{bmatrix} | \rho \rangle\rangle \right| \\ &= |\langle\langle O_0 | \rho \rangle\rangle| \cdot |\hat{f}^{-1}f - 1| \\ &\leq \|O_0\|_\infty \cdot |\hat{f}^{-1}f - 1|, \end{aligned} \quad (\text{B18})$$

where  $O_0 = O - [\text{Tr}(O)/d]I$  is the traceless part of  $O$ . Now we want to upper bound  $|\hat{f}^{-1}f - 1|$  by some  $\varepsilon > 0$ . Suppose that, with high probability, the estimator in Protocol 1 satisfies  $|\hat{f} - f| \leq \gamma$  for some  $0 \leq \gamma \leq |f|$ . Then we have

$$|\hat{f}^{-1}f - 1| = |\hat{f}^{-1}| \cdot |\hat{f} - f| \leq \frac{\gamma}{|\hat{f}|} \leq \frac{\gamma}{|f| - \gamma}, \quad (\text{B19})$$

where the last inequality follows by the triangular inequality. Now if we have

$$\frac{\gamma}{|f| - \gamma} \leq \varepsilon \iff \gamma \leq \frac{\varepsilon|f|}{1 + \varepsilon} \quad (\text{B20})$$

then we obtain the bound  $|\hat{f}^{-1}f - 1| \leq \varepsilon$  with high success probability. Now is the time to calculate the number of rounds  $R$  in order to bound  $|\hat{f} - f|$  into a small interval with high probability. As noted before, we use the median of means estimator [79,80] in order to get a preferable scaling with respect to the failing probability. Similar techniques are also applied in Ref. [38]. Specifically, we conduct  $R = KN$  rounds of the procedure in Protocol 1, calculate  $K$  estimators, each of which is the average of  $N$  single-round estimators  $\hat{f}$ , and take the median of these  $K$  estimators as our final estimator  $\hat{f}$ . That is,

$$\begin{aligned} \bar{f}^{(k)} &:= \frac{1}{N} \sum_{r=(k-1)N+1}^{kN} \hat{f}^{(r)}, \quad k = 1, 2, \dots, K. \\ \hat{f} &:= \text{median}\{\bar{f}^{(1)}, \bar{f}^{(2)}, \dots, \bar{f}^{(K)}\}. \end{aligned} \quad (\text{B21})$$

The performance of this estimator is given in the following lemma.

**Lemma 4** (Refs. [79,80], rephrased by Huang *et al.* [38]). *For the estimator described by Eq. (B21), where  $\hat{f}^{(r)}$  is an identical and independent sample of  $f$ , if  $N =$*

$34\text{Var}(\hat{f})/\gamma^2$  for any given  $\gamma > 0$  then

$$\Pr(|\hat{f} - \mathbb{E}\hat{f}| \geq \gamma) \leq 2 \exp(-K/2). \quad (\text{B22})$$

Furthermore, by taking  $K = 2 \ln(2\delta^{-1})$  for any  $\delta > 0$ , we have

$$\Pr(|\hat{f} - \mathbb{E}\hat{f}| \geq \gamma) \leq \delta. \quad (\text{B23})$$

Thanks to this lemma and the above discussion, we reach the following theorem that summarizes the trade-off between precision and the sample complexity of our main protocol. This theorem is the rigorous version of Theorem 1 in the main text.

**Theorem 7.** *Given  $\varepsilon, \delta > 0$ ,*

$$R = 136 \ln(2\delta^{-1}) \frac{(1 + \varepsilon^2)(1 + 1/d)^2}{\varepsilon^2(F_Z - 1/d)^2} \quad (\text{B24})$$

*rounds of calibration in Protocol 1 are enough for the asymptotic error of the subsequent estimation procedure to satisfy*

$$\begin{aligned} & \left| \langle\langle O | \widehat{\mathcal{M}}^{-1} \widetilde{\mathcal{M}} | \rho \rangle\rangle - \langle\langle O | \rho \rangle\rangle \right| \leq \varepsilon \|O_0\|_\infty \\ & \text{for all } O \in \text{Herm}(2^n) \text{ and all } \rho \in \mathcal{D}(2^n) \end{aligned} \quad (\text{B25})$$

*with a success probability of at least  $1 - \delta$ , where  $F_Z \equiv F_Z(\Lambda)$  is the Z-basis average fidelity of noise channel  $\Lambda$ .*

*Proof.* Construct the median of means estimator  $\hat{f}$  with  $K = 2 \ln(2\delta^{-1})$  and  $N = 34\text{Var}(\hat{f})/\gamma^2$ , where  $\gamma = \varepsilon|f|/(1 + \varepsilon)$  as Eq. (B19) suggests. Use Proposition 1 to get

$$\text{Var}(\hat{f}) \leq \frac{2}{(d-1)^2}, \quad |f| = \frac{dF_Z - 1}{d^2 - 1}. \quad (\text{B26})$$

Then Lemma 4 guarantees that

$$R = KN = 136 \ln(2\delta^{-1}) \frac{(1 + \varepsilon^2)(d+1)^2}{\varepsilon^2(dF_Z - 1)^2}. \quad (\text{B27})$$

This completes the proof. ■

Theorem 7 provides an upper bound on the necessary number of rounds that scales as

$$R = O\left(\frac{1}{\varepsilon^2(F_Z - 1/d)^2}\right). \quad (\text{B28})$$

## 2. Estimation procedure: global

Till now, we have proved the efficiency of the calibration procedure, but have not addressed the efficiency of the estimation procedure. In the noiseless case, the performance of the standard quantum shadow estimation protocol has been characterized in Ref. [38]. Here, we extend their methods to show the performance of the **RShadow** estimation procedure.

For any set of observables  $\{O_i\}_{i=1}^M$  and an unknown state  $\rho$ , the single-round estimation and the final estimation of  $o_i := \text{Tr}(O_i \rho)$  are denoted by  $\hat{o}_i^{(r)}$  and  $\hat{o}_i$ , respectively, given by Algorithm 2. The deviation of  $\mathbb{E}(\hat{o}_i^{(r)})$  from  $o_i$  has been bounded by Theorem 7. Now we want to bound  $\text{Var}(\hat{o}_i^{(r)})$ . We first introduce the following lemma.

**Lemma 5.** *For any  $O \in \text{Herm}(2^n)$  and an unknown state  $\rho \in \mathcal{D}(2^n)$ , the single-round estimator  $\hat{o}^{(r)}$  given by the **RShadow** protocol using either  $\text{Cl}(2^n)$  or  $\text{Cl}_2^{\otimes n}$  satisfies*

$$\text{Var}(\hat{o}^{(r)}) \leq \|O_0\|_{\text{shadow}, \Lambda}^2, \quad (\text{B29})$$

where  $O_0 \equiv O - [\text{Tr}(O)/2^n]I$ . The function  $\|\cdot\|_{\text{shadow}, \Lambda}$  depends on the noise channel and the unitary group being used:

$$\|O\|_{\text{shadow}, \Lambda} := \max_{\sigma \in \mathcal{D}(2^n)} \left( \mathbb{E}_{U \sim \mathbb{G}} \sum_{b \in \{0,1\}^n} \langle b | \Lambda(U\sigma U^\dagger) | b \rangle \times \langle b | U\tilde{\mathcal{M}}^{-1}(O)U^\dagger | b \rangle^2 \right)^{1/2}. \quad (\text{B30})$$

When  $\Lambda = \text{id}$ , the function  $\|\cdot\|_{\text{shadow}, \Lambda}$  degrades to the norm  $\|\cdot\|_{\text{shadow}}$  defined in Ref. [38].

*Proof.* First observe that the variance of  $\hat{o}^{(r)}$  from Algorithm 2 depends only on the traceless part of  $O$ , i.e.,

$$\begin{aligned} \hat{o}^{(r)} - \mathbb{E}(\hat{o}^{(r)}) &= \langle \langle O | \widehat{\mathcal{M}}^{-1} U^\dagger | b \rangle \rangle - \langle \langle O | \widehat{\mathcal{M}}^{-1} \tilde{\mathcal{M}} | \rho \rangle \rangle \\ &= \langle \langle O_0 | \widehat{\mathcal{M}}^{-1} U^\dagger | b \rangle \rangle - \langle \langle O_0 | \widehat{\mathcal{M}}^{-1} \tilde{\mathcal{M}} | \rho \rangle \rangle \\ &= \hat{o}_0^{(r)} - \mathbb{E}(\hat{o}_0^{(r)}), \end{aligned} \quad (\text{B31})$$

because  $\widehat{\mathcal{M}}$  is diagonal in the Pauli-transfer matrix representation, and  $\tilde{\mathcal{M}}$  is a trace-preserving map. Therefore,

$$\begin{aligned} \text{Var}(\hat{o}^{(r)}) &= \mathbb{E}\{[\hat{o}^{(r)} - \mathbb{E}(\hat{o}^{(r)})]^2\} \\ &= \mathbb{E}\{[\langle \langle O_0 | \widehat{\mathcal{M}}^{-1} U^\dagger | b \rangle \rangle - \langle \langle O_0 | \widehat{\mathcal{M}}^{-1} \tilde{\mathcal{M}} | \rho \rangle \rangle]^2\} \\ &\leq \mathbb{E}\langle \langle O_0 | \widehat{\mathcal{M}}^{-1} U^\dagger | b \rangle \rangle^2 \\ &= \mathbb{E}_{U \sim \mathbb{G}} \sum_{b \in \{0,1\}^n} \langle \langle b | \Lambda U | \rho \rangle \rangle \langle \langle b | U \widehat{\mathcal{M}}^{-1} | O_0 \rangle \rangle^2 \\ &\leq \max_{\sigma \in \mathcal{D}(2^n)} \mathbb{E}_{U \sim \mathbb{G}} \sum_{b \in \{0,1\}^n} \langle \langle b | \Lambda U | \sigma \rangle \rangle \langle \langle b | U \widehat{\mathcal{M}}^{-1} | O_0 \rangle \rangle^2 \\ &= \|O_0\|_{\text{shadow}, \Lambda}^2. \end{aligned} \quad (\text{B32})$$

This completes the proof.  $\blacksquare$

In the special case that  $\mathbb{G} := \text{Cl}(2^n)$ , we can obtain the following bound on the shadow norm  $\|\cdot\|_{\text{shadow}, \Lambda}$ .

**Lemma 6.** *For **RShadow** using  $\text{Cl}(2^n)$ , if the calibration procedure guarantees  $\hat{f} \geq \delta f$  for some  $\delta > 0$ , and we assume that  $F_Z(\Lambda) \geq 1/d$ , then we have*

$$\|O_0\|_{\text{shadow}, \Lambda}^2 \leq \delta^{-2} \left( F_Z - \frac{1}{d} \right)^{-2} 3 \text{Tr}(O_0^2) \quad (\text{B33})$$

for any observable  $O$ .

