arXiv:2209.02093v3 [quant-ph] 25 May 2023

Scalable and Flexible Classical Shadow Tomography with Tensor Networks

A. A. Akhtar[1], Hong-Ye Hu[1,2], and Yi-Zhuang You[1]

[1]Department of Physics, University of California San Diego, La Jolla, CA 92093, USA
[2]Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 02138, USA

Classical shadow tomography is a powerful randomized measurement protocol for predicting many properties of a quantum state with few measurements. Two classical shadow protocols have been extensively studied in the literature: the single-qubit (local) Pauli measurement, which is well suited for predicting local operators but inefficient for large operators; and the global Clifford measurement, which is efficient for low-rank operators but infeasible on near-term quantum devices due to the extensive gate overhead. In this work, we demonstrate a scalable classical shadow tomography approach for generic randomized measurements implemented with finite-depth local Clifford random unitary circuits, which interpolates between the limits of Pauli and Clifford measurements. The method combines the recently proposed locally-scrambled classical shadow tomography framework with tensor network techniques to achieve scalability for computing the classical shadow reconstruction map and evaluating various physical properties. The method enables classical shadow tomography to be performed on shallow quantum circuits with superior sample efficiency and minimal gate overhead and is friendly to noisy intermediate-scale quantum (NISQ) devices. We show that the shallow-circuit measurement protocol provides immediate, exponential advantages over the Pauli measurement protocol for predicting quasi-local operators. It also enables a more efficient fidelity estimation compared to the Pauli measurement.

1 Introduction

Quantum technology is advancing rapidly. A central task is to characterize and exploit the features of many-qubit quantum states created in the lab [9, 15, 23, 34]. To fully determine the density matrix ρ of a quantum system of n qubits, exponential ($\sim 4^n$) amount of repeated measurements and classical processing is needed [18, 24, 42]. Therefore full quantum state tomography is not scalable to large systems. However, for many purposes (e.g. estimating physical observables on the quantum state), a far less complete description is adequate [1, 3], and the amount of measurement and classical processing can be drastically reduced. Much of the recent progress has been made by exploiting the randomized measurement strategy [16, 41, 43], particularly through the *classical shadow tomography* [10, 13, 25, 26, 30–32, 35, 38, 48, 49, 54].

A. A. Akhtar: a1akhtar@ucsd.edu
Hong-Ye Hu: hongyehu@g.harvard.edu
Yi-Zhuang You: yzyou@physics.ucsd.edu

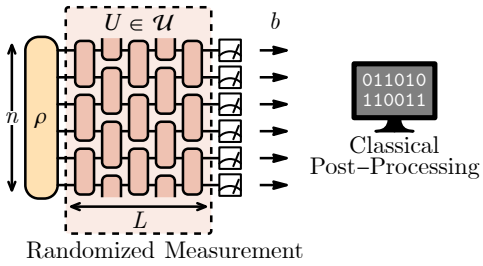


Figure 1: Classical shadow tomography in a system of n qubits by randomized measurement via a finite-depth local random unitary circuit of L layers.

Classical shadow tomography converts quantum states to classical data with a superior sample efficiency. It can estimate expectation values of M observables using only $\sim (\log M)\|O\|^2_{\text{shd}}$ independent randomized measurements [32, 45], saturating the theoretical optimal bound on sample efficiency. Nevertheless, the constant coefficient $\|O\|^2_{\text{shd}}$, known as the *operator shadow norm*, does depend on the type of observable O and the randomized measurement scheme. For example, the single-qubit (local) Pauli measurements are efficient for predicting local observables, while the global Clifford measurements are efficient in estimating certain global properties such as quantum fidelity. However, an efficient and scalable approach to interpolate the local and global limits is still missing in classical shadow tomography.

As illustrated in Fig. 1, the randomized measurement protocol in classical shadow tomography is generally realized by first transforming the quantum state ρ of interest by a random unitary circuit U sampled from a random unitary ensemble $\mathcal{U}$, then performing a computational basis measurement and collecting the

measurement outcomes b (as a bit-string). Much of the literature has focused on two measurement protocols: the single-qubit Pauli measurements (where $\mathcal{U} = \mathrm{Cl}(2)^{\otimes n}$) and the global Clifford measurements (where $\mathcal{U} = \mathrm{Cl}(2^n)$). In terms of the depth L of the random unitary circuit $U \in \mathcal{U}$, we may think of these two randomized measurement schemes as the $L \to 0$ and $L \to \infty$ limits of randomized measurement protocols defined on more general finite-depth unitary ensembles (see Fig. 1). This work aims to explore the intermediate measurement schemes based on finite-depth circuits between these two limits. For simplicity, we focus on the collection of unitary ensembles defined by the brick wall arrangement of two-local random unitary gates as shown in Fig. 1, but our approach is directly applicable to other circuit structures as well.

The ability to adjust the circuit depth L (or even the circuit structure) provides great flexibility to modify the measurement scheme adaptively so as to optimize the tomography efficiency for the given target observables. This can be particularly useful when the target observables are not neatly local or low-rank (such that neither Pauli nor Clifford measurement is optimal). The shallow circuit implementation of randomized measurements is also friendly to near-term quantum devices. Motivated by these objectives, Ref. [29] develops the foundation for shallow-circuit classical shadow tomography based on the theory of *locally scrambled* quantum dynamics and the *entanglement feature* formalism [4, 17, 28, 36, 52, 53]. The key finding is that for a large class of randomized measurements, the classical shadow reconstruction map only depends on the entanglement properties of the classical snapshot states $U^\dagger|b\rangle$, which can be computed by solving a quantum entanglement dynamics problem.

However, the reconstruction algorithm demonstrated in Ref. [29] is based on a brute-force computation, which is not scalable to larger systems (the complexity for classical post-processing will be exponential in system size). The largest system size achieved in Ref. [29] was nine qubits. Facing this challenge, in this work, we further develop efficient numerical methods based on *matrix product state* (MPS) [14, 19, 27, 44] for calculating entanglement features and solving reconstruction coefficients. The key idea is to pack the entanglement feature (purities in different entanglement regions) of the classical snapshot state $U^\dagger|b\rangle$ into a fictitious quantum many-body state and represent this entanglement feature state by an MPS. This enables efficient computation of the entanglement dynamics and finding the classical shadow reconstruction map. Our tomography procedure is outlined in Fig. 2. Using this approach, we can perform the classical shadow post-processing on 22 qubits ($n = 22$) for the randomized measurement scheme based on three-layer random Clifford circuits ($L = 3$)

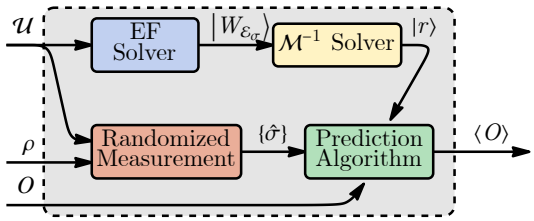


Figure 2: We outline the procedure for predicting $\langle O\rangle = \mathrm{Tr}(\rho O)$ using the shallow circuit classical shadow tomography approach. The ingredients are the state ρ, the observable O, and the unitary ensemble $\mathcal{U}$ defining the measurement protocol. The entanglement feature $W_{\mathcal{E}_\sigma}$ is totally determined by the second moment of the unitary ensemble $\mathcal{U}$. The snapshot states $\hat{\sigma}$ is collected from randomized measurements given $\mathcal{U}$ and ρ. The reconstruction map $\mathcal{M}^{-1}$ is determined by the entanglement features only, and is encapsulated by the reconstruction coefficients r. These reconstruction coefficients r, alongside the snapshot states $\hat{\sigma}$ and the observable O are used to make predictions. Importantly, the entire algorithm only involves matrix product states and stabilizer states. Both are efficient representations of quantum states that can be processed on classical computers.

with superior sample efficiency. The primary advantages of this approach are its scalability and flexibility, i.e. the classical post-processing can be performed for large system sizes with *polynomial complexity* (given a fixed amount of measurements), while the method applies to a wide range of random Clifford measurement schemes.

The paper is organized as follows. In Sec. 2, we review the basic setup of classical shadow tomography and the key concepts of entanglement feature formalism. In Sec. 3, we describe the general algorithm/procedure for computing the expectation values of generic observables and analyzing the corresponding shadow norm. In Sec. 4, we demonstrate the shallow-circuit classical shadow tomography on the Greenberger-Horne-Zeilinger (GHZ) state and the cluster state. The simulation confirms an unbiased tomographic reconstruction with favorable sample efficiency compared to that of Pauli measurements. In Sec. 5, we suggest some future lines of inquiry and summarize the key results.

2 Background

2.1 General Framework of Classical Shadow Tomograhy

Classical shadow tomography is an efficient protocol for predicting features of a quantum state based on a few measurements [32]. The trick is to use entanglement generated from a random unitary ensemble $\mathcal{U}$ to form approximate classical "shadows" $\hat{\rho}$ of the original state ρ which are then used for predicting the desired observable. Different unitary ensembles pro-

duce different collections of shadows, which are suited for predicting different kinds of observables.

The protocol consists of two steps: randomized measurement and classical post-processing, as illustrated in Fig. 1. Starting with the initial state ρ, apply a randomly sampled unitary $U \in \mathcal{U}$ so that $\rho \to U\rho U^\dagger$. Then perform a projective measurement on the transformed state $U\rho U^\dagger$ in the computational basis (Z-basis on each qubit independently). The resulting bit-string state $|b\rangle$ (labeled by the bit-string b of measurement outcomes), as well as the performed unitary transformation U, will be recorded. These two pieces of information define a *snapshot state* $\hat{\sigma} = U^\dagger|b\rangle\langle b|U$. The randomized measurement protocol can be formulated as a quantum channel $\mathcal{M}$, which maps the initial state ρ to the average snapshot state σ by

$$\sigma = \mathcal{M}(\rho) := \mathop{\mathbb{E}}_{\hat{\sigma}\in\mathcal{E}_{\sigma|\rho}} \hat{\sigma} = 2^n \mathop{\mathbb{E}}_{\hat{\sigma}\in\mathcal{E}_{\sigma}} \hat{\sigma}\,\mathrm{Tr}(\hat{\sigma}\rho), \quad (1)$$

where $\mathcal{E}_\sigma = \{U^\dagger|b\rangle\langle b|U \mid b \in \{0,1\}^{\times n}, U \in \mathcal{U}\}$ denotes the *prior* ensemble of all possible classical snapshots. In the prior snapshot ensemble, the joint probability to sample a (b, U) pair is $P(b, U) = 2^{-n}P(U)$, where $P(U)$ is the probability to sample U and 2^{-n} is the probability to sample b uniformly (independent of ρ). Given the observation of ρ, the *posterior* snapshot ensemble $\mathcal{E}_{\sigma|\rho}$ is defined by deforming the joint probability distribution to $P_\rho(b, U) = \langle b|U\rho U^\dagger|b\rangle P(U)$. The measurement channel $\mathcal{M}$ can be formulated based on either of the snapshot ensembles equivalently, as stated in Eq. (1).

Furthermore, if this measurement channel $\mathcal{M}$ is tomographically complete, i.e. distinct states ρ can be distinguished by different measurement outcomes (b, U), then there is an inverse map $\mathcal{M}^{-1}$ called the *reconstruction map* such that $\rho = \mathcal{M}^{-1}(\sigma)$. The image of a given snapshot state $\hat{\sigma}$ under the reconstruction map is a *classical shadow* $\hat{\rho} = \mathcal{M}^{-1}(\hat{\sigma})$, which provides a single-shot classical representation of the initial quantum state ρ. The initial state ρ can be restored as the ensemble expectation of classical shadows $\rho = \mathbb{E}_{\hat{\sigma}\in\mathcal{E}_{\sigma|\rho}}\mathcal{M}^{-1}(\hat{\sigma})$, which enables us to predict physical observables of the initial state based on the classical snapshots collected from randomized measurements,

$$\langle O\rangle = \mathrm{Tr}(O\rho) = \mathop{\mathbb{E}}_{\hat{\sigma}\in\mathcal{E}_{\sigma|\rho}} \mathrm{Tr}(O\mathcal{M}^{-1}(\hat{\sigma})). \quad (2)$$

This computation will be carried out on a classical computer. It will be efficient if the snapshot states are stabilizer states. Since the reconstruction map $\mathcal{M}^{-1}$ is not a quantum channel (because it is not a positivity-preserving map), the resulting shadow states $\hat{\rho}$ may not be positive semi-definite. Therefore, one may encounter unphysical single-shot estimation values of O for certain individual samples. Nevertheless, the ensemble average of all single-shot estimations is physical and unbiased.

