Classical shadow tomography with locally scrambled quantum dynamics

Hong-Ye Hu [1,2] Soonwon Choi,[3,4] and Yi-Zhuang You [1,*]

[1]*Department of Physics, University of California San Diego, La Jolla, California 92093, USA*
[2]*Department of Physics, Harvard University, 17 Oxford Street, Cambridge, Massachusetts 02138, USA*
[3]*Department of Physics, University of California, Berkeley, California 94720, USA*
[4]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

We generalize the classical shadow tomography scheme to a broad class of finite-depth or finite-time local unitary ensembles, known as locally scrambled quantum dynamics, where the unitary ensemble is invariant under local-basis transformations. In this case, the reconstruction map for the classical shadow tomography depends only on the average entanglement feature of classical snapshots. We provide an unbiased estimator of the quantum state as a linear combination of reduced classical snapshots in all subsystems, where the combination coefficients are solely determined by the entanglement feature. We also bound the number of experimental measurements required for the tomography scheme, so-called sample complexity, by formulating the operator shadow norm in the entanglement feature formalism. We numerically demonstrate our approach for finite-depth local unitary circuits and finite-time local-Hamiltonian generated evolutions. The shallow-circuit measurement can achieve a lower tomography complexity compared to the existing method based on Pauli or Clifford measurements. Our approach is also applicable to approximately locally scrambled unitary ensembles with a controllable bias that vanishes quickly. Surprisingly, we find a single instance of time-dependent local Hamiltonian evolution is sufficient to perform an approximate tomography as we numerically demonstrate it using a paradigmatic spin chain Hamiltonian modeled after trapped ion or Rydberg atom quantum simulators. Our approach significantly broadens the application of classical shadow tomography on near-term quantum devices.

I. INTRODUCTION

Quantum state tomography [1–3] is an essential task in many quantum technology applications. It seeks to reconstruct a quantum state from experimental data of repeated measurements. While reconstructing the full density matrix of a many-body system quickly becomes unfeasible with increasing system size due to the curse of dimensionality [4,5], predicting a collection of (possibly exponentially many) properties of the quantum system can still be efficiently achieved with an only polynomial number of state copies, which was the idea of *shadow tomography* proposed by Aaronson [6,7]. The idea is further improved by the recent work [8] to propose the *classical shadow tomography*, which significantly reduces the demand on the quantum hardware and enables efficient classical postprocessing.

Given a copy of an unknown quantum state ρ of N qubits, the classical shadow tomography protocol (see Fig. 1) first transforms the state $\rho \rightarrow \rho' = U\rho U^\dagger$ by a unitary U, which is randomly sampled (independently each time) from some probability distribution $P(U)$, then measures the trans-

formed state ρ' in the computational basis $\rho' \rightarrow |b\rangle\langle b|$, which collapses the system to a product state $|b\rangle$ labeled by a bit string $b \in \{0, 1\}^{\times N}$ of measurement outcomes $b = (b_1, \ldots, b_N)$ with the probability $P(b|\rho') = \langle b|\rho'|b\rangle$. Based on the observed bit string b and the classical description of the unitary U, a classical snapshot $\hat{\sigma}_{U,b} = U^\dagger|b\rangle\langle b|U$ can be constructed in principle, which essentially encodes the measurement outcomes together with their basis choice (pulled back through the unitary evolution). Repeating such measurements on independent and identical copies of ρ for a few times, a collection of classical snapshots $\mathcal{E}_{\sigma|\rho} = \{\hat{\sigma}_{U,b}\}$ can be obtained (which correlates with ρ). Reference [9] showed that as long as the unitary ensemble is expressive enough (i.e., tomographically complete), there exists a linear reconstruction map $\mathcal{M}^{-1}$ such that the density matrix ρ can be formally recovered as $\rho = \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \mathcal{M}^{-1}[\hat{\sigma}]$. This also enables the prediction of many properties of ρ, like the expectation value of any physical observable O as $\langle O \rangle = \text{Tr}(O\rho) = \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \text{Tr}(O\mathcal{M}^{-1}[\hat{\sigma}])$. The construction of classical snapshots $\hat{\sigma}_{U,b}$ and the computation of their associated properties are performed on a classical computer.

However, the existing methods [6–8,10] have limitations in applying to near-term quantum devices. First, depending on the type of observables O that we are interested in, one needs to employ different strategies to design the unitary circuit U. Two limiting cases have been analyzed in Ref. [8]: (i) if the observable is low rank (such as many-body overlap fidelity), it is most efficient to adopt deep circuits, such that U effectively forms a global Haar random ensemble; (ii) if the observable is

*yzyou@physics.ucsd.edu

high rank and quasilocal, it would be more efficient to adopt shallow circuits (e.g., the onsite Haar random). Otherwise the sample complexity will be high. However, the flexibility to interpolate between these two limits has not been available yet, such that the tomography protocol can not adjust to the target observables in a more adaptive manner. Second, more importantly, in existing quantum simulation platforms, applying random unitary circuits is very challenging because it requires high degrees of sophisticated quantum controls. In particular, for programmable quantum simulators of large systems based on trapped ions or Rydberg atom systems [11–13], a certain set of entangling unitary evolution is much more favorable to implement than typical random unitaries that require fine-tuned control. Therefore, it is desirable to develop a method applicable for systems with limited controls.

In this work, we address these challenges by generalizing the classical shadow tomography methods to a broad class of unitary ensembles. In our approach, the specific details of the unitary ensemble are not important as long as the ensemble generates *locally scrambled* quantum dynamics [14]. Rigorously speaking, the probability distribution $P(U)$ of evolution unitaries is invariant under local-basis transformations, i.e., $\forall\ V \in \mathrm{U}(d)^N : P(U) = P(UV) = P(VU)$ where $V = \prod_i V_i$ is a product of local unitary operator V_i on each qudit. This basically means that the unitary evolution U is efficient in scrambling local quantum information, such that the initial local-basis choice is quickly "forgotten" under the quantum dynamics. Examples of locally scrambled quantum dynamics includes random unitary circuits (including random Clifford circuit at the three-design level) [15–20] and quantum Brownian dynamics [21–25]. As the unitary ensemble does not care about local-basis choice, the only information that matters will be the quantum entanglement that the unitary dynamics can create in the quantum system. Therefore, for locally scrambled quantum dynamics, the reconstruction map only depends on the entanglement property of the classical snapshots. The density matrix ρ can be reconstructed as a linear superposition of the classical snapshot $\hat{\sigma}$ reduced in different subsystems. The combination coefficient can be calculated from the entanglement feature [26,27] of the classical snapshots, which is simply the collection of average purities of classical snapshots in all possible subregions.

Since our method is applicable to a broad class of quantum dynamics, it is natural to consider an ensemble of realistic Hamiltonian evolutions that are readily available in near-term quantum devices. To this end, we introduce an *approximate* classical shadow tomography (with a nonvanishing but small bias) applicable to an ensemble of time-dependent Hamiltonian evolution that generates approximately locally scrambled dynamics. We numerically demonstrate this idea by using a simple spin chain Hamiltonian modeled after programmable trapped ions or Rydberg atom array systems. We introduce the local frame potential to characterize the bias and we show the bias decreases rapidly for the initial short period of time, and reaches a vanishingly small plateau value for the proposed Hamiltonian. Surprisingly, we find even a single instance from an ensemble of Hamiltonian evolution suffices to perform an approximate tomography, implying that our method is hardware efficient for existing quantum devices [28,29].