*Proof.* From the definition of the noisy shadow norm and using the Weingarten functions from Eq. (A32), we have

$$\begin{aligned} \|O_0\|_{\text{shadow}, \Lambda}^2 &= \max_{\sigma \in \mathcal{D}(2^n)} \mathbb{E}_{U \sim \text{Cl}(2^n)} \sum_{b \in \{0,1\}^n} \hat{f}^{-2} \text{Tr}\{(U\sigma U^\dagger \\ &\quad \otimes UO_0U^\dagger \otimes UO_0U^\dagger)[\Lambda^\dagger(|b\rangle\langle b|) \otimes |b\rangle\langle b| \\ &\quad \otimes |b\rangle\langle b|]\} \\ &= \max_{\sigma \in \mathcal{D}(2^n)} \sum_{b \in \{0,1\}^n} \hat{f}^{-2} \text{Tr}\{\Phi_{\text{Haar}}^{(3)}(\sigma \otimes O_0 \otimes O_0) \\ &\quad \times [\Lambda^\dagger(|b\rangle\langle b|) \otimes |b\rangle\langle b| \otimes |b\rangle\langle b|]\} \\ &= \max_{\sigma \in \mathcal{D}(2^n)} \sum_{b \in \{0,1\}^n} \hat{f}^{-2} \sum_{\pi, \xi \in \mathcal{S}_3} c_{\pi, \xi} \\ &\quad \text{Tr}[W_\pi(\sigma \otimes O_0 \otimes O_0)] \\ &\quad \times \text{Tr}\{W_\xi[\Lambda^\dagger(|b\rangle\langle b|) \otimes |b\rangle\langle b| \otimes |b\rangle\langle b|]\}, \end{aligned} \quad (\text{B34})$$

where in the last line we have used the Weingarten function to expand the Haar integral [see Eq. (A32)]. Now using Eq. (B8) to compute the traces appearing above, we have

$$\begin{aligned} \text{Tr}[\tilde{W}(\sigma \otimes O_0 \otimes O_0)] \\ = [0, 0, 0, \text{Tr}(O_0^2), \text{Tr}(\sigma O_0^2), \text{Tr}(\sigma O_0^2)]. \end{aligned} \quad (\text{B35})$$

Recall that  $F_Z(\Lambda)$  is the  $Z$ -basis average fidelity of  $\Lambda$  as defined in Proposition 1, and we denote it simply as  $F_Z$  in the following. Then we also have

$$\begin{aligned} \sum_{b \in \{0,1\}^n} \text{Tr}(\tilde{W}(\Lambda^\dagger(|b\rangle\langle b|) \otimes |b\rangle\langle b| \otimes |b\rangle\langle b|)) \\ = \sum_{b \in \{0,1\}^n} [\text{Tr}(\Lambda^\dagger(|b\rangle\langle b|)), \langle b | \Lambda^\dagger(|b\rangle\langle b|) | b \rangle, \\ \langle b | \Lambda^\dagger(|b\rangle\langle b|) | b \rangle, \text{Tr}(\Lambda^\dagger(|b\rangle\langle b|)), \\ \langle b | \Lambda^\dagger(|b\rangle\langle b|) | b \rangle, \langle b | \Lambda^\dagger(|b\rangle\langle b|) | b \rangle] \\ = d * [1, F_Z(\Lambda), F_Z(\Lambda), 1, F_Z(\Lambda), F_Z(\Lambda)]. \end{aligned} \quad (\text{B36})$$

Again,  $\vec{W}$  is a vectorization of  $S_3$  defined in Eq. (A35), just for the notational simplicity. Inserting the above two equations and the value of the Weingarten matrix from Eq. (A37) into Eq. (B34),

$$\begin{aligned}
 & \|O_0\|_{\text{shadow},\Lambda}^2 \\
 &= \max_{\sigma \in \mathcal{D}(2^n)} \hat{f}^{-2} \\
 &\quad \times \frac{\text{Tr}(O_0^2)(d - 2F_Z + 1) + 2 \text{Tr}(\sigma O_0^2)(dF_Z - 1)}{(d + 2)(d^2 - 1)} \\
 &\leq \hat{f}^{-2} \frac{2dF_Z + d - 2F_Z - 1}{(d + 2)(d^2 - 1)} \text{Tr}(O_0^2) \\
 &= \frac{f^2}{\hat{f}^2} \left( \frac{d^2 - 1}{dF_Z - 1} \right)^2 \frac{2dF_Z + d - 2F_Z - 1}{(d + 2)(d^2 - 1)} \text{Tr}(O_0^2) \\
 &\leq \frac{f^2}{\hat{f}^2} \left( F_Z - \frac{1}{d} \right)^{-2} 3 \text{Tr}(O_0^2), \tag{B37}
 \end{aligned}$$

where the first inequality follows from the fact that  $\text{Tr}(\sigma O_0^2) \leq \|O_0^2\|_\infty \leq \text{Tr}(O_0^2)$  and the assumption  $F_Z \geq 1/d$ , and in the second equality we have used the expression for  $f$  from Proposition 1. ■

Compared to Proposition 1 of [38] that states that  $\|O_0\|_{\text{shadow}}^2 \leq 3 \text{Tr}(O_0^2)$ , we conclude the following. As long as noise channel  $\Lambda$  has a  $Z$ -basis fidelity that is not too low and the noise calibration procedure is conducted with sufficiently many rounds, then the estimation procedure of our **RShadow** protocol using  $\text{Cl}(2^n)$  is as efficient as the noiseless standard quantum shadow estimation protocol [38] up to a small multiplicative factor. That is, the expectation value of any observable  $O$  that has small Hilbert-Schmidt norm can be efficiently estimated by **RShadow**.

To complete the discussion, we give the following theorem as a rigorous version of Theorem 2 in the main text.

**Theorem 8.** *For **RShadow** with  $\text{Cl}(2^n)$ , given that the noise channel satisfies  $F_Z(\Lambda) \geq 1/d$ , if the number of calibration samples  $R_C$  and the number of estimation samples  $R_E$  satisfy*

$$\begin{aligned}
 R_C &= 136 \ln(2\delta_1^{-1}) \frac{(1 + \varepsilon_1^2)(1 + 1/d)^2}{\varepsilon_1^2(F_Z - 1/d)^2}, \\
 R_E &= \frac{204}{\varepsilon_2^2} \ln\left(\frac{2M}{\delta_2}\right) (1 + \varepsilon_1)^2 \left(F_Z - \frac{1}{d}\right)^{-2}, \tag{B38}
 \end{aligned}$$

respectively, then the protocol can estimate  $M$  arbitrary linear functions  $\text{Tr}(O_1\rho), \dots, \text{Tr}(O_M\rho)$  such that  $\text{Tr}(O_i^2) \leq 1$ , up to accuracy  $\varepsilon_1 + \varepsilon_2$  with success probability at least  $1 - \delta_1 - \delta_2$ .

*Proof.* First, according to Theorem 7, for the given number of samples  $R_C$ , we have

$$|\mathbb{E}(\hat{\rho}_i^{(r)}) - \text{Tr}(O_i\rho)| \leq \varepsilon_1. \tag{B39}$$

Meanwhile, from the proof of Theorem 7 [see Eq. (B19)], we also have

$$|\hat{f}^{-1}f - 1| \leq \varepsilon_1 \implies \hat{f} \geq (1 + \varepsilon_1)^{-1}f. \tag{B40}$$

Both of the above equations hold simultaneously with probability at least  $1 - \delta_1$ .

Now, by Lemmas 5 and 6, the single-round estimators in the estimation procedure satisfy

$$\text{Var}(\hat{\rho}_i^{(r)}) \leq 3(1 + \varepsilon_1)^2 \left(F_Z - \frac{1}{d}\right)^{-2}. \tag{B41}$$

So we set the median of mean estimators  $\hat{\rho}_i$  of the estimation procedure with the parameters

$$N = \frac{34}{\varepsilon_2^2} \times 3(1 + \varepsilon_1)^2 \left(F_Z - \frac{1}{d}\right)^{-2}, \quad K = 2 \ln\left(\frac{2M}{\delta_2}\right). \tag{B42}$$

Then it follows from Lemma 4 combined with the union bound that the following statement holds for all  $i$  with probability at least  $1 - \delta_2$ :

$$|\hat{\rho}_i - \mathbb{E}(\hat{\rho}_i^{(r)})| \leq \varepsilon_2. \tag{B43}$$

Combining Eqs. (B39) and (B43) using the triangular inequality gives

$$|\hat{\rho}_i - \text{Tr}(O_i\rho)| \leq \varepsilon_1 + \varepsilon_2, \tag{B44}$$

which holds with probability at least  $1 - \delta_1 - \delta_2$ . This completes the proof. ■

### APPENDIX C: SAMPLE COMPLEXITY OF RSHADOW WITH THE LOCAL CLIFFORD GROUP

The result in Appendix B is based on the  $n$ -qubit Clifford group, which is challenging to implement in the experiment. In this section, we analyze the protocol using the  $n$ -qubit local Clifford group, denoted as  $\text{Cl}_2^{\otimes n}$ , which is the  $n$ -fold direct product of the single-qubit Clifford group. Such unitaries are all single-qubit operations, and thus much easier to implement in the experiment.

### 1. Calibration procedure: local

Being twirled by the local Clifford group, channel  $\tilde{\mathcal{M}}$  becomes a Pauli channel that is symmetric among the  $X, Y, Z$  indices, whose Pauli-Liouville representation is [57]

$$\tilde{\mathcal{M}} = \mathbb{E}_{U \sim \text{Cl}_2^{\otimes n}} U^\dagger M_z \Lambda U = \sum_{z \in \{0,1\}^n} f_z \Pi_z, \quad (\text{C1})$$

where  $\Pi_z = \bigotimes_{i=1}^n \Pi_{z_i}$ ,

$$\Pi_{z_i} = \begin{cases} |\sigma_0\rangle\langle\sigma_0|, & z_i = 0, \\ I - |\sigma_0\rangle\langle\sigma_0|, & z_i = 1, \end{cases} \quad (\text{C2})$$

and  $f_z$  is the Pauli fidelity. In the noiseless case, one can obtain  $f_z = 3^{-|z|}$ , where  $|z|$  is the number of 1s in  $z$ .

*Notation.* For any string  $m \in \{0,1\}^n$ , we define  $|m\rangle\rangle$  to be the Liouville representation of the computational basis state  $|m\rangle$ , while  $|\sigma_m\rangle\rangle$  stands for the normalized Pauli operator corresponding to  $P_m := \bigotimes_{i=1}^n P_Z^{m_i}$ . On the other hand, the notation of  $z$  in this section consistently stands for an  $n$ -bit string and should not be confused with the Pauli- $Z$  index.

The **RShadow** protocol using local Clifford group can be written as follows.

#### Protocol 2 (RShadow with $\text{Cl}_2^{\otimes n}$ ).

1. Prepare  $|0\rangle \equiv |0\rangle^{\otimes n}$ . Sample  $U$  uniformly from  $\text{Cl}_2^{\otimes n}$  and apply it to  $|0\rangle$ .
2. Measure the above state in the computational basis. Denote the outcome state vector as  $|b\rangle$ .
3. Calculate the single-round Pauli fidelity estimator  $\hat{f}_z^{(r)} = \langle\langle b | \mathcal{U} | P_z \rangle\rangle$  for all  $z \in \{0,1\}^n$ .
4. Repeat steps 1–3 for  $R = NK$  rounds. Then the final estimation of  $f_z$  is given by a median of means estimator  $\hat{f}_z$  constructed from the single-round estimators  $\{\hat{f}_z^{(r)}\}_{r=1}^R$  with parameters  $N, K$  [see Eq. (B21)].
5. After the above steps, apply the standard shadow estimation protocol of Ref. [38] on  $\rho$ , with the inverse channel  $\tilde{\mathcal{M}}^{-1}$  replaced by

$$\widehat{\mathcal{M}}^{-1} = \sum_{z \in \{0,1\}^n} \hat{f}_z^{-1} \Pi_z. \quad (\text{C3})$$

Of course, it is unaffordable in classical computational resources to compute all  $\hat{f}_z^{(r)}$  in a single round. In practice, we only need to compute those  $f_z$  of interest. For example, if we only want to predict  $k$ -local properties then only  $\hat{f}_z^{(r)}$  such that  $|z| \leq k$  need to be computed. If we are only interested in nearby qubits then the number of necessary  $\hat{f}_z^{(r)}$  can be further reduced.

Now we show that the single-round estimators  $\{\hat{f}_z^{(r)}\}$  are unbiased and their variance is bounded.

**Proposition 2.** *The single-round Pauli fidelity estimator  $\hat{f}_z^{(r)}$  satisfies*

$$\mathbb{E}(\hat{f}_z^{(r)}) = f_z = 3^{-|z|} \Gamma_\Lambda(z), \quad \text{Var}(\hat{f}_z^{(r)}) \leq 3^{-|z|}, \quad (\text{C4})$$

where  $\Gamma_\Lambda(z) := (1/2^n) \sum_{x,b \in \{0,1\}^n} (-1)^{z \cdot (x \oplus b)} \langle\langle b | \Lambda | x \rangle\rangle$ .