In practice, quantum devices can have a significant amount of noise which can introduce randomness into meaurement outcomes. One solution to deal with noise is to incorporate the noise into the measurement channel itself. The idea is that the prior snapshot ensemble is composed of several layers of channels acting on an initial product state $\mathcal{E}_\sigma = \mathcal{E}_1 \circ \mathcal{E}_2 \ldots \mathcal{E}_L(|b\rangle\langle b|)$ which are averaged over to form the measurement channel. Incorporating noise amounts to simply introducing new noise layers resulting in a new measurement channel $\mathcal{M}'$ based on the statistical properties of the noise. If we assume that the noise is local and does not have a preferential basis, then the approach outlined in this paper would still be applicable and the resulting reconstruction map $(\mathcal{M}')^{-1}$ should have an efficient MPO description in the shallow circuit limit.

For any finite-sized classical shadow ensemble, the estimation exhibits statistical fluctuations around the true expectation value $\langle O\rangle$. The variance in the single-shot estimation of any observable O can be quantified through its *operator shadow norm* $\|O\|^2_{\mathrm{shd}}$, introduced in Ref. [32]. The estimated expectation value obtained from averaging over M samples will be within the error bound of $\|O\|^2_{\mathrm{shd}}/M$ with high probability. For a given class of observables, it is possible to choose the random unitary ensemble $\mathcal{U}$ wisely so as to reduce the shadow norm and to optimize the sample efficiency [29].

2.2 Locally-Scrambled Classical Shadow Tomography

Restricting the randomized measurement scheme to the scope of locally-scrambled ensembles proves to be advantageous [29]. Locally-scrambled ensembles [36] are random unitary ensembles which are invariant under both left and right local basis transformations, i.e. the probability distribution for all $U \in \mathcal{U}$ satisfies $P(U) = P(UV) = P(VU)$ where $V = \bigotimes_i V_i$ with $V_i \in \mathrm{U}(2)$ being any local basis transformation. Many typically studied random circuit models are in fact, local basis invariant, such as random Haar circuits [6, 12, 17, 39, 40, 56] and quantum Brownian circuits [11, 20, 37, 51, 55]. Furthermore, since the reconstruction map only depends on the second moment of the unitary ensemble, Clifford circuits, which are typically used for classical shadow tomography because of their classical simulability, are equivalent in terms of their tomographic properties to the random Haar circuit. Therefore, the formulation of locally-scrambled classical shadow tomography in Ref. [29] is applicable to generic random Clifford circuits of any depth and any gate arrangement.

In the locally-scrambled case, it was shown that the reconstruction map is determined entirely by the average purity (exponential of 2nd Rényi entropy) of the (prior) classical snapshot states [29] in *all* possible sub-regions, which are also known as the *entangle-*

ment features [4, 17, 36, 52]. One might be daunted by the fact that there are exponentially many entanglement features to keep track of, one for every possible sub-region of the system. Thus the classical post-processing will not be scalable. This paper aims to answer the challenge. In particular, two questions will be addressed:

1. Is there any hope of representing the entanglement features efficiently?

2. Furthermore, can we utilize such representation to determine and apply the reconstruction map efficiently to large systems?

The answer to the first question is yes: average sub-region purities of locally-scrambled systems have efficient representations as matrix product states (MPS) due to the fact that purities are completely positive and many-body states of positive wave functions are typically area-law entangled [22], such that the MPS bond dimension is a constant that does not scale with the system size n. The key idea is that the average purity of a n-qubit quantum system over all sub-regions can be compactly encoded in a fictitious many-body state vector, called the *entanglement feature state*. Given an ensemble $\mathcal{E}$ of random states (random density matrices) of the system, the entanglement feature state $|W_\mathcal{E}\rangle$ associated with $\mathcal{E}$ is a vector whose components are

$$(W_\mathcal{E})_A = \langle A|W_\mathcal{E}\rangle := \mathop{\mathbb{E}}_{\rho\in\mathcal{E}} \mathrm{Tr}_A(\mathrm{Tr}_{\bar{A}}\,\rho)^2, \tag{3}$$

where A denotes a sub-region of the system (as a subset of qubits) and $\bar{A}$ is the complement region of A. For each state ρ in the ensemble $\mathcal{E}$, $\rho_A = \mathrm{Tr}_{\bar{A}}\,\rho$ is the reduced density matrix of ρ in the region A, by tracing out qubits in the region $\bar{A}$. The entanglement feature $(W_\mathcal{E})_A$ in the region A is simply the purity of ρ_A averaged over all random states ρ in the ensemble $\mathcal{E}$. The entanglement feature state $|W_\mathcal{E}\rangle$ is a superposition of the basis states $|A\rangle$ labeled by sub-regions, and the superposition coefficients are the average purities in the corresponding sub-regions. Note that the entanglement feature state is not a physical state of the quantum system. It merely encodes the entanglement properties of the underlying physical quantum states. For example, the ensemble of *pure product states* has the following entanglement feature state

$$|W_\mathrm{prod}\rangle = \sum_A |A\rangle, \tag{4}$$

because the purity is unity in any sub-region of any pure product state. In this way, we can efficiently encode the purity information over all sub-regions in a single entanglement feature state $|W_\mathcal{E}\rangle$.

The advantage of the entanglement feature formalism lies in the fact that there exists an efficient numerical method to calculate the *time evolution* of the entanglement feature state as the underlying quantum state evolves in time together. Let $\mathcal{U}$ be a locally-scrambled random unitary ensemble and $\mathcal{E}$ be a random state ensemble. One can define a new random state ensemble $\mathcal{E}' = \{U\rho U^\dagger \mid \rho\in\mathcal{E}, U\in\mathcal{U}\}$, which can be interpreted as the ensemble of the random states evolved by the random unitaries. Then the entanglement features of $\mathcal{E}'$ and $\mathcal{E}$ are related by

$$|W_{\mathcal{E}'}\rangle = \hat{W}_\mathcal{U}\hat{W}_\mathbb{1}^{-1}|W_\mathcal{E}\rangle, \tag{5}$$

where $\hat{W}_\mathcal{U}$ is called the *entanglement feature operator* associated with the random unitary ensemble $\mathcal{U}$, whose matrix components are given by $\langle A|\hat{W}_\mathcal{U}|B\rangle = \mathbb{E}_{U\in\mathcal{U}}\,\mathrm{Tr}_{A,B}(\mathrm{Tr}_{\bar{A},\bar{B}}\,U)^2$ with A and B being the sub-regions on the future and the past sides of U respectively. $\hat{W}_\mathbb{1}$ is the entanglement feature operator for the identity operator, and its inverse operator is denoted as $\hat{W}_\mathbb{1}^{-1}$. Therefore, each step of the unitary evolution of the physical state ensemble can be mapped to a step of transfer matrix evolution of the entanglement feature state.

An efficient numerical approach to compute the entanglement feature dynamics Eq. (5) has been developed [4, 17] based on the MPS representation of entanglement feature states.

$$(W_\mathcal{E})_A = \langle A|W_\mathcal{E}\rangle = \mathrm{Tr}\prod_i(\delta_{i\notin A}\mathsf{W}^0_{\mathcal{E},i} + \delta_{i\in A}\mathsf{W}^1_{\mathcal{E},i}). \tag{6}$$

Here $\mathsf{W}^0_{\mathcal{E},i}$ and $\mathsf{W}^1_{\mathcal{E},i}$ are two sets of matrices parametrizing the MPS, where i labels the qubit/site. If the state ensemble $\mathcal{E}$ is translation invariant, we may choose to drop the index i on the MPS matrices $\mathsf{W}^0_{\mathcal{E},i}, \mathsf{W}^1_{\mathcal{E},i}$. Furthermore, for the pure state entanglement feature, the MPS matrices must satisfy various constraints. For example, for pure states, the purity of the whole system is one i.e. $\mathrm{Tr}(\mathsf{W}^0_\mathcal{E})^n = \mathrm{Tr}(\mathsf{W}^1_\mathcal{E})^n = 1$. In addition, because the purity of complementary regions is the same, i.e. $(W_\mathcal{E})_A = (W_\mathcal{E})_{\bar{A}}$, the entanglement feature state has a $\mathbb{Z}_2 : A \to \bar{A}$ symmetry, so the MPS tensor (formed by combining the two matrices into a single tensor) must carry definite $\mathbb{Z}_2$ representation. These symmetries greatly constrain the form of the MPS matrices (see Ref. [4] for more discussions).

As a fictitious many-body state, the entanglement feature state is typically area-law entangled due to its completely positive sign structure [22]. Thus it generally admits efficient MPS representations if the system is one-dimensional. For example, the entanglement feature state of pure product states Eq. (4) can be represented as a trivial MPS (a product state) with bond dimension one. Most typical entanglement feature states will have a low bond dimension even as the underlying physical state ensemble is of volume-law entanglement. For nearly all locally-scrambled dynamics with various entanglement production rates, it has been numerically observed [4] that a bond dimension $D_W = 2$ MPS state could capture the purity evolution in all sub-regions to high accuracy. The

accuracy can be further improved with a larger bond dimension. The evolution of the entanglement feature state, for example, by applying alternating layers of brick wall unitaries, can also be performed at the MPS level using the entanglement feature transfer matrix following Eq. (5). To implement that, we use a time-evolved block decimation (TEBD) approach and truncate at a fixed bond dimension D_W (see Appendix C of Ref. [17] or Appendix A, C of Ref. [4] for further details). With these previously developed tools, one can efficiently calculate the evolution of the average purity given a unitary ensemble. Furthermore, given the same circuit geometry, all unitary two-designs have the same entanglement feature dynamics, and so the Clifford circuit entanglement feature transfer matrix is equivalent to the random Haar one.

For the purpose of classical shadow tomography, what we need to know is the entanglement feature of the classical snapshots. Given any randomized measurement scheme (as specified by the random unitary ensemble $\mathcal{U}$), all possible classical snapshots form a random state ensemble, denoted as

$$\mathcal{E}_\sigma = \{U^\dagger|b\rangle\langle b|U \mid b \in \{0,1\}^{\times n}, U \in \mathcal{U}\}. \tag{7}$$

This is also the prior snapshot ensemble defined below Eq. (1). Following the definition in Eq. (3), we can define the entanglement feature state $|W_{\mathcal{E}_\sigma}\rangle$ for the classical snapshot ensemble $\mathcal{E}_\sigma$, whose components are

$$(W_{\mathcal{E}_\sigma})_A = \langle A|W_{\mathcal{E}_\sigma}\rangle := \mathop{\mathbb{E}}_{\hat\sigma\in\mathcal{E}_\sigma} \mathrm{Tr}_A(\mathrm{Tr}_{\bar A}\,\hat\sigma)^2. \tag{8}$$

Note that the classical snapshot state $\hat\sigma = U^\dagger|b\rangle\langle b|U$ is always constructed from a (reversed) unitary evolution $U^\dagger$ on a pure product state $|b\rangle$. So to obtain $|W_{\mathcal{E}_\sigma}\rangle$, we only need to take the MPS representation of $|W_{\mathrm{prod}}\rangle$ in Eq. (4) as the initial state and use the TEBD algorithm to evolve the entanglement feature MPS step by step following the entanglement feature dynamics in Eq. (5), as the underlying snapshot state gets evolved by the (reversed) Clifford circuit $U^\dagger$ layer by layer. Computing the entanglement feature $|W_{\mathcal{E}_\sigma}\rangle$ from the random unitary ensemble $\mathcal{U}$ is denoted as "EF Solver" in Fig. 2. The algorithm complexity is linear in the circuit depth L and independent of the number of qubit n (assuming translation symmetry).