In the following, we will first establish the general theoretical framework to calculate the reconstruction map in Sec. II A and to bound the sample complexity in Sec. II B. We also provide a two-qudit toy model to analytically demonstrate our construction in Sec. II C. We comment on how to carry out the computation efficiently in Sec. II D. Then we apply our construction for local unitary circuits and numerically demonstrate its accuracy in quantum fidelity and Pauli observable estimation tasks in Sec. III A, as well as their scaling of sample complexity in Sec. III B. Finally, we show in Sec. III D that our approach can be extended to broader classes of unitary ensembles that are approximately locally scrambled. We propose a frame potential to characterize the level of approximation, which serves as a powerful indicator to design nearly locally scrambled unitary ensembles that are available for existing analog quantum simulators [28,29]. We summarize our classical postprocessing protocol and outline a few interesting future applications in Sec. IV

II. THEORETICAL FRAMEWORK

A. Reconstruction map from entanglement features

To be general, we consider a quantum system consists of N qudits, where each qudit has the Hilbert space dimension d (where $d = 2$ corresponds to the qubit system). The protocol of classical shadow tomography describes a process that first measures the unknown quantum state ρ in a random basis specified by the unitary transformation U and then prepares the classical snapshot $\hat{\sigma}_{U,b} \equiv U^\dagger |b\rangle\langle b| U$ based on the measurement outcome b. The randomness involved in the process includes (i) sampling U from the distribution $P(U)$ and (ii) obtaining the measurement outcome b conditioned on the evolved state $\rho' = U\rho U^\dagger$ with the probability $P(b|\rho') = \langle b|\rho'|b\rangle = \mathrm{Tr}(\hat{\sigma}_{U,b}\rho)$. Inspired by the discussion in Ref. [30], we define

$$\mathcal{E}_{\sigma|\rho} = \{\hat{\sigma}_{U,b} \mid P(\hat{\sigma}_{U,b}|\rho) = \mathrm{Tr}(\hat{\sigma}_{U,b}\rho)P(U)\} \qquad (1)$$

as the *posterior snapshot ensemble*, as it is conditioned on the observation of ρ. The posterior snapshot ensemble reduces to the *prior snapshot ensemble*

$$\mathcal{E}_\sigma = \{\hat{\sigma}_{U,b} \mid P(\hat{\sigma}_{U,b}) = d^{-N}P(U)\}, \qquad (2)$$

when there is no knowledge contained in ρ, i.e., $\rho = d^{-N}\mathbb{1}$. For the prior distribution $P(\hat{\sigma}_{U,b})$, the outcome b is uniformly drawn from all possible outcomes in $\{0, 1, \ldots, d - 1\}^{\times N}$ (independent of U, ρ). The prior snapshot ensemble $\mathcal{E}_\sigma$ only depends on the unitary ensemble $\mathcal{E}_U = \{U|P(U)\}$.

With the notation introduced above, the expected classical snapshot σ can be expressed as

$$\sigma \equiv \mathop{\mathbb{E}}_{\hat{\sigma}\in\mathcal{E}_{\sigma|\rho}} \hat{\sigma} = \mathop{\mathbb{E}}_{\hat{\sigma}\in\mathcal{E}_\sigma} \hat{\sigma}\ \mathrm{Tr}(\hat{\sigma}\rho)d^N = \mathcal{M}[\rho], \qquad (3)$$

which is related to the original state ρ by a quantum channel $\mathcal{M}$, called the *measurement channel*. It is easy to check that the measurement channel $\mathcal{M}$ is trace preserving, completely positive, and self-adjoint. It is generally difficult to obtain an explicit expression of $\mathcal{M}$ for generic unitary ensemble $\mathcal{E}_U$ (or for generic prior snapshot ensemble $\mathcal{E}_\sigma$). Results of $\mathcal{M}$ are known for global and onsite two-design unitaries [9,31,32]

(possibly with noise [33,34]), fermionic Gaussian unitaries [35], and many-body Gaussian unitaries [10].

We can make progress in computing the measurement channel $\mathcal{M}$ (and its inverse) for yet another class of unitary ensemble, namely, the *locally scrambled unitaries* [14], for which $P(U)$ obeys the local-basis invariance condition

$$\forall\, V \in \mathrm{U}(d)^N : P(U) = P(UV) = P(VU), \qquad (4)$$

where the local scrambling unitary V is an element in the group $\mathrm{U}(d)^N$ [the tensor product of the onsite unitary group $\mathrm{U}(d)$ of each qudit]. This condition is sufficient to ensure the prior ensemble $\mathcal{E}_\sigma$ of snapshot states $\hat{\sigma}$ to be invariant under $\hat{\sigma} \to V^\dagger \hat{\sigma} V$:

$$\forall\, V \in \mathrm{U}(d)^N : P(\hat{\sigma}) = P(V^\dagger \hat{\sigma} V). \qquad (5)$$

In this case, we say that $\mathcal{E}_\sigma$ is a *locally scrambled ensemble*. In fact, our following derivation only requires the weaker condition (5) at the state level, instead of Eq. (4) at the channel level, though it will be practically more straightforward to design unitary circuits that satisfy Eq. (4) by assembling locally scrambled unitary gates.

Nevertheless, as long as the states $\hat{\sigma}$ are locally scrambled (even if the unitaries U may or may not be locally scrambled), we will be able to insert local-basis transformations V in Eq. (3), and average V over any ensemble of our choice:

$$\sigma = \mathop{\mathbb{E}}_{V \in \mathrm{U}(d)^N} \mathop{\mathbb{E}}_{\hat{\sigma} \in \mathcal{E}_\sigma} V^\dagger \hat{\sigma} V \ \mathrm{Tr}(V^\dagger \hat{\sigma} V \rho) d^N. \qquad (6)$$

We can choose the ensemble of $V = \prod_i V_i$ to be such that every V_i is independently a local two-design unitary. With this choice, the ensemble average of V can be evaluated by averaging every V_i over the Haar unitary measure following Refs. [36,37], and the result can be written as (see Appendix A for derivation)

$$\sigma = \sum_{B,C \in 2^{\Omega_N}} d^{2N-|B|} \rho_B \mathsf{Wg}_{B,C} W^{(2)}_{\mathcal{E}_\sigma,C}, \qquad (7)$$

with B, C summing over all possible subregions of the N qudit system, where each subregion is labeled by a subset of $\Omega_N = \{1, \ldots, N\}$ (as an element in the power set 2^{Ω_N}). $|B|$ denotes the size (cardinality) of the region B. $\rho_B = (\mathrm{Tr}_{\bar{B}} \rho) \otimes (\mathbb{1}_{\bar{B}}/d^{|\bar{B}|})$ is the reduced density matrix of ρ in region B embedded back into the total Hilbert space. $\bar{B}$ denotes the complement of region B. Note that B and $\bar{B}$ do not need to be consecutive regions in the space, and they can intertwine with each other in general. $\mathsf{Wg}_{B,C} = (d^2-1)^{-N}(-1/d)^{|B \ominus C|}$ is the Weingarten function of regions B and C, where $B \ominus C$ denotes the subregions that belong to either B or C but not both,

$$W^{(2)}_{\mathcal{E}_\sigma,C} \equiv \mathop{\mathbb{E}}_{\hat{\sigma} \in \mathcal{E}_\sigma} \mathrm{Tr}_C(\mathrm{Tr}_{\bar{C}} \hat{\sigma})^2 = \mathop{\mathbb{E}}_{\hat{\sigma} \in \mathcal{E}_\sigma} e^{-S^{(2)}_C(\hat{\sigma})} \qquad (8)$$

is the second *entanglement feature* [26,27] of the prior snapshot ensemble $\mathcal{E}_\sigma$, where $S^{(2)}_C(\hat{\sigma})$ denotes the second Rényi entanglement entropy of the state $\hat{\sigma}$ in region C. The entanglement feature $W^{(2)}_{\mathcal{E}_\sigma,C}$ is merely a property of the unitary ensemble $\mathcal{E}_U$ (which determines $\mathcal{E}_\sigma$). It describes how the unitary channel entangles a product state in general. It depends on neither the underlying state ρ to be reconstructed nor

any particular snapshot state $\hat{\sigma}$ collected in the tomography process.