*Proof.* To begin with, we show that  $\hat{f}_z^{(r)}$  is an unbiased estimator of  $f_z$ . From the definition of  $\hat{f}_z^{(r)}$  in Protocol 2, the expectation value over the experiments is given by

$$\begin{aligned} \mathbb{E}(\hat{f}_z^{(r)}) &= \mathbb{E}_{U \sim \text{Cl}_2^{\otimes n}} \sum_b \langle\langle P_z | \mathcal{U}^\dagger | b \rangle\rangle \langle\langle b | \Lambda \mathcal{U} | 0 \rangle\rangle \\ &= \langle\langle P_z | \tilde{\mathcal{M}} | 0 \rangle\rangle \\ &= f_z \langle\langle P_z | 0 \rangle\rangle \\ &= f_z. \end{aligned} \quad (\text{C5})$$

To derive the expression for  $\hat{f}_z$  that depends on the noise channel  $\Lambda$ , we can alternatively expand the expectation as

$$\begin{aligned} \mathbb{E}(\hat{f}_z^{(r)}) &= \mathbb{E}_{U \sim \text{Cl}_2^{\otimes n}} \sum_b \langle b | \Lambda(U|0\rangle\langle 0|U^\dagger) | b \rangle \\ &\quad \times \text{Tr}[U^\dagger | b\rangle\langle b| U P_z] \\ &= \sum_b \text{Tr}\{\mathbb{E}_{U \sim \text{Cl}_2^{\otimes n}}(U|0\rangle\langle 0|U^\dagger \otimes U P_z U^\dagger)\} \\ &\quad \times \Lambda^\dagger(|b\rangle\langle b| \otimes |b\rangle\langle b|). \end{aligned} \quad (\text{C6})$$

To evaluate this expression, we first consider the single-qubit case. By direct calculation we obtain

$$\begin{aligned} \mathbb{E}_{U \sim \text{Cl}_2}(U|0\rangle\langle 0|U^\dagger \otimes U P_I U^\dagger) &= \frac{1}{2}I, \\ \mathbb{E}_{U \sim \text{Cl}_2}(U|0\rangle\langle 0|U^\dagger \otimes U P_Z U^\dagger) &= \frac{2}{3}P_{\text{sym}^2} - \frac{1}{2}I. \end{aligned} \quad (\text{C7})$$

Hence, for any  $X \in \text{Herm}(2)$  and  $b \in \{0,1\}$ , by Lemma 3,

$$\begin{aligned} &\text{Tr}[\mathbb{E}_{U \sim \text{Cl}_2}(U|0\rangle\langle 0|U^\dagger \otimes U P_I U^\dagger)(X \otimes |b\rangle\langle b|)] \\ &= \frac{1}{2}(\langle b | X | b \rangle + \langle b \oplus 1 | X | b \oplus 1 \rangle), \\ &\text{Tr}[\mathbb{E}_{U \sim \text{Cl}_2}(U|0\rangle\langle 0|U^\dagger \otimes U P_Z U^\dagger)(X \otimes |b\rangle\langle b|)] \\ &= \frac{1}{6}(\langle b | X | b \rangle - \langle b \oplus 1 | X | b \oplus 1 \rangle). \end{aligned} \quad (\text{C8})$$

Applying this to the  $n$ -qubit case, one can then verify that

$$\begin{aligned} \mathbb{E}(\hat{f}_z^{(r)}) &= \frac{1}{3^{|z|}} \frac{1}{2^n} \sum_{x,b} (-1)^{z \cdot (x \oplus b)} \langle x | \Lambda^\dagger(|b\rangle\langle b|) | x \rangle \\ &= \frac{1}{3^{|z|}} \Gamma_\Lambda(z). \end{aligned} \quad (\text{C9})$$

To compute the variance, we compute

$$\begin{aligned} \mathbb{E}(\hat{f}_z^{(r)^2}) &= \mathbb{E}_{U \sim \mathcal{C}_2^{\otimes n}} \sum_b \langle b | \Lambda(U|0\rangle\langle 0|U^\dagger) | b \rangle \\ &\quad \times \text{Tr}[U^\dagger | b \rangle \langle b | U P_z]^2 \\ &= \sum_b \text{Tr}\{\mathbb{E}_{U \sim \mathcal{C}_2^{\otimes n}} (U|0\rangle\langle 0| U^\dagger \otimes U P_z U^\dagger \\ &\quad \otimes U P_z U^\dagger) [\Lambda^\dagger(|b\rangle\langle b|) \otimes |b\rangle\langle b| \otimes |b\rangle\langle b|]\}. \end{aligned} \quad (\text{C10})$$

Again, first consider the single-qubit case. One can verify that

$$\begin{aligned} \mathbb{E}_{U \sim \mathcal{C}_2}(U|0\rangle\langle 0| U^\dagger \otimes U P_I U^\dagger \otimes U P_I U^\dagger) &= \frac{1}{2} I, \\ \mathbb{E}_{U \sim \mathcal{C}_2}(U|0\rangle\langle 0| U^\dagger \otimes U P_Z U^\dagger \otimes U P_Z U^\dagger) \\ &= \frac{1}{2} P_{\text{sym}^3} + \frac{1}{3} (P_{\text{sym}^2}^{(2,3)} - P_{\text{sym}^2}^{(1,2)} - P_{\text{sym}^2}^{(1,3)}). \end{aligned} \quad (\text{C11})$$

Hence, for any  $X \in \text{Herm}(2)$  and  $b \in \{0, 1\}$ , by Lemma 3,

$$\begin{aligned} \text{Tr}[\mathbb{E}_{U \sim \mathcal{C}_2}(U|0\rangle\langle 0| U^\dagger \otimes U P_I U^\dagger \otimes U P_I U^\dagger)(X \otimes |b\rangle\langle b| \\ \otimes |b\rangle\langle b|)] &= \frac{1}{2} \text{Tr}(X), \\ \text{Tr}[\mathbb{E}_{U \sim \mathcal{C}_2}(U|0\rangle\langle 0| U^\dagger \otimes U P_Z U^\dagger \\ \otimes U P_Z U^\dagger)(X \otimes |b\rangle\langle b| \otimes |b\rangle\langle b|)] &= \frac{1}{6} \text{Tr}(X). \end{aligned} \quad (\text{C12})$$

One can also verify these equations using the Weingarten matrix. Applying to the  $n$ -qubit case, one can verify that

$$\begin{aligned} \mathbb{E}(\hat{f}_z^{(r)^2}) &= \frac{1}{2^n} \frac{1}{3^{|z|}} \sum_b \text{Tr}[\Lambda^\dagger(|b\rangle\langle b|)] \\ &= \frac{1}{2^n} \frac{1}{3^{|z|}} \sum_{x,b} \langle b | \Lambda(|x\rangle\langle x|) | b \rangle \\ &= \frac{1}{3^{|z|}}. \end{aligned} \quad (\text{C13})$$

Since  $\mathbb{E}(\hat{f}_z^{(r)^2})$  serves as an upper bound for  $\text{Var}(\hat{f}_z^{(r)})$ , this completes the proof of Proposition 2.  $\blacksquare$

Based on Proposition 2, we can now bound the sample complexity of Protocol 2. First, we set the median of mean estimator  $\hat{f}_z$  according to Lemma 4 as

$$\bar{f}_z^{(t)} := \frac{1}{N} \sum_{r=(t-1)N+1}^{tN} \hat{f}_z^{(r)}, \quad t = 1, 2, \dots, K, \quad (\text{C14})$$

$$\hat{f}_z := \text{median}\{\bar{f}_z^{(1)}, \bar{f}_z^{(2)}, \dots, \bar{f}_z^{(K)}\}, \quad (\text{C15})$$

with  $N$  and  $K$  to be specified. The following theorem gives the performance of Protocol 2.

**Theorem 9.** *Given  $\varepsilon, \delta > 0$ , the number of qubits  $n \geq 2$ , and an integer  $k \leq n$ ,*

$$R = \mathcal{O}\left(\frac{3^k(k \ln n + \ln \delta^{-1})}{\varepsilon^2 \min_{|z| \leq k} \Gamma_\Lambda^2(z)}\right) \quad (\text{C16})$$

*samples for the calibration procedure are enough for the subsequent shadow estimation procedure to estimate any  $k$ -local observable for any state to the precision*

$$\begin{aligned} |\langle\langle O | \widehat{\mathcal{M}}^{-1} \widetilde{\mathcal{M}} | \rho \rangle\rangle - \langle\langle O | \rho \rangle\rangle| &\leq \varepsilon 2^k \|O\|_\infty \\ \text{for all } k\text{-local } O \in \text{Herm}(2^n) \text{ and all } \rho \in \mathcal{D}(2^n) \end{aligned} \quad (\text{C17})$$

*with a success probability of at least  $1 - \delta$ .*

Here, An operator  $O$  is called  $k$  local if it only nontrivially acts on a  $k$ -qubit subspace, i.e.,  $O = \tilde{O}_S \otimes I_{[n] \setminus S}$  for some index set  $S \subset [n]$  and  $|S| = k$ .

*Proof of Theorem 9.* We first note that

$$\begin{aligned} &|\langle\langle O | \widehat{\mathcal{M}}^{-1} \widetilde{\mathcal{M}} | \rho \rangle\rangle - \langle\langle O | \rho \rangle\rangle| \\ &= \left| \sum_{a \in \mathbb{Z}_2^{2n}} (\hat{f}_{z(a)}^{-1} f_{z(a)} - 1) \langle\langle O | \sigma_a \rangle\rangle \langle\langle \sigma_a | \rho \rangle\rangle \right| \\ &\leq \max_{|z| \leq k} |\hat{f}_z^{-1} f_z - 1| \cdot \sum_{a \in \mathbb{Z}_2^{2n}} |\langle\langle O | \sigma_a \rangle\rangle| \cdot |\langle\langle \sigma_a | \rho \rangle\rangle| \\ &\leq \max_{|z| \leq k} |\hat{f}_z^{-1} f_z - 1| \cdot \sum_{a \in \mathbb{Z}_2^{2n}} \frac{1}{2^n} |\langle\langle O | P_a \rangle\rangle|, \end{aligned} \quad (\text{C18})$$

where the first equality follows by expanding the Pauli transfer basis and defining the mapping  $z$  as

$$z : \mathbb{Z}_2^{2n} \rightarrow \{0, 1\}^n, \quad z(p)_i = \begin{cases} 0, & (P_p)_i = I, \\ 1, & (P_p)_i \neq I, \end{cases} \quad (\text{C19})$$

and the first inequality uses the fact that  $O$  is  $k$  local. Now we bound the second factor of the above equation. Without loss of generality, suppose that  $O$  acts nontrivially on the first  $k$  qubits, i.e.,  $O = \tilde{O} \otimes I_{2^{n-k}}$ , and that  $\tilde{O}$  can be decomposed as

$$\tilde{O} = \sum_{\tilde{a} \in \mathbb{Z}_2^{2k}} \alpha_{\tilde{a}} P_{\tilde{a}}. \quad (\text{C20})$$

Then we naturally have

$$O = \tilde{O} \otimes I_{2^{n-k}} = \sum_{\tilde{a} \in \mathbb{Z}_2^{2k}} \alpha_{\tilde{a}} P_{\tilde{a}} \otimes P_I^{\otimes(n-k)}. \quad (\text{C21})$$

So,

$$\begin{aligned} \sum_{a \in \mathbb{Z}_2^{2^n}} \frac{1}{2^n} |\langle O | P_a \rangle| &= \sum_{\tilde{a} \in \mathbb{Z}_2^{2^k}} |\alpha_{\tilde{a}}| \leq \sqrt{4^k} \sqrt{\sum_{\tilde{a} \in \mathbb{Z}_2^{2^k}} \alpha_{\tilde{a}}^2} \\ &= 2^k \sqrt{\frac{\text{Tr}(\tilde{O}^2)}{2^k}} \leq 2^k \|\tilde{O}\|_\infty = 2^k \|O\|_\infty, \end{aligned} \quad (\text{C22})$$

where the first inequality follows by the Cauchy-Schwarz inequality. Combining the above results, we have

$$|\langle O | \widehat{\mathcal{M}}^{-1} \widetilde{\mathcal{M}} | \rho \rangle - \langle O | \rho \rangle| \leq \max_{|z| \leq k} |\hat{f}_z^{-1} f_z - 1| \cdot 2^k \|O\|_\infty. \quad (\text{C23})$$

For any  $z \in \{0, 1\}^n$ , suppose that  $|\tilde{f}_z - f_z| \leq \gamma_z$ . Then we have

$$|1 - \hat{f}_z^{-1} f_z| \leq \frac{|\hat{f}_z - f_z|}{|\hat{f}_z|} \leq \frac{\gamma_z}{|f_z| - \gamma_z}. \quad (\text{C24})$$

By setting  $\gamma_z = \varepsilon |f_z| / (1 + \varepsilon)$ , the above equation is upper bounded by  $\varepsilon$ . Therefore, if we set

$$N = 34 \text{Var}(\hat{f}_z) / \gamma_z^2, \quad K = 2 \ln(2\delta^{-1})$$

for the median of mean estimator in Eqs. (C14) and (C15), by Lemma 4 we have  $|1 - \hat{f}_z^{-1} f_z| \leq \varepsilon$  with a success probability of at least  $1 - \delta$ . Now we want all  $z \in \{0, 1\}^n$  such that  $|z| \leq k$  to satisfy this inequality. The number of such strings is no larger than  $n^k$ , so we set

$$N = \max_{|z| \leq k} 34 \text{Var}(\hat{f}_z) / \gamma_z^2 \leq 34 \times 3^k \frac{(1 + \varepsilon)^2}{\varepsilon^2} \max_{|z| \leq k} \Gamma_\Lambda^{-2}(z), \quad (\text{C25})$$

$$K = 2 \ln[2(\delta/n^k)^{-1}], \quad (\text{C26})$$

and apply the union bound. Now we have  $|1 - \hat{f}_z^{-1} f_z| \leq \varepsilon$  for all  $|z| \leq k$  with probability at least  $1 - \delta$ . Our final upper bound of the sample complexity is

$$R = NK \leq 68 \times 3^k \frac{(1 + \varepsilon)^2}{\varepsilon^2} (k \ln n + \ln 2\delta^{-1}) \max_{|z| \leq k} \Gamma_\Lambda^{-2}(z), \quad (\text{C27})$$

which completes the proof.  $\blacksquare$

The quantity  $\Gamma_\Lambda(z)$  can be lower bounded when  $\Lambda$  is close to an identity channel, as shown by the following lemma.