According to Ref. [29], the entanglement feature of classical snapshots $|W_{\mathcal{E}_\sigma}\rangle$ is all we need to obtain the classical shadow reconstruction map $\mathcal{M}^{-1}$ for the corresponding randomized measurement scheme, as long as the random unitary ensemble $\mathcal{U}$ is locally scrambled (which applies to random Clifford circuits). The steps to obtain $\mathcal{M}^{-1}$ are as follows:

First, one solves for the *reconstruction coefficients* $|r\rangle$ (or $r_A = \langle A|r\rangle$) from $|W_{\mathcal{E}_\sigma}\rangle$ through the following consistency equation derived in Ref. [29] (which is essentially requiring $\mathcal{M}^{-1}\mathcal{M} = \mathbb{1}$ at the quantum channel level)

$$\sum_{A,C} r_A f_{A,B,C} (W_{\mathcal{E}_\sigma})_C = \delta_{B=\Omega}. \tag{9}$$

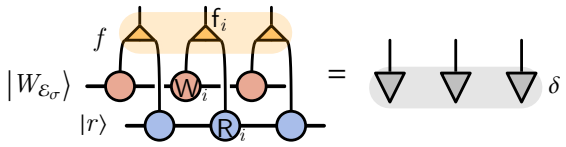


Figure 3: Illustration of the MPS equation Eq. (9). The fusion tensor f_i is defined in Tab. 1. The reconstruction coefficient $|r\rangle$ can be determined from the entanglement feature $|W_{\mathcal{E}_\sigma}\rangle$ by solving this MPS equation.

Here, A, C are summed over all sub-regions of the system (as subsets of qubits). Ω stands for the full system (as the full set of qubits) and $\delta_B = 1$ if $B = \Omega$ otherwise, $\delta_B = 0$. Also, r_A and $(W_{\mathcal{E}_\sigma})_C$ are components of the corresponding vectors $|r\rangle$ and $|W_{\mathcal{E}_\sigma}\rangle$, and $f_{A,B,C}$ is a fusion tensor that can be factorized to each site as $f = \bigotimes_{i=1}^n \mathsf{f}_i$, where components of the on-site tensor f_i are given by Tab. 1.

f_i	$i \notin A$		$i \in A$	
	$i \notin C$	$i \in C$	$i \notin C$	$i \in C$
$i \notin B$	2	0	$\frac{8}{3}$	$-\frac{4}{3}$
$i \in B$	0	0	$-\frac{2}{3}$	$\frac{4}{3}$

Table 1: Components of the f_i tensor (for qubit systems). The systematic formula for f_i in generic qudit systems can be found in Ref. [29].

Using MPS representations for both $|r\rangle$ and $|W_{\mathcal{E}_\sigma}\rangle$, and the tensor product structure of f, the left hand side of Eq. (9) may be represented as an MPS equation, as shown in Fig. 3. The left-hand side of Eq. (9) is made up of an MPS of bond dimension $D_r \times D_W$, whereas the right-hand side of Eq. (9) is a trivial MPS (a product state). We may then solve for the MPS tensors describing $|r\rangle$ by minimizing the following loss function

$$\begin{aligned}\mathcal{L} &= \|\mathcal{M}^{-1}\mathcal{M} - \mathbb{1}\|^2 \\ &= \sum_B \Big(\sum_{A,C} r_A f_{A,B,C} (W_{\mathcal{E}_\sigma})_C - \delta_{B,\Omega}\Big)^2.\end{aligned} \tag{10}$$

The optimization is achieved by the gradient descent algorithm using the PyTorch package [46] for auto-differentiation. The loss function can be suppressed to $\mathcal{L} \lesssim 10^{-3}$ with a modest bond dimension $D_r = 6$ for circuit depths $L \le 5$. In summary, for shallow circuits, we numerically observe that the reconstruction coefficients can be well described by a low bond-dimension MPS. Solving the reconstruction coefficient $|r\rangle$ from the entanglement feature $|W_{\mathcal{E}_\sigma}\rangle$ is denoted as the "$\mathcal{M}^{-1}$ Solver" in Fig. 2. The algorithm complexity is *linear* in the number n of qubits and *independent* of the circuit depth L.

Then, the reconstruction map is a linear functional of the reconstruction coefficients.

$$\mathcal{M}^{-1}(\sigma) = 2^n \sum_A r_A \sigma_A \tag{11}$$

Each reconstruction coefficient is weighted by a partial trace of the input σ embedded back into the Hilbert space so that the sum is over normalized density matrices on the full Hilbert space. More specifically, $\sigma_A = 2^{|A|-n}(\mathrm{Tr}_{\bar{A}}\,\sigma)\otimes \mathbb{1}^{\otimes(n-|A|)}$, which is a linear functional on σ. In practice, one works with individual snapshot states as opposed to their ensemble average. For Clifford circuit-based randomized measurement, each individual snapshot state is a stabilizer state, which admits efficient post-processing on a classical computer.

We would like to mention that the analytic expression for the reconstruction coefficient $|r\rangle$ is known from [8] by solving Eq. (9)

$$r_A = 2^{-n}(-1)^{|A|}\sum_{C|C\supseteq A}\frac{3^{|C|}}{\sum_{B|B\subseteq C}(-2)^{|B|}(W_{\mathcal{E}_\sigma})_B}. \tag{12}$$

Nevertheless, it does not seem to help our purpose, as we do not know how to encode this expression as a MPS efficiently. Directly evaluating Eq. (12) for every sub-region A is not scalable. Therefore, we will still follow our approach to solve Eq. (12) by MPS optimization.

In the next section, we will answer the second question: how can we leverage the scalability and efficiency of matrix product states for prediction using classical shadow tomography?

3 Applications for scalable prediction

3.1 Matrix Product State Representation of Reconstruction Coefficients

We consider that the quantum system is made of n qubits arranged on a one-dimensional lattice. We assume that there exists an efficient MPS representation for the reconstruction coefficients r_A. The existence of an efficient representation for $|r\rangle$ is implied from the tensor network structure of the self-consistency equation and the efficient MPS representation for the entanglement feature. Let $\mathsf{R}_i^j, j=0,1, i\in\{1\cdots n\}$ be the MPS matrices for $|r\rangle$ at site i of type j, i.e. R_i^j is a $D_r\times D_r$ real matrix, satisfying

$$r_A = \langle A|r\rangle = \mathrm{Tr}\left(\mathsf{B}\prod_{i=1}^{n}(\delta_{i\notin A}\mathsf{R}_i^0+\delta_{i\in A}\mathsf{R}_i^1)\right), \tag{13}$$

where B is a $D_r\times D_r$ real matrix serving as a twisted periodic boundary condition for the MPS. Here the site index $i=1,2,\cdots,n$ labels the qubits on the one-dimensional lattice. For $L=0,\infty$, we may drop the site index altogether because there is translation invariance. For $0<L<\infty$, there is only a two-site translation invariance, so only the parity of the site is important. For example, consider the deep circuit limit ($L=\infty$), which corresponds to random Clifford measurements. In this case, $r_\emptyset=-1$ and $r_\Omega = 1+2^{-n}$, where 2 is the local Hilbert space dimension and n is the number of qubits. One choice of tensors here is

$$\mathsf{R}^j = \begin{cases}\begin{pmatrix}1&0\\0&0\end{pmatrix} & j=0\\ \begin{pmatrix}0&0\\0&r_\Omega^{1/n}\end{pmatrix} & j=1\end{cases} \qquad \mathsf{B}=\begin{pmatrix}-1&0\\0&1\end{pmatrix}. \tag{14}$$

As the MPS matrix R_i^j is translationally invariant, we will drop the site index i and denote it as R^j. Introducing the boundary tensor can help simplify calculations. For example, without inserting the boundary tensor (or equivalently setting $\mathsf{B}=\mathbb{1}$), we have that $\lambda_1^n+\lambda_2^n=-1$ for the $\emptyset$ component, where λ_1,λ_2 are the eigenvalues of R^0. When n is even, the equation has no real solutions, so the eigenvalues must be complex. In practice, this makes solving for the R matrices more difficult, as one has to optimize over complex tensors; in addition, one has to perform complex tensor network contractions, which can also significantly slow down calculations. This can be avoided, as suggested above, through a boundary tensor.

On the other hand, in the shallow circuit limit $L=0$, the circuit reduces to random Pauli measurements. In this case, a trivial product state MPS suffices with

$$\mathsf{R}^j = \begin{cases}-1 & j=0\\ 3/2 & j=1\end{cases} \tag{15}$$

and $\mathsf{B}=1$. Note that translation invariance is restored is both limits because $\mathsf{R}^0,\mathsf{R}^1$ does not depend on the site index i explicitly. For finite-depth circuits, $0<L<\infty$, the brick wall structure breaks the translation invariance into the subgroup of two-site translations.

Also, as expected, the MPS representation of $|r\rangle$ both the shallow and deep circuit limits have low bond dimensions, although the underlying snapshot states have very different entanglement scalings (from area-law in $L=0$ to volume-law in $L\to\infty$). The MPS representation of the reconstruction coefficients r_A allows us to apply the reconstruction map Eq. (11) in prediction algorithms in a scalable way, which we will see next. Even simply knowing the reconstruction coefficients exactly does not solve this problem because there are an extensive number of them. Therefore, the MPS representation of the reconstruction coefficients is key for classical shadow tomography (beyond the Pauli measurement) to be feasible on near-term quantum devices with large numbers of qubits; otherwise, we cannot reconstruct classical shadows efficiently.

Given the MPS representation of the reconstruction coefficient $|r\rangle$, the reconstruction map $\mathcal{M}^{-1}$ may be viewed as a matrix product operator (MPO). The local tensors are a generalized dephasing channel $\mathcal{D}_{\mathsf{R}_i}$ defined in terms of the MPS tensor for the reconstruction coefficients at a particular site i. It acts on a

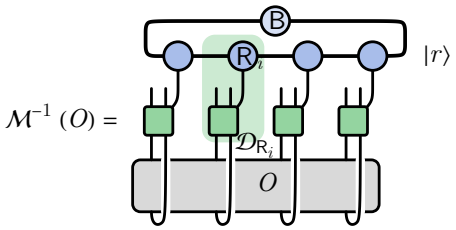


Figure 4: Tensor network representation of the reconstruction map $\mathcal{M}^{-1}(O)$ acting on a generic operator O, as defined in Eq. (17). Thick lines represent virtual bonds in the MPS. Each green box represents a dephasing channel $\mathcal{D}_{\mathsf{R}_i}$ at the qubit i in Eq. (16). The dephasing channels are correlated together by the MPS to form the reconstruction map.

generic operator O as

$$\mathcal{D}_{\mathsf{R}_i}(O) := \mathsf{R}_i^0 \otimes (\mathbb{1}\,\mathrm{Tr}_i(O)/2) + \mathsf{R}_i^1 \otimes O \tag{16}$$

Then the reconstruction map on an arbitrary operator can be given as a super-operator defined in terms of the MPS tensors for the reconstruction coefficients, see Fig. 4,

$$\mathcal{M}^{-1} = 2^n\,\mathrm{Tr}_{\mathrm{virtual}}\left(\mathsf{B}\prod_{i=1}^{n}\mathcal{D}_{\mathsf{R}_i}\right) \tag{17}$$

where the trace and product are over the virtual indices of the MPS. When acting on a single-qubit Pauli operator P_i, where $P_i \in \{\mathbb{1}, X, Y, Z\}$, the generalized dephasing channel evaluates to

$$\mathcal{D}_{\mathsf{R}_i}(P_i) = (\delta_{P_i=\mathbb{1}}\mathsf{R}_i^0 + \mathsf{R}_i^1) \otimes P_i. \tag{18}$$

Hence, single-qubit Pauli operators are eigen-operators of the generalized dephasing channel, and Pauli strings are eigen-operators of the reconstruction map. On a arbitrary n-qubit Pauli string $P = \bigotimes_{i=1}^n P_i$, where $P_i \in \{\mathbb{1}, X, Y, Z\}$ the reconstruction map evaluates to

$$\mathcal{M}^{-1}(P) = 2^n\,\mathrm{Tr}\left(\mathsf{B}\prod_i(\delta_{P_i=\mathbb{1}}\mathsf{R}_i^0 + \mathsf{R}_i^1)\right)P. \tag{19}$$

The above formula is most useful for designing prediction algorithms. As we will see below, we can use the self-adjoint property of the reconstruction map to apply it to the operator whose expectation value we want to predict. This then allows us to take the expectation value of the transformed observable under the snapshot state, which has an efficient representation as a stabilizer state.