Given the entanglement feature $W^{(2)}_{\mathcal{E}_\sigma,C}$, Eq. (7) spells out how the expected classical snapshot σ is written as a linear combination of reduced density matrices ρ_B in all regions, which explicitly specifies the measurement channel $\mathcal{M}$ as a linear map $\sigma = \mathcal{M}[\rho]$ from ρ to σ. Therefore, any reduced classical snapshot σ_A must also be a linear combination of reduced density matrices ρ_B, which implies that the measurement channel can be represented as a matrix $\mathcal{M}_{AB}$ such that $\sigma_A = \sum_B \mathcal{M}_{AB} \rho_B$. Suppose the map $\mathcal{M}$ is invertible (i.e., the unitary ensemble is tomographically complete), the inverse map $\mathcal{M}^{-1}$ (the reconstruction map) must also be a linear map that combines all reduced classical snapshots σ_A to reconstruct $\rho_B = \sum_A (\mathcal{M}^{-1})_{BA} \sigma_A$. In particular, we are most interested to reconstruct the full density matrix ρ (because all reduced density matrices follow from its partial trace), which must also be a linear combination of σ_A with some coefficients $r_A \in \mathbb{R}$,

$$\rho = \mathcal{M}^{-1}[\sigma] = d^N \sum_{A \in 2^{\Omega_N}} r_A \sigma_A, \qquad (9)$$

where $\sigma_A = (\mathrm{Tr}_{\bar{A}} \sigma) \otimes (\mathbb{1}_{\bar{A}}/d^{|\bar{A}|})$ follows the same definition as the reduced density matrix. The reconstruction map $\mathcal{M}^{-1}$ is not a physical channel because the reconstruction coefficients r_A may not be positive definite in general. Nevertheless, $\mathcal{M}^{-1}$ is still trace preserving and self-adjoint. Since $\mathcal{M}^{-1}$ is linear, we have $\rho = \mathcal{M}^{-1}[\mathbb{E}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \hat{\sigma}] = \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \mathcal{M}^{-1}[\hat{\sigma}]$, which enables us to reconstruct the underlying state ρ from the ensemble of classical snapshots. The collection of $\hat{\rho} = \mathcal{M}^{-1}[\hat{\sigma}]$ is also called the *classical shadow* [8] of ρ, which can then be used to predict many properties of ρ efficiently.

Now the key problem is to compute r_A from $W^{(2)}_{\mathcal{E}_\sigma,C}$. For a system of N qudits, there will be 2^N many reconstruction coefficients r_A. To determine them, we substitute Eq. (7) to Eq. (9) and find

$$\rho = \sum_{A,B,C \in 2^{\Omega_N}} f_{A,B,C}\, r_A \rho_B W^{(2)}_{\mathcal{E}_\sigma,C}, \qquad (10)$$

with the fusion coefficient $f_{A,B,C}$ given by

$$f_{A,B,C} = \sum_{D \in 2^{\Omega_N}} \delta_{B,A \cap D} d^{2N+|A|-|B|+|\bar{A} \cap \bar{D}|} \mathsf{Wg}_{D,C}$$

$$= \left(\frac{d^3}{d^2-1}\right)^N \sum_{D \in 2^{\Omega_N}} \delta_{B,A \cap D} d^{-|D|} \left(-\frac{1}{d}\right)^{|C \ominus D|}, \quad (11)$$

which is universally determined by the qudit dimension d. Here $\delta_{A,B}$ denotes the Kronecker delta of two regions A and B, such that $\delta_{A,B} = 1$ (or 0) if $A = B$ (or $A \neq B$). Equation (10) will hold for any choice of ρ if and only if

$$\sum_{A,C \in 2^{\Omega_N}} r_A f_{A,B,C} W^{(2)}_{\mathcal{E}_\sigma,C} = \delta_{B,\Omega_N}, \qquad (12)$$

where $\Omega_N = \{1, \ldots, N\}$ is the full set that labels the full system of N qudits. By solving this linear equation, we can determine the reconstruction coefficients r_A in terms of of the entanglement feature $W^{(2)}_{\mathcal{E}_\sigma,C}$, such that the reconstruction map $\mathcal{M}^{-1}$ can be constructed according to Eq. (9).

In conclusion, we provide a general framework to compute the reconstruction map for the classical shadow tomography with locally scrambled quantum dynamics. The protocol is summarized as follows:

---

(1) Given the prior snapshot ensemble $\mathcal{E}_\sigma$, first calculate its entanglement feature by

$$W^{(2)}_{\mathcal{E}_\sigma, C} = \underset{\hat{\sigma} \in \mathcal{E}_\sigma}{\mathbb{E}} \operatorname{Tr}_C(\operatorname{Tr}_{\bar{C}} \hat{\sigma})^2.$$

(2) Solve for the reconstruction coefficient r_A by

$$\sum_{A, C \in 2^{\Omega_N}} r_A f_{A,B,C} W^{(2)}_{\mathcal{E}_\sigma, C} = \delta_{B, \Omega_N}.$$

(3) Then the reconstruction map is given by

$$\rho = \mathcal{M}^{-1}[\sigma] = d^N \sum_{A \in 2^{\Omega_N}} r_A \sigma_A.$$

All computations are supposed to be carried out on a classical computer in the postprocessing procedure. Although solving for r_A may be computationally demanding for large systems, it only needed to be done once and its result can be applied to process all classical snapshots collected from all possible states ρ to be learned.

B. Variance estimation and sample complexity

Given the ensemble $\mathcal{E}_{\sigma|\rho}$ of classical snapshots collected from measuring the unknown state ρ, we can use the reconstruction map $\mathcal{M}^{-1}$ to predict properties of ρ. For example, let O be a traceless Hermitian operator representing a physical observable. Its expectation value $\langle O \rangle \equiv \operatorname{Tr}(O\rho)$ can be predicted via

$$\langle O \rangle = \underset{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}}{\mathbb{E}} \operatorname{Tr}(O\mathcal{M}^{-1}[\hat{\sigma}]) = \underset{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}}{\mathbb{E}} \operatorname{Tr}(\mathcal{M}^{-1}[O]\hat{\sigma}), \quad (13)$$

where we have used the self-adjoint property of $\mathcal{M}^{-1}$ to transpose its action from $\hat{\sigma}$ to O. We can interpret $\hat{o}(\hat{\sigma}) \equiv \operatorname{Tr}(\mathcal{M}^{-1}[O]\hat{\sigma})$ as the single-shot estimation of the observable (based on a particular classical snapshot $\hat{\sigma}$), such that $\langle O \rangle = \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \hat{o}(\hat{\sigma})$.