**Lemma 7.** *if the Z-basis average fidelity of  $\Lambda$  satisfies  $F_Z(\Lambda) \geq 1 - c$  for some  $0 \leq c \leq 1$  then  $\Gamma_\Lambda(z) \geq 1 - 2c$  for all  $z \in \{0, 1\}^n$ .*

*Proof.* We have

$$\begin{aligned} \Gamma_\Lambda(z) &= \frac{1}{2^n} \sum_{x, \delta \in \{0, 1\}^n} (-1)^{z \cdot \delta} \langle x \oplus \delta | \Lambda | x \rangle \\ &\geq \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} \left( \langle x | \Lambda | x \rangle - \sum_{\delta \in \{0, 1\}^n, |\delta| \neq 0} \langle x \oplus \delta | \Lambda | x \rangle \right) \\ &= \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} (2 \langle x | \Lambda | x \rangle - 1) \\ &= 2F_Z(\Lambda) - 1 \\ &\geq 1 - 2c, \end{aligned} \quad (\text{C28})$$

where the second equality follows from the fact that  $\Lambda$  is trace preserving, and hence  $\sum_{b \in \{0, 1\}^n} \langle b | \Lambda | x \rangle = 1$ .  $\blacksquare$

Specifically, if we substitute the bound for  $\Gamma_\Lambda(z)$  from Lemma 7 into the above theorem, we get Theorem 3 in the main text. We conclude that our Protocol 2 can mitigate the noise in the computation of the expectation of any  $k$ -local observable efficiently, given that  $k$  is small and the noise is weak.

## 2. Estimation procedure: local

Now we consider the **RShadow** estimation procedure using  $\mathbf{C}_2^{\otimes n}$ . Thanks to Lemma 5, we only need to characterize  $\|\cdot\|_{\text{shadow}, \Lambda}^2$ . Because of technical difficulties, we are currently not able to bound  $\|\cdot\|_{\text{shadow}, \Lambda}^2$  for the most general noise channel  $\Lambda$ , but we do have results for local noise channel  $\Lambda$  (hence also for any separable  $\Lambda$  by linearity). Suppose that  $\Lambda \equiv \bigotimes_{i=1}^n \Lambda_i$ , and denote the Z-basis fidelity of the qubit channels  $\Lambda_i$  as  $F_{Z,i}$ . Furthermore, assume that  $O$  is  $k$  local, which means that it is nontrivially supported on only  $k$  qubits. We have

$$\begin{aligned} &\|O\|_{\text{shadow}, \Lambda}^2 \\ &= \max_{\sigma \in \mathcal{D}(2^n)} \mathbb{E}_{U \sim \mathbf{C}_2^{\otimes n}} \sum_{b \in \{0, 1\}^n} \\ &\quad \times \text{Tr}[\{\sigma \otimes \widehat{\mathcal{M}}^{-1}(O) \otimes \widehat{\mathcal{M}}^{-1}(O)\} U^{\dagger \otimes 3} \\ &\quad \times [\Lambda^\dagger(|b\rangle\langle b|) \otimes |b\rangle\langle b| \otimes |b\rangle\langle b|] U^{\otimes 3}\}. \end{aligned} \quad (\text{C29})$$

Consider the single-qubit case. We have

$$\begin{aligned}
 \Phi_i &:= \mathbb{E}_{U \sim \text{Cl}_2} \sum_{b=0,1} U^{\dagger \otimes 3} [\Lambda_i^\dagger (|b\rangle \langle b|) \otimes |b\rangle \langle b| \otimes |b\rangle \langle b|] U^{\otimes 3} \\
 &= \sum_{b=0,1} \Phi_{\text{Haar}}^{(3)} [\Lambda_i^\dagger (|b\rangle \langle b|) \otimes |b\rangle \langle b| \otimes |b\rangle \langle b|] \\
 &= \sum_{b=0,1} \sum_{\pi, \xi \in S_3} c_{\pi, \xi} W_\pi \text{Tr}\{W_\xi [\Lambda_i^\dagger (|b\rangle \langle b|) \otimes |b\rangle \langle b| \otimes |b\rangle \langle b|]\} \\
 &= \frac{1}{12} [(3 - 2F_{Z,i})(W_0 + W_{(2,3)}) + (2F_{Z,i} - 1)(W_{(1,2)} + W_{(1,3)} + W_{(1,2,3)} + W_{(1,3,2)})], \tag{C30}
 \end{aligned}$$

where we have used the Weingarten function to expand the Haar integral [see Eq. (A32)], and the value of the Weingarten matrix is from Eq. (A38).

For any  $X \in \text{Herm}(2^n)$  and single-qubit Pauli operators  $P_p, P_q$ , we want to calculate the quantity  $\text{Tr}[(X \otimes P_p \otimes P_q)\Phi_i]$ . By direct calculation using Eq. (C30), one can verify that there are four different cases:

$$\begin{aligned}
 \text{Tr}[(X \otimes P_p \otimes P_q)\Phi_i] &= \text{Tr}(XP_pP_q) \\
 &\begin{cases} 1, & P_p = P_q = I, \\ \frac{1}{3}, & P_p = P_q \neq I, \\ \frac{2F_{Z,i} - 1}{3}, & (P_p = I, P_q \neq I) \text{ or } (P_p \neq I, P_q = I), \\ 0, & \text{otherwise.} \end{cases} \tag{C31}
 \end{aligned}$$

This indicates that the value  $\text{Tr}[(X \otimes P_p \otimes P_q)\Phi_i]$  is nonzero if and only if the two single-qubit Pauli operators  $P_p$  and  $P_q$  commute.

Now we return to the evaluation of Eq. (C29). Our strategy is similar to that of Ref. [38]. We first decompose  $O$  into the Pauli operator basis (note that we use unnormalized Pauli operators here)

$$O \equiv \sum_{p \in \mathbb{Z}_2^{2n}} \alpha_p P_p \quad \text{for } \alpha_p \in \mathbb{R}. \tag{C32}$$

Since  $O$  is  $k$  local, we have  $\alpha_p = 0$  for all  $|p| > k$ , where, for any  $p \in \mathbb{Z}_2^{2n}$ , we denote the Pauli weight of  $P_p$  by  $|p|$ . Also, recall from Eq. (C3) that

$$\widehat{\mathcal{M}} = \sum_{p \in \mathbb{Z}_2^{2n}} \hat{f}_{z(p)} |\sigma_p\rangle \langle \sigma_p|, \tag{C33}$$

where we define  $z$  as the mapping

$$\begin{aligned}
 z : \mathbb{Z}_2^{2n} &\rightarrow \{0, 1\}^n, \quad z(p)_i = 0 \text{ if and only if } (P_p)_i = I \\
 &\times \text{ for all } i \in [n]. \tag{C34}
 \end{aligned}$$

The intuition is that after twirling over the local Clifford group the Pauli  $X, Y, Z$  indexes are symmetrized.

Now we can calculate Eq. (C29) as

$$\begin{aligned}
 \|O\|_{\text{shadow}, \Lambda}^2 &= \max_{\sigma \in \mathcal{D}(2^n)} \sum_{p, q \in \mathbb{Z}_2^{2n}} \hat{f}_{z(p)}^{-1} \hat{f}_{z(q)}^{-1} \alpha_p \alpha_q \text{Tr} \left[ (\sigma \otimes P_p \otimes P_q) \left( \bigotimes_{i=1}^n \Phi_i \right) \right] \\
 &= \max_{\sigma \in \mathcal{D}(2^n)} \sum_{p, q \in \mathbb{Z}_2^{2n}} \hat{f}_{z(p)}^{-1} \hat{f}_{z(q)}^{-1} \alpha_p \alpha_q \delta(p, q) \text{Tr}(\sigma P_p P_q) \frac{\prod_{i \in [n]: (P_{p,i} = I, P_{q,i} \neq I) \vee (P_{p,i} \neq I, P_{q,i} = I)} (2F_{Z,i} - 1)}{3^{|p \vee q|}} \\
 &= \left\| \sum_{p, q \in \mathbb{Z}_2^{2n}} \hat{f}_{z(p)}^{-1} \hat{f}_{z(q)}^{-1} \alpha_p \alpha_q \delta(p, q) P_p P_q \frac{\prod_{i \in [n]: (P_{p,i} = I, P_{q,i} \neq I) \vee (P_{p,i} \neq I, P_{q,i} = I)} (2F_{Z,i} - 1)}{3^{|p \vee q|}} \right\|_\infty \\
 &\leq \sum_{p, q \in \mathbb{Z}_2^{2n}} \left| \hat{f}_{z(p)}^{-1} \hat{f}_{z(q)}^{-1} \alpha_p \alpha_q \delta(p, q) \frac{\prod_{i \in [n]: (P_{p,i} = I, P_{q,i} \neq I) \vee (P_{p,i} \neq I, P_{q,i} = I)} (2F_{Z,i} - 1)}{3^{|p \vee q|}} \right| \\
 &\leq \sum_{p, q \in \mathbb{Z}_2^{2n}} \delta(p, q) 3^{|p \wedge q|} |\alpha_p| |\alpha_q| \frac{|\hat{f}_{z(p)}^{-1} \hat{f}_{z(q)}^{-1}|}{3^{|p|} 3^{|q|}} \\
 &\leq \left( \sum_{p, q \in \mathbb{Z}_2^{2n}} \delta(p, q) 3^{|p \wedge q|} |\alpha_p| |\alpha_q| \right) \cdot \left( \max_{z \in \{0, 1\}^n: |z| \leq k} \frac{\hat{f}_z^{-2}}{3^{2|z|}} \right). \tag{C35}
 \end{aligned}$$



Here, for the second equality, we apply the single-qubit result from Eq. (C31). The functional  $\delta(p, q)$  equals 1 if  $P_{p_i}$  commutes with  $P_{q_i}$  for all  $i \in [n]$  and equals 0 otherwise, and we have the definitions

$$\begin{aligned} |p \vee q| &:= \text{No. } \{i \in [n] : P_{p,i} \neq I \text{ or } P_{q,i} \neq I\}, \\ |p \wedge q| &:= \text{No. } \{i \in [n] : P_{p,i} \neq I \text{ and } P_{q,i} \neq I\}. \end{aligned} \quad (\text{C36})$$

The third equality follows from the dual characterization of the operator norm. The first inequality follows from the fact that the operator norm of a Pauli operator is 1. The second inequality follows by relaxing  $F_{Z,i}$  to 1 and noting that  $|p \wedge q| = |p \vee q| - |p| - |q|$ . The last inequality uses the  $k$ -local property of  $O$ .