3.2 Pauli Estimation

Suppose we wish to estimate a Pauli observable $P = \bigotimes_{i=1}^n P_i$, where $P_i \in \{\mathbb{1}, X, Y, Z\}$. Efficient classical shadow tomography schemes for the prediction of such operators only exist using random Pauli measurements and become quickly infeasible as the weight of the Pauli string grows. Here, we show how the scheme can be extended to finite-depth circuits given the reconstruction coefficients. In practice, the expectation value is given by an average over the snapshot states $\hat{\sigma}_j$, $j = 1 \cdots M$.

$$\langle P\rangle = \frac{1}{M}\sum_{m=1}^{M}\mathrm{Tr}(P\mathcal{M}^{-1}(\hat{\sigma}_m)). \tag{20}$$

Since the measurement channel $\mathcal{M}$ defined in Eq. (1) is self-adjoint, so as its inverse

$$\mathrm{Tr}(P\mathcal{M}^{-1}(\hat{\sigma})) = \mathrm{Tr}(\mathcal{M}^{-1}(P)\hat{\sigma}), \tag{21}$$

we may move $\mathcal{M}^{-1}$ onto P, and apply Eq. (19) to simplify $\mathcal{M}^{-1}(P)$, then Eq. (20) becomes

$$\langle P\rangle = 2^n\,\mathrm{Tr}\left(\mathsf{B}\prod_i(\delta_{P_i=\mathbb{1}}\mathsf{R}_i^0 + \mathsf{R}_i^1)\right)\langle P\rangle_\sigma, \tag{22}$$

where $\langle P\rangle_\sigma = \frac{1}{M}\sum_{m=1}^{M}\mathrm{Tr}(\hat{\sigma}_m P)$. This is a key result of this work. It enables us to extend classical shadow tomography beyond the Pauli measurement limit without losing the classical post-processing efficiency. Every term above is efficiently computable — the trace is over a product of n low-dimensional matrices and can be thought of as an inner product over MPS states, and the expectation value of the Pauli observable P on a stabilizer state $\hat{\sigma}_j$ can also be computed efficiently according to the Gottesmann-Knill theorem [2, 21], which is implemented in [30].

3.3 Generic Observable and Fidelity Estimation

What about more general operators? To effectively utilize the results of the previous section for scalable prediction, we need local tensor network descriptions of operators. One approach is to view the general operator O as a state in the Pauli basis i.e.

$$(O)_P = \langle P|O\rangle = \mathrm{Tr}(PO). \tag{23}$$

The operator O may be viewed as a state in the 4^n dimensional Hilbert space with basis states labeled by the different Pauli strings. Given this state description, we can now assign onsite matrices O_i^j where i labels the site and the j labels the Pauli at site i in the matrix-product expansion of the operator. If O has little operator entanglement, then the corresponding matrices O_i^j will also have a small bond dimension, and the decomposition can be considered efficient.

$$(O)_P = \mathrm{Tr}\left(\prod_i(\mathsf{O}_i^0\delta_{P_i=\mathbb{1}} + \mathsf{O}_i^1\delta_{P_i=X} + \mathsf{O}_i^2\delta_{P_i=Y} + \mathsf{O}_i^3\delta_{P_i=Z})\right). \tag{24}$$

One context in which more general operator estimation is useful is when estimating fidelity. The fidelity $F(\rho,\rho')$ provides a way to characterize the closeness of one state ρ to another ρ'. It is defined as $F(\rho,\rho') = (\mathrm{Tr}\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}})^2 \in [0,1]$. The greater the value of F, the closer the two states are. If any of the states ρ or ρ' is pure, their fidelity simplifies to

$$F(\rho,\rho') = \mathrm{Tr}(\rho\rho') \tag{25}$$

By viewing $F(\rho,\rho')$ as the expectation value of the operator $O=\rho'$ on the state ρ, the fidelity estimation amounts to predicting a low-rank, non-local observable $O=\rho'$ using the means described in this paper.

For fidelity estimation, one might be worried that the operator entanglement of the reference state ρ' is too large to represent efficiently as MPS. Though this may seem like an obstacle, ground states, low-entangled states, and stabilizer states of shallow, local circuits, all have an efficient MPS description of the form above. The last one, in particular, is useful in fidelity estimation, since we will use the MPS description of snapshot states to convert the trace into a tensor network. We will lay out the general formula given the matrix-product representation of ρ' per Eq. (24), and leave details about how to construct the MPS for $\hat{\sigma}$ in the Appendix A. A straightforward application of Eq. (19) gives the desired formula.

$$F(\rho,\rho') = \mathbb{E}_{\hat{\sigma}}\left[\sum_P (\hat{\sigma})_P(\rho')_P \,\mathrm{Tr}\left(\mathsf{B}\prod_i(\delta_{P_i=\mathbb{1}}\mathsf{R}_i^0 + \mathsf{R}_i^1)\right)\right] \tag{26}$$

Using the MPS representation for the Pauli coefficients of $\hat{\sigma}$ and ρ', the summation above can be viewed as a tensor network, as shown in Fig. 5. The idea is to take the MPS matrix for $\hat{\sigma},\rho'$ and r and contract them along an appropriate vertex tensor $u = \bigotimes_{i=1}^n \mathsf{u}_i$. If we label the indices of u_i as $(\mathsf{u}_i)_{P_i,P_i',j}$, where P,P' are Pauli strings in the basis expansion of $\hat{\sigma},\rho'$, respectively, and j labels the local reconstruction MPS matrix R_i^j, then

$$(\mathsf{u}_i)_{P_i,P_i',j} = \delta_{P_i=P_i'}(\delta_{P_i\neq\mathbb{1}}\delta_{j=1} + \delta_{P_i=\mathbb{1}}). \tag{27}$$

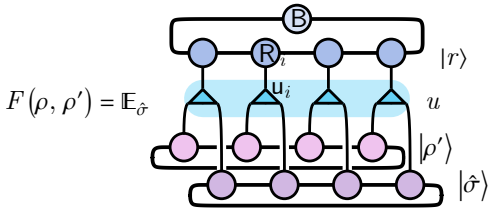


Figure 5: Tensor network representation of the fidelity estimator $F(\rho,\rho')$ in Eq. (26). The vertex tensor u_i is defined in Eq. (27). $|\rho'\rangle$ (or $|\hat{\sigma}\rangle$) is the vector representation of the density matrix ρ' (or the snapshot state $\hat{\sigma}$) in the Pauli operator basis.

Note that this procedure applies also for predicting generic local operators O, by expanding the operator in its Pauli basis representation first. One only needs to replace ρ' by O in Eq. (26), then $\langle O\rangle = \mathrm{Tr}(\rho O)$ can be evaluated by a similar tensor network of the same structure in Fig. 5. The procedure outlined above only relies on an efficient Pauli basis representation for the state or operator in question.

3.4 Locally-Scrambled Shadow Norm

Apart from estimating the expectation value of an observable O from the classical shadow data, we also want to bound the variance of our estimation. This is provided by the operator shadow norm $\|O\|_{\mathrm{shd}}^2$ introduced in Ref. [32]. In the following, we will focus on the case that O is a *traceless* operator, i.e. $\mathrm{Tr}\,O = 0$. Let $\mathrm{var}\,\langle O\rangle$ be the variance of the estimation $1/M\sum_{m=1}^M \mathrm{Tr}(O\mathcal{M}^{-1}(\hat{\sigma}_m))$ obtained from averaging M samples (for finite M). The variance of the mean decays as

$$\mathrm{var}\,\langle O\rangle \lesssim \|O\|_{\mathrm{shd}}^2/M, \tag{28}$$

where $\|O\|_{\mathrm{shd}}^2 = \max_\rho \mathbb{E}_{\hat{\sigma}\in\mathcal{E}_\sigma} 2^n\,\mathrm{Tr}(\hat{\sigma}\rho)(\mathrm{Tr}(\hat{\sigma}\mathcal{M}^{-1}(O)))^2$ is the shadow norm. A more precise statement of Eq. (28) and its proof can be found in Ref. [32].

The conventional shadow norm $\|O\|_{\mathrm{shd}}^2$ is defined as a maximum over all possible states ρ of the quantum system. This definition requires information about the third moment of the snapshot state ensemble $\mathcal{E}_\sigma$, which could take some effort to compute in general. In the context of locally-scrambled ensembles, Ref. [29] introduces an alternative version of the shadow norm, called the *locally-scrambled shadow norm*, denoted as $\|O\|_{\mathcal{E}_\sigma}^2$. It only requires the second moment of $\mathcal{E}_\sigma$, which can be conveniently read out from the entanglement feature $|W_{\mathcal{E}_\sigma}\rangle$ computed already. The key modification in the definition is to replace the maximization $\max_\rho$ by an average of ρ over its locally scrambled ensemble $\mathcal{E}_\rho = \{V\rho V^\dagger | V = \bigotimes_i V_i \in \mathrm{U}(2)^{\otimes n}\}$ with contains all states related to ρ by local basis transformations V. The locally-scrambled shadow norm is defined as

$$\begin{aligned}\|O\|_{\mathcal{E}_\sigma}^2 :&= \mathop{\mathbb{E}}_{\rho\in\mathcal{E}_\rho}\mathop{\mathbb{E}}_{\hat{\sigma}\in\mathcal{E}_\sigma} 2^n\,\mathrm{Tr}(\hat{\sigma}\rho)(\mathrm{Tr}(\hat{\sigma}\mathcal{M}^{-1}(O)))^2\\ &= \mathop{\mathbb{E}}_{\hat{\sigma}\in\mathcal{E}_\sigma}(\mathrm{Tr}(\hat{\sigma}\mathcal{M}^{-1}(O)))^2,\end{aligned} \tag{29}$$

where the dependence on ρ drops from the result due to the locally scrambled property of $\mathcal{E}_\rho$. So $\|O\|_{\mathcal{E}_\sigma}^2$ is only a function of the randomized measurement scheme specified by the prior snapshot ensemble $\mathcal{E}_\sigma$ (defined in Eq. (7)) and the observable O in question.