The variance of the single-shot estimation is defined as $\operatorname{Var} \hat{o} \equiv \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \hat{o}(\hat{\sigma})^2 - (\mathbb{E}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \hat{o}(\hat{\sigma}))^2$, which can be bounded by (the first term in $\operatorname{Var} \hat{o}$)

$$\operatorname{Var} \hat{o} \leqslant \|O\|^2_{\mathcal{E}_{\sigma|\rho}} \equiv \underset{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}}{\mathbb{E}} \hat{o}(\hat{\sigma})^2$$

$$= \underset{\hat{\sigma} \in \mathcal{E}_\sigma}{\mathbb{E}} (\operatorname{Tr} \mathcal{M}^{-1}[O]\hat{\sigma})^2 \operatorname{Tr}(\hat{\sigma}\rho)d^N. \quad (14)$$

The bound $\|O\|_{\mathcal{E}_{\sigma|\rho}}$ can be considered as a generalized ρ-dependent notion of the (squared) *shadow norm* [8] of an operator O (whereas the shadow norm originally defined in Ref. [8] further maximizes over all possible underlying states ρ to remove the dependence on ρ). Assuming $\mathcal{E}_\sigma$ is locally scrambled, following the same approach of inserting and averaging local-basis transformations as in Eq. (6), the bound in

Eq. (14) becomes

$$\|O\|^2_{\mathcal{E}_{\sigma|\rho}} = \sum_{g,h \in S_3^N} \|O\|^2_{\rho, g} \mathsf{Wg}_{g,h} W^{(3)}_{\mathcal{E}_\sigma, h}, \quad (15)$$

where g, h are group elements in the S_3^N (product of threefold permutation groups over N qudits). $\mathsf{Wg}_{g,h}$ is the Weingarten function of permutations g and h, which is equivalent to traditional Weingarten function as $\mathsf{Wg}_{g,h} = \mathsf{Wg}(gh^{-1}, d)$, where d is the local Hilbert dimension of qudit. $\|O\|^2_{\rho, g}$ is a generalized operator norm for O, which is defined as

$$\|O\|^2_{\rho, g} \equiv d^N \operatorname{Tr}[(\mathcal{M}^{-1}[O]^{\otimes 2} \otimes \rho)\chi_g], \quad (16)$$

where χ_g is the representation of the S_3^N permutation g in the threefold Hilbert space. $W^{(3)}_{\mathcal{E}_\sigma, h}$ is the third *entanglement feature* of the ensemble $\mathcal{E}_\sigma$, defined as

$$W^{(3)}_{\mathcal{E}_\sigma, h} \equiv \underset{\hat{\sigma} \in \mathcal{E}_\sigma}{\mathbb{E}} \operatorname{Tr}(\hat{\sigma}^{\otimes 3}\chi_h). \quad (17)$$

Note that the second entanglement feature previously defined in Eq. (8) can be consistently cast into the form of Eq. (17) in terms of permutation operators (see Ref. [27]).

In practice, the expectation value $\langle O \rangle$ is always estimated based on a finite collection of the snapshot states. Let M be the number of samples of ρ measured in the data acquisition stage (each sample results in a snapshot state $\hat{\sigma}_k$). The finite-average estimation $\bar{o} = \frac{1}{M} \sum_{k=1}^M \hat{o}(\hat{\sigma}_k)$ will fluctuate around the true expectation value $\langle O \rangle$ with a variance that scales as $(\operatorname{Var} \hat{o})/M$. By the Chebyshev inequality, the probability for $\bar{o}$ to deviate from $\langle O \rangle$ by more than ϵ amount is bounded by

$$\operatorname{Pr}(|\bar{o} - \langle O \rangle| \geqslant \epsilon) \leqslant \frac{\operatorname{Var} \hat{o}}{\epsilon^2 M} \leqslant \frac{\|O\|^2_{\mathcal{E}_{\sigma|\rho}}}{\epsilon^2 M}. \quad (18)$$

Therefore, to control the failure probability within a threshold δ, i.e., $\operatorname{Pr}(|\bar{o} - \langle O \rangle| \geqslant \epsilon) \leqslant \delta$, sufficient number of samples is required

$$M \geqslant \frac{\|O\|^2_{\mathcal{E}_{\sigma|\rho}}}{\epsilon^2 \delta}. \quad (19)$$

A larger (smaller) shadow norm $\|O\|^2_{\mathcal{E}_{\sigma|\rho}}$ indicates that more (less) samples are needed.

However, the ρ-dependent shadow norm $\|O\|^2_{\mathcal{E}_{\sigma|\rho}}$ is generally complicated to evaluate. If we are not interested in the shadow norm for a specific state ρ, but rather the expectation of the shadow norm over an ensemble of states $\{V\rho V^\dagger\}$ that are similar to ρ by local-basis transformations $V \in \mathrm{U}(d)^N$, we can actually define a ρ-independent shadow norm by averaging over V:

$$\|O\|^2_{\mathcal{E}_\sigma} \equiv \underset{V \in \mathrm{U}(d)^N}{\mathbb{E}} \|O\|^2_{\mathcal{E}_{\sigma|V\rho V^\dagger}}. \quad (20)$$

The expected shadow norm can be expressed purely in terms of the entanglement features of $\mathcal{E}_\sigma$ and $\mathcal{E}_O$ (see Appendix B for derivation),

$$\|O\|^2_{\mathcal{E}_\sigma} = \sum_{A,B,C,D \in 2^{\Omega_N}} v_{A,B,C,D} W^{(2)}_{\mathcal{E}_\sigma, A \cap B \cap C} W^{(2)}_{\mathcal{E}_O, D}, \quad (21)$$

where the coefficient $v_{A,B,C,D}$ is given by

$$v_{A,B,C,D} = r_A r_B \left(\frac{d^2}{d^2 - 1}\right)^N d^{|A \cap B \cap C| - |C|} \left(-\frac{1}{d}\right)^{|C \ominus D|}, \quad (22)$$
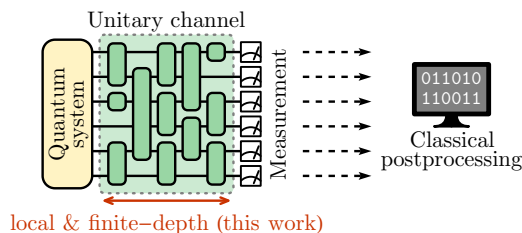
FIG. 1. Illustration of classical shadow tomography protocol. This work focuses on the case when the unitary channel is of finite depth and respects locality.

and $\mathcal{E}_O = \{V^\dagger O V | V \in \mathrm{U}(d)^N\}$ denotes the locally scrambled ensemble [or known as $\mathrm{U}(d)^N$ twirling] associated with the observable O in question.

In conclusion, given a traceless Hermitian operator O, its expected shadow norm $\|O\|_{\mathcal{E}_\sigma}^2$ provides a typical lower bound for the number of samples needed

$$M \gtrsim \frac{\|O\|_{\mathcal{E}_\sigma}^2}{\epsilon^2 \delta}, \qquad (23)$$

in order to control the error of the prediction $\bar{o}$ given by the classical shadow tomography within the probability bound $\Pr(|\bar{o} - \langle O \rangle| \geqslant \epsilon) \leqslant \delta$. Here we have only analyzed the sample complexity for a single linear observable. For the analysis of multiple and/or nonlinear observables, we refer to the original paper of Ref. [8]. Their result applies to our case simply by replacing the shadow norm with our version.

