Shadow tomography based on informationally complete positive operator-valued measure

Atithi Acharya,[1] Siddhartha Saha,[1] and Anirvan M. Sengupta[1,2]
[1]*Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA*
[2]*Center for Computational Mathematics and Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA*

(Received 3 June 2021; accepted 13 September 2021; published 16 November 2021)

Recently introduced shadow tomography protocols use "classical shadows" of quantum states to predict many target functions of an unknown quantum state. Unlike full quantum state tomography, shadow tomography does not insist on accurate recovery of the density matrix for high rank mixed states. Yet, such a protocol makes multiple accurate predictions with high confidence, based on a moderate number of quantum measurements. One particular influential algorithm, proposed by Huang *et al.* [Huang, Kueng, and Preskill, Nat. Phys. **16**, 1050 (2020)], requires additional circuits for performing certain random unitary transformations. In this paper, we avoid these transformations but employ an arbitrary informationally complete positive operator-valued measure and show that such a procedure can compute k-bit correlation functions for quantum states reliably. We also show that, for this application, we do not need the median of means procedure of Huang *et al.* Finally, we discuss the contrast between the computation of correlation functions and fidelity of reconstruction of low rank density matrices.

DOI: 10.1103/PhysRevA.104.052418

I. INTRODUCTION

Recent advances in quantum information processing often require characterizing quantum states prepared during various stages of a procedure. As a result, the problem of characterizing a quantum state, more specifically, a density matrix, from measurements on an ensemble of identical states, known as quantum state tomography, has seen a surge of interest [1–3]. One of the key challenges is that, for n-qubit quantum systems, the density matrix is of size $2^n \times 2^n$. As the number of qubits become large, inferring the density matrix from a limited number of measurements becomes difficult.

Can we get by without fully characterizing the quantum state, but by constructing an approximate classical description that predicts many different functions of the state accurately? Shadow tomography [4] precisely aims to do this, namely, predict a power-law number of observations in number of qubits, n, from $O(n)$ copies of the density matrix ρ. This idea was taken further by Huang *et al.* [2], who have constructed such a description of low sample complexity via classical shadows ($\hat{\rho}$), related to states without any entanglement in the appropriate basis, corresponding to each copy of ρ.

Quantum measurement requires specifying a set of positive operator-valued measures (POVMs) [5], which is a generalization of a complete set of projection operators. The work by Huang *et al.* [2] involves measurements via projection operators. Since projection operators are not informationally complete (see Sec. II), Huang *et al.* employ a set of random unitary transformations before taking measurements. In this paper, we directly employ a complete or overcomplete POVM system and perform shadow tomography. We also gain some insight into how the choice of POVM affects the efficiency of the method.

II. GENERALIZED MEASUREMENTS

A projective measurement is described by an observable, A, a Hermitian operator on the state space of the system being observed. The observable has a spectral decomposition, $A = \sum_a aP_a$, where P_a is the projector onto the eigenspace of A with eigenvalue a. The possible outcomes of the measurement corresponding to the eigenvalues, a, of the observable and the outcome probability are $p(a) = \langle\psi|P_a|\psi\rangle$. Projection valued measures (PVMs) are a special case of general measurements, where the measurement operators are Hermitian and orthogonal projectors. A set of POVMs [5] forms a generalization of PVMs. The index a in the POVM element M_a refers to the measurement outcomes that may occur in the experiment. The probability of the measurement outcome is given by $p(a) = \mathrm{tr}(\rho M_a)$ and the postmeasurement density matrix can be written as $\frac{K_a \rho K_a^\dagger}{\mathrm{tr}(K_a \rho K_a^\dagger)}$, where $\{K_a\}$ are the Kraus operators [5] corresponding to the POVM, with $K_a^\dagger K_a = M_a$. The operators $\{M_a\}$ form a complete set of Hermitian non-negative operators. Namely, they satisfy Hermiticity, $M_a = M_a^\dagger$; positivity, $\langle\psi|M_a|\psi\rangle \geqslant 0$, for any vector $|\psi\rangle$; and completeness, $\sum_a M_a = \mathbb{I}$. Such a POVM could be thought of as a partition of unity by non-negative operators.

Informational completeness

The density matrix, ρ, is a Hermitian and unit trace operator. If we have a d-dimensional system, ρ will be a complex square matrix represented by $d^2 - 1$ real parameters. The operator space for this d-dimensional operator will be however spanned by d^2 linearly independent basis operators. Note that PVMs only have d projection operators. They are capable of providing only the diagonal elements of ρ in a particular orthonormal basis, leaving out potential entanglement-related

 ©2021 American Physical Society

information from the off-diagonal elements. Thus, PVMs are examples of POVMs that are informationally undercomplete.

If the number of outcomes k satisfies $k \geqslant d^2$, and we can form exactly d^2 linearly independent operators by linearly combining the set of POVMs, such POVMs will be called informationally complete (IC). However, in the most common terminology, informationally complete actually refers to the minimally complete POVM ($k = d^2$). If we proceed to reconstruct the density matrix for an informationally complete POVM, we can expand ρ as

$$\rho = \sum_{a=0}^{d^2-1} \xi_a M_a. \tag{1}$$

If $k = d^2$ we have an informationally complete (minimally complete) basis set. However, if we have $k > d^2 \Rightarrow$ it forms an informationally overcomplete set [6]. IC POVMs have been used for entanglement detection [7] and for individual elements of the density matrix [8].

We start out by giving the example of a rather simple overcomplete set in the single-qubit Hilbert space, $\boldsymbol{M}_{\text{Pauli-6}}$. The Pauli-6 POVM has six outcomes $\boldsymbol{M}_{\text{Pauli-6}} = \{M_0 = \frac{1}{3} \times |0\rangle\langle 0|0, M_1 = \frac{1}{3} \times |1\rangle\langle 1|1, M_2 = \frac{1}{3} \times |+\rangle\langle +|+, M_3 = \frac{1}{3} \times |-\rangle\langle -|-, M_4 = \frac{1}{3} \times |l\rangle\langle l|l, M_5 = \frac{1}{3} \times |r\rangle\langle r|r$ where $\{|0\rangle, |1\rangle\}$, $\{|+\rangle, |-\rangle\}$, and $\{|r\rangle, |l\rangle\}$ stand for the eigenbases of the Pauli operators σ^z, σ^x, and σ^y, respectively. Experimentally, it can be implemented directly by first randomly choosing x, y, or z, and then measuring the respective Pauli operator, which justifies the $1/3$ factor. However, other probabilities will also be valid for this example of an overcomplete POVM.

Now, let us give an example of a minimally complete POVM, the Pauli-4 POVM: $\boldsymbol{M}_{\text{Pauli}-4} = \{M_0 = \frac{1}{3} \times |0\rangle\langle 0|0, M_1 = \frac{1}{3} \times |+\rangle\langle +|+, M_2 = \frac{1}{3} \times |l\rangle\langle l|l, M_3 = \frac{1}{3} \times (|1\rangle\langle 1|1 + |-\rangle\langle -|- + |r\rangle\langle r|r)\}$. As a sanity check for the completeness relation, one can see $\sum_a M_a = 1/3(|0\rangle\langle 0|0 + |1\rangle\langle 1|1 + |+\rangle\langle +|+ + |-\rangle\langle -|- + |l\rangle\langle l|l + |r\rangle\langle r|r) = \mathbb{I}$. The experimental procedure will be similar to that of the Pauli-6 POVM, with an additional step where three different outcomes of Pauli-6 are identified as the single element of Pauli-4, M_3. Thus, this set contains an element which is not a rank-1 projector.

The third one is the tetrahedral POVM $\boldsymbol{M}_{\text{tetra}} = \{M_a = \frac{1}{4}(\mathbb{I} + \boldsymbol{s}_a \cdot \boldsymbol{\sigma})\}_{a\in\{0,1,2,3\}}$, the outcomes of which correspond to subnormalized rank-1 projectors along the directions $\boldsymbol{s}_0 = (0, 0, 1)$, $\boldsymbol{s}_1 = (\frac{2\sqrt{2}}{3}, 0, -\frac{1}{3})$, $\boldsymbol{s}_2 = (-\frac{\sqrt{2}}{3}, \sqrt{\frac{2}{3}}, -\frac{1}{3})$, and $\boldsymbol{s}_3 = (-\frac{\sqrt{2}}{3}, -\sqrt{\frac{2}{3}}, -\frac{1}{3})$ in the Bloch sphere. Since the tetrahedron formed is regular, it forms an example of a symmetric informationally complete POVM. The experimental implementation of $\boldsymbol{M}_{\text{tetra}}$ relies on Neumark's dilation theorem. The theorem implies that $\boldsymbol{M}_{\text{tetra}}$ can be physically realized by coupling the system qubit to an ancillary qubit and performing a von Neumann measurement on the two qubits (see [1,9] for explicit constructions).