Using the definition of the measurement channel $\mathcal{M}$ in Eq. (1) and the reconstruction map $\mathcal{M}^{-1}$ in Eq. (11), Eq. (29) can be further reduced to

$$\begin{aligned}\|O\|_{\mathcal{E}_\sigma}^2 &= 2^{-n}\,\mathrm{Tr}(O\mathcal{M}^{-1}(O))\\ &= \sum_A 2^{|A|-n} r_A(W_O)_A,\end{aligned} \tag{30}$$

where r_A is the reconstruction coefficient given by the solution of Eq. (9). $(W_O)_A$ is the entanglement feature of the operator O, which follows the definition of the entanglement features of a state ensemble in Eq. (3), with the density matrix ρ simply replaced by the operator O:

$$(W_O)_A = \langle A|W_O\rangle := \mathrm{Tr}_A(\mathrm{Tr}_{\bar{A}}\, O)^2. \tag{31}$$

It is also possible to represent $|W_O\rangle$ as an MPS, similar to the state entanglement feature in Eq. (6),

$$(W_O)_A = \mathrm{Tr}\prod_i(\delta_{i\notin A}\mathsf{W}^0_{O,i} + \delta_{i\in A}\mathsf{W}^1_{O,i}). \tag{32}$$

This defines the MPS tensor $\mathsf{W}^j_{O,i}$ for any operator O. Hence, for a generic operator O, the shadow norm $\|O\|^2_{\mathcal{E}_\sigma}$ in Eq. (30) can be evaluated as a simple overlap given the MPS representations of two objects: the MPS matrices $\mathsf{R}^0, \mathsf{R}^1$ for the reconstruction coefficients $|r\rangle$ is given in Eq. (13); and the MPS matrices $\mathsf{W}^0_O, \mathsf{W}^1_O$ for $|W_O\rangle$ as in Eq. (32). Fig. 6 illustrates the tensor network contraction, which can be computed efficiently with complexity linear in the system size.

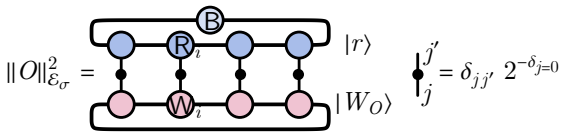


Figure 6: Tensor network representation of the locally-scrambled shadow norm $\|O\|^2_{\mathcal{E}_\sigma}$ defined in Eq. (30).

The locally-scrambled shadow norm is upper bounded by the conventional shadow norm as $\|O\|^2_{\mathcal{E}_\sigma} \le \|O\|^2_{\mathrm{shd}}$. It plays a similar role in variance estimation,

$$\mathrm{var}\,\langle O\rangle \simeq \|O\|^2_{\mathcal{E}_\sigma}/M. \tag{33}$$

The shadow norm $\|O\|^2_{\mathcal{E}_\sigma}$ has two complementary physical meanings: (i) it characterizes the estimation variance $\mathrm{var}\,\langle O\rangle$ given a fixed number of samples M (operators with larger shadow norms will have proportionally higher variances), (ii) it determines the sample complexity given the variance level (i.e. $M \sim \|O\|^2_{\mathcal{E}_\sigma}/\epsilon^2$ samples are needed in order to control the variance below a given threshold $\mathrm{var}\,\langle O\rangle \le \epsilon^2$ set by a small ϵ).

For Pauli estimation $O = P$, the operator entanglement feature $|W_P\rangle$ can be computed easily from its support. We define the support of a Pauli operator P as the subset of sites on which it acts non-trivially, denoted as $\mathsf{supp}\ P = \{i|P_i \neq \mathbb{1}\}$. Define the *operator weight* $k = |\mathsf{supp}\ P|$ as the size of the support of P. We find that the operator entanglement feature of a Pauli string P is simply given by

$$(W_P)_A = \delta_{\mathsf{supp}\ P\subseteq A}2^{2n-|A|}. \tag{34}$$

The corresponding MPS matrices for Pauli observables $\mathsf{W}^0_{P,i}, \mathsf{W}^1_{P,i}$ take different values depending on if site i is inside the support of P or not. Given the general MPS form in Eq. (32), it is easy to check that

$$\mathsf{W}^0_{P,i} = 4\delta_{i\notin\mathsf{supp}\ P}, \qquad \mathsf{W}^1_{P,i} = 2. \tag{35}$$

Nonetheless, the matrices have bond dimension one because of the tensor product structure of P.

Plugging in (34) into (30), we obtain the shadow norm of a Pauli string as a function of the reconstruction coefficients,

$$\|P\|^2_{\mathcal{E}_\sigma} = 2^n \sum_{A|A\supseteq\mathsf{supp}\ P} r_A. \tag{36}$$

Using the solution of $|r\rangle$ in Eq. (12), we have

$$\|P\|^2_{\mathcal{E}_\sigma} = \frac{(-3)^k}{\sum_{B|B\subseteq\mathsf{supp}\ P}(-2)^{|B|}(W_{\mathcal{E}_\sigma})_B}, \tag{37}$$

where $k = |\mathsf{supp}\ P|$. This result shows how the shadow norm of a Pauli operator P depends on both its support and the entanglement feature of the snapshot state ensemble.

In the following, we will directly evaluate the locally-scrambled shadow norm $\|O\|^2_{\mathcal{E}_\sigma}$ for several choices of the circuit depth L following Eq. (30). For notational convenience, we will treat $|W_O\rangle$ as translation-invariant unless evaluating the shadow norm for specific operator types that explicitly break translation symmetry, e.g. Pauli operators.

3.4.1 $L = 0$

Here, the unitary ensemble is simply Pauli measurement. The vector $|r\rangle$ is a product states, and the reconstruction coefficients are given in (15) as $\mathsf{R}^0 = -1, \mathsf{R}^1 = 3/2$. In terms of the matrices for $|W_O\rangle$, the expression for the shadow norm reduces the trace of a matrix power:

$$\|O\|^2_{L=0} = \mathrm{Tr}\left[\left(\frac{-\mathsf{W}^0_O + 3\mathsf{W}^1_O}{2}\right)^n\right]. \tag{38}$$

We can see that in this limit, the shadow norm is closely related to the locality of the operator, since if O is the identity on some site i, then the corresponding contribution to the shadow norm is a factor of 1, i.e. it doesn't affect the shadow norm.

Consider the case $O = P$. In the translation-symmetry breaking version of Eq. (38), the power of n breaks into individual factors for each site, with each term in the product depending on the operator entanglement feature matrix at that site:

$$\|P\|^2_{L=0} = \prod_{i=1}^n\left(\frac{-\mathsf{W}^0_{P,i} + 3\mathsf{W}^1_{P,i}}{2}\right). \tag{39}$$

Using Eq. (35), the product contributes a factor of 1 where P_i is identity and a factor of 3 where P_i is not.

Hence we see that the locally-scrambled shadow norm for P grows exponentially in the weight $k = |\mathsf{supp}\ P|$,

$$\|P\|^2_{L=0} = 3^k. \tag{40}$$

This result agrees with the conventionally defined shadow norm. It can also be seen to agree with (37) since the entanglement feature state of the unitary ensemble $|W_{\mathcal{E}_\sigma}\rangle$ is a trivial product state in the $L = 0$ case.

Furthermore, although the locally scrambled shadow norm is less than the conventionally defined shadow norm, for locally-scrambled unitary ensembles, it provides accurate bounds on the single-shot variance of Pauli observables [29]. Support for this claim is given in Sec. 4.

3.4.2 $L = 1$

Another important case we may check is $L = 1$. In this case, the ensemble is a product of two-local, two-design unitary gates. Each gate effectively scrambles site $2i$ with its neighbor at $2i + 1$, and generates no entanglement outside this unit cell. Within each unit cell, the unitary is Clifford random. Thus the MPS for reconstruction coefficients in each block for the $L = 1$ case is given by (14). Therefore,

$$\|O\|^2_{L=1} = \mathrm{Tr}\left[\left(\frac{-(\mathsf{W}^0_O)^2 + 5(\mathsf{W}^1_O)^2}{4}\right)^{n/2}\right]. \tag{41}$$

Consider next the case $O = P$, and suppose n is even for simplicity. Then the translation-symmetry breaking version of (41) for a Pauli operator is

$$\|P\|^2_{L=1} = \prod_{i=1}^{n/2}\left(\frac{-\mathsf{W}^0_{P,2i-1}\mathsf{W}^0_{P,2i} + 5\mathsf{W}^1_{P,2i-1}\mathsf{W}^1_{P,2i}}{4}\right). \tag{42}$$

There are two types of factors depending on the support of the operator P in that unit cell. If the Pauli operator P is trivial in the unit cell, then the factor is one. Otherwise, if the Pauli operator P has any support in the unit cell, then the factor is five. Hence, for a contiguous Pauli operator of weight k, there is a (staggered) exponential growth in the shadow norm due to the unit cell structure of the brick wall circuit i.e.

$$\|P\|^2_{L=1} = 5^{\lceil k/2\rceil}. \tag{43}$$

Thus, as the operator weight k grows, the shadow norm, and hence the number of samples required for reliable prediction, is already exponentially smaller for $L = 1$ than $L = 0$, since $\frac{\|P\|^2_{L=1}}{\|P\|^2_{L=0}} \sim (\frac{\sqrt{5}}{3})^k$. We find that this pattern, that longer operators can be better estimated by deeper circuits, continues for $L > 1$ in Sec. 4. For a given operator weight k, there is an ideal short circuit depth $L^* \geq 0$ such that the shadow norm and, therefore, the singles-shot variance are minimized. This shows that classical shadow tomography beyond Pauli measurements has practical applications since we can use circuits of varying depths to minimize the number of samples one has to take.

When $L \geq 2$, the shadow norm is more difficult to evaluate analytically. This is because both the entanglement features of the state ensemble and the reconstruction coefficients are more complicated. While the cases $L = 0, 1, \infty$ can be helpful, for $1 < L < \infty$, we resort to numerical evaluation of the shadow norm. See Sec. 4. We find that the shadow norm is a reliable indicator of the estimation variance in the case where O is a tensor product of local operators.

3.4.3 $L = \infty$

When $L \to \infty$, there are only two non-zero reconstruction coefficients $r_\emptyset = -1$ and $r_\Omega = 1 + 2^{-n}$, which correspond to the empty set and the total system, respectively. Then the shadow norm $\|O\|^2_{L=\infty}$ for a traceless Hermitian operator reduces to

$$\|O\|^2_{L=\infty} = (1 + 2^{-n})\,\mathrm{Tr}(O^2). \tag{44}$$

This is reminiscent of the conventional shadow norm introduced in [32], as both are proportional to $\mathrm{Tr}(O^2)$, indicating that the locally-scrambled shadow norm in the deep circuit limit also depends strongly on the rank of the operator. However, the locally-scrambled shadow norm has a smaller proportionality factor since the conventional shadow norm is defined by taking a maximum over all state ρ. We also note that afterwards in [5], some closed form analytic expressions were developed for the reconstruction map in brickwall circuits that are consistent with these results.

4 Numerical Demonstrations

In this section, we demonstrate the prediction algorithms for various system sizes based on numerical simulation[1]. We evolve the system using random two-local Clifford gates arranged in a brick wall pattern with periodic boundary conditions. The depth L of the circuit refers to the number of layers of Clifford gates. For example, for $L = 2$, we apply one even layer and then one odd layer of independently random two-qubit local Clifford gates. After the Clifford circuit evolution, we then measure the qubits and store the resulting snapshot state $\hat{\sigma} = U^\dagger|b\rangle\langle b|U$ for later use in predicting various quantities. For demonstration purposes, we collect the classical snapshots by numerically simulating the randomized measurement. Suppose the underlying original state ρ is a stabilizer state; the simulation can be efficiently carried out on

[1] Most of the numerical codes used in this paper can be found at this github repository.

a classical computer using the stabilizer table algorithm, thanks to the Gottesmann-Knill theorem.

In our numerical demonstrations, we consider two example initial states: the cluster state ρ_{ZXZ} and the Greenberger-Horne-Zeilinger (GHZ) state ρ_{GHZ}. They can be specified by their stabilizer groups,

$$\mathcal{G}_{\mathrm{ZXZ}} = \langle Z_1 X_2 Z_3, \cdots, Z_n X_1 Z_2 \rangle, \tag{45}$$
$$\mathcal{G}_{\mathrm{GHZ}} = \langle Z_1 Z_2, \cdots, Z_{n-1} Z_n, \textstyle\prod_{i=1}^{n} X_i \rangle, \tag{46}$$

Each stabilizer state $\rho = |\mathcal{G}|^{-1} \sum_{g\in\mathcal{G}} g$ is defined as the invariant state of the corresponding stabilizer group $\mathcal{G}$.