C. A toy example of two-qudit system

To demonstrate our framework and to gain some analytical intuition, we present a toy example to compute the reconstruction map in a two-qudit ($N = 2$) system. We assume that the two-qudit system always evolves under a locally scrambled quantum dynamics, which can be modeled (for example) by a finite-time Brownian evolution[1] driven by random Hamiltonians. Every classical snapshot $\hat{\sigma}_{U,b} = U^\dagger |b\rangle\langle b| U$ is generated by the reversed evolution from the product state $|b\rangle\langle b|$. In the long-time limit [Fig. 2(a)], the entanglement feature $W_{\mathcal{E}_\sigma}^{(2)} = (1, \frac{2d}{d^2+1}, \frac{2d}{d^2+1}, 1)$ follows from that of Page states, where the subregion basis is arranged in the order of $\{\cdot\}, \{1\}, \{2\}, \{1, 2\}$. This is because the evolution of entanglement feature under any locally scrambled quantum dynamics always converges to the Page state, regardless of the initial state, as proven in Ref. [14]. In the short-time limit [Fig. 2(b)], $\hat{\sigma}$ remains as a product state, therefore, the entanglement entropy vanishes for all regions, which translates to $W_{\mathcal{E}_\sigma}^{(2)} = (1, 1, 1, 1)$. In general, for any intermediate time, the entanglement feature should take the form of

$$W_{\mathcal{E}_\sigma}^{(2)} = (1, w, w, 1), \qquad (24)$$

with w varies between $\frac{2d}{d^2+1}$ (the long-time limit) and 1 (the short-time limit). The physical meaning of w is the average single-qudit purity in the snapshot state $\hat{\sigma}$.

Given $W_{\mathcal{E}_\sigma}^{(2)}$ in Eq. (24), Eq. (12) reads as

$$d^2 \begin{bmatrix} 1 & \frac{d(d-w)}{d^2-1} & \frac{d(d-w)}{d^2-1} & \frac{d^2(d^2-2dw+1)}{(d^2-1)^2} \\ 0 & \frac{dw-1}{d^2-1} & 0 & \frac{d(d^2w-2d+w)}{(d^2-1)^2} \\ 0 & 0 & \frac{dw-1}{d^2-1} & \frac{d(d^2w-2d+w)}{(d^2-1)^2} \\ 0 & 0 & 0 & \frac{d^2-2dw+1}{(d^2-1)^2} \end{bmatrix} \begin{bmatrix} r_{\{\}} \\ r_{\{1\}} \\ r_{\{2\}} \\ r_{\{1,2\}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \qquad (25)$$

By solving this linear equation, the reconstruction coefficient r_A can be obtained:

$$r = \begin{bmatrix} r_{\{\cdot\}} \\ r_{\{1\}} \\ r_{\{2\}} \\ r_{\{1,2\}} \end{bmatrix} = \begin{bmatrix} \frac{d^3w - 3d^2 + 3dw - 2w^2 + 1}{(dw-1)(d^2-2dw+1)} \\ \frac{-d^4w + 2d^3 - 2d + w}{d(dw-1)(d^2-2dw+1)} \\ \frac{-d^4w + 2d^3 - 2d + w}{d(dw-1)(d^2-2dw+1)} \\ \frac{(d^2-1)^2}{d^2(d^2-2dw+1)} \end{bmatrix}. \qquad (26)$$

The behavior of r_A as a function of w is shown in Fig. 2(c), which continuously interpolates the two limits.

In the short-time limit, $w = 1$ and Eq. (26) reduces to $r_A = (1, -(d+1)/d, -(d+1)/d, (d+1)^2/d^2)$, corresponding to the reconstruction map

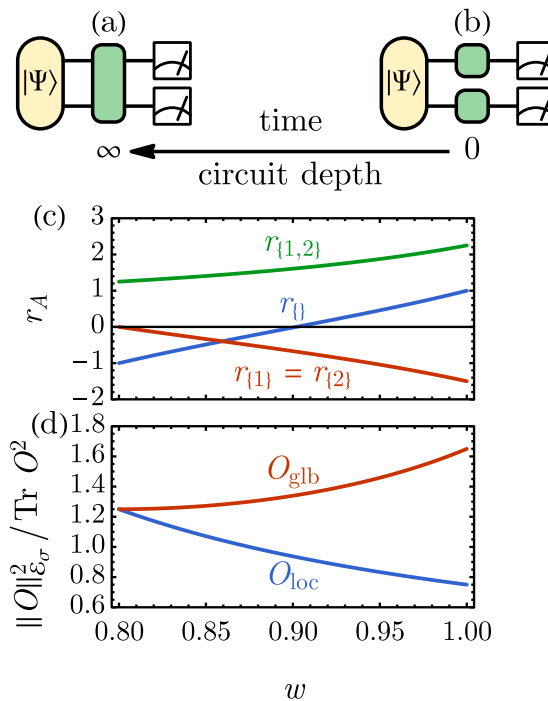$$\mathcal{M}^{-1}[\sigma] = \bigotimes_{i=1,2} [(d+1)\sigma_i - \mathbb{1}_i], \qquad (27)$$

FIG. 2. Two-qudit unitary channel in (a) the long-time (Page state) limit and (b) the short-time (product state) limit. (c) Reconstruction coefficients r_A and (d) the shadow norm $\|O\|_{\mathcal{E}_\sigma}^2 / \mathrm{Tr}\, O^2$ vs the single-qudit purity w, for $d = 2$. w varying from 1 to $\frac{4}{5}$ effectively models the circuit depth (or evolution time) growing from 0 to ∞.

---

[1] The Brownian unitary evolution is a product of a sequence of infinitesimal time evolution $U = \prod_t e^{-iH_t \, \delta t}$, but the Hamiltonian H_t at each time step is independent drawn from a random Hamiltonian ensemble (unlike the coherent quantum dynamics, where the same Hamiltonian drives the dynamics through all time.)

matching the result of onsite two-design circuits [31,32]. In the long-time limit, $w = \frac{2d}{d^2+1}$ and Eq. (26) reduces to $r_A = (-1, 0, 0, (d^2 + 1)/d^2)$, corresponding to the reconstruction map

$$\mathcal{M}^{-1}[\sigma] = (d^2 + 1)\sigma - \mathbb{1}, \tag{28}$$

matching the result of global two-design circuits [31,32]. The general result in Eq. (26) provides the reconstruction map that interpolates these two limits, which allows us to perform classical shadow tomography for intermediate unitary channels that are neither onsite nor global two-design.

To investigate the sample complexity of the tomography scheme in the two-qudit system, we consider a traceless Hermitian operator O (i.e., $\mathrm{Tr}\, O = 0$) and define two parameters k_1 and k_2 to parametrize the purity:

$$k_1 = \mathrm{Tr}_{\{1\}}(\mathrm{Tr}_{\{2\}}\, O)^2 / \mathrm{Tr}\, O^2,$$
$$k_2 = \mathrm{Tr}_{\{2\}}(\mathrm{Tr}_{\{1\}}\, O)^2 / \mathrm{Tr}\, O^2. \tag{29}$$

Then the entanglement feature of the observable O can be arranged as the following vector:

$$W_{\mathcal{E}_O}^{(2)} = (0, k_1, k_2, 1)\, \mathrm{Tr}\, O^2, \tag{30}$$

with the same choice of region basis as in Eq. (24). Given r, $W_{\mathcal{E}_O}^{(2)}$, and $W_{\mathcal{E}_\sigma}^{(2)}$, we have all the information needed to calculate the shadow norm, according to Eq. (21),

$$\|O\|_{\mathcal{E}_\sigma}^2 = \frac{d^2 - 1}{d^3}\left(\frac{k_{\mathrm{tot}}}{dw - 1} + \frac{(d^2 - 1)(d - k_{\mathrm{tot}})}{d^2 - 2dw + 1}\right)\mathrm{Tr}\, O^2, \tag{31}$$

where $k_{\mathrm{tot}} = k_1 + k_2$.