III. CLASSICAL SHADOWS WITH POVMs

Aaronson introduced the idea of "pretty good tomography" [10], with the focus on predicting many observations accurately, based on N copies of the density matrix. This idea parallels the "learnability" of quantum states in a probably approximately correct sense [11]. Proceeding along this line, he later introduced the concept of shadow tomography [4], where from N copies of the density matrix ρ, we want to predict L different linear target functions $\text{tr}(O_1\rho), \text{tr}(O_2\rho) \ldots \text{tr}(O_L\rho)$ up to an additive error less than ϵ.

Huang *et al.* [2] build their methods on the idea of shadow tomography [4]. They repeatedly perform a measurement procedure, i.e., apply a random unitary to rotate the state ($\rho \mapsto U\rho U^\dagger$) and perform a computational-basis measurement. Then, after the measurement, they apply the inverse of U to the resulting computational basis state. This procedure collapses ρ to a snapshot $U^\dagger|\hat{b}\rangle\langle\hat{b}|U$, producing a quantum channel $\mathcal{M}$, which depends on the ensemble of (random) unitary transformations.

If the collection of unitaries is defined to be tomographically complete, namely, if the condition, i.e., for each $\sigma \neq \rho$ there exist $U \in \mathcal{U}$ and b such that $\langle b|U\sigma U^\dagger|b\rangle \neq \langle b|U\rho U^\dagger|b\rangle$, is met, then $\mathcal{M}$—viewed as a linear map—has a unique inverse $\mathcal{M}^{-1}$. Huang *et al.* [2] set

$$\hat{\rho} = \mathcal{M}^{-1}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U) \quad \text{(classical shadow)}. \tag{2}$$

Although the inverted channel $\mathcal{M}^{-1}$ is not physical (it is not completely positive), one can still apply $\mathcal{M}^{-1}$ to the (classically stored) measurement outcome $U^\dagger|\hat{b}\rangle\langle\hat{b}|U$ in a completely classical postprocessing step. Even if an individual sample of $\hat{\rho}$ is not a density matrix, the expectation of $\hat{\rho}$'s is the original density matrix ρ. One can use this property to get a good prediction of measurements performed on ρ.

If, instead of working with the computational basis measurements, we decide to use an IC POVM (Sec. II), we can avoid dealing with particular random unitary ensembles. The only thing we need to make sure of is that the resulting channel $\mathcal{M}$ is invertible.

A. POVMs for the the *n*-qubit system

From single qubit POVMs $\{M_a\}$, we introduce k^n operators by taking tensor products and form POVMs for the n-qubit system: $\boldsymbol{M} = \{M_{a_1} \otimes M_{a_2} \otimes \ldots \otimes M_{a_n}\}_{a_1,\ldots a_n}$. The outcomes of this measurements in this system are of the form $\vec{\mathbf{a}} = (a_1, a_2, \ldots a_n)$. Now, we discuss how to form shadows from such an observation.

B. A synthetic measurement channel

Here is our scheme for generating a pure state in response to a measurement, a scheme that generalizes the projective-measurement-induced collapse in Ref. [2]. Let the POVM elements be diagonalized as follows: $M_a = \sum_i \lambda_i^a |i, a\rangle\langle i, a|i, a$. We have $\lambda_i^a \geqslant 0$, $\forall i, a$, since $M_a \succeq 0$. The probability of outcome a is given as

$$p(a) = \text{tr}(\rho M_a). \tag{3}$$

Each time we perform a measurement and get an outcome a, we construct a pure output state $|i, a\rangle\langle i, a|i, a$ with probability

$p(i|a)$. We will discuss the exact form of $p(i|a)$ shortly. This process is a quantum channel. Although this is a synthetic channel, we will refer to it as the measurement channel, in analogy with the case where $\{M_a\}$ are projections. Thus, the measurement channel, for a single qubit, is given by

$$\tilde{\rho} = \mathcal{M}(\rho) = \sum_a p(a) \sum_i p(i|a)|i,a\rangle\langle i,a|i,a. \tag{4}$$

For simplicity, in our paper, we will take this probability $p(i|a)$ to be concentrated on the leading eigenvectors of M_a. More precisely, if we have a set of orthogonal states associated with the highest eigenvalue of M_a, we will choose among them with equal probability. When the highest eigenvalue of each M_a is nondegenerate, which is the case for the POVMs we will use, we represent the leading eigenvector by $|\psi_a\rangle$. Here, the measurement channel is just

$$\tilde{\rho} = \mathcal{M}(\rho) = \sum_a \mathrm{tr}(\rho M_a)|\psi_a\rangle\langle\psi_a|\psi_a. \tag{5}$$

This is one of many schemes that could be used. Our algorithm works, as long as $\mathcal{M}$ remains invertible. For a discussion of how the nature of the probability distribution $p(i|a)$ affects the performance of the method, see Appendix 3 for more detail.

In the formalism developed in Ref. [2], the channel and its inversion were related to the ensemble of (random) unitary transformations (e.g., the Clifford unitary ensemble). The condition of tomographical completeness depended on the existence of a unitary transformation in the chosen ensemble to distinguish different density matrices [2]. However, with our reformulation of the measurement channel, we need to use an informationally complete set of POVMs (e.g., Pauli-6, see Sec. II).

In the example of a single qubit measured using the six projectors coming from the three Pauli matrices, i.e., the Pauli-6 POVM, the channel and its inverse can be explicitly computed. Similar to the classical shadows built out of random Pauli measurements [2], we get a depolarizing channel, i.e., a channel that contracts a pure state (lying on the surface of the Bloch sphere) towards the "center" of the Bloch sphere, namely, the maximally mixed state $\rho = \mathbb{I}_2/2$. The inverse (a nonphysical map) can be computed, which can map a point inside the Bloch ball to the outside.

C. Multiqubit system

For local measurements (not necessarily the depolarizing channel), the inverse channel for the n-qubit system can be written as

$$\mathcal{M}_n^{-1} = \bigotimes_{j=1}^{n} \mathcal{M}_1^{-1}. \tag{6}$$

We can now reformulate the shadows with our overcomplete POVM set and its corresponding channel. For instance, when we work with the Pauli-6 POVM, we will get

$$\hat{\rho} = \bigotimes_{j=1}^{n} \mathcal{M}_1^{-1}(|\psi_{a,j}\rangle\langle\psi_{a,j}|\psi_{a,j}) \quad \text{(classical shadow)}, \tag{7}$$

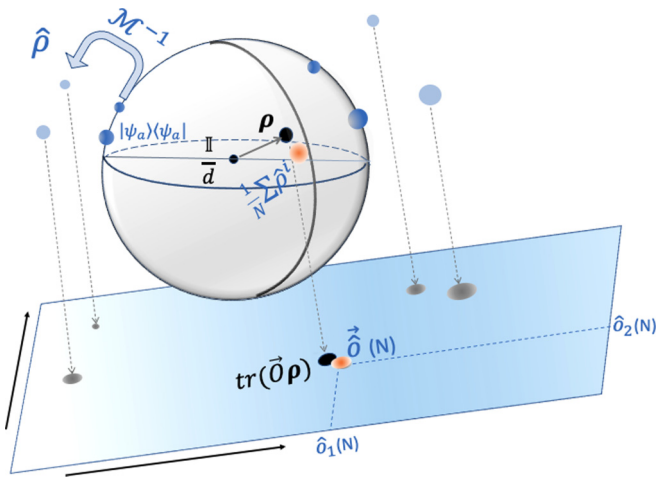

FIG. 1. The convex region in the figure is the set of admissible density matrices. We schematically describe the process of forming classical shadows from N copies of ρ. For the ith observation with outcome a, the inverse of the channel, $\mathcal{M}^{-1}$, acts on the projectors $|\psi_a\rangle\langle\psi_a|$ to construct the shadow $\hat{\rho}^i$. The sample mean of the shadows cast by ρ, i.e., $\frac{1}{N}\sum_i \hat{\rho}^i$, fluctuates around the true ρ and could be outside the convex region. However, while measuring L k-local observables [2] $\vec{O} = (O_1, \ldots O_L)$, the convergence of the sample averages $\vec{\delta}(N)$ to the true expected values can be guaranteed with a number of samples $O(\log L)$. See Theorem 1.

where $\mathcal{M}_1^{-1}(X) = 3X - \mathrm{tr}(X)\mathbb{I}$ (see Appendix 2). Note that the $2^n \times 2^n$ matrix $\hat{\rho}$ need not be constructed explicitly. We just need to store $|\psi_{a,j}\rangle$ for each qubit j.

Since the inverted channel $\mathcal{M}^{-1}$ is not physical (it is not completely positive), the $\hat{\rho}$ in Eq. (7) need not be physical. In other words, there is no guarantee the output of the inverse channel is positive semidefinite. See Fig. 1 for a schematic description. We recover the true density matrix only in expectation. However, if the shadow matrix is forced to be positive semidefinite, we can see how the observations such as fidelity change (see Sec. IV C).