The first factor of Eq. (C35) can be bounded using the same method as in Ref. [38]. We reproduce their proof here for the convenience of the reader. Without loss of generality, suppose that  $O$  is supported on the first  $k$  qubits, and hence can be written as  $O = \tilde{O} \otimes I_{2^{n-k}}$ . The decomposition of  $\tilde{O}$  is denoted as

$$\tilde{O} = \sum_{p \in \mathbb{Z}_2^{2k}} \tilde{\alpha}_p P_p. \quad (\text{C37})$$

For any two  $q, s \in \mathbb{Z}_2^{2n}$ , we write  $q \triangleright s$  if one can obtain  $P_q$  from  $P_s$  by replacing some single-qubit Pauli operators of  $P_s$  with  $I$ . Then

$$\begin{aligned} & \sum_{p, q \in \mathbb{Z}_2^{2n}} \delta(p, q) 3^{|p \wedge q|} |\alpha_p| |\alpha_q| \\ &= \sum_{p, q \in \mathbb{Z}_2^{2k}} \delta(p, q) 3^{|p \wedge q|} |\tilde{\alpha}_p| |\tilde{\alpha}_q| \\ &= \frac{1}{3^k} \sum_{P_s \in \{P_X, P_Y, P_Z\}^{\otimes k}} \left( \sum_{q: q \triangleright s} 3^{|q|} |\tilde{\alpha}_q| \right)^2 \\ &\leq \frac{1}{3^k} \sum_{P_s \in \{P_X, P_Y, P_Z\}^{\otimes k}} \left( \sum_{q: q \triangleright s} 3^{|q|} \right) \left( \sum_{q: q \triangleright s} 3^{|q|} |\tilde{\alpha}_q|^2 \right) \\ &= 4^k \sum_{P_s \in \{P_X, P_Y, P_Z\}^{\otimes k}} \sum_{q: q \triangleright s} 3^{|q|-k} |\tilde{\alpha}_q|^2 \\ &= 4^k \sum_{q \in \mathbb{Z}_2^{2k}} |\tilde{\alpha}_q|^2 \\ &= 2^k \text{Tr}(\tilde{O}^2) \\ &\leq 4^k \|\tilde{O}\|_\infty^2 \\ &= 4^k \|O\|_\infty^2, \end{aligned} \quad (\text{C38})$$

where in the first equality we restricted attention to the first  $k$  qubits, the second equality can be verified by checking the coefficients of every  $|\tilde{\alpha}_p| |\tilde{\alpha}_q|$ , the first inequality follows from the Cauchy-Schwarz inequality, and the

third and fourth equalities follow from simple combinatoric arguments. The penultimate equality follows from the definition of  $\tilde{O}$ , the last inequality follows from the relationship between the Hilbert-Schmidt norm and the operator norm, and the last equality follows from the fact that the largest eigenvalue of  $O$  equals that of  $\tilde{O}$ .

On the other hand, suppose that the preceding calibration procedure guarantees that  $\hat{f}_z \geq \delta f_z$  for all  $|z| \leq k$  for some positive number  $\delta$  close to 1. Then the second term of Eq. (C35) can be bounded by Proposition 2 as

$$\max_{|z| \leq k} \frac{\hat{f}_z^{-2}}{3^{2|z|}} \leq \delta^{-2} \max_{|z| \leq k} \frac{f_z^{-2}}{3^{2|z|}} = \delta^{-2} \max_{|z| \leq k} \Gamma_\Lambda(z)^{-2}. \quad (\text{C39})$$

Since  $\Lambda$  is assumed to be local noise, we have the following bound for  $\Gamma_\Lambda(z)$ , which could be better than Lemma 7.

**Lemma 8.** *Suppose that  $\Lambda := \bigotimes_{i=1}^n \Lambda_i$  and satisfies  $F_Z(\Lambda_i) \geq 1 - \xi$  for all  $i \in [n]$  and some  $0 \leq \xi < 1/2$ . Then*

$$\Gamma_\Lambda(z) \geq (1 - 2\xi)^{|z|} \quad \text{for all } z \in \{0, 1\}^n. \quad (\text{C40})$$

*Proof.* We have

$$\begin{aligned} \Gamma_\Lambda(z) &= \frac{1}{2^n} \sum_{x, \delta \in \{0, 1\}^n} (-1)^{z \cdot \delta} \langle x \oplus \delta | \Lambda | x \rangle \\ &= \frac{1}{2^n} \prod_{i=1}^n \sum_{x, \delta \in \{0, 1\}} (-1)^{z_i \cdot \delta} \langle x \oplus \delta | \Lambda_i | x \rangle \\ &= \frac{1}{2^{|z|}} \prod_{i: z_i=1} \sum_{x, \delta \in \{0, 1\}} (-1)^\delta \langle x \oplus \delta | \Lambda_i | x \rangle \\ &= \prod_{i: z_i=1} \left( \sum_{x \in \{0, 1\}} \langle x | \Lambda_i | x \rangle - 1 \right) \\ &= [2F_Z(\Lambda_i) - 1]^{|z|} \\ &\geq (1 - 2\xi)^{|z|}, \end{aligned} \quad (\text{C41})$$

where the third equality follows from fact that  $\Lambda_i$  is trace preserving, and hence  $\sum_{x, \delta \in \{0, 1\}} \langle x \oplus \delta | \Lambda_i | x \rangle = 2$ , so we can eliminate those indexes  $i$  such that  $z_i = 0$ . ■

Combining Lemma 8 with Eq. (C39) yields the following lemma. (Note that we substitute  $O$  with its traceless part  $O_0$  in order to use Lemma 5 later.)

**Lemma 9.** *For **RShadow** using  $\text{Cl}_2^{\otimes n}$ , suppose that the noise is local, i.e.,  $\Lambda := \bigotimes_{i=1}^n \Lambda_i$ , and satisfies  $F_Z(\Lambda_i) \geq 1 - \xi$  for all  $i \in [n]$  and some  $0 \leq \xi < 1/2$ . Then, if the calibration procedure guarantees that  $\hat{f}_z \geq \delta f_z$  for all*

$|z| \leq k$  and some  $\delta > 0$ , we have

$$\|O_0\|_{\text{shadow},\Lambda}^2 \leq \delta^{-2}(1 - 2\xi)^{-2k} 4^k \|O\|_\infty^2 \quad (\text{C42})$$

for any  $k$ -local observable  $O$ .

Compared to Proposition 2 of Ref. [38] that states that  $\|O_0\|_{\text{shadow}}^2 \leq 4^k \|O\|_\infty^2$ , we conclude that, when the separable noise channel  $\Lambda$  has not too low  $Z$ -basis fidelity per qubit and the noise calibration procedure is conducted with sufficiently many rounds, the estimation procedure of our **RShadow** protocol using  $\text{Cl}_2^{\otimes n}$  is as efficient as the noiseless standard quantum shadow estimation protocol [38] up to a small multiplicative factor. That is, the expectation value of any observable  $O$  located on a  $k$ -qubit subsystem can be efficiently estimated.

To complete the discussion, we give the following theorem as a rigorous version of Theorem 4 in the main text.

**Theorem 10.** For **RShadow** with  $\text{Cl}_2^{\otimes n}$ , suppose that the noise is local, i.e.,  $\Lambda := \bigotimes_{i=1}^n \Lambda_i$ , and satisfies  $F_Z(\Lambda_i) \geq 1 - \xi$  for all  $i \in [n]$  and some  $0 \leq \xi < 1/2$ . Then, if the number of calibration samples  $R_C$  and the number of estimation samples  $R_E$  satisfy

$$\begin{aligned} R_C &= 68 \times 3^k \left(1 + \frac{2^k}{\varepsilon_1}\right)^2 (k \ln n + \ln 2\delta^{-1})(1 - 2\xi)^{-2k}, \\ R_E &= \frac{34}{\varepsilon_2^2} \times 4^k \ln\left(\frac{2M}{\delta_2}\right) (1 + \varepsilon_1)^2 (1 - 2\xi)^{-2k}, \end{aligned} \quad (\text{C43})$$

respectively, then the protocol can estimate  $M$  arbitrary linear functions  $\text{Tr}(O_1\rho), \dots, \text{Tr}(O_M\rho)$  such that  $\|O_i\|_\infty \leq 1$  and that  $O_i$  is  $k$  local, up to accuracy  $\varepsilon_1 + \varepsilon_2$  with success probability at least  $1 - \delta_1 - \delta_2$ .

*Proof.* First, according to Theorem 9, for the given number of samples  $R_C$ , we have

$$|\mathbb{E}(\hat{o}_i^{(r)}) - \text{Tr}(O_i\rho)| \leq \varepsilon_1. \quad (\text{C44})$$

Note that we apply the bound for  $\Gamma_\Lambda(z)$  from Lemma 8.

Meanwhile, from the proof of Theorem 9 [see Eq. (B19)], we also have

$$\begin{aligned} |\hat{f}_z^{-1} f_z - 1| &\leq \varepsilon_1 \\ \implies \hat{f}_z &\geq (1 + \varepsilon_1)^{-1} f_z \quad \text{for all } |z| \leq k. \end{aligned} \quad (\text{C45})$$

Both equations hold simultaneously with probability at least  $1 - \delta_1$ .

Now, by Lemmas 5 and 9, the single-round estimators in the estimation procedure satisfy

$$\text{Var}(\hat{o}_i^{(r)}) \leq 4^k (1 + \varepsilon_1)^2 (1 - 2\xi)^{-2k}. \quad (\text{C46})$$

So we set the median of mean estimators  $\hat{o}_i$  of the estimation procedure with the parameters

$$N = \frac{34}{\varepsilon_2^2} \times 4^k (1 + \varepsilon_1)^2 (1 - 2\xi)^{-2k}, \quad K = 2 \ln\left(\frac{2M}{\delta_2}\right). \quad (\text{C47})$$

Then it follows from Lemma 4 combined with the union bound that the following statement holds for all  $i$  with probability at least  $1 - \delta_2$ :

$$|\hat{o}_i - \mathbb{E}(\hat{o}_i^{(r)})| \leq \varepsilon_2. \quad (\text{C48})$$

Combining Eqs. (C44) and (C48) using the triangular inequality gives

$$|\hat{o}_i - \text{Tr}(O_i\rho)| \leq \varepsilon_1 + \varepsilon_2, \quad (\text{C49})$$

which holds with probability at least  $1 - \delta_1 - \delta_2$ . This completes the proof. ■

Specifically, if  $\xi \ll 1/2$  then  $(1 - 2\xi)^{-2k} = [(1 - 2\xi)^{-1/2\xi}]^{4k\xi} \approx e^{4k\xi}$ . That is how we get the bound in Theorem 4.

#### APPENDIX D: THE EFFECT OF STATE PREPARATION NOISE

In this section, we prove Theorems 5 and 6 in the main text establishing the robustness of **RShadow** against state preparation noise in the calibration procedure. Let us first fix the notation. We assume that  $|\mathbf{0}\rangle$  is experimentally prepared as some other state  $\rho_0$  that is fixed over time, and we use a subscript ‘‘SP’’ to denote the state preparation noisy version of our estimators. For example,  $\hat{\mathcal{M}}_{\text{SP}} = \sum_{\lambda \in R_C} \hat{f}_{\lambda, \text{SP}} \Pi_\lambda$  is our estimation for the physical channel  $\hat{\mathcal{M}} := \sum_{\lambda \in R_C} f_\lambda \Pi_\lambda$  when the calibration process suffers from state preparation error.

##### 1. Robustness of RShadow with the global Clifford group

**Lemma 10.** For **RShadow** using  $\text{Cl}(2^n)$ , if the state preparation fidelity satisfies

$$F(|\mathbf{0}\rangle\langle\mathbf{0}|, \rho_0) \geq 1 - \varepsilon_{\text{SP}} \quad (\text{D1})$$

then the SP-noisy single-round estimator  $\hat{f}_{\text{SP}}^{(r)}$  satisfies

$$\begin{aligned} f &\geq \mathbb{E}(\hat{f}_{\text{SP}}^{(r)}) \geq (1 - 2\varepsilon_{\text{SP}})f, \\ \text{Var}(\hat{f}_{\text{SP}}^{(r)}) &\leq \frac{6d}{(d - 1)^3}. \end{aligned} \quad (\text{D2})$$

*Proof.* According to the calibration procedure described in Algorithm 2 or Protocol 1 of Appendix B, we have

$$\begin{aligned}\mathbb{E}(\hat{F}_{\text{SP}}^{(r)}) &= \mathbb{E}_{U \sim \text{Cl}(2^n)} \sum_b \langle \langle \mathbf{0} | U^\dagger | b \rangle \rangle \langle \langle b | \Lambda U | \rho_0 \rangle \rangle \\ &= \langle \langle \mathbf{0} | [\sigma_0] \rangle \rangle \langle \langle \sigma_0 | + f(I - |\sigma_0\rangle \langle \sigma_0|) | \rho_0 \rangle \rangle \\ &= \frac{1}{d} + f \left( \langle \langle \mathbf{0} | \rho_0 | \mathbf{0} \rangle \rangle - \frac{1}{d} \right),\end{aligned}\quad (\text{D3})$$

$$\begin{aligned}\mathbb{E}(\hat{f}_{\text{SP}}^{(r)}) &= \frac{d\mathbb{E}(\hat{F}_{\text{SP}}^{(r)}) - 1}{d-1} \\ &= \frac{d \langle \langle \mathbf{0} | \rho_0 | \mathbf{0} \rangle \rangle - 1}{d-1} f \\ &\geq \left( 1 - \varepsilon_{\text{SP}} \frac{d}{d-1} \right) f.\end{aligned}\quad (\text{D4})$$

One can immediately conclude that  $f \geq \mathbb{E}(\hat{f}_{\text{SP}}^{(r)}) \geq (1 - 2\varepsilon_{\text{SP}})f$ .