In the following figures, the initial state will be indicated using squares (for ρ_{ZXZ}) and triangles (for ρ_{GHZ}), respectively. The cluster state ρ_{ZXZ} is the ground state of a gapped, local, stabilizer Hamiltonian, and therefore carries strictly area-law entanglement. The GHZ state ρ_{GHZ} is the symmetric ground state of the Ising model, which contains long-range entanglement and is more challenging for prediction based on the randomized measurement on shallow local circuits. For example, the extensive generators (i.e. $\prod_i X_i$ in the $\mathcal{G}_{\mathrm{GHZ}}$ stabilizer group) will be hard to probe by local circuits of a finite depth L, because the generators of the snapshot state are at most $\sim 2L$ in length.

We then use the reconstruction coefficients, together with the algorithms outlined above, to perform prediction using the snapshot states. The reconstruction coefficients are calculated using PyTorch, specifically the AdamW optimizer, by minimizing the loss function $\mathcal{L}$ in Eq. (10). The minimization procedure is continued until $\mathcal{L} \leq 10^{-3}$.

We would like to briefly comment on the flexibility of this approach with respect to different circuit evolutions. As explained in Fig. 1, the input for determining the reconstruction coefficients is the entanglement features of the underlying unitary ensemble. One might be wondering if the brick-wall Clifford unitary circuit is the only specialized case which admits efficient entanglement feature dynamics. For example, one may be interested doing tomography using other locally-scrambled unitary ensembles, such as random Hamiltonian-generated evolution, which was proposed in [29]. It was found in prior work that locally-scrambled dynamics generated by continuously scrambling gates is also well described by a simple $D = 2$ MPS ansatz [4]. Similar efficiency was found in [17]. It is believed that this locally-scrambled purity dynamics is generically efficiently representable as MPS away from finely tuned points like the measurement-induced entanglement transition. Therefore, the shallow-circuit tomography scheme presented in this paper can be adapted to other circuits which are also locally-scrambled at the two-design level. We will leave the analysis of the tomography scheme in these more general contexts to future work.

Lastly, a similar approach using MPS to do efficient shallow circuit tomography was developed concurrently to ours in [7]. In this paper, the Pauli weight instead of the entanglement feature is computed. Since these quantities are related by local basis transformation in the entanglement-feature Hilbert space, we suspect that the efficiency of their approach will be comparable to ours.

4.1 Pauli Estimation

To demonstrate the efficiency of our approach in estimating Pauli observables, we will take $Z^{\otimes k} := \bigotimes_{i=1}^{k} Z_i$ as the target observable without loss of generality. The operator weight (operator size) k can be adjusted. Since the randomized measurement schemes are invariant under local Clifford transformations, any Pauli string P of the same weight k will exhibit the same sample efficiency (as the shadow norm $\|P\|^2_{\hat{\mathcal{E}}_\sigma}$ only depends on the support of P). Nevertheless, the expectation value $\langle P \rangle$ will still depend on the choice of the observable P and the underlying state. On the GHZ state, the expectation value of $Z^{\otimes k}$ is 1 or 0 depending on if k is even or odd, respectively. On the cluster state, it is 0 for all $k \geq 1$.

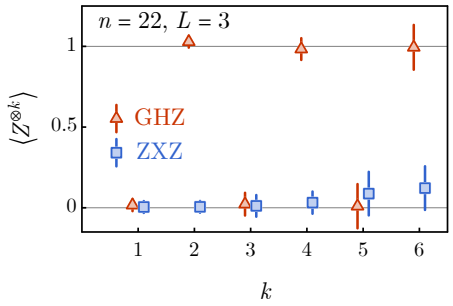


Figure 7: The estimated Pauli observable $Z^{\otimes k} := \bigotimes_{i=1}^{k} Z_i$ of different weight k with the underlying state ρ being the cluster (squares) and GHZ (triangles) states. Each point is based on 50000 measurement samples collected from a brick-wall circuit of depth $L = 3$ measurement protocol. The system size is $n = 22$ qubits. The error bar indicates two standard deviations of the estimated value.

In Fig. 7, we demonstrate that our method can reliably predict the expectation value of the Pauli string observable $Z^{\otimes k}$ on a system of $n = 22$ qubits via randomized measurements on the random Clifford circuit of $L = 3$ layers. Our numerical result is obtained by first simulating the randomized measurement on a classical computer and then processing the data using the algorithm described in Sec. 3.2. Our classical post-processing approach provides unbiased estimations of $Z^{\otimes k}$ on both the cluster state (blue squares) and the GHZ state (red triangles) for various operator weights k, although the estimation variance is growing with k (which will be analyzed soon). Here we only demonstrate the case of circuit depth $L = 3$, but

our method is applicable for other circuit depths, and the quality of estimation remains similar.

The variance of the estimation is associated with the sample complexity. A larger variance (a larger shadow norm) means more samples are needed to reduce the variance to the desired accuracy level. The traditional classical shadow tomography with Pauli measurements ($L = 0$) quickly requires an extensive number of samples to estimate a weight-k Pauli string observable (as the shadow norm $\|P\|^2_{L=0} = 3^k$ grows exponentially with k). However, with shallow circuits $L \geq 1$, we can easily achieve the same accuracy with far fewer samples. We have analytically proved this statement in Sec. 3.4.2 for $L = 1$ by calculating the locally-scrambled shadow norm. For large L, it is difficult to obtain a closed-form analytic expression for the shadow norm, but we can numerically investigate the shadow norm as a function of L according to Eq. (30) using tensor network techniques. Our results are presented in Fig. 8.

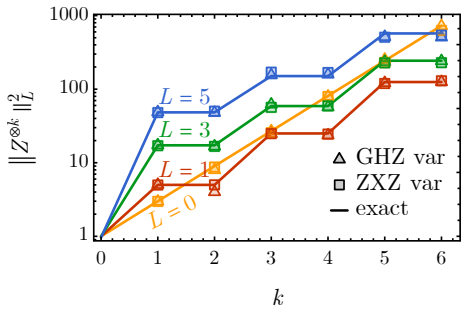


Figure 8: The locally-scrambled shadow norms for circuit depth $L = 0, 1, 3, 5$ (yellow, red, green, and blue lines, respectively), with $n = 22$ qubits. We also plot the corresponding variances over all samples for the cluster (squares) and GHZ (triangles) states. These agree with the shadow norms. The staggered behavior of the shadow norm for $L > 0$ is because these circuits break translation invariance.

We first verify that the locally-scrambled shadow norm $\|Z^{\otimes k}\|^2_L$ computed from Eq. (30) correctly captures the variance $\mathrm{var}\, Z^{\otimes k}$ of the estimated expectation value observed in our numerical simulation. Based on Eq. (33), we expect the variance and the shadow norm to match each other, as $\mathrm{var}\,\langle Z^{\otimes k}\rangle = \|Z^{\otimes k}\|^2_L/M$, where M is the number of samples. This is indeed the case as shown in Fig. 8. The numerically simulated variance and the computed shadow norm are in good agreement as we vary L and k, even as k approaches a quarter of the system size. This confirms the correctness of our approach to computing the shadow norm. The shadow norm is useful as it can give us an idea of the number of samples needed to control the estimation variance to the desired level, even before we start performing the classical shadow tomography experiment.

As shown in Fig. 8, the shadow norm $\|Z^{\otimes k}\|^2_L$ always grows with the operator weight k, meaning that we always need to perform more measurements to estimate longer Pauli strings accurately. However, the growth rates of shadow norms are different for different circuit depths L. For example, the $L = 1$ shadow norm increases with k slower than that of $L = 0$, meaning that the $L = 1$ circuit will be more advantageous for probing Pauli string operators of sufficiently large weight k.

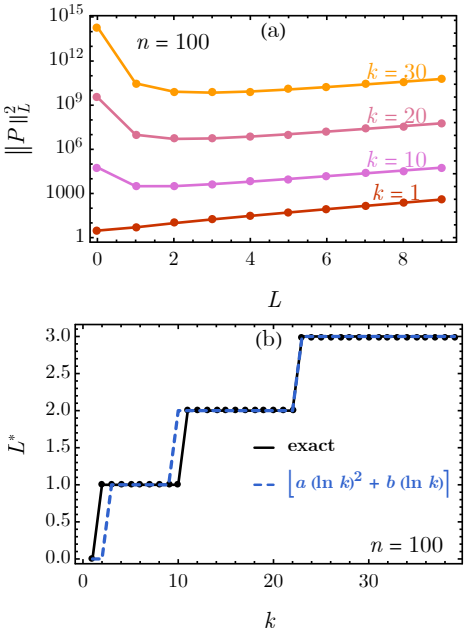


Figure 9: (a) The locally-scrambled shadow norm for a general Pauli operator P of size k v.s. the circuit depth L of the randomized measurement scheme. (b) The optimal value of circuit depth L^* to predict a contiguous Pauli operator of weight k. We use a system size of $n = 100$ qubits here. The results do not change as we further increase n.

To study this trend more quantitatively, we use (37) to compute the shadow norm of Pauli operators for various weights k under various circuit depths L. The result is presented in Fig. 9(a). Given the Pauli operator size k, it is clear that there is an optimal circuit depth L^* that minimizes the shadow norm. We conjecture that L^* can be fitted by an empirical formula

$$L^* \sim a(\ln k)^2 + b(\ln k). \tag{47}$$

The formula fits the numerically found L^* nicely with parameters $a = 0.14, b = 0.35$, as shown in Fig. 9(b), where the rounded value of the fitting function is shown since L^* must always be an integer. This confirms the sub-polynomial growth in the optimal circuit depth L^* as proposed in Eq. (47). The result indicates that a relatively shallow circuit can efficiently optimize the sample complexity for classical shadow tomography of large Pauli strings, which demonstrates the advantage of our shallow-circuit measurement approach. The logarithmic scaling of the optimal circuit depth was confirmed afterwards in [33]. In this work, the authors were able to derive a bound on the

shadow norm by studying the evolution of the weight distribution of contiguous operators under the random brick-wall circuit.

We would like to point out that for estimating the Pauli string with a fixed operator weight k, the shadow norm remains constant with the system size n, meaning that classical shadow tomography is scalable for this task. Our approach of using shallow-circuit measurements enables further improvement in the sample complexity compared to the Pauli measurement without losing the scalability and feasibility on near-term quantum devices.

4.2 Fidelity Prediction

Fidelity estimation is another important task in quantum information. We can also compare the fidelity of ρ_{GHZ} and ρ_{ZXZ} against various states using classical shadow tomography on shallow circuits. A high fidelity $F = 1$ ensures that the state produced in the lab is close to the desired state, whereas a low fidelity means that the state produced in the lab has very little overlap with the desired state. To demonstrate our approach, we simulate the shallow-circuit randomized measurement on a classical computer and then use the algorithm outlined in Sec. 3.3 to process the measurement data collected from the simulation. We estimate the fidelity of the classical shadow reconstructed state with the underlying original state and show that our approach successfully reconstructs the original quantum state from measurement outcomes with high fidelity. Although the shallow-circuit classical shadow tomography still has exponential sample complexity for the fidelity estimation task, increasing the circuit depth a bit can significantly reduce the base of this exponential complexity, which makes our approach useful for quantum state tomography on near-term quantum devices.

First, we compare the fidelity of the cluster state with itself as a function of the qubit number n for $n = 6, 10, 14, 18, 22$, using a circuit depth of $L = 3$, using 50000 samples. The result is shown in Fig. 10(a). We find the reconstructed states have close-to-one fidelities with negligible error bars up to 22 qubits. We also observe a (rather slow) exponential growth in the variance of the fidelity, in agreement with Ref. [29]. In that paper, a scaling form is proposed with var $F \propto \exp(cn/(L+1)^{\alpha})$. Fixing $\alpha = 0.72$, per Ref. [29], we can fit the parameter c based on our data. We find that the fitted value $c = 0.4$ agrees with the value $c = 0.47 \pm 0.08$ previously observed in Ref. [29]. This empirical relation suggests that although the variance of fidelity estimation (and hence the sample complexity) grows exponentially with the number of qubits n, increasing the circuit depth L in the shallow circuit regime can significantly reduce the variance (by making the exponential growth very slow).