D. Noisy shadow

Earlier, we defined our measurement channel, Eq. (5). However, we can also let each of our qubits pass through a previously characterized noise channel $\mathcal{E}_1$ and then take the measurements [12]. The combined channel $M_{\mathcal{E},1}$ is given by

$$\tilde{\rho} = \mathcal{M}_{\mathcal{E},1}(\rho) = \sum_a \mathrm{tr}(\mathcal{E}_1(\rho)M_a)|\psi_a\rangle\langle\psi_a|\psi_a. \tag{8}$$

We used IC POVMs to ensure that the measurement channel was invertible. As long as the action of the noise channel $\mathcal{E}_1$ itself is invertible, $\mathcal{M}_{\mathcal{E},1}$ is also invertible. We will work with an n-qubit noise channel of the form $\mathcal{E}_n = \bigotimes_{j=1}^{n} \mathcal{E}_1$. Thus, we can still write the inverse of the new noisy measurement channel for the n-qubit system in terms of the single-qubit inverse shadow channel $\mathcal{M}_{\mathcal{E},1}^{-1}$:

$$\mathcal{M}_{\mathcal{E},n}^{-1} = \bigotimes_{j=1}^{n} \mathcal{M}_{\mathcal{E},1}^{-1}. \tag{9}$$

If we choose an amplitude damping channel with damping parameter γ, one of the Kraus operator representations can be given as

$$\mathcal{E}_{AD}(\rho) = K_0 \rho K_0^\dagger + K_1 \rho K_1^\dagger, \tag{10}$$

where $K_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{(1-\gamma)} \end{bmatrix}$, $K_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}$.

The inverse of the noisy shadow channel $\mathcal{M}_{AD}^{-1}(X)$ is given in Eq. (A6). Its action on $\mathbb{I}$ and $\sigma_{x,y,z}$ is given as $\mathcal{M}_{AD}^{-1}(\mathbb{I}) = \mathbb{I} - \frac{\gamma}{1-\gamma}\sigma_z$, $\mathcal{M}_{AD}^{-1}(\sigma_{x,y}) = \frac{3}{\sqrt{1-\gamma}}\sigma_{x,y}$, and $\mathcal{M}_{AD}^{-1}(\sigma_z) = \frac{3}{1-\gamma}\sigma_z$. See Appendix 2 for a general description on the inversion of a noisy shadow channel. Here, we will construct the shadows (noisy) with the following definition:

$$\hat{\rho} = \bigotimes_{j=1}^{n} \mathcal{M}_{AD}^{-1}(|\psi_{a,j}\rangle\langle\psi_{a,j}|\psi_{a,j}). \tag{11}$$

E. Predicting linear functions with classical shadows

Using the statistical properties of a single shadow, we can predict linear functions in the unknown state ρ as

$$o = \mathrm{tr}(O\rho) = \mathbb{E}[\hat{o}], \quad \text{where} \quad \hat{o} = \mathrm{tr}(O\hat{\rho}). \tag{12}$$

In practice, using an array of shadows (i.e., N snapshots), we can estimate the expectation o. Given an array of N independent classical snapshots [each defined as in Eq. (7)],

$$\mathsf{S}(\rho; N) = \{\hat{\rho}^{(1)}, \hat{\rho}^{(1)}, \ldots, \hat{\rho}^{(N)}\}. \tag{13}$$

The sample mean is $\bar{o} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{tr}(O\hat{\rho}^{(j)})$. This sample mean will fluctuate around the true prediction, with $\mathbb{E}(\bar{o}) = o$.

F. The algorithm and the guarantee of performance

We want to predict the expected value of multiple k-local observables $O_1, \ldots O_L$ based on shadows using the two algorithms below.

Algorithm 1: Generating Shadows with POVMs.

Input: IC POVM with k outcomes, $\rho \in \mathcal{C}^{2^n}$ (N copies of the unknown density matrix)
1 Compute the measurement channel $\mathcal{M}_1$ and its inverse $\mathcal{M}_1^{-1}$ for the chosen IC POVM. (See Appendix 2);
2 **for** $i = 1, \ldots N$ **do**
3 Perform measurements using the POVM elements M_a to get outcomes $a_{ji} \in \{1, \ldots, k\}$;
4 Construct shadows $\hat{\rho}_i = \bigotimes_{j=1}^{n} \mathcal{M}_1^{-1}(|\psi_{a_{ji}}\rangle\langle\psi_{a_{ji}}|\psi_{a_{ji}})$ (See Sec. III B and Appendix 3 for the general version)
5 **end**
Output: $\hat{\rho}_1, \hat{\rho}_2 \ldots \hat{\rho}_N$

Algorithm 2: Predicting many properties using mean as an estimate.

Input: A POVM set, N copies of unknown density matrix ρ, L different k-local Pauli observables $O_1, O_2, \ldots O_L$ and error parameters ϵ, δ
1 Find bounds on the local observables $B(\{O\}, \mathcal{M})$. (See Appendix 4 for details).
2 Using Algorithm I, collect $N \geqslant \frac{B(\{O\},\mathcal{M})\log(\frac{2L}{\delta})}{2\epsilon^2}$ shadows.
3 Compute means $\hat{o}_i = \frac{1}{N}\sum_{j=1}^{N} \mathrm{tr}(O_i\hat{\rho}^{(j)})$
Output: $\bar{o}_1, \bar{o}_2 \ldots \bar{o}_L$

The existence of the bound is guaranteed by the following theorem.

Theorem 1. With $N \geqslant \frac{B(\{O\},\mathcal{M})\log(\frac{2L}{\delta})}{2\epsilon^2}$ samples of ρ, we can predict L different linear target functions $\mathrm{tr}(O_1\rho)$, $\mathrm{tr}(O_2\rho), \ldots, \mathrm{tr}(O_L\rho)$ up to additive error ϵ with maximum failure probability δ.

The constant bound $B(\{O\}, \mathcal{M})$ will depend on the measurement channel $\mathcal{M}$ (which depends on the choice of POVM) and on the operator set $\{O\} = \{O_1, O_2, .., O_L\}$. The important thing is that $B(\{O\}, \mathcal{M}$ is bounded for so-called k-local operators, as defined in Ref. [2].

For instance, if we choose Pauli-6, the bound is given as $\hat{o}_i^{(j)} \in [-3^k, 3^k]$, in which case $B(k\text{-local, Pauli-6}) = 4 \times 9^k$ (see Appendix 4). See Sec. III B for the algorithm, including the construction of the measurement channel. In the Appendix section on sample complexity (Appendix 4), the details of the proof are provided.

IV. NUMERICAL RESULTS

For many quantum systems in condensed-matter physics, one of the objects of interest is the two-point correlation function. Two-point correlators could be efficiently estimated using classical shadows based on the Pauli-6 POVM. The predictions of two-point functions $\langle\sigma_i^Z\sigma_j^Z\rangle$ for the Greenberger-Horne-Zeilinger (GHZ) states with varying degree of noise are shown in Fig. 2. For GHZ states, two-point correlations between any two sites are equivalent.

We can write the action of the single-qubit depolarizing noise [5] on an arbitrary ρ written in the Bloch sphere representation:

$$\rho' = \left(1 - \tfrac{4}{3}p\right)\rho + \tfrac{4}{3}p\mathbb{I}. \tag{14}$$

Applying this channel to every qubit, we generate a noisy GHZ state [13] from a pure one. The expected two-point correlations $\langle\sigma_i^Z\sigma_j^Z\rangle$ vary as $(1 - 4p/3)^2$ with the noise parameter p.

While predicting multiple $1, \ldots, L$, two-point, or k-point correlations, we monitor the maximum possible error among all the observables. This measure of error is expected to go down with increasing number of samples. This scaling, as seen in Fig. 3, gives us some idea of the appropriateness of a POVM set for a particular task.

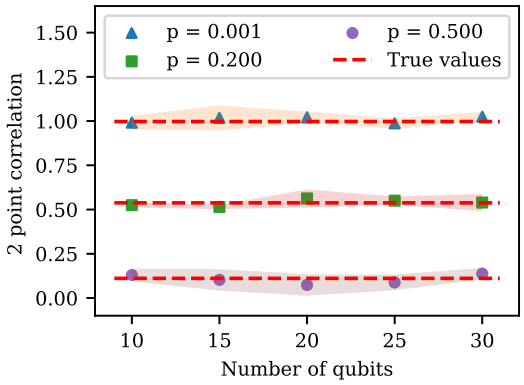

FIG. 2. Prediction of two-point correlations, $\langle \sigma_0^z \sigma_{n-2}^z \rangle$, for noisy n-qubit GHZ target states using classical shadows for Pauli-6, with one standard deviation band. The standard deviations are estimated over ten independent runs, each of which involved $N = 5000$ samples. The theoretical error bound for this two-point correlation is 0.3115, for a failure probability $\delta = 0.1$. This bound is much larger than the observed error. See Appendix 4 for a discussion. The parameter p, representing the local depolarizing noise strength, is described in Eq. (14).