The second moment of  $\hat{F}_{\text{SP}}^{(r)}$  can be written as [see Eq. (B16)]

$$\begin{aligned}\mathbb{E}(\hat{F}_{\text{SP}}^{(r)2}) &= \sum_{b \in \{0,1\}^n} \text{Tr}\{\mathbb{E}_{U \sim \text{Cl}(2^n)}(U \rho_0 U^\dagger \otimes U | \mathbf{0} \rangle \langle \mathbf{0} | U^\dagger \\ &\quad \otimes U | \mathbf{0} \rangle \langle \mathbf{0} | U^\dagger) [\Lambda^\dagger(|b\rangle \langle b|) \otimes |b\rangle \langle b| \otimes |b\rangle \langle b|]\} \\ &= \sum_{b \in \{0,1\}^n} \text{Tr}\{\Phi_{\text{Haar}}^{(3)}(\rho_0 \otimes | \mathbf{0} \rangle \langle \mathbf{0} | \otimes | \mathbf{0} \rangle \langle \mathbf{0} |) \\ &\quad [\Lambda^\dagger(|b\rangle \langle b|) \otimes |b\rangle \langle b| \otimes |b\rangle \langle b|]\} \\ &= \sum_{b \in \{0,1\}^n} \sum_{\pi, \sigma \in S_3} c_{\pi, \sigma} \text{Tr}[W_\pi(\rho_0 \otimes | \mathbf{0} \rangle \langle \mathbf{0} | \otimes | \mathbf{0} \rangle \langle \mathbf{0} |)] \\ &\quad \times \text{Tr}\{W_\sigma[\Lambda^\dagger(|b\rangle \langle b|) \otimes |b\rangle \langle b| \otimes |b\rangle \langle b|]\} \\ &= \frac{2(d - 2F_Z - 2F_0 + 2dF_Z F_0)}{(d^2 - 1)(d + 2)} \\ &\leq \frac{6d}{(d^2 - 1)(d + 2)},\end{aligned}\quad (\text{D5})$$

where we have defined  $F_0 := \langle \langle \mathbf{0} | \rho_0 | \mathbf{0} \rangle \rangle$  and  $F_Z := F_Z(\Lambda)$ . Therefore,

$$\begin{aligned}\text{Var}(\hat{f}_{\text{SP}}^{(r)}) &= \frac{d^2}{(d-1)^2} \text{Var}(\hat{F}_{\text{SP}}^{(r)}) \leq \frac{d^2}{(d-1)^2} \mathbb{E}(\hat{F}_{\text{SP}}^{(r)2}) \\ &\leq \frac{6d}{(d-1)^3}.\end{aligned}\quad (\text{D6})$$

This completes the proof.  $\blacksquare$

The following theorem is a more detailed formalization of Theorem 5 in the main text.

**Theorem 11.** For *RShadow* using  $\text{Cl}(2^n)$ , if the state preparation fidelity satisfies

$$F(|\mathbf{0}\rangle \langle \mathbf{0}|, \rho_0) \geq 1 - \varepsilon_{\text{SP}} \quad (\text{D7})$$

then, with  $R = \tilde{\mathcal{O}}(\varepsilon^{-2} F_Z^{-2})$  calibration samples, the subsequent estimation procedure with high probability satisfies

$$|\mathbb{E}(\hat{\delta}^{(r)}) - \text{Tr}(O\rho)| \leq (\varepsilon + 2\varepsilon_{\text{SP}}) \|O\|_\infty \quad (\text{D8})$$

up to first orders in  $\varepsilon$  and  $\varepsilon_{\text{SP}}$  for any observable  $O$ . We have assumed that  $F_Z := F_Z(\Lambda) \gg 1/d$ .

*Proof.* First note that the target function can be upper bounded as

$$\begin{aligned}|\mathbb{E}(\hat{\delta}^{(r)}) - \text{Tr}(O\rho)| &= |\langle \langle O | \widehat{\mathcal{M}}_{\text{SP}}^{-1} \widetilde{\mathcal{M}} - 1 | \rho \rangle \rangle| \\ &= |\langle \langle O_0 | \widehat{\mathcal{M}}_{\text{SP}}^{-1} \widetilde{\mathcal{M}} - 1 | \rho \rangle \rangle| \\ &\leq |\langle \langle O_0 | \rho \rangle \rangle| \cdot |\hat{f}_{\text{SP}}^{-1} f - 1| \\ &\leq \|O\|_\infty \cdot |\hat{f}_{\text{SP}}^{-1} f - 1|.\end{aligned}\quad (\text{D9})$$

According to Lemma 4, by taking the parameters of the median of mean estimators as

$$\begin{aligned}N &= 34 \text{Var}(\hat{f}_{\text{SP}}^{(r)}) \varepsilon^{-2} f^{-2}, \\ K &= 2 \ln(2\delta^{-1}),\end{aligned}\quad (\text{D10})$$

the following holds with probability at least  $1 - \delta$ :

$$|\hat{f}_{\text{SP}} - \mathbb{E}(\hat{f}_{\text{SP}}^{(r)})| \leq \varepsilon f. \quad (\text{D11})$$

We also have, from Lemma 10,

$$|\mathbb{E}(\hat{f}_{\text{SP}}^{(r)}) - f| \leq 2\varepsilon_{\text{SP}} f. \quad (\text{D12})$$

Therefore, our final bound is, as claimed,

$$\begin{aligned}|\mathbb{E}(\hat{\delta}^{(r)}) - \text{Tr}(O\rho)| &\leq \|O\|_\infty \times \frac{|f - \hat{f}_{\text{SP}}|}{|\hat{f}_{\text{SP}}|} \\ &\leq \|O\|_\infty \times \frac{\varepsilon + 2\varepsilon_{\text{SP}}}{1 - \varepsilon - 2\varepsilon_{\text{SP}}} \\ &= \|O\|_\infty \cdot [\varepsilon + 2\varepsilon_{\text{SP}} + o(\varepsilon + 2\varepsilon_{\text{SP}})].\end{aligned}\quad (\text{D13})$$

The sample complexity is

$$\begin{aligned}R = NK &\leq 2 \ln(2\delta^{-1}) \cdot 204\varepsilon^{-2} \left( F_Z - \frac{1}{d} \right)^{-2} \\ &= \frac{(d+1)^2}{d(d-1)} = \tilde{\mathcal{O}}(\varepsilon^{-2} F_Z^{-2})\end{aligned}\quad (\text{D14})$$

for  $F_Z := F_Z(\Lambda) \gg 1/d$ . Here we have used Lemma 10 and Proposition 1 to bound  $\text{Var}(\hat{f}_{\text{SP}}^{(r)})$  and  $f$ , respectively.  $\blacksquare$

## 2. Robustness of RShadow with the local Clifford group

Note that we consider a local state preparation noise model for the results in this section, i.e., no crosstalk between qubits.

**Lemma 11.** *For RShadow using  $\text{Cl}_2^{\otimes n}$ , if the prepared state is in a product form, i.e.,  $\rho_0 = \bigotimes_{i=1}^n \rho_{0,i}$ , and the single-qubit state preparation fidelity satisfies*

$$F(|0\rangle\langle 0|, \rho_{0,i}) \geq 1 - \xi_{\text{SP}} \quad \text{for all } i \in [n] \quad (\text{D15})$$

for some  $\xi_{\text{SP}} < 1/2$ , then the SP-noisy single-round estimator  $\hat{f}_{z,\text{SP}}^{(r)}$  satisfies

$$\begin{aligned} f_z &\geq \mathbb{E}(\hat{f}_{z,\text{SP}}^{(r)}) \geq (1 - 2\xi_{\text{SP}}|z|)f_z, \\ \text{Var}(\hat{f}_{z,\text{SP}}^{(r)}) &\leq 3^{-|z|} \quad \text{for all } z \in \{0, 1\}^n. \end{aligned} \quad (\text{D16})$$

*Proof.* According to the calibration procedure described in Algorithm 2 or Protocol 2 of Appendix C, we have

$$\begin{aligned} \mathbb{E}(\hat{f}_{z,\text{SP}}^{(r)}) &= \mathbb{E}_{U \sim \text{Cl}_2^{\otimes n}} \sum_b \langle\langle P_z | \mathcal{U}^\dagger | b \rangle\rangle \langle\langle b | \Lambda \mathcal{U} | \rho_0 \rangle\rangle \\ &= \langle\langle P_z | \sum_{m \in \{0,1\}^n} f_m \Pi_m | \rho_0 \rangle\rangle \\ &= f_z \langle\langle P_z | \rho_0 \rangle\rangle \\ &= f_z \prod_{i:z_i=1} (2 \langle 0 | \rho_{0,i} | 0 \rangle - 1) \\ &\geq (1 - 2|z| \xi_{\text{SP}})f_z. \end{aligned} \quad (\text{D17})$$

One can immediately conclude that  $f_z \geq \mathbb{E}(\hat{f}_{z,\text{SP}}^{(r)}) \geq (1 - 2|z| \xi_{\text{SP}})f_z$ .

To calculate the second moment,

$$\begin{aligned} \mathbb{E}(\hat{f}_{z,\text{SP}}^{(r)2}) &= \sum_b \text{Tr} \{ \mathbb{E}_{U \sim \text{Cl}_2^{\otimes n}} (U \rho_0 U^\dagger \otimes U P_z U^\dagger \\ &\quad \otimes U P_z U^\dagger) [\Lambda^\dagger (|b\rangle\langle b| \otimes |b\rangle\langle b| \otimes |b\rangle\langle b|)], \end{aligned} \quad (\text{D18})$$

we can first investigate the single-qubit case:

$$\begin{aligned} \mathbb{E}_{U \sim \text{Cl}_2} (U \rho_{0,i} U^\dagger \otimes U P_i U^\dagger \otimes U P_i U^\dagger) &= \frac{1}{2} I_2^{\otimes 3}, \\ \mathbb{E}_{U \sim \text{Cl}_2} (U \rho_{0,i} U^\dagger \otimes U P_z U^\dagger \otimes U P_z U^\dagger) \\ &= \Phi_{\text{Haar}}^{(3)}(\rho_{0,i} \otimes P_z \otimes P_z). \end{aligned} \quad (\text{D19})$$

To further simplify the second expressions, one can verify that

$$\text{Tr}[\vec{W}(\rho_{0,i} \otimes P_z \otimes P_z)] = [0, 0, 0, 2, 1, 1], \quad (\text{D20})$$

where  $\vec{W}$  is defined in Eq. (A35). Calculating the Haar integral using Eq. (A32), one immediately notes that the form

of  $\rho_{0,i}$  has nothing to do with the result. So we can safely replace all  $\rho_{0,i}$  with  $|0\rangle\langle 0|$  and retrieve the result with no state preparation error:  $\mathbb{E}(\hat{f}_{z,\text{SP}}^{(r)2}) = \mathbb{E}(f_z^{(r)2}) = 3^{-|z|}$ ; hence,  $\text{Var}(\hat{f}_{z,\text{SP}}^{(r)}) \leq 3^{-|z|}$ . ■

The following theorem is a more detailed formalization of Theorem 6 in the main text.