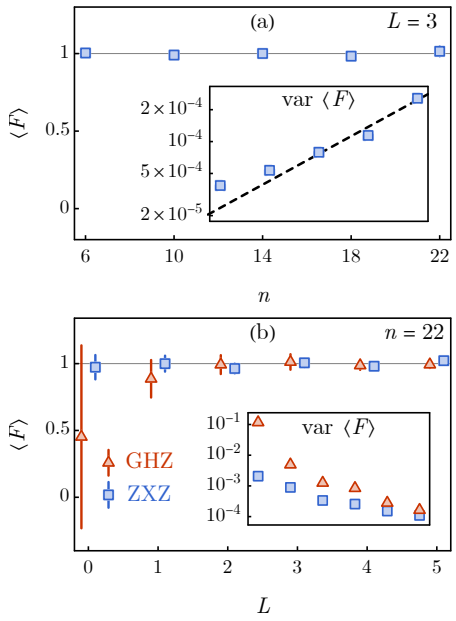


Figure 10: (a) The fidelity and its variance as a function of system size, for systems of size $n = 6, 10, 14, 18, 22$ for a fixed circuit depth of size $L = 3$. We use the cluster state as the reference state. Each point is the average of 50000 measurement samples. The mean fidelity is tightly centered around one, which is expected. The growth of the sample variance is indicated in the inset. (b) The fidelity and its variance versus circuit depth L for a system of $n = 22$ qubits. The two states are the cluster state (squares) and the GHZ state (triangles). We use a median-of-means average to estimate the fidelity, using 50000 total samples broken up into 12 groups. The error bar corresponds to two standard deviations. The inset shows that the variance decays with L rapidly.

Next, we study the efficiency of the protocol as a function of the circuit depth L for a system of size $n = 22$, as shown in Fig. 10(b). We plot the fidelity of both the cluster state and the GHZ states with themselves and find that both are centered around one, as expected. For both states, we observe a rapid decline in the variance of the sample fidelity as we increase circuit depth. However, we notice that the GHZ state has a much higher sample variance. This is likely due to the fact that the GHZ state has long-range correlations, unlike the cluster state which only has finite-range correlations. Thus quasi-local measurements will not be very efficient in probing the GHZ state. For example, at $L = 0$ (Pauli measurements), the reconstructed state has very low fidelity $F < 0.5$ and very large sample variance var$F \sim 1$ for the GHZ state. We see a sharp (more than an order of magnitude) decline in the sample variance for $L = 1$, and a continued exponential decline as we increase L. By $L = 3$, the fidelity approaches one for both states within a variance less than 10^{-3}. This is because a deeper scrambling circuit will enable the final computational basis measurements to probe more complex (higher weight) operators in the original basis, which benefits the fidelity estimation. This demonstrates the great advantage of entanglement-assisted shallow-circuit measurements. The rapid decline of the variance in fidelity versus circuit depth L makes it promising to use classical shadow tomography for fidelity estimation with shallow circuits achievable on near-term quantum devices.

5 Conclusion and Open Questions

In this paper, we demonstrate a novel approach to predicting expectation values of observables using shallow-depth circuit classical shadow tomography. Our work extends the well-known Pauli measurement and Clifford measurement protocols to the case of finite-depth brick wall circuits, which are available on near-term devices on many qubits. This allows us to use shallow circuit measurement to "approximate" the Clifford measurement limit to do fidelity estimation efficiently. It also allows us to use the scrambling properties of circuits with $L > 0$ to more efficiently probe large-weight operators, e.g. long Pauli strings. The approach that we take is based on our entanglement features formalism, which allows us to simply represent the purity data as MPS and consequently determine the reconstruction coefficients. Given the reconstruction coefficients and that they can be represented with a low-bond dimension MPS, we propose a scalable approach to predict many properties of quantum states.

There are still several questions that remain and that we wish to explore in future works. Firstly, is there an exact and efficient representation of the reconstruction coefficients as MPS? In our paper, we determine the reconstruction coefficients by learning their MPS tensors via gradient descent on the consistency equation Eq. (9). This procedure only needs to be done once, but it is bound to produce some error that can propagate through the prediction algorithm. In Ref. [8], the exact, individual reconstruction coefficients were determined. In practice, we want to be able to organize all 2^n coefficients efficiently — is there some way to represent the exact individual coefficients scalably? Such an approach would also give us insight into how the bond dimension or complexity of the reconstruction map evolves with circuit depth. We know that it is small in the shallow circuit limit and also small at $L = \infty$. Is it possible that the bond dimension remains small and finite for all values of the circuit depth L? Furthermore, is there some way to generalize the results in this paper to higher dimensions? For example, do the entanglement feature states and reconstruction coefficients now a low-dimensional PEPS structure?

Moreover, the prediction procedures described in Sec. 3 assume that the operator whose expectation value we want to predict has an efficient Pauli basis expansion. This means that the operator with a large amount of operator entanglement will be difficult to predict. Furthermore, this procedure is limited by the growth in operator entanglement of the snapshot states over time. The snapshot states' bond dimension grows exponentially in L, which can be seen from the fact that the typical size of the generators of the snapshot states grows linearly in L. This means this procedure will be costly for deep circuits. What approach can we take in this case?

In addition, we only considered measuring quantities that depend on the first moment of ρ i.e. $\mathrm{Tr}(O\rho)$. We could also consider quantities that depend on higher moments of ρ, e.g. $\mathrm{Tr}(O\rho \otimes \rho)$, where O is an operator on the doubled Hilbert space. How can we generalize the approach developed here to expectation values based on higher moments of ρ?

In this paper, we only considered a particular class of circuit ensembles formed by varying the circuit depth L. However, the procedure we outlined is independent of the particular circuit geometry. How might other geometries be useful in other contexts? For example, instead of considering depth L circuits, we could also consider circuits composed of adjacent k-qubit unitaries. Such circuits should also be feasible on near-term devices, provided that k is small. More ambitiously, one could also consider circuits with holographic geometries. Such circuit geometries may be optimal for prediction on gapless or scale-invariant states. We leave these cases to future research.

There are still many conceptual questions that have yet to be answered. For example, what is the physical meaning of the locally-scrambled shadow norm? The norm is closely related to the scrambling properties of the unitary ensemble, which begs the question of how

it can be related to other information-theoretic and thermodynamic quantities.

Our setup assumes a qubit system, where the local degrees of freedom are finite and commuting. How do we apply our approach in Sec. 3 to the fermionic context or other experimental contexts which may not be easily described in terms of qubits?

One may also consider different types of locally-scrambled or approximately locally-scrambled ensembles than the Clifford ensemble. For example, one may consider instead quantum Brownian motion ensembles, where the evolution is generated by a random Hamiltonian coupling nearby sites. The entanglement features of such ensembles have been extensively studied in prior work [4, 36, 52]. What are the reconstruction coefficients for such ensembles? What are their prediction properties? Classical shadow tomography is a rapidly developing field with many interesting applications and open questions. We hope that our paper provides some insights that can motivate further interesting developments.

Acknowledgements

We acknowledge the discussions with Christian Bertoni, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, Hakop Pashayan, who were working on similar topics concurrently [7]. We thank Soonwon Choi for the helpful discussion and previous collaboration. We also thank Xun Gao for the helpful discussion on tensor networks. The authors are supported by the UCSD Hellman Fellowship. This research was done using services provided by the OSG Consortium [47, 50], which is supported by the National Science Foundation awards 2030508 and 1836650.

References

[1] Scott Aaronson. Shadow Tomography of Quantum States. *arXiv e-prints*, art. arXiv:1711.01053, November 2017.

[2] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Phys. Rev. A*, 70:052328, Nov 2004. DOI: 10.1103/PhysRevA.70.052328. URL https://link.aps.org/doi/10.1103/PhysRevA.70.052328.

[3] Scott Aaronson and Guy N. Rothblum. Gentle Measurement of Quantum States and Differential Privacy. *arXiv e-prints*, art. arXiv:1904.08747, April 2019.

[4] A. A. Akhtar and Yi-Zhuang You. Multiregion entanglement in locally scrambled quantum dynamics. *Phys. Rev. B*, 102(13):134203, October 2020. DOI: 10.1103/PhysRevB.102.134203.

[5] Mirko Arienzo, Markus Heinrich, Ingo Roth, and Martin Kliesch. Closed-form analytic expressions for shadow estimation with brickwork circuits, 2022. URL https://arxiv.org/abs/2211.09835.

[6] Yimu Bao, Soonwon Choi, and Ehud Altman. Theory of the phase transition in random unitary circuits with measurements. *Phys. Rev. B*, 101 (10):104301, March 2020. DOI: 10.1103/PhysRevB.101.104301.

[7] Christian Bertoni, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, and Hakop Pashayan. Shallow shadows: Expectation estimation using low-depth random Clifford circuits. *arXiv e-prints*, art. arXiv:2209.12924, September 2022.

[8] Kaifeng Bu, Dax Enshan Koh, Roy J. Garcia, and Arthur Jaffe. Classical shadows with pauli-invariant unitary ensembles, 2022. URL https://arxiv.org/abs/2202.03272.

[9] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. Unknown quantum states: The quantum de Finetti representation. *Journal of Mathematical Physics*, 43(9):4537–4559, September 2002. DOI: 10.1063/1.1494475.

[10] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. *PRX Quantum*, 2:030348, Sep 2021. DOI: 10.1103/PRXQuantum.2.030348. URL https://link.aps.org/doi/10.1103/PRXQuantum.2.030348.

[11] Xiao Chen and Tianci Zhou. Quantum chaos dynamics in long-range power law interaction systems. *Phys. Rev. B*, 100(6):064305, August 2019. DOI: 10.1103/PhysRevB.100.064305.

[12] Soonwon Choi, Yimu Bao, Xiao-Liang Qi, and Ehud Altman. Quantum Error Correction in Scrambling Dynamics and Measurement-Induced Phase Transition. *Phys. Rev. Lett.*, 125(3):030505, July 2020. DOI: 10.1103/PhysRevLett.125.030505.

[13] Ze-Pei Cian, Hossein Dehghani, Andreas Elben, Benoît Vermersch, Guanyu Zhu, Maissam Barkeshli, Peter Zoller, and Mohammad Hafezi. Many-body chern number from statistical correlations of randomized measurements. *Phys. Rev. Lett.*, 126:050501, Feb 2021. DOI: 10.1103/PhysRevLett.126.050501. URL https://link.aps.org/doi/10.1103/PhysRevLett.126.050501.

[14] J. Ignacio Cirac, David Pé rez-García, Norbert Schuch, and Frank Verstraete. Matrix product states and projected entangled pair states: Concepts, symmetries, theorems. *Reviews of Modern Physics*, 93(4), dec 2021. DOI: 10.1103/revmodphys.93.045003. URL https://doi.org/10.1103%2Frevmodphys.93.045003.

[15] G. M. D'Ariano and P. Perinotti. Optimal Data Processing for Quantum Measurements. *Phys.*

Rev. Lett., 98(2):020403, January 2007. DOI: 10.1103/PhysRevLett.98.020403.

[16] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *arXiv e-prints*, art. arXiv:2203.11374, March 2022.

[17] Ruihua Fan, Sagar Vijay, Ashvin Vishwanath, and Yi-Zhuang You. Self-organized error correction in random unitary circuits with measurement. *Phys. Rev. B*, 103(17):174309, May 2021. DOI: 10.1103/PhysRevB.103.174309.

[18] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, September 2012. DOI: 10.1088/1367-2630/14/9/095022.