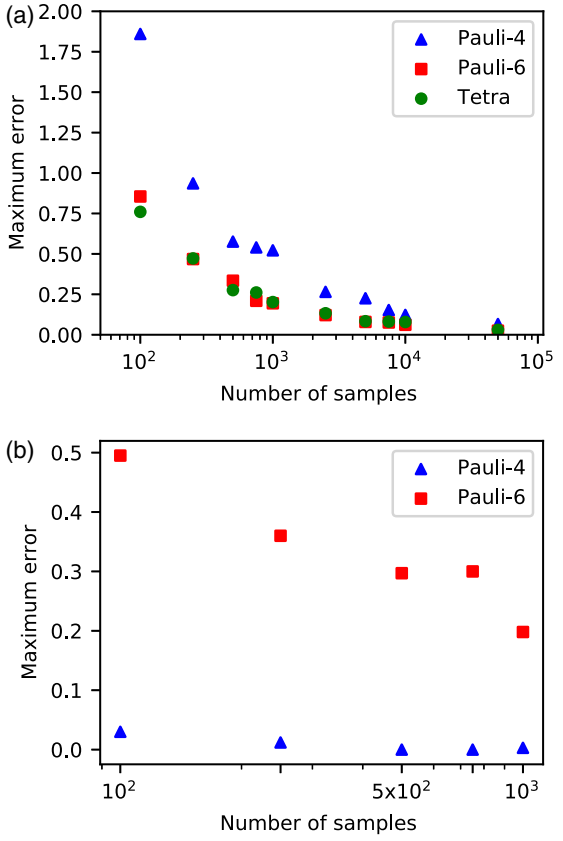

FIG. 3. Maximum error in two-point correlators. (a) Scaling of maximum error among all two-point correlations in a 30 qubit pure GHZ state, plotted against different numbers of samples for different choices of POVMs: Pauli-6, Pauli-4, and tetrahedral. (b) Scaling of maximum error for all spin-down states with Pauli-4 and Pauli-6 POVM. Pauli-4 ensures a much better scaling. See Appendix 4 for details.

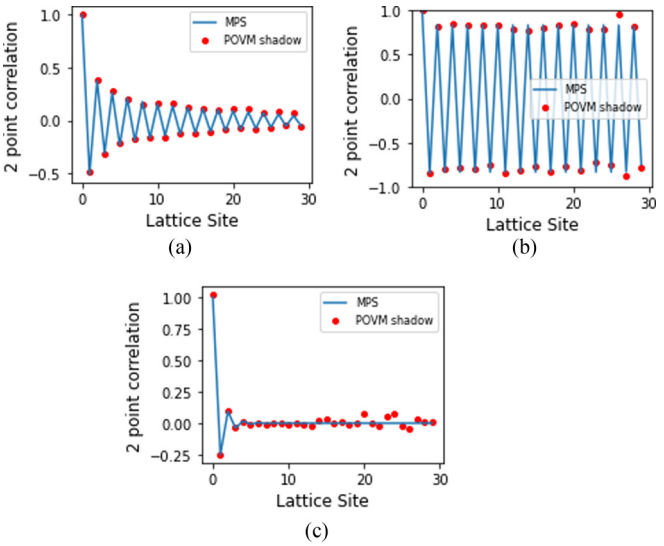

FIG. 4. Two-point functions $\langle \sigma_0^z \sigma_i^z \rangle$ for ground states of the antiferromagnetic one-dimensional (1D) transverse field Ising model using Pauli-6 POVM based shadows and the true value, as computed using matrix product states. The correlations are plotted against the lattice separation. The lattice size is 30 and the number of samples used is 5000. (a) Critical ($J = h = 1$) antiferromagnetic 1D TFIM, showing signatures of power-law correlation. (b) Ordered state ($J > h, J = 1, h = 0.5$), where correlation saturates with increasing lattice separation. (c) The paramagnetic state ($J < h, J = 1, h = 0.5$), displaying exponential decay of the correlations.

A. One-dimensional transverse field Ising model

We take the antiferromagnetic [$J > 0$ in Eq. (15)] transverse field Ising model in one dimension:

$$H = J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z + h \sum_i \sigma^x. \tag{15}$$

The quantum critical point at $h/J = 1$ will be exhibited by the power-law decay of the correlations. See Fig. 4 for results in the three regimes: critical, ordered, and paramagnetic. The exact numerical correlations are plotted using the matrix product representations of the ground states [14]. in Refs. [1,15], POVM-based measurements, followed by a neural-network-centric approach for constructing the ground state and computing the resulting two-point correlations, were presented for the same system.

B. One-dimensional disordered Heisenberg model

The Hamiltonian for the one-dimensional (1D) disordered Heisenberg model is given by

$$\hat{H} = -\frac{1}{2} \sum_{j=1}^{N} \left(J_j^x \sigma_j^x \sigma_{j+1}^x + J_j^y \sigma_j^y \sigma_{j+1}^y + J_j^z \sigma_j^z \sigma_{j+1}^z + h \sigma_j^z \right). \tag{16}$$

The properties of spin-$\frac{1}{2}$ antiferromagnetic chains with various types of random exchange coupling have been studied in an exact decimation renormalization-group (strong-disorder) scheme, some of which involve generalization or modifications of the scheme introduced by Dasgupta and Ma [16]. The numerical studies done by Bhatt and Lee [17] indicate that the

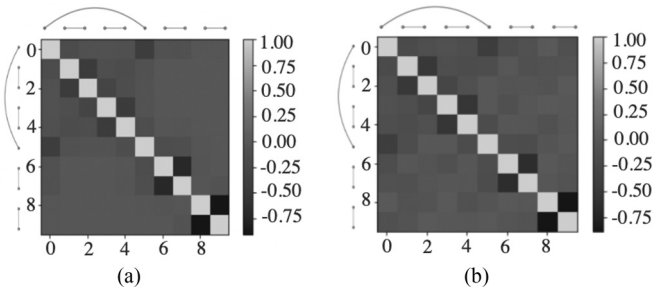

FIG. 5. Two-point functions $\langle \sigma_i^z \sigma_j^z \rangle$ for ground states of a disordered 1D Heisenberg spin chain with length 10 and open boundary conditions. (a) Exact diagonalization results. (b) Results from using Pauli-6 POVM based shadows using 5000 samples.

system could be in a random-singlet phase. In such a phase, each spin is paired with another spin that may be far away on the lattice. We perform exact diagonalization, obtain the ground state, and then compute two-point quantum correlations. The two-dimensional plot of the correlation matrix will also inform us about the locations of the singlet formations in the chain. We can also reconstruct these behaviors of a ground state corresponding to one particular disorder realization of the XXZ Heisenberg model Eq. (16) ($J_x = J_y = 2J_z$, $h = 0$) with sufficient number of shadows. See Fig. 5, where the singlet formations are indicated by the schematics drawn on the axes of the matrix visualization plots and the results from the two methods are compared.

C. Exploring quantum fidelity

In our approach to construct shadows using local POVMs, we ensure prediction of local observables. However, we can also explore nonlocal observables such as fidelity. Using the sample mean as an estimator, we can construct a hypothesis state (σ):

$$\sigma = \frac{1}{N} \sum_i^N \hat{\rho}. \tag{17}$$

When our target state is pure, say $|\psi\rangle\langle\psi|\psi$, and we have a physical density matrix $\hat{\sigma}$, we can express quantum fidelity between the two states as a linear prediction of an observable $O = |\psi\rangle\langle\psi|\psi$. In other words, the quantum fidelity can be written as

$$F = \mathrm{tr}(\hat{\sigma} O) = \langle\psi|\hat{\sigma}|\psi\rangle.$$

On the other hand, the measure $\mathrm{tr}(\sigma O)$ is equivalent to fidelity, only when σ is a physical density matrix, i.e., when $\sigma \succeq 0$. That property is likely to hold only when the number of samples is large. We expect $\mathrm{tr}(\sigma O)$ to fluctuate around its mean value 1, as seen in Fig. 6, even when the typical σ is not a physical state, meaning it is not positive semidefinite. Also, the fluctuation around this mean keeps on growing exponentially with the number of qubits (see Fig. 6). This growth cannot be dealt with even by the median of means (MoM) procedure [2] within the shadow formalism. Numerical computations using MoM also show no advantage over sample means here.