**Theorem 12.** *For RShadow using  $\text{Cl}_2^{\otimes n}$ , if the state is prepared as some product state  $\rho_0 = \bigotimes_{i=1}^n \rho_{0,i}$  and the single-qubit state preparation fidelity satisfies*

$$F(|0\rangle\langle 0|, \rho_{0,i}) \geq 1 - \xi_{\text{SP}} \quad \text{for all } i \in [n] \quad (\text{D21})$$

then, with  $R = \tilde{O}(3^k \varepsilon^{-2} F_z^{-2})$  calibration samples, the subsequent estimation procedure with high probability satisfies

$$|\mathbb{E}(\hat{o}^{(r)}) - \text{Tr}(O\rho)| \leq (\varepsilon + 2k\xi_{\text{SP}})2^k \|O\|_\infty \quad (\text{D22})$$

up to first orders in  $\varepsilon$  and  $k\xi_{\text{SP}}$  for any  $k$ -local observable  $O$ .

*Proof.* Suppose that  $O$  is a  $k$ -local observable for some  $k$ . Following exactly the same procedure as in the proof of Theorem 9 [see Eq. (C23)], we can bound our target function as

$$|\mathbb{E}(\hat{o}^{(r)}) - \text{Tr}(O\rho)| \leq 2^k \|O\|_\infty \times \max_{|z| \leq k} |\hat{f}_{z,\text{SP}}^{-1} f_z - 1|. \quad (\text{D23})$$

According to Lemma 4, by taking the parameters of the median of mean estimators as

$$\begin{aligned} N &= \max_{|z| \leq k} 34 \text{Var}(\hat{f}_{z,\text{SP}}^{(r)}) \varepsilon^{-2} f_z^{-2}, \\ K &= 2 \ln[2(\delta/n^k)^{-1}], \end{aligned} \quad (\text{D24})$$

the following holds with probability at least  $1 - \delta/n^k$  for any  $z$  whose weight is no larger than  $k$ , and hence simultaneously holds for all such  $z$  with probability at least  $1 - \delta$  by the union bound:

$$|\hat{f}_{z,\text{SP}} - \mathbb{E}(\hat{f}_{z,\text{SP}}^{(r)})| \leq \varepsilon f_z \quad \text{for all } z \in \{0, 1\}^n : |z| \leq k. \quad (\text{D25})$$

We also have, from Lemma 11,

$$|\mathbb{E}(\hat{f}_{z,\text{SP}}^{(r)}) - f_z| \leq 2\xi_{\text{SP}}|z|f_z \quad \text{for all } z \in \{0, 1\}^n. \quad (\text{D26})$$

Therefore, our final bound is, as claimed,

$$\begin{aligned}
 & |\mathbb{E}(\hat{\delta}^{(r)}) - \text{Tr}(O\rho)| \\
 & \leq 2^k \|O\|_\infty \times \max_{|z| \leq k} \frac{|f_z - \hat{f}_{z,\text{SP}}|}{|\hat{f}_{z,\text{SP}}|} \\
 & \leq 2^k \|O\|_\infty \times \frac{\varepsilon + 2k\xi_{\text{SP}}}{1 - \varepsilon - 2k\xi_{\text{SP}}} \\
 & = 2^k \|O\|_\infty \cdot [\varepsilon + 2k\xi_{\text{SP}} + o(\varepsilon + 2k\xi_{\text{SP}})]. \quad (\text{D27})
 \end{aligned}$$

The sample complexity is

$$\begin{aligned}
 R &= NK \leq 2 \ln(2\delta^{-1}n^k) \cdot 34 \cdot 3^k \varepsilon^{-2} \Gamma_z(\Lambda)^{-2} \\
 &\leq 2 \ln(2\delta^{-1}n^k) \cdot 34 \cdot 3^k \varepsilon^{-2} F_Z(\Lambda)^{-2} \\
 &= \tilde{O}(3^k \varepsilon^{-2} F_Z^{-2}), \quad (\text{D28})
 \end{aligned}$$

where we have used Lemma 11 and Proposition 2 to bound  $\text{Var}(\hat{f}_{z,\text{SP}}^{(r)})$  and  $f_z$ , respectively. The second inequality follows from Lemma 7. We remark that one can alternatively use a stronger bound given in Lemma 8 when the noise model is assumed to be local. ■

## APPENDIX E: MORE NUMERICAL RESULTS

### 1. Coherent and correlated noise with the local Clifford group

Here, we present more numerical results to show the performance of **RShadow** in the task of estimating the (average) two-point  $ZZ$  correlation function of the GHZ state using the local Clifford group. More precisely, the quantity we want to estimate can be expressed as

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \langle \text{GHZ}_n | Z_i Z_{i+1} | \text{GHZ}_n \rangle, \quad (\text{E1})$$

the true value of which is obviously 1. The aim of this appendix is to close one gap between the theory we derived in the main text and practical needs: whether **RShadow** using the local Clifford group is still sample efficient against qubitwise-correlated noises. Our numerical results in Fig. 8 give an affirmative answer. Here, we work with a five-qubit GHZ state. We perform the estimation task under two different noise models: single-qubit  $X$ -axis rotation and two-qubit  $XX$  crosstalk noises. When the single-qubit  $X$ -axis rotation error happens, each qubit will experience a coherent rotation after the implementation of the random unitary gate  $R_X(\theta) = e^{-i\theta X}$ , where  $\theta = k\pi/40$ ,  $k = 0, 1, 2, 3, 4, 5$ . When the two-qubit  $XX$  crosstalk noise happens, each two adjacent qubits will experience a coherent rotation after the implementation of the random unitary gate  $R_{XX}(\theta) = e^{-i\theta XX}$ , where  $\theta = 3k\pi/100$ ,  $k = 0, 1, 2, 3, 4, 5$ . For clarity, we estimate the

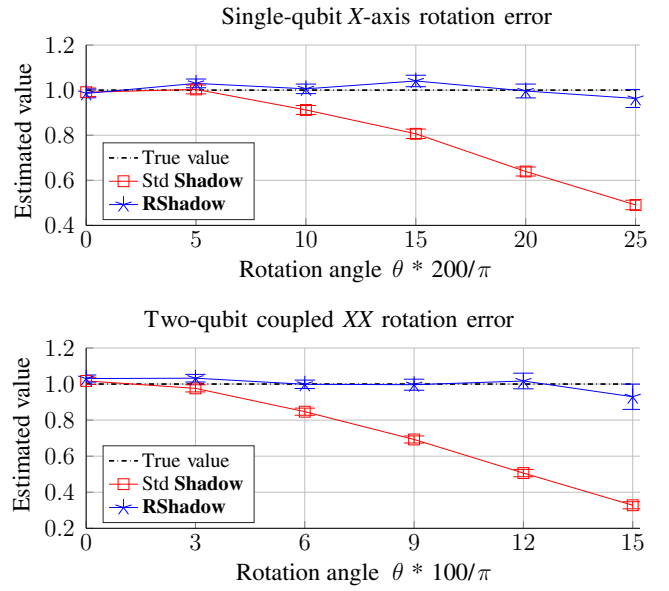


FIG. 8. Five-qubit GHZ two-point correlation function estimation under the coherent noise models, including single-qubit  $X$ -axis rotation and two-qubit  $XX$ -coupling noise.

two-point correlation function of the fifth qubit with every other qubit and plot the average values.

### 2. Gate-dependent noise: more details

Here, we present more details about the gate-dependent noise simulations in Sec. VIII of the main text.

*Gate decomposition.* In our simulations, we decompose each single-qubit Clifford gate using the three generators

$$\left\{ R_P\left(\frac{\pi}{2}\right) = \exp\left(-i\frac{\pi}{4}P\right), P = X, Y, Z \right\}. \quad (\text{E2})$$

To see how this works, note that each single-qubit Clifford gate can be uniquely specified by its conjugation on Pauli  $Z$  and Pauli  $X$  operators. Any single-qubit Clifford gate  $C$  can be equivalently described by the notation

$$\{\text{st}[C] := CZC^\dagger, \text{de}[C] := CXC^\dagger\}, \quad (\text{E3})$$

where  $\text{st}[C]$  and  $\text{de}[C]$  are both one of  $\{\pm X, \pm Y, \pm Z\}$ . They are usually called the stabilizer and destabilizer of  $C$ ,

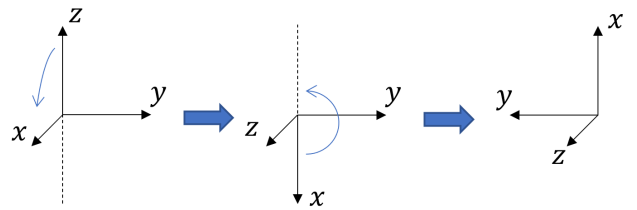


FIG. 9. Rotational decomposition of the Hadamard gate.

respectively. Another way to say this is that  $C$  rotates the 3D Cartesian coordinates so that the  $+X$  ( $+Z$ ) direction is now at the  $\text{st}[C]$  ( $\text{de}[C]$ ) direction.

Now, we can decompose  $C$  into two subsequent rotations. First rotate the  $+Z$  direction into the direction specified by  $\text{st}[C]$  along either the  $X$  or  $Y$  axis. Then rotate the *new*  $+X$  direction into  $\text{de}[C]$  along the  $\text{st}[C]$  axis. Each rotation is of angle  $\{0, \pi/2, \pi, 3\pi/2\}$ , so it can be implemented by concatenating up to three generators.

As an example mentioned in the main text, the Hadamard gate has the stabilizer and destabilizer

$$\{\text{st}[H] = +X, \text{de}[H] = +Z\},$$

so it can be decomposed as  $H = R_X^2(\pi/2)R_Y(\pi/2)$ , as shown in Fig. 9.

*Gate dependence of the noise model.* We claim in the main text that both *pulse miscalibration* and *random over-rotation* are gate-dependent noise models. Here we explain this claim in more detail. The noisy generators for the *pulse miscalibration* error are

$$\left\{ \tilde{R}_P\left(\frac{\pi}{2}\right) = \exp\left[-i\frac{1}{2}\left(\frac{\pi}{2}P + \Delta_0\right)\right], P = X, Y, Z \right\}.$$

By the Baker-Campbell-Hausdorff formula, this can be expanded as

$$\begin{aligned} \tilde{R}_P\left(\frac{\pi}{2}\right) &= \exp\left(-i\frac{\pi}{4}P\right) \exp\left(-i\frac{1}{2}\Delta_0\right) \\ &\times \exp\left(\frac{\pi}{16}[P, \Delta_0]\right) \dots \end{aligned}$$

As long as  $[P, \Delta_0]$  is different for different  $P$ , the noise channel is different for different generators. We also note that the gate-dependent effect only appears in higher-order terms. For the *random over-rotation* error, the noisy

TABLE I. Directions of  $\Delta_0$  that are uniformly randomly sampled from a sphere.

Label of samples	$\mathbf{v}$ (direction of $\Delta_0$ )
1	(−0.550, +0.288, +0.784)
2	(−0.086, +0.987, −0.136)
3	(+0.652, −0.396, +0.647)
4	(+0.589, +0.183, +0.787)
5	(+0.543, −0.781, −0.309)
6	(−0.666, −0.720, −0.196)
7	(+0.174, −0.512, −0.841)
8	(+0.908, +0.417, +0.034)
9	(−0.183, +0.935, −0.304)
10	(+0.599, −0.704, −0.382)

generators are

$$\tilde{R}_P = \exp\left[-i\frac{1}{2}\left(\frac{\pi}{2} + \delta\right)P\right].$$

For a generator  $R_P$ , the noise channel for this noise model can be written as

$$\begin{aligned} \Lambda_P(\rho) &= \int d\delta p(\delta; \sigma) R_P(\delta) \rho R_P^\dagger(\delta) \\ &= \int d\delta p(\delta; \sigma) e^{-i\delta P/2} \rho e^{i\delta P/2}, \end{aligned} \quad (\text{E4})$$

where  $p(\delta; \sigma) = (1/\sigma\sqrt{2\pi}) \exp(-\delta^2/2\sigma^2)$  is the Gaussian distribution. Let  $|P\pm\rangle$  denote the  $\pm 1$  eigenvectors of  $P$ , and carry out the Gaussian integral. Then

$$\begin{aligned} \Lambda_P(\rho) &= |P+\rangle \langle P+| (P+|\rho|P+) \\ &+ |P-\rangle \langle P-| (P-|\rho|P-) \\ &+ e^{-\sigma^2/2} |P+\rangle \langle P-| (P+|\rho|P-) \\ &+ e^{-\sigma^2/2} |P-\rangle \langle P+| (P-|\rho|P+). \end{aligned} \quad (\text{E5})$$

This is a dephasing channel on the  $|P\pm\rangle$  basis. Since this basis is different for different Pauli  $P$ , this noise model is gate dependent by definition.