The operator locality crucially affects k_{tot}. Consider modeling a local operator O_{loc} by a random operator drawn from the Gaussian unitary ensemble (GUE) and acting on the first qudit only, we have

$$W_{\mathcal{E}_{O_{\mathrm{loc}}}}^{(2)} = (0, d, 0, 1)\, \mathrm{Tr}\, O_{\mathrm{loc}}^2, \tag{32}$$

hence $k_{\mathrm{tot}} = d$. On the other hand, for a global operator O_{glb} modeled by a global GUE random operator acting on both qudits simultaneously, we have

$$W_{\mathcal{E}_{O_{\mathrm{glb}}}}^{(2)} = \left(0, \frac{d}{d^2 + 1}, \frac{d}{d^2 + 1}, 1\right)\mathrm{Tr}\, O_{\mathrm{glb}}^2, \tag{33}$$

hence $k_{\mathrm{tot}} = \frac{2d}{d^2+1}$. In these two cases, the shadow norm in Eq. (31) becomes

$$\|O_{\mathrm{loc}}\|_{\mathcal{E}_\sigma}^2 = \frac{d^2 - 1}{d^2(dw - 1)}\, \mathrm{Tr}\, O_{\mathrm{loc}}^2,$$
$$\|O_{\mathrm{glb}}\|_{\mathcal{E}_\sigma}^2 = \frac{d^2 - 1}{d^2(d^2 + 1)}\left(\frac{(d^2 - 1)^2}{d^2 - 2dw - 1} + \frac{2}{dw - 1}\right)\mathrm{Tr}\, O_{\mathrm{glb}}^2. \tag{34}$$

Their dependence in w is plotted in Fig. 2(d). In the short-time limit ($w = 1$), $\|O_{\mathrm{loc}}\|_{\mathcal{E}_\sigma}^2 < \|O_{\mathrm{glb}}\|_{\mathcal{E}_\sigma}^2$, meaning that the shallow circuit is more efficient in predicting local observables. In the long-time limit ($w = \frac{2d}{d^2+1}$), $\|O_{\mathrm{loc}}\|_{\mathcal{E}_\sigma}^2 = \|O_{\mathrm{glb}}\|_{\mathcal{E}_\sigma}^2 = (1 + d^{-2})\, \mathrm{Tr}\, O^2$, such that there is no difference in predicting both local and global observables in terms of the sample efficiency, because all operators are equally scrambled in this limit.
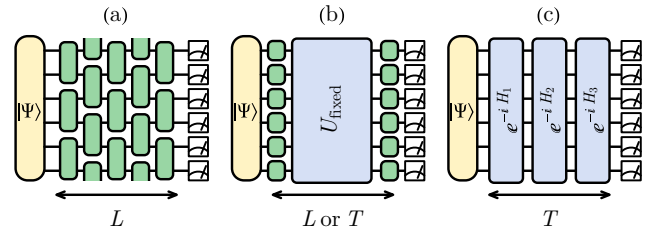

FIG. 3. Classical shadow tomography with (a) finite-depth random unitary and Clifford circuits (of L layers), (b) a fixed unitary twirled by single-qubit random Clifford gates, and (c) discrete-time Hamiltonian dynamics (of T steps).

D. Additional remarks on computational methods and future directions

Efficient numerical methods have been developed [20,38] to calculate the evolution of entanglement feature $W_{\mathcal{E}_\sigma}^{(k)}$ under locally scrambled quantum dynamics by solving the corresponding entanglement dynamics equation (without simulating the quantum dynamics using brute force). However, we will leave this approach for future exploration. In this work, we will compute the entanglement feature beforehand based on the definition (8), by direct sampling from the prior snapshot ensemble $\mathcal{E}_\sigma$. For experimentally generated random unitaries whose distribution is *a priori* unknown, it is also possible to estimate the entanglement feature efficiently from Rényi entropy measurements [39–41] following the definition (8).

As shown in Ref. [38], for one-dimensional quantum systems, the entanglement feature vector $W_{\mathcal{E}_\sigma}^{(k)}$ admits efficient matrix product state (MPS) representation, even if snapshot states in $\mathcal{E}_\sigma$ are volume-law entangled. Combining the MPS representation of $W_{\mathcal{E}_\sigma}^{(2)}$ with the fact that $f_{A,B,C}$ is factorizable to every qudit, one can develop efficient MPS-based numerical approach to find the solution of r_A (also as a MPS). However, we will defer the development of this approach to future work. In the following numerical demonstrations, we will directly solve Eq. (12) for small systems as a proof of concept.

The MPS representations for r, $W_{\mathcal{E}_\sigma}^{(2)}$, and $W_{\mathcal{E}_O}^{(2)}$ also enable us to calculate the shadow norm $\|O\|_{\mathcal{E}_\sigma}^2$ efficiently by a four-way MPS contraction. In Appendix C, we discuss the detail of the efficient classical postprocessing algorithm with the tensor network method. The ability to compute the shadow norm efficiently will be particularly useful if we want to design optimal unitary channels to minimize the sample complexity for a given set of designated observables. It is possible to apply machine-learning approaches (such as deep reinforcement learning) to perform the circuit structure optimization. Therefore, our construction provides the flexibility to allow the classical shadow tomography to adapt to designated observables, which has not been possible before. We will also leave this promising direction to future research.

III. NUMERICAL DEMONSTRATIONS

To demonstrate the effectiveness of our approach, we consider three types of unitary ensembles for the unitary channel in the data acquisition protocol, as illustrated in Fig. 3.

A. Classical shadow tomography with shallow random unitary and Clifford circuits

We first consider using random unitary circuits (RUCs) [15] for the unitary channel. As illustrated in Fig. 3(a), the unitary circuit consists of two-qubit local unitary gates arranged in the brick-wall pattern with a periodic boundary condition. Each gate in the circuit is independently drawn from the Haar random unitary ensemble. The depth L of the circuit can be adjusted. Obviously, RUCs are locally scrambled, as any local-basis transformation (from both left and right) can be absorbed by the Haar random unitary gates in the circuit. Therefore, we expect our reconstruction map to work perfectly in this case for any choice of the circuit depth L.

For illustration purpose, we start with a Greenberger-Horne-Zeilinger (GHZ) state $\rho = |\Psi\rangle\langle\Psi|$, where $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\ldots0\rangle + |11\ldots1\rangle)$. For every given circuit depth L, we first calculate the entanglement feature $W^{(2)}_{\mathcal{E}_\sigma, C}$ to determine the reconstruction map $\mathcal{M}^{-1}$. This calculation is done for once and stored in the classical memory for future reference. In our numerical simulation of the data acquisition process, we sample the RUC, apply it to the GHZ state $|\Psi\rangle$, and perform the computational basis measurement. We generate a collection of classical snapshots $\mathcal{E}_{\sigma|\rho} = \{\hat{\sigma}\}$ of size M by repeated measurements. We then estimate the fidelity F of the reconstructed state by

$$F = \sqrt{\frac{1}{M} \sum_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \langle\Psi|\mathcal{M}^{-1}[\hat{\sigma}]|\Psi\rangle}. \tag{35}$$

Following the philosophy of classical shadow tomography, one should view Eq. (35) as a prediction task. If the shadow tomography is successful, then the estimated fidelity should converge to $F = 1$. This estimation can be achieved accurately by a few measurements, even though the full density matrix estimation $\text{avg}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \mathcal{M}^{-1}[\hat{\sigma}]$ may still have large fluctuations. In addition, when the reconstruction is biased, for example, the experimental channel $\mathcal{M}$ does not match the theoretical assumption of the unitary ensemble, then the fidelity estimation will deviate from one. Figure 4(a) shows that the entanglement-feature-based reconstruction map $\mathcal{M}^{-1}_{\text{EF}}$ indeed gives unbiased estimation of fidelity F for different circuit depths L and for different system sizes N. Furthermore, when the GHZ state is prepared with Z errors, our method can give the correct fidelity estimation that decreases linearly with the probability of Z error, which is challenging for the current state-of-art machine-learning quantum state tomography method [8,42] (see Appendix D for more discussions).