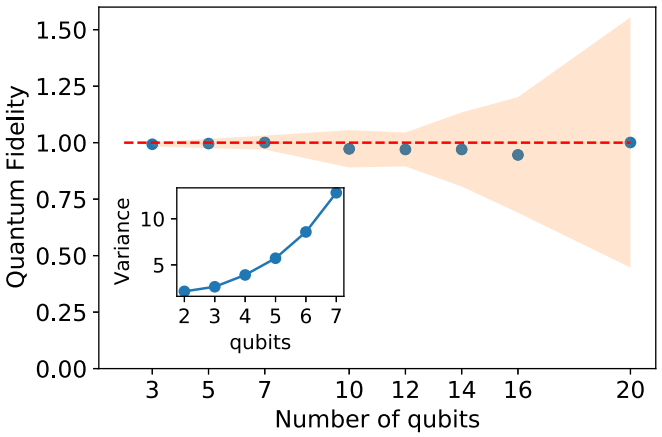

FIG. 6. Quantum fidelity predicted for the pure GHZ state using the sample mean of shadows constructed on 10^4 samples. The shaded regions are the standard deviation over ten independent runs. The inset shows the scaling of the variance of fidelity which grows exponentially with number of qubits.

Hence, we need a procedure to find the "closest" physical state to σ. The trace condition $\mathrm{tr}(\hat{\sigma}) = 1$ ensures that, once $\hat{\sigma} \not\succeq 0$, some of the eigenvalues will be greater than 1 to compensate for the negative eigenvalues. Thus, we cannot just throw away the negative eigenvalues, as would be done for projecting a Hermitian matrix to the space of positive semidefinite matrices.

We define the the convex set of physical states to be $C = \{\rho | \rho \succeq 0, \mathrm{tr}(\rho) = 1\}$. Our nonlinear projection to C is

$$\Pi_C(\sigma) = \arg \min_{\rho \in C} \mathrm{tr}[(\rho - \sigma)^2]. \tag{18}$$

We achieve this by diagonalizing σ, projecting the eigenvalues λ_i of σ onto a canonical simplex $\Delta = \{(\lambda_1^p, \ldots, \lambda_D^p) \mid \lambda_i^p \geqslant 0, \sum_{i=1}^D \lambda_i^p = 1\}$, using the recipe from [18], while leaving the eigenvectors untouched. Here, $D = 2^n$ where n is the total number of qubits. The projected state is a biased estimator. We can hope that the price paid by accepting some bias comes with the benefit of reduced variance. This expectation seems to be born out in Fig. 7. However, as the number of qubits increases, the bias itself reduces fidelity. To compensate for this effect, we need larger sample sizes (N). Figure 7 shows all these trends.

V. DISCUSSIONS

We provide an approach to predict expectations of local observables without having to apply random unitary transformations, which sometimes require complex circuits of their own and can become a practical bottleneck. We show that this can rather be done using an IC POVM. For illustrations, we show faithful reconstruction properties of low-energy states coming from different many-body Hamiltonians relevant to near-term applications of quantum devices. When we have additional information about the possible noisy channels we also adapt the shadow channel as a composition of the noise channel and the measurement channel. The invertibility becomes straightforward in the proposed framework. We also comment on why the mean as an estimator is sufficient throughout our

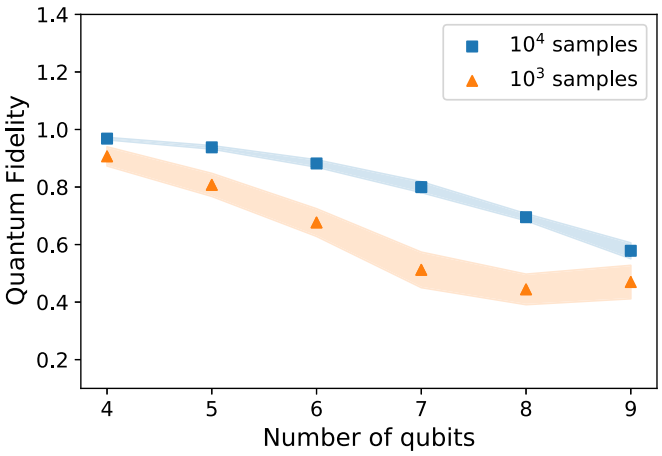

FIG. 7. Quantum fidelity of the projected shadows (onto the physical positive definite space) with the noiseless GHZ state. As we increase the number of samples, from 10^3 to 10^4, the quantum fidelity improves. The shaded regions indicate one-standard-deviation bands, estimated over ten independent runs.

discussion. And as long as we are dealing with local observables, we can provide efficient sample complexity using Hoeffding's inequality directly.

We provide instances where the choice of POVM impacts the sample complexity for predicting two-point correlators in certain quantum states for fixed maximum error. We note that the different POVMs work better for different states. It is an exciting endeavor to understand which sets of POVM would be ideal for different classes of quantum states and observables.

Although an exploration, we attempt to reconstruct fidelity using the locally built shadows and show that we cannot benefit from median of means as an estimator, since variance of fidelity becomes exponential in number of qubits. Additionally, when presented with few samples we raise the issue of unphysical, i.e., not positive semidefinite, $\hat{\rho}$ and then provide a projection technique, similar to [19,20], to estimate fidelity. Unfortunately, the estimator no longer remains unbiased. Addressing this issue would require methods to deal with nonlocal observables.

We did not provide an effective analog of the global Clifford unitary transformation-based method in Ref. [2]. There has been work which provides description of global alternatives using stabilizer states [20]. Whether there can be a scheme based on such states that is competitive with the classical shadows method [2] remains to be seen.

The use of generalized measurement to unambiguously discriminate nonorthogonal states with lower failure probability is well known [5,21,22]. Efficient prediction of expectations of local observables combined with the generalized measurement scheme to obtain the shadows can be used as an optimal framework in the discrimination of nonorthogonal states. It is a promising future direction of exploration.

ACKNOWLEDGMENT

We would like to thank Shagesh Sridharan, James Stokes, Miles Stoudenmire, and Emina Soljanin for insightful discussions.

APPENDIX

1. A simple measurement channel

We can take the simple rank-1 Pauli-6 POVMs to see the action of a measurement channel:

$$\tilde{\rho} = \mathcal{M}(\rho) = \sum_a \mathrm{tr}\left(\frac{1}{2}(\mathbb{I} + \mathbf{r}\cdot\sigma)M_a\right)|\psi_a\rangle\langle\psi_a|\psi_a, \quad \text{(A1)}$$

where we use the Bloch representation $\rho = \frac{1}{2}(\mathbb{I} + \mathbf{r}\cdot\sigma)$.

The contribution of the first two POVM elements of Pauli-6 only gets contribution from $\mathbb{I}$ and $r_z\sigma_z$, generating

$$\mathrm{tr}\left(\tfrac{1}{2}(\mathbb{I} + r_z\sigma_z)M_0\right)|0\rangle\langle 0|0 + \mathrm{tr}\left(\tfrac{1}{2}(\mathbb{I} + r_z\sigma_z)M_1\right)|1\rangle\langle 1|1.$$

Using $M_0 = \frac{1}{3}\times|0\rangle\langle 0|0$ and $M_1 = \frac{1}{3}\times|1\rangle\langle 1|1$ this expression becomes $\frac{1}{6}\mathbb{I} + \frac{1}{3}r_z\sigma_z$ Following similar steps for pairs M_2, M_3 and M_4, M_5, we get that

$$\mathcal{M}(\rho) = \mathcal{M}\left(\tfrac{1}{2}(\mathbb{I} + \mathbf{r}\cdot\sigma)\right) = \tfrac{1}{2}\left(\mathbb{I} + \tfrac{1}{3}\mathbf{r}\cdot\sigma\right),$$

making $\mathcal{M}$ a depolarizing channel.

2. Inverse of the measurement channel

Given any single-qubit channel, the inverse can be easily computed using the *Bloch-sphere* representation. We can write any two-dimensional (single-qubit) quantum operation (X) as $X = (x_0\mathbb{I} + \vec{r}.\vec{\sigma})$. Any arbitrary trace-preserving quantum operation is given as $\mathcal{E}(X) = (x_0\mathbb{I} + \vec{r}\,'.\vec{\sigma})$. The map $\vec{r} \xrightarrow{\mathcal{E}} \vec{r}\,'$ is equivalent to

$$\vec{r}\,' = T\vec{r} + x_0\vec{c}, \quad T_{\alpha,\beta} = \tfrac{1}{2}\mathrm{tr}(\sigma_\alpha\mathcal{E}(\sigma_\beta)). \quad \text{(A2)}$$

The component of displacement $(\vec{c})$ is given as $c_\alpha = \frac{1}{2}\mathrm{tr}(\sigma_\alpha\mathcal{E}(\mathbb{I}))$. Also, $\{\alpha, \beta\} \in \{1, 2, 3\}$. The *affine map* between the Bloch sphere and itself is given by T, and its meaning is understood better by doing a singular value decomposition, i.e., $T = O_1DO_2'$ where O_1 and O_2 are orthogonal matrices. The singular values capture the deformation of the Bloch sphere about its principal axes. A superoperator $\hat{T}_{4\times4}$ can be defined as

$$\hat{T}\begin{bmatrix} x_0 \\ \vec{r} \end{bmatrix} = \begin{bmatrix} x_0 \\ \vec{r}\,' \end{bmatrix}, \quad \hat{T}_{4\times4} = \begin{bmatrix} 1 & 0 \\ \vec{c} & T_{3\times3} \end{bmatrix}. \quad \text{(A3)}$$

Computing the inverse of the channel is equivalent to writing $(x_0, \vec{r})$ from $(x_0, \vec{r}\,')$, i.e., computing $\hat{T}^{-1}$.