*Different directions of  $\Delta_0$ .* In the numerical results of Fig. 7, we fixed a random direction for  $\Delta_0$  and only varied its magnitude. Here we show that there is nothing special about the direction we picked. We uniformly sample ten unit vectors  $\mathbf{v}$  and set  $\Delta_0$  to be  $0.1\pi \times (\mathbf{v}_1 X + \mathbf{v}_2 Y + \mathbf{v}_3 Z)$ . For each of these noise settings, we use

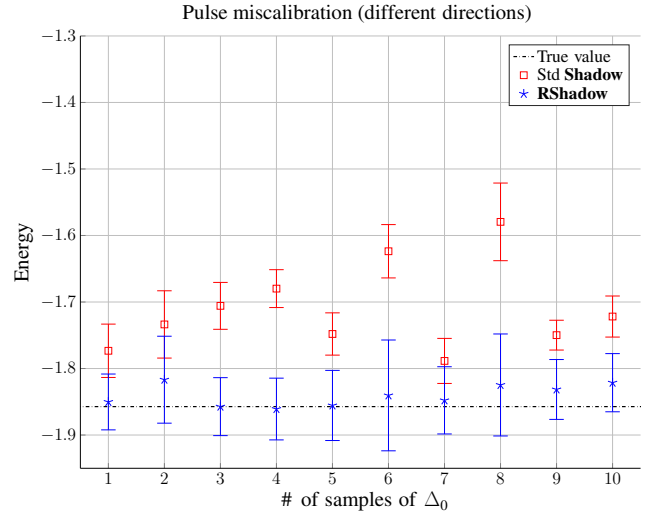


FIG. 10. Ground-state energy estimation of  $\text{H}_2$  with pulse miscalibration noise. Here we choose  $\Delta_0 = 0.1\pi \times (\mathbf{v}_1 X + \mathbf{v}_2 Y + \mathbf{v}_3 Z)$ , where each different label of the  $x$  axis corresponds to a different sample of  $\mathbf{v}$  given in Table I.

$R = 30\,000$  ( $N = 3000, K = 10$ ) calibration samples and  $R = 10\,000$  ( $N = 1000, K = 10$ ) estimation samples for **RShadow**, and  $R = 10\,000$  ( $N = 1000, K = 10$ ) samples for standard **Shadow**. The performance is shown in Fig. 10 in which the average and standard deviation are calculated over ten independent runs. For all cases, **RShadow** gives a more accurate estimate than standard **Shadow**.

- 
- [1] L. Amico, R. Fazio, A. Osterloh, and V. Vedral, Entanglement in many-body systems, *Rev. Mod. Phys.* **80**, 517 (2008).
- [2] X.-L. Qi, Does gravity come from quantum information?, *Nat. Phys.* **14**, 984 (2018).
- [3] P. W. Shor, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994), p. 124.
- [4] L. K. Grover, Quantum Mechanics Helps in Searching for a Needle in a Haystack, *Phys. Rev. Lett.* **79**, 325 (1997).
- [5] S. Lloyd, Universal quantum simulators, *Science* **273**, 1073 (1996).
- [6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, New York, NY, USA, 2011), 10th ed.
- [7] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing* (India, 1984), p. 175.
- [8] A. K. Ekert, Quantum Cryptography Based on Bell's Theorem, *Phys. Rev. Lett.* **67**, 661 (1991).
- [9] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels, *Phys. Rev. Lett.* **70**, 1895 (1993).
- [10] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, Spin squeezing and reduced quantum noise in spectroscopy, *Phys. Rev. A* **46**, R6797 (1992).
- [11] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum Metrology, *Phys. Rev. Lett.* **96**, 010401 (2006).
- [12] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical Characterization of Quantum Devices Without Tomography, *Phys. Rev. Lett.* **107**, 210404 (2011).
- [13] S. T. Flammia and Y.-K. Liu, Direct Fidelity Estimation from Few Pauli Measurements, *Phys. Rev. Lett.* **106**, 230501 (2011).
- [14] F. G. S. L. Brandão, Quantifying entanglement with witness operators, *Phys. Rev. A* **72**, 022310 (2005).
- [15] D. A. Abanin and E. Demler, Measuring Entanglement Entropy of a Generic Many-Body System with a Quantum Switch, *Phys. Rev. Lett.* **109**, 020504 (2012).
- [16] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [17] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [18] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [19] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, *Phys. Rev. X* **8**, 031027 (2018).
- [20] F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec, Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography, *Quantum* **4**, 257 (2020).
- [21] S. Bravyi, S. Sheldon, A. Kandala, D. C. McKay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, *Phys. Rev. A* **103**, 042605 (2021).
- [22] M. McKague, T. H. Yang, and V. Scarani, Robust self-testing of the singlet, *J. Phys. A: Math. Theor.* **45**, 455304 (2012).
- [23] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014).
- [24] O. Gühne and G. Tóth, Entanglement detection, *Phys. Rep.* **474**, 1 (2009).
- [25] G. M. D'Ariano and P. Perinotti, Optimal Data Processing for Quantum Measurements, *Phys. Rev. Lett.* **98**, 020403 (2007).
- [26] M. Guta, J. Kahn, R. J. Kueng, and J. A. Tropp, Fast state tomography with optimal error bounds, *J. Phys. A: Math. Theor.* **53**, 204001 (2020).
- [27] 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).
- [28] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, *New J. Phys.* **14**, 095022 (2012).
- [29] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Permutationally Invariant Quantum Tomography, *Phys. Rev. Lett.* **105**, 250403 (2010).
- [30] T. Moroder, P. Hyllus, G. Tóth, C. Schwemmer, A. Niggelbaum, S. Gaile, O. Gühne, and H. Weinfurter, Permutationally invariant state reconstruction, *New J. Phys.* **14**, 105001 (2012).
- [31] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, *Nat. Commun.* **1**, 1 (2010).
- [32] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, Scalable Reconstruction of Density Matrices, *Phys. Rev. Lett.* **111**, 020401 (2013).
- [33] J. Cotler and F. Wilczek, Quantum Overlapping Tomography, *Phys. Rev. Lett.* **124**, 100401 (2020).
- [34] X. Bonet-Monroig, R. Babbush, and T. E. O'Brien, Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States, *Phys. Rev. X* **10**, 031064 (2020).
- [35] T. J. Evans, R. Harper, and S. T. Flammia, Scalable bayesian hamiltonian learning, [arXiv:1912.07636](https://arxiv.org/abs/1912.07636) [quant-ph] (2019).
- [36] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, *Nat. Mach. Intell.* **1**, 155 (2019).
- [37] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018 (Association for Computing Machinery, New York, NY, USA, 2018), p. 325.

- [38] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [39] M. Painsi, A. Kalev, D. Padilha, and B. Ruck, Estimating expectation values using approximate quantum states, *Quantum* **5**, 413 (2021).
- [40] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, *J. Opt. B* **7**, S347 (2005), [quant-ph/0503243](#).
- [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. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Randomized Benchmarking and Process Tomography for Gate Errors in a Solid-State Qubit, *Phys. Rev. Lett.* **102**, 090502 (2009).
- [43] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [44] J. Helsen, X. Xue, L. M. Vandersypen, and S. Wehner, Soundness and completeness of quantum root-mean-square errors, *Npj Quantum Inf.* **5**, 1 (2019).
- [45] W. Fulton and J. Harris, *Representation Theory: a First Course* (Springer Science & Business Media, New York, NY, USA, 2013), Vol. 129.
- [46] E. Knill, Quantum computing with realistically noisy devices, *Nature* **434**, 39 (2005), [arXiv:quant-ph/0410199](#).
- [47] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, *Nat. Phys.* **16**, 1184 (2020).
- [48] S. T. Flammia and J. J. Wallman, Efficient estimation of pauli Channels, *ACM Transactions on Quantum Computing* **1**, 1 (2020).
- [49] R. Harper, W. Yu, and S. T. Flammia, Fast estimation of sparse quantum Noise, Randomized benchmarking with gate-dependent noise, *PRX Quantum* **2**, 010322 (2021).
- [50] J. J. Wallman, *Quantum* **2**, 47 (2018).
- [51] S. T. Merkel, E. J. Pritchett, and B. H. Fong, Randomized benchmarking as convolution: Fourier analysis of gate dependent errors, [arXiv:1804.05951 \[quant-ph\]](#) (2018).
- [52] Z. Webb, The Clifford Group Forms a Unitary 3-Design, *Quantum Info. Comput.* **16**, 1379 (2016).
- [53] H. Zhu, Multiqubit clifford groups are unitary 3-designs, *Phys. Rev. A* **96**, 062336 (2017).
- [54] R. Kueng and D. Gross, [arXiv:1510.02767](#) (2015).
- [55] D. Gottesman, Ph.D. thesis, Caltech 1997, [quant-ph/9705052](#).
- [56] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, *Phys. Rev. A* **70**, 052328 (2004).
- [57] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Characterization of Addressability by Simultaneous Randomized Benchmarking, *Phys. Rev. Lett.* **109**, 240504 (2012).
- [58] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Rev. Mod. Phys.* **92**, 015003 (2020).
- [59] H. Abraham *et al.*, Qiskit: An open-source framework for quantum computing (2019).
- [60] B. Efron, in *Breakthroughs in statistics* (Springer, New York, NY, USA, 1992), p. 569.
- [61] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, [arXiv:2006.15788](#) (2020).
- [62] H.-Y. Huang, R. Kueng, and J. Preskill, [arXiv:2103.07510](#) (2021).
- [63] S. B. Bravyi and A. Y. Kitaev, Fermionic quantum Computation, *Ann. Phys. (N. Y.)* **298**, 210 (2002).
- [64] One might object to the fact that we have taken four times as many total samples using **RShadow** compared with **Shadow**. While increasing the number of samples in **Shadow** would indeed improve the precision, it would not impact the *accuracy*, which is where **RShadow** outperforms **Shadow** in these numerical simulations.
- [65] Note that the error bars shown here are obtained in a different manner from those in Sec. VII. Here, we also take into account the deviation for the calibration procedure of **RShadow**, so the error bars here look longer than those of Sec. VII.
- [66] R. Takagi, [arXiv:2006.12509](#) (2020).
- [67] Y. Suzuki, S. Endo, and Y. Tokunaga, [arXiv:2010.03887](#) (2020).
- [68] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, Circuit quantum electrodynamics, *Rev. Mod. Phys.* **93**, 025005 (2021).
- [69] D. E. Koh and S. Grewal, Classical shadows with noise, [arXiv:2011.11580](#) (2020).
- [70] E. v. d. Berg, Z. K. Mineev, and K. Temme, [arXiv:2012.09738](#) (2020).
- [71] B. Collins and I. Nechita, Random matrix techniques in quantum information theory, *J. Math. Phys.* **57**, 015215 (2016).
- [72] H. Zhu, R. Kueng, M. Grassl, and D. Gross, [arXiv:1609.08172](#) (2016).
- [73] D. Weingarten, Asymptotic behavior of group integrals in the limit of infinite rank, *J. Math. Phys.* **19**, 999 (1978).
- [74] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability, *Int. Math. Res. Notices* **2003**, 953 (2003).
- [75] B. Collins and P. Śniady, Integration with respect to the haar measure on unitary, orthogonal and symplectic Group, *Commun. Math. Phys.* **264**, 773 (2006).
- [76] D. A. Roberts and B. Yoshida, Chaos and complexity by design, *J. High Energy Phys.* **2017**, 121 (2017).
- [77] P. Zinn-Justin, Jucys–Murphy elements and weingarten Matrices, *Lett. Math. Phys.* **91**, 119 (2010).
- [78] D. Gross, F. Kraemer, and R. Kueng, A partial derandomization of phaseLift using spherical Designs, *J. Fourier Anal. Appl.* **21**, 229 (2015).
- [79] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani, Random generation of combinatorial structures from a uniform distribution, *Theor. Comput. Sci.* **43**, 169 (1986).
- [80] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization* (Chichester: Wiley, New York, NY, USA, 1983).