[19] Chenhua Geng, Hong-Ye Hu, and Yijian Zou. Differentiable programming of isometric tensor networks. *Machine Learning: Science and Technology*, 3(1):015020, jan 2022. DOI: 10.1088/2632-2153/ac48a2. URL https://doi.org/10.1088/2632-2153/ac48a2.

[20] Hrant Gharibyan, Masanori Hanada, Stephen H. Shenker, and Masaki Tezuka. Onset of random matrix behavior in scrambling systems. *Journal of High Energy Physics*, 2018(7):124, July 2018. DOI: 10.1007/JHEP07(2018)124.

[21] Daniel Gottesman. The heisenberg representation of quantum computers. 1998. DOI: 10.48550/ARXIV.QUANT-PH/9807006. URL https://arxiv.org/abs/quant-ph/9807006.

[22] Tarun Grover and Matthew P. A. Fisher. Entanglement and the sign structure of quantum states. *Physical Review A*, 92(4), oct 2015. DOI: 10.1103/physreva.92.042308. URL https://doi.org/10.1103%2Fphysreva.92.042308.

[23] Madalin Guta, Jonas Kahn, Richard Kueng, and Joel A. Tropp. Fast state tomography with optimal error bounds. *arXiv e-prints*, art. arXiv:1809.11162, September 2018.

[24] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *arXiv e-prints*, art. arXiv:1508.01797, August 2015.

[25] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows. *arXiv e-prints*, art. arXiv:2006.15788, June 2020.

[26] Guang Hao Low. Classical shadows of fermions with particle number symmetry. *arXiv e-prints*, art. arXiv:2208.08964, August 2022.

[27] Markus Hauru, Maarten Van Damme, and Jutho Haegeman. Riemannian optimization of isometric tensor networks. *SciPost Phys.*, 10:40, 2021. DOI: 10.21468/SciPostPhys.10.2.040. URL https://scipost.org/10.21468/SciPostPhys.10.2.040.

[28] Hong-Ye Hu and Yi-Zhuang You. Hamiltonian-driven shadow tomography of quantum states. *Physical Review Research*, 4(1):013054, January 2022. DOI: 10.1103/PhysRevResearch.4.013054.

[29] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical Shadow Tomography with Locally Scrambled Quantum Dynamics. *arXiv e-prints*, art. arXiv:2107.04817, July 2021.

[30] Hong-Ye Hu, Ryan LaRose, Yi-Zhuang You, Eleanor Rieffel, and Zhihui Wang. Logical shadow tomography: Efficient estimation of error-mitigated observables. *arXiv e-prints*, art. arXiv:2203.07263, March 2022.

[31] Hongye Hu. *Efficient representation and learning of quantum many-body states*. PhD thesis, UC San Diego, 2022.

[32] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, June 2020. DOI: 10.1038/s41567-020-0932-7.

[33] Matteo Ippoliti, Yaodong Li, Tibor Rakovszky, and Vedika Khemani. Operator relaxation and the optimal depth of classical shadows, 2023.

[34] Daniel F. V. James, Paul G. Kwiat, William J. Munro, and Andrew G. White. Measurement of qubits. *Physical Review A*, 64(5):052312, November 2001. DOI: 10.1103/PhysRevA.64.052312.

[35] Dax Enshan Koh and Sabee Grewal. Classical Shadows With Noise. *Quantum*, 6:776, August 2022. ISSN 2521-327X. DOI: 10.22331/q-2022-08-16-776. URL https://doi.org/10.22331/q-2022-08-16-776.

[36] Wei-Ting Kuo, A. A. Akhtar, Daniel P. Arovas, and Yi-Zhuang You. Markovian entanglement dynamics under locally scrambled quantum evolution. *Phys. Rev. B*, 101(22):224202, June 2020. DOI: 10.1103/PhysRevB.101.224202.

[37] Nima Lashkari, Douglas Stanford, Matthew Hastings, Tobias Osborne, and Patrick Hayden. Towards the fast scrambling conjecture. *Journal of High Energy Physics*, 2013:22, April 2013. DOI: 10.1007/JHEP04(2013)022.

[38] Ryan Levy, Di Luo, and Bryan K. Clark. Classical Shadows for Quantum Process Tomography on Near-term Quantum Computers. *arXiv e-prints*, art. arXiv:2110.02965, October 2021.

[39] Adam Nahum, Jonathan Ruhman, Sagar Vijay, and Jeongwan Haah. Quantum Entanglement Growth under Random Unitary Dynamics. *Physical Review X*, 7(3):031016, July 2017. DOI: 10.1103/PhysRevX.7.031016.

[40] Adam Nahum, Sagar Vijay, and Jeongwan Haah. Operator Spreading in Random Unitary Circuits. *Physical Review X*, 8(2):021014, April 2018. DOI: 10.1103/PhysRevX.8.021014.

[41] Simone Notarnicola, Andreas Elben, Thierry Lahaye, Antoine Browaeys, Simone Montangero, and Benoit Vermersch. A randomized measurement toolbox for rydberg quantum technologies, 2021. URL https://arxiv.org/abs/2112.11046.

[42] Ryan O'Donnell and John Wright. Efficient quantum tomography. *arXiv e-prints*, art. arXiv:1508.01907, August 2015.

[43] M. Ohliger, V. Nesme, and J. Eisert. Efficient and feasible state tomography of quantum many-body systems. *New Journal of Physics*, 15 (1):015024, January 2013. DOI: 10.1088/1367-2630/15/1/015024.

[44] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, oct 2014. DOI: 10.1016/j.aop.2014.06.013. URL https://doi.org/10.1016%2Fj.aop.2014.06.013.

[45] Marco Paini and Amir Kalev. An approximate description of quantum states. *arXiv e-prints*, art. arXiv:1910.10543, October 2019.

[46] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Curran Associates Inc., Red Hook, NY, USA, 2019.

[47] Ruth Pordes, Don Petravick, Bill Kramer, Doug Olson, Miron Livny, Alain Roy, Paul Avery, Kent Blackburn, Torre Wenaus, Frank Würthwein, Ian Foster, Rob Gardner, Mike Wilde, Alan Blatecky, John McGee, and Rob Quick. The open science grid. In *J. Phys. Conf. Ser.*, volume 78 of *78*, page 012057, 2007. DOI: 10.1088/1742-6596/78/1/012057.

[48] Stefan H. Sack, Raimel A. Medina, Alexios A. Michailidis, Richard Kueng, and Maksym Serbyn. Avoiding barren plateaus using classical shadows. *PRX Quantum*, 3:020365, Jun 2022. DOI: 10.1103/PRXQuantum.3.020365. URL https://link.aps.org/doi/10.1103/PRXQuantum.3.020365.

[49] Alireza Seif, Ze-Pei Cian, Sisi Zhou, Senrui Chen, and Liang Jiang. Shadow Distillation: Quantum Error Mitigation with Classical Shadows for Near-Term Quantum Processors. *arXiv e-prints*, art. arXiv:2203.07309, March 2022.

[50] Igor Sfiligoi, Daniel C Bradley, Burt Holzman, Parag Mhashilkar, Sanjay Padhi, and Frank Wurthwein. The pilot way to grid resources using glideinwms. In *2009 WRI World Congress on Computer Science and Information Engineering*, volume 2 of *2*, pages 428–432, 2009. DOI: 10.1109/CSIE.2009.950.

[51] Shenglong Xu and Brian Swingle. Locality, Quantum Fluctuations, and Scrambling. *Physical Review X*, 9(3):031048, July 2019. DOI: 10.1103/PhysRevX.9.031048.

[52] Yi-Zhuang You and Yingfei Gu. Entanglement features of random Hamiltonian dynamics. *Phys. Rev. B*, 98(1):014309, July 2018. DOI: 10.1103/PhysRevB.98.014309.

[53] Yi-Zhuang You, Zhao Yang, and Xiao-Liang Qi. Machine learning spatial geometry from entanglement features. *Phys. Rev. B*, 97(4): 045153, February 2018. DOI: 10.1103/PhysRevB.97.045153.

[54] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. Fermionic partial tomography via classical shadows. *Phys. Rev. Lett.*, 127:110504, Sep 2021. DOI: 10.1103/PhysRevLett.127.110504. URL https://link.aps.org/doi/10.1103/PhysRevLett.127.110504.

[55] Tianci Zhou and Xiao Chen. Operator dynamics in a Brownian quantum circuit. *Phys. Rev. E*, 99(5):052212, May 2019. DOI: 10.1103/PhysRevE.99.052212.

[56] Tianci Zhou and Adam Nahum. Emergent statistical mechanics of entanglement in random unitary circuits. *Phys. Rev. B*, 99(17):174205, May 2019. DOI: 10.1103/PhysRevB.99.174205.

A MPS-representation of a finite-depth stabilizer states

Here, we outline an algorithm for constructing the MPS representation of a stabilizer state, given the generators g_i of the corresponding stabilizer group $G = \langle g_1 \cdots g_n \rangle$. In the classical shadow protocol, we start with a trivial product state $|b\rangle = \otimes_{i=1}^n |b_i\rangle$ which represents a measurement outcome in the computational basis, and evolve by a two-local Clifford circuit. The initial stabilizers therefore can transform into more complicated Pauli group elements.

$$(-1)^{b_i} Z_i \to g_i \tag{48}$$

For a Clifford circuit of depth $L > 0$, the size of g_i is at most $2L$. The state ρ is then a sum over stabilizer group elements. Similarly, we can also represent the state as an MPS in the Pauli-basis, described by coefficients $c_{\vec{s}}$ where $\vec{s} \in \{0 \cdots 3\}^n$ denotes a Pauli string.

$$\rho = \frac{1}{2^n} \sum_{g \in G} g = \frac{1}{2^n} \sum_{\vec{b} \in \Omega^n} \prod_{i=1}^n g_i^{b_i} = \frac{1}{2^n} \sum_{\vec{s}} c_{\vec{s}} P^{\vec{s}} \tag{49}$$

For example, the GHZ state on two sites has two generators, $g_1 = ZZ, g_2 = XX$, and four nonzero components in the Pauli basis, corresponding to II, XX, YY, ZZ. Therefore, there are four nonzero components for c, namely $c_{00} = c_{11} = c_{33} = -c_{22} = 1/4$, and all other components are zero.

We describe a procedure to determine $c_{\vec{s}}$ as an MPS, given a set of generators $g_1 \cdots g_n$ and corresponding phases $(-1)^{s_1} \cdots (-1)^{s_n}$. The idea is that the local MPS tensor at site x should transmit through its virtual indices which of the stabilizers that have support on x are included in the product (i.e. have $b_i = 1$). First, we need to determine a contiguous set of sites that contain the support of each generator. Let us call this the *extent* of an operator. Given the extent of each generator, let $P_i = \mathbb{I} + (-1)^{s_i} g_i$ be an projection operator (or state in the Pauli basis) defined on the extent. Schematically, the local MPS tensor $[P_i^j]$ at site j is

$$[P_i^j] = \begin{cases} (\mathbb{I}, (-1)^{s_i} g_i^j) & j = 1 \\ \begin{pmatrix} \mathbb{I} & 0 \\ 0 & g_i^j \end{pmatrix} & 1 < j < k \\ (\mathbb{I}, g_i^j)^T & j = k \end{cases} \tag{50}$$

where k is the size of the extent, and g_i^j is the Pauli operator at site j in g_i. Given the MPS P_i, we only need one more ingredient to determine the MPS representation of the stabilizer state: Pauli algebra fusion tensor $V_{a,b}^c$. This tensor implements the product of Pauli operators, and evaluates to the correct phase associated to the product e.g.

$$V_{a,0}^c = \delta_{a,c}, V_{1,2}^3 = i = -V_{2,1}^3, \text{etc.} \tag{51}$$

Then, to determine the MPS tensor for the stabilizer state at site j, first collect all the generators whose extent is on j. Then, combine their MPS along the physical dimension using the fusion tensor. If there is more than two generators with support at j, then we can use multiple fusion tensors.