In the main text, the Pauli measurement channel $(\mathcal{E} = \mathcal{M}_1)$ turns out to be a depolarizing channel, and its inverse that acts on the local qubit is given as

$$\mathcal{M}_1^{-1}(X) = 3X - \mathrm{tr}(X)\mathbb{I}.$$

We take a more general example following our definition of a measurement channel:

$$\tilde{\rho} = \mathcal{M}(\rho) = \sum_a \mathrm{tr}(\rho M_a)|\psi_a\rangle\langle\psi_a|\psi_a.$$

If a particular POVM element is not rank 1, $|\psi_a\rangle$ can be taken as the eigenvector corresponding to the highest eigenvalue of M_a. For Pauli-4, except for $M_3 = \frac{1}{3}(|1\rangle\langle 1|1 + |-\rangle\langle -|- + |r\rangle\langle r|r) = \mathbb{I}$, all other elements are rank 1. Since M_3 is rank 2, when the outcome is 3, we take the eigenvector $|t\rangle$ corresponding to eigenvalue $\frac{1}{2}(1 + \frac{1}{\sqrt{3}})$ instead of the other

corresponding to $\frac{1}{2}(1-\frac{1}{\sqrt{3}})$. Rewriting Eq. (A2), we get

$$\tilde{\rho} = \mathcal{M}_1(\rho) = \mathrm{tr}(\rho M_0)|0\rangle\langle 0|0 + \mathrm{tr}(\rho M_1)|+\rangle\langle +|+ \\ + \mathrm{tr}(\rho M_2)|l\rangle\langle l|l + \mathrm{tr}(\rho M_3)|t\rangle\langle t|t. \tag{A4}$$

The inverse of the channel can be written as

$$\mathcal{M}_1^{-1}(X) = 6X - \left(\frac{x_s}{\sqrt{3}} - x_0(\sqrt{3}-1)\right)\left(\sum_i \sigma_\alpha\right) - 5x_0\mathbb{I}, \\ x_0 = \frac{\mathrm{tr}(X)}{2}, \quad x_s = \frac{3\left(\mathrm{tr}\left(X\left(\sum_\alpha \sigma_\alpha\right)\right) + (\sqrt{3}-1)x_0\right)}{\sqrt{3}+1}. \tag{A5}$$

When we are working with a known noise channel $\mathcal{E}$, the inverse is given as $\hat{T}^{-1}_{\mathcal{M}_\mathcal{E}} = \hat{T}^{-1}_{\mathcal{E}}\hat{T}^{-1}_{\mathcal{M}}$. If we choose an amplitude damping channel with a damping parameter γ, the inverse can be given as

$$\mathcal{M}_\mathcal{E}^{-1}(X) = \frac{3\left(1-\frac{1}{2}\mathrm{tr}(X\sigma_z)\right)}{\sqrt{(1-\gamma)}}X + \frac{3\mathrm{tr}(\sigma_z X)}{2(1-\gamma)}X \\ + \left(\frac{1}{2} - \frac{3}{2\sqrt{(1-\gamma)}}\right)\mathrm{tr}(X)\mathbb{I} + \frac{\gamma\,\mathrm{tr}(X)}{2(\gamma-1)}\sigma_z. \tag{A6}$$

3. More general schemes for the measurement channel

In Sec. III B, we provided a specific scheme for outputting vectors for measurement outcomes a, coming from our set of POVMs($\{M_a\}$). The probability of the outcome a is given as

$$p(a) = \mathrm{tr}(\rho M_a). \tag{A7}$$

For a more general scheme, each time we perform a measurement and get an outcome a, the state "collapses" to $|i,a\rangle\langle i,a|i,a$ with probability

$$p(i|a) = \frac{f\left(\lambda_i^a\right)}{\sum_j f\left(\lambda_j^a\right)}. \tag{A8}$$

We choose $f : \mathcal{R}^+ \to \mathcal{R}^+$ to be a monotonic function which is defined over the eigenvalues of the POVM elements ($\mathcal{R}^+$ since $M_a \succeq 0$), i.e., $M_a = \sum_i \lambda_i^a |i,a\rangle\langle i,a|i,a, \quad \forall i, \lambda_i^a \geqslant 0$. The function is nonzero for at least one value of λ_i^a for each a, so that the denominator of the expression in Eq. (A8) is nonzero.

As before, the measurement channel can be defined as

$$\tilde{\rho} = \mathcal{M}(\rho) = \sum_a p(a) \sum_i p(i|a)|i,a\rangle\langle i,a|i,a. \tag{A9}$$

One such function is $f(\lambda) = \lambda^m$. To be more precise, we should think of

$$p(i|a) = \lim_{\eta\to 0+} \frac{\left(\lambda_i^a + \eta\right)^m}{\sum_j \left(\lambda_j^a + \eta\right)^m} \tag{A10}$$

to take care of the case where some λ_i^a's are zero. In the large positive m limit, this particular choice is interesting since it is closer to the projective measurement case, while making sure we have an invertible map. This results in choosing the leading eigenvector.

However, choosing the eigenvector corresponding to the lowest eigenvector will also keep $\mathcal{M}$ invertible. This corresponds to $f(\lambda) = \lambda^m$ and taking m to be negative but with a large absolute value. In fact, our computation, using the lowest eigenvector, is as effective for shadow tomography as choosing the leading eigenvector (data not shown). The troublesome case is $m = 0$, where the information about the original measurements gets lost and all eigenvectors are chosen with equal probability.

To understand this phenomenon, we need to characterize the statistical properties of our estimates. Let X be a matrix-valued random variable. It takes the value $\mathcal{M}_1^{-1}(|i,a\rangle\langle i,a|)$ with probability $\mathrm{tr}(\rho M_a)p(i|a)$, as explained above. We know that $\bar{X} = \mathbb{E}[X] = \rho$. Let O be an observable. We define the corresponding estimated observable value to be $\mathrm{tr}(XO)$, which is a random variable. Its expectation is the answer we are after:

$$\mathbb{E}[\mathrm{tr}(XO)] = \mathrm{tr}(\bar{X}O) = \mathrm{tr}(\rho O).$$

We would be interested in sampling fluctuations of this quantity. The key metric to assess is the variance

$$\mathrm{Var}[\mathrm{tr}(XO)] = \mathrm{Var}[\mathrm{tr}((X-\bar{X})O)] \\ = \mathbb{E}[(\mathrm{tr}((X-\bar{X})O))^2] \quad \text{since} \quad \bar{X} = \mathbb{E}[X].$$

Now, using the Cauchy-Schwarz inequality, applied to Hermitian matrices, we can bound this variance as follows:

$$\mathrm{Var}[\mathrm{tr}((X-\bar{X})O)] \leqslant \mathbb{E}\left[\,||X-\bar{X}||_F^2\,\right] ||O||_F^2,$$

where the Frobenius norm $||.||_F$ is defined as $||A||_F = \sqrt{\mathrm{tr}(AA^\dagger)}$.

In absence of any prior information on the observable O, we can still bound the variance by simply looking at the statistics of $X - \bar{X}$. The role of function f can be better understood by monitoring the quantity $\mathbb{E}[||X-\bar{X}||_F^2]$. Of course, this expectation depends upon the density matrix ρ. We need to further take an average over an ensemble of ρ's, to not depend on the peculiarity of one particular state. We will call this ρ ensemble averaged expectation, $\mathbb{E}_\rho\mathbb{E}[\,||X-\bar{X}||_F^2\,]$, the average Frobenius norm square.

In particular, for a single qubit, we choose this ensemble of ρ's to consist of the six matrices $\{\frac{1}{2}(\mathbb{I} \pm \sigma_i)|i = 1,2,3\}$ with equal probability. Using the Pauli-4 POVM, the average Frobenius norm square of the difference between a single snapshot and the average, i.e., ρ, is plotted in Fig. 8 against various choices of the parameter m, which defines the function $f(\lambda) = \lambda^m$. Note that $\mathbb{E}[||X-\bar{X}||_F^2] = 7.085$ is the lowest value of the bound, attained for the eigenvector corresponding to both the highest and the lowest eigenvalues.