To compare with the existing classical shadow tomography method [8], we consider the reconstruction maps $\mathcal{M}^{-1}_{\text{GH}}[\sigma] = (d^N + 1)\sigma - \mathbb{1}$ and $\mathcal{M}^{-1}_{\text{LH}}[\sigma] = \bigotimes_i ((d+1)\sigma_i - \mathbb{1}_i)$, where $\mathcal{M}^{-1}_{\text{GH}}$ (or $\mathcal{M}^{-1}_{\text{LH}}$) assumes the unitary ensemble is global (or onsite local) Haar random. They can be viewed as special limits where circuit depth L tends to infinity and zero, respectively. Although they can also achieve an unbiased estimation of quantum fidelity, the tomography efficiency differs. In Fig. 4(b), the error bar shows how the (3-times) standard deviation of the estimated fidelity scales with the number of qubits N at 5000 sample size. As we can see [both from the error
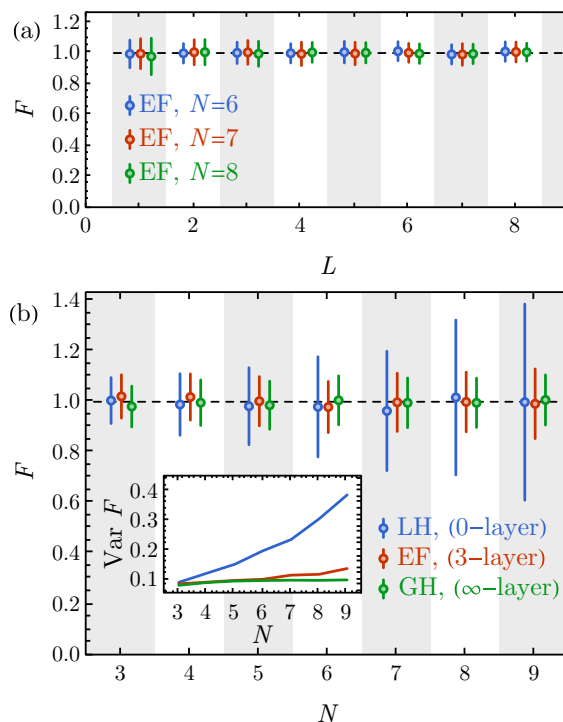

FIG. 4. (a) Fidelity estimation of GHZ state with RUC of different circuit depth L using entanglement-feature-based reconstruction $\mathcal{M}^{-1}_{\text{EF}}$ (denoted by EF) over different number N of qubits. (b) Fidelity estimation of GHZ state using shallow RUC (three layer, with $\mathcal{M}^{-1}_{\text{EF}}$, denoted by EF), random onsite (local Haar) gates (zero layer, with $\mathcal{M}^{-1}_{\text{LH}}$, denoted by LH), and global Haar unitary (∞ layer, with $\mathcal{M}^{-1}_{\text{GH}}$, denoted by GH). The inset shows the variance Var F of the predicted fidelity as a function of system size N. In both subfigures, the sample size is 5000. Error bar indicates 3 standard deviations estimated by the bootstrap method. Points are split horizontally to avoid the overlap of markers.

bar and from the inset of Fig. 4(b)], the variance of (onsite) local Haar estimation increases drastically as N increases, which implies an increasingly high sample complexity for large systems.

In the other limit, the variance of global Haar estimation is independent of system size, achieving the optimal sample complexity as advocated in Ref. [8]. However, to realize the global Haar ensemble, the circuit depth needs to be at least of order $O(N)$, which is quite demanding for quantum devices. If we approximate the global Haar ensemble with finite-depth circuits and use the reconstruction map $\mathcal{M}^{-1}_{\text{GH}}$ on data collected from finite-depth circuit measurements [9], this will yield systematically biased predictions for physical quantities when the circuit is not deep enough. In Fig. 5(a), we show that the biased prediction tends to overestimate the fidelity, leading to the unphysical result of $F > 1$ (the correct behavior is $F = 1$). This occurs because, when the measurement channel $\mathcal{M}$ in data acquisition protocol disagrees with the reconstruction channel $\mathcal{M}^{-1}$ in classical postprocessing protocol, the reconstructed density matrix $\frac{1}{M} \sum_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \mathcal{M}^{-1}[\hat{\sigma}]$ may not be positive definite (see Appendix E for detailed discussions), resulting in the unphysical fidelity estimation. This bias gets worse for larger system size. In Fig. 5(b), we also show the
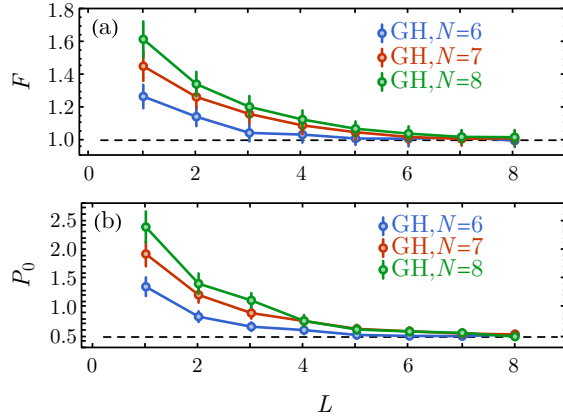
FIG. 5. (a) Fidelity estimation of the reconstructed GHZ state with RUC of finite depth L. (b) Estimation of observable $P_0 = |\langle \Psi |00 \dots 0\rangle|^2$ (the projection operator to the $|00 \dots 0\rangle$ state) on the reconstructed GHZ state with RUC of finite depth L. In both cases, the reconstruction uses the global Haar reconstruction map. The sample size is 5000. Error bar indicates 3 standard deviations estimated by the bootstrap method.

estimation of $P_0 = |\langle \Psi |00 \dots 0\rangle|^2$. For GHZ state, the correct behavior is $P_0 = 0.5$, and we still see significant bias when applying $\mathcal{M}_{\text{GH}}^{-1}$ for shallow circuits.

However, with the entanglement-feature-based reconstruction map $\mathcal{M}_{\text{EF}}^{-1}$, as demonstrated in Fig. 4(b), we are able to achieve an *unbiased* fidelity estimation with a three-layer shallow circuit, approaching *similar variance level* (i.e., similar sample efficiency) as global Haar ensemble while keeping a *low-circuit complexity*. This clearly demonstrates the advantage of our approach.

B. Scaling of variance and tomography complexity

The above discussion motivates us to define the *tomography complexity* as $\mathcal{C} = (L + 1)M$, where L is the *circuit complexity* (the number of layers in the quantum circuit), and M is the *sample complexity* (the number of samples needed). M will be proportional to the single-shot variance $\text{Var}\, \hat{o}$. Suppose applying each layer of quantum gates and performing measurements both take a unit of time on the quantum device, then $\mathcal{C}$ is roughly the total amount of time needed to collect the classical shadow from M copies of the quantum state, which characterizes the complexity of the data acquisition protocol. This notion of complexity is consistent with the quantum algorithmic measurement (QUALM) complexity introduced in Ref. [43] (with LM and M being their gate and query complexities, respectively). In the following, we will investigate the scaling of single-shot variance $\text{Var}\, \hat{o}$ as a function of circuit depth L and system size N for both low-rank operators (such as fidelity) and full-rank operators (such as Pauli operators), and show how the tomography complexity $\mathcal{C}$ can guide us to find the optimal circuit depth L_*.