It is in fact possible to perform shadow tomography for other m choices as well, at the cost of bigger sampling noise (data not shown). However, making no preferred choice, i.e., choosing i given a with equal probability, will lead to a non-invertible measurement channel. In other words, the channel $\mathcal{M}$, thought of as a linear operator, develops a zero eigenvalue at $m = 0$. When m is nonzero but small, this eigenvalue is small as well. The variance bound is affected by eigenvalues of $\mathcal{M}^{-1}$, and the small eigenvalue leads to large variances. This behavior can be seen in Fig. 8, where the variance increases rapidly near $m = 0$.

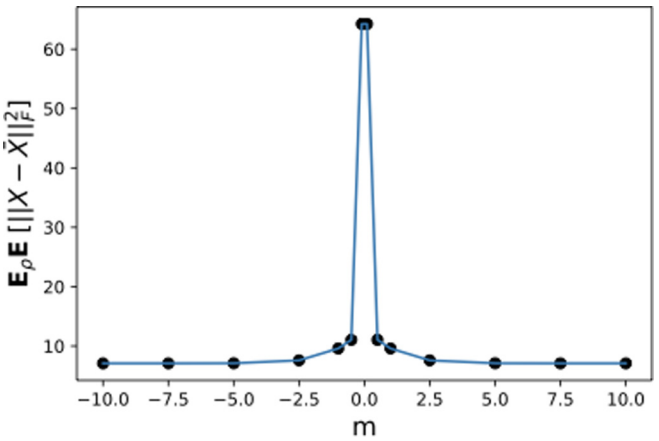

FIG. 8. The average Frobenius norm square of the difference between a single snapshot and ρ, $\mathbb{E}_\rho\mathbb{E}[||X-\bar{X}||_F^2]$, against m [various choices of the function $f(\lambda)=\lambda^m$].

4. Sample complexity

a. Variance of the estimate for a single observable

Given an array of N independent, classical snapshots [each defined as Eq. (7)],

$$\mathsf{S}(\rho;N)=\{\hat{\rho}^{(1)},\hat{\rho}^{(1)},\ldots,\hat{\rho}^{(N)}\}. \tag{A11}$$

The sample mean is $\hat{o}=\frac{1}{N}\sum_{j=1}^{N}\mathrm{tr}(O\hat{\rho}^{(j)})$. The bound on probability of deviation of the sample mean is given by Chebyshev's inequality:

$$\Pr(|\hat{o}-\mathbb{E}[\hat{o}]|\geqslant\epsilon)\leqslant\frac{\mathrm{Var}[\hat{o}]}{\epsilon^2} \tag{A12}$$

where $\mathbb{E}[\hat{o}]=\mathrm{tr}(O\rho)$ where ρ is the true density matrix. Fluctuations of $\hat{o}$ around this desired expectation are controlled by the variance. $\mathrm{Var}[\hat{o}]=\frac{1}{N}\mathrm{Var}[\mathrm{tr}(O\hat{\rho}^{(j)})]=\frac{\mathrm{Var}[o^{(j)}]}{N}$. However, since the classical shadows are unit trace by construction, the variance depends only on the traceless part of the observable, i.e., $O_0=O-\frac{\mathrm{tr}(O)}{2^n}\mathbb{I}$. The minimum number of samples needed to assure a maximum failure probability (δ) using Eq. (A12) is

$$N\geqslant\frac{\mathrm{Var}[o^{(j)}]}{\epsilon^2\delta}. \tag{A13}$$

b. Dependence on POVM

Given a measurement channel and an observable, we can bound the variance of its estimator, using familiar maneuvers with superoperators [2]:

$$\begin{aligned}\mathrm{Var}[o^{(j)}]&=\mathbb{E}((o^{(j)})^2)-(\mathbb{E}(o^{(j)}))^2\leqslant\mathbb{E}((o^{(j)})^2)\\&=\sum_{a_1,..,a_n}\Pr(a_1..a_n)\langle a_1,.,a_n|\left[\mathcal{M}_n^{-1}\right]^\dagger(O_0)|a_1,.,a_n\rangle^2\\&\text{where}\quad\Pr(a_1,..,a_n)=\mathrm{tr}(\rho M^{a_1}\otimes M^{a_1}..\otimes M^{a_n}).\end{aligned}$$

We broadly define a k-local Pauli observable as an operator which acts nontrivially only on k qubits. Traceless k local operators can be expressed as linear combinations of tensor products of identity matrices and k or less Pauli matrices. Hence, we need to focus only on the special class of k-local operators.

Let P_i be a traceless operator acting on the ith qubit. We can choose it to be any one of the three Pauli matrices $\{\sigma_1,\sigma_2,\sigma_3\}$. We focus on tensor products like $O_0=P_1\otimes P_2\otimes..\otimes P_k\otimes\mathbb{I}^{\otimes(n-k)}$, where, without loss of generality, we assume that the operator acts nontrivially on only the first k qubits.

For the Pauli-6 POVM, the inverse of the measurement channel is a self-adjoint map, and thus one can verify its action as

$$\left[\mathcal{M}_1^{-1}\right]^\dagger(\sigma_\alpha)=\mathcal{M}_1^{-1}(\sigma_\alpha)=3\sigma_\alpha,$$

where σ_α denotes a Pauli matrix and $[\mathcal{M}_1^{-1}]^\dagger(\mathbb{I})=\mathcal{M}_1^{-1}(\mathbb{I})=\mathbb{I}$. Given a k-local observable, we can further compute the bound on variance:

$$\begin{aligned}\mathrm{Var}[o^{(j)}]&\leqslant\sum_{a_1,..,a_n}\Pr(a_1..a_n)\prod_{i=1}^{k}\langle a_i|3P_i|a_i\rangle^2\\&=\sum_{a_1,..,a_n}\mathrm{tr}(\rho M^{a_1}\otimes M^{a_2}..\otimes M^{a_n})\prod_{i=1}^{k}\langle a_i|3P_i|a_i\rangle^2\\&=\mathrm{tr}\left[\rho\sum_{a_1,..,a_n}\left(M^{a_1}\otimes M^{a_1}..\otimes M^{a_n}\prod_{i=1}^{k}\langle a_i|3P_i|a_i\rangle\right)^2\right]\\&=\mathrm{tr}\left[\rho\bigotimes_{i=1}^{k}\underbrace{\sum_{a_i}M^{a_i}\langle a_i|3P_i|a_i\rangle^2}_{3\mathbb{I}}\bigotimes\mathbb{I}^{\otimes(n-k)}\right]=3^k.\end{aligned}$$

Now, we take up the Pauli-4 POVM. One can verify the action of $[\mathcal{M}_1^{-1}]^\dagger$ as

$$\left[\mathcal{M}_1^{-1}\right]^\dagger(P_\alpha)=(2-\sqrt{3})\mathbb{I}+(3+\sqrt{3})P_\alpha-\sum_{\beta\neq\alpha}(3-\sqrt{3})P_\beta,$$

where P_α denotes a Pauli matrix. Using the fact that $\mathcal{M}_1^{-1}$ is a trace-preserving map, one can say its adjoint has to be unital, i.e., $[\mathcal{M}_1^{-1}]^\dagger(\mathbb{I})=\mathbb{I}$. Given a k-local Pauli observable, one can again compute the bound on variance:

$$\begin{aligned}&\mathrm{Var}[o^{(j)}]\\&\leqslant\sum_{a_1,..,a_n}\Pr(a_1..a_n)\prod_{i=1}^{k}\langle a_i|\left[\mathcal{M}_1^{-1}\right]^\dagger(P_i)|a_i\rangle^2\\&=\sum_{a_1,..,a_n}\mathrm{tr}(\rho M^{a_1}\otimes M^{a_2}..\otimes M^{a_n})\prod_{i=1}^{k}\langle a_i|\left[\mathcal{M}_1^{-1}\right]^\dagger(P_i)|a_i\rangle^2\\&=\mathrm{tr}\left[\rho\sum_{a_1,..,a_n}(M^{a_1}\otimes M^{a_1}..\otimes M^{a_n}\prod_{i=1}^{k}\langle a_i|\left[\mathcal{M}_1^{-1}\right]^\dagger(P_i)|a_i\rangle^2\right]\\&=\mathrm{tr}\left[\rho\bigotimes_{i=1}^{k}\underbrace{\sum_{a_i}M^{a_i}\langle a_i|\left[\mathcal{M}_1^{-1}\right]^\dagger(P_i)|a_i\rangle^2}_{5\mathbb{I}+4P_i}\bigotimes\mathbb{I}^{\otimes(n-k)}\right]\\&=\mathrm{tr}\left[\rho\bigotimes_{i=1}^{k}(5\mathbb{I}+4P_i)\bigotimes\mathbb{I}^{\otimes(n-k)}\right].\end{aligned}$$

Clearly, the above bound on variance is dependent on the state ρ, unlike the bound we obtained using the Pauli-6 POVM. Since ρ is a density matrix and the operator $5\mathbb{I} + 4P_i$ is a positive semi-definite operator, one gets the minimum value for the bound when ρ is of the form

$$\rho = \left(\bigotimes_{i=1}^{k} |p_i\rangle\langle p_i|\right) \bigotimes \tilde{\rho}_{n-k}, \tag{A14}$$

where $|p_i\rangle\langle p_i|$ is the projector into the eigenvector corresponding to the lowest eigenvalue of the operator $5\mathbb{I} + 4P_i$, and $\tilde{\rho}_{n-k}$ is a valid density matrix in the Hilbert space of $n-k$ qubits on which the k-local Pauli observable acts trivially. For the above ρ, it is simple to verify that the value of the variance bound is 1 (independent of k). Thus, for example, if the unknown state ρ is the all spin-down state, then the Pauli-4 POVM works better than the Pauli-6 POVM in predicting two-point correlators $\langle \sigma_i^Z \sigma_j^Z \rangle$, since the variance is higher in the latter.

c. Improved bound using Hoeffding's inequality

Furthermore, we can use Hoeffding's inequality to provide theoretical bounds when we are dealing with the k-local Pauli observable, since we are working with bounded random variables. If $\hat{o} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{tr}(O\hat{\rho}^{(j)})$, $\hat{o}^{(j)} \in [a, b]$ for all j, where $-\infty < a \leqslant b \leqslant \infty$, we can write

$$\Pr(|\hat{o} - \mathbb{E}[\hat{o}]| \geqslant \epsilon) \leqslant 2e^{\left(\frac{-2N\epsilon^2}{(b-a)^2}\right)}. \tag{A15}$$

The minimum number of samples needed to assure a maximum failure probability (δ) is

$$N \geqslant \ln\left(\frac{2}{\delta}\right)\frac{(b-a)^2}{2\epsilon^2}. \tag{A16}$$

$B(\{O\}, \mathcal{M}) = (b-a)^2$ depends on the locality of the observable and the maximum eigenvalue $\lambda_{\max}$ of the inverse channel acting on the observable. The bound on random variable $\hat{o}^{(j)}$ can be found as the range of the Rayleigh quotient of the inverse of the measurement channel, acting on the observable over all possible states. For instance, if we choose Pauli-6, the bounds can be shown to lie within $\hat{o}^{(j)} \in [-3^k, 3^k]$ in which case $B(k\text{-local, Pauli-6}) = 4 \times 9^k$. Using the action of $[\mathcal{M}^{-1}]^\dagger$, one can verify that the value of the random variable $\mathrm{tr}([\mathcal{M}^{-1}]^\dagger(P_i)|a_i\rangle\langle a_i|)$ belongs to the set $\{5, -1\}$ for any Pauli matrix P_i when $|a_i\rangle\langle a_i|)$ is the inferred state for Pauli-4. Thus, the random variable $\hat{o}$ is contained in the range $\{-5^k, 5^k\}$ which is exponential on the locality rather than the number of qubits.

d. Guarantee of performance for multiple observables

We have L different k-local Pauli observables $\hat{o}_1, \ldots, \hat{o}_i, \ldots, \hat{o}_L$ with the sample mean corresponding to the observable i defined as $\hat{o}_i = \frac{1}{N}\sum_{j=1}^{N} \mathrm{tr}(O_i\hat{\rho}^{(j)})$. If $\hat{o}_i = \frac{1}{N}\sum_{j=1}^{N} \mathrm{tr}(O_i\hat{\rho}^{(j)})$, $\hat{o}_i^{(j)} \in [a, b]$ for all j, where $-\infty < a \leqslant b < \infty$, we can combine the union bound with Hoeffding's inequality to write

$$\Pr\left(\max_{1 \leqslant i \leqslant L} |\hat{o}_i - \mathbb{E}[\hat{o}_i]| \geqslant \epsilon\right) \leqslant 2Le^{\left(\frac{-2N\epsilon^2}{(b-a)^2}\right)}. \tag{A17}$$

The minimum number of samples needed to assure a maximum failure probability (δ) among all the observables using Eq. (A17) is

$$N \geqslant \ln\left(\frac{2L}{\delta}\right)\frac{(b-a)^2}{2\epsilon^2}. \tag{A18}$$

The scaling is logarithmic in the number of observables L, instead of the linear behavior we get using Chebyshev's inequality. We do not need to use MoM procedure [2], which would have been necessary if we were dealing with estimator distributions with long tails (unlike the bounded estimates for k-local Pauli observables).

For example, in the numerical experiment corresponding to Fig. 2, we used these inequalities to estimate the error in prediction. Substituting $L = 1$ in Eq. (A18), the theoretical estimate for the error in prediction using $N = 5000$ samples and allowing a maximum failure probability $\delta = 0.1$ is 0.3115. Since we are predicting two-point correlators using the Pauli-6 POVM, the bound $B = (b-a)^2 = 4 \times 9^2$.

5. Numerical computations

The computation for GHZ states has been carried out using matrix product state (MPS) representations for noiseless states and matrix product operator (MPO) representations for the noisy states. The details on the simulation of mixed states using MPOs have been shown in Refs. [1,23]. The datasets corresponding to the ground states of spin Hamiltonians such as the transverse field Ising model have been generated using the density-matrix renormalization group (DMRG). The library used is mpnum (a matrix product representation library for PYTHON) [24]. Given a particular spin model, with the Hamiltonian expressed as a MPO, the DMRG algorithm attempts to find the optimal MPS with the lowest energy. However, for the visualization of correlations in the disordered 1D Heisenberg spin chain, the computations are made without using the DMRG framework. This is because the number of sites was low for this particular numerical experiment.

[1] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, Nat. Mach. Intelligence **1**, 155 (2019).

[2] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050 (2020).

[3] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nature Physics **14**, 447 (2018).

[4] S. Aaronson, Shadow tomography of quantum states, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, 2018* (unpublished), pp. 325–338.

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University, Cambridge, England, 2011).

[6] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Symmetric informationally complete quantum measurements, J. Math. Phys. **45**, 2171 (2004).

[7] A. Dimić and B. Dakić, Single-copy entanglement detection, npj Quant. Info. **4**, 11 (2018).

[8] J. Morris and B. Dakić, Selective quantum state tomography, arXiv:1909.05880.

[9] G. N. M. Tabia, Experimental scheme for qubit and qutrit symmetric informationally complete positive operator-valued measurements using multiport devices, Phys. Rev. A **86**, 062107 (2012).

[10] S. Aaronson, The learnability of quantum states, Proc. R. Soc. A **463**, 3089 (2007).

[11] L. G. Valiant, A theory of the learnable, Communications of the ACM **27**, 1134 (1984).

[12] D. E. Koh and S. Grewal, Classical shadows with noise, arXiv:2011.11580.

[13] D. M. Greenberger, M. A. Horne, and A. Zeilinger, Going beyond Bell's theorem, arXiv:0712.0921.

[14] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, Ann. Phys. (NY) **349**, 117 (2014).

[15] I. A. Luchnikov, A. Ryzhov, P.-J. Stas, S. N. Filippov, and H. Ouerdane, Variational autoencoder reconstruction of complex many-body physics, Entropy **21**, 1091 (2019).

[16] C. Dasgupta and S.-K. Ma, Low-temperature properties of the random Heisenberg antiferromagnetic chain, Phys. Rev. B **22**, 1305 (1980).

[17] R. N. Bhatt and P. A. Lee, Scaling Studies of Highly Disordered Spin-$\frac{1}{2}$ Antiferromagnetic Systems, Phys. Rev. Lett. **48**, 344 (1982).

[18] W. Wang and M. Á. Carreira-Perpiñán, Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application, arXiv:1309.1541.

[19] J. A. Smolin, J. M. Gambetta, and G. Smith, Efficient Method for Computing the Maximum-Likelihood Quantum State from Measurements with Additive Gaussian Noise, Phys. Rev. Lett. **108**, 070502 (2012).

[20] G. I. Struchalin, Ya. A. Zagorovskii, E. V. Kovlakov, S. S. Straupe, and S. P. Kulik, Experimental estimation of quantum state properties from classical shadows, PRX Quantum **2**, 010307 (2021).

[21] S. M. Barnett and S. Croke, Quantum state discrimination, Adv. Opt. Photon. **1**, 238 (2009).

[22] A. Chefles, Quantum state discrimination, Contemp. Phys. **41**, 401 (2000).

[23] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems, Phys. Rev. Lett. **93**, 207204 (2004).

[24] D. Suess and M. Holzäpfel, mpnum: A matrix product representation library for Python, J. Open Source Software **2**, 465 (2017).