For low-rank operators, we will focus on quantum fidelity, which is important in many quantum information applications, such as (variational) state preparation. We will define the zero-depth limit ($L \to 0$) of the RUC to be a single layer of onsite Haar-random gates because even if there is no two-
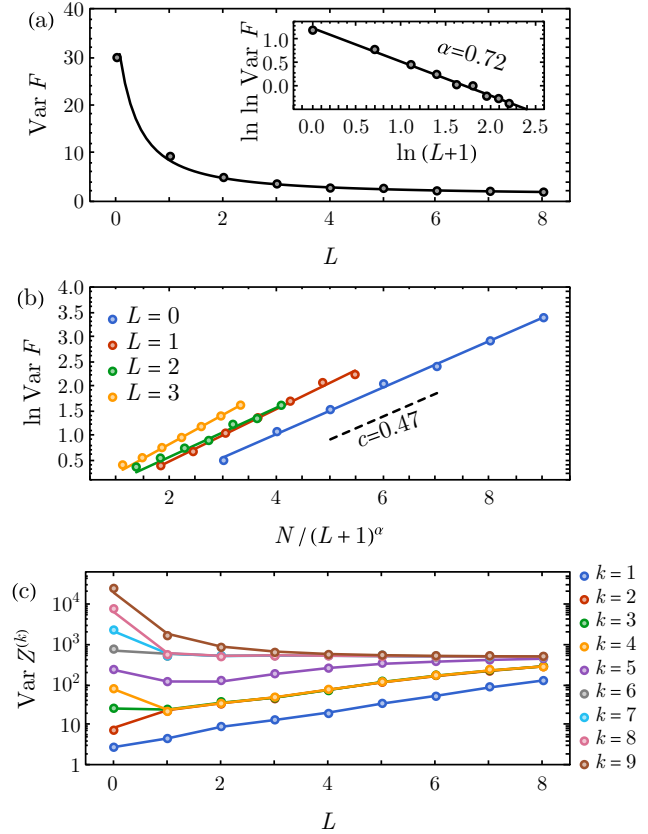

FIG. 6. (a) Single-shot variance of estimated fidelity vs circuit depth L for a nine-qubit GHZ state. (b) Single-shot variance of estimated fidelity as a function of the effective system size N_{eff}. The best fit for $\text{Var}\, F \propto \exp(c \frac{N}{(L+1)^\alpha})$ gives $c = 0.47 \pm 0.08$ and $\alpha = 0.72 \pm 0.1$. (c) Variance of full-rank operator estimation on a nine-qubit GHZ state. The full-rank operators are Pauli-Z operators of the form $Z^{(k)} = Z^{\otimes k} I^{\otimes (N-k)}$ with different support k. The dots are experimental results from simulation, and the lines are theoretical prediction using operator shallow norm by Eq. (21). They match perfectly.

qubit gate in the "zero-depth" circuit, we still assume that the unitary ensemble is locally scrambled such that onsite scrambling unitaries continue to persist. In this limit, the single-shot variance $\text{Var}\, F$ of fidelity estimation scales exponentially with the number of qubits N. On the other hand, in the deep circuit limit ($L \to \infty$), RUCs will approach the global Haar unitary ensemble, and the variance $\text{Var}\, F$ will be independent of system size. We are interested to investigate how $\text{Var}\, F$ behaves in the shallow circuit regime. In Fig. 6(a), we calculated $\text{Var}\, F$ numerically using the bootstrap method for the nine-qubit GHZ state. It shows the variance $\text{Var}\, F$ will decrease quickly in the shallow circuit regime. Interestingly, we found that an empirical formula $\text{Var}\, F \propto \exp(c \frac{N}{(L+1)^\alpha})$ fits the data well in the shallow circuit regime, with $\alpha = 0.7 \pm 0.1$. In Fig. 6(b), we plot $\text{Var}\, F$ as a function of $\frac{N}{(L+1)^\alpha}$ for different fixed circuit depth L. We find curves with different choices of circuit depth L all collapse together with the same coefficient $c = 0.47 \pm 0.08$.

The physical intuition behind the empirical formula has to do with the operator growth in RUCs. If the quantum circuit is

very shallow, then the computational basis measurement will only probe local information in the original basis. If the circuit becomes deeper, computational basis measurement can probe information in larger regions in the original basis because the measurement operator has grown under the (backward) circuit evolution. Suppose the size of the measurement operator grows in a power-law manner $\sim (L+1)^{\alpha}$,[2] with respect to the depth L of the RUC, the relative size of the system will effectively shrink to $N_{\text{eff}} = \frac{N}{(L+1)^{\alpha}}$, such that $\text{Var}\,F$ should scale universally with N_{eff}, as proposed in the empirical formula. We might expect $\alpha = \frac{1}{2}$ (or $\alpha = 1$), if the operator grew diffusively (or ballisticaly). However, the best fit of our numerical result seems to indicate an effective operator growth between the diffusive and ballistic limits. Due to the limited system size in this study, we are unable to determine whether our observation persists to the thermodynamic limit. We will leave this intriguing scaling behavior for further investigation in the future. Nevertheless, for any α, the variance decreases faster than exponential with L in the shallow circuit regime, which already speaks for the advantage of applying shallow circuits in classical shadow tomography.

For full-rank operators, we mainly focus on consecutive strings of Pauli operators of the form

$$Z^{(k)} = Z^{\otimes k} I^{\otimes(N-k)} = \underbrace{ZZ\ldots Z}_{k}\underbrace{II\ldots I}_{N-k}, \qquad (36)$$

where Z is the Pauli-Z operator, and I is the identity operator. We define the locality of the Pauli string operator by its length k. In the shallow circuit limit ($L \to 0$), the variance of estimation for $Z^{(k)}$ scales $\text{Var}\,Z^{(k)} \propto 4^k$. So the shallow circuit is only efficient for predicting the local observables, and becomes inefficient for nonlocal observables. In the deep circuit limit ($L \to \infty$), as the unitary ensemble becomes globally Haar, there is no difference between local and nonlocal operators in this limit, and $\text{Var}\,Z^{(k)} \propto 2^N$. A simple comparison indicates when $k \gtrsim N/2$, $\text{Var}\,Z^{(k)}$ will decrease with L, thus deep circuits will have lower sample complexity; when $k \lesssim N/2$, $\text{Var}\,Z^{(k)}$ will increase with L, thus shallow circuits will have lower sample complexity. In Fig. 6(c), the dots show the variance $\text{Var}\,Z^{(k)}$ as a function of circuit depth L for different support k. The trend agrees with our simple argument. For nonlocal operators, their variance will quickly decrease with the circuit depth L, while the variance for local operators will mildly increase with L. The behavior is theoretically described by how the operator shadow norm $\|O\|_{\mathcal{E}_\sigma}^2$ depends on both the circuit depth L and the operator locality k, which are separately encoded in the entanglement features of $\mathcal{E}_\sigma$ and $\mathcal{E}_O$. We calculate the shadow norm $\|Z^{(k)}\|_{\mathcal{E}_\sigma}^2$ based on the entanglement feature formalism using Eq. (21), and plot the result as lines in Fig. 6(c). The theoretical calculation agrees perfectly with the numerical results, which also indicates that the shadow norm bounds the single-shot variance (and hence the sample complexity) quite tightly.

Based on the discussion in Sec. II B, the sample complexity M is proportional to the single-shot variance. Given the

---

[2]In the $L \to 0$ limit, the measurement operator is still of at least size 1, which motivates the "+1" regularization in $(L+1)^{\alpha}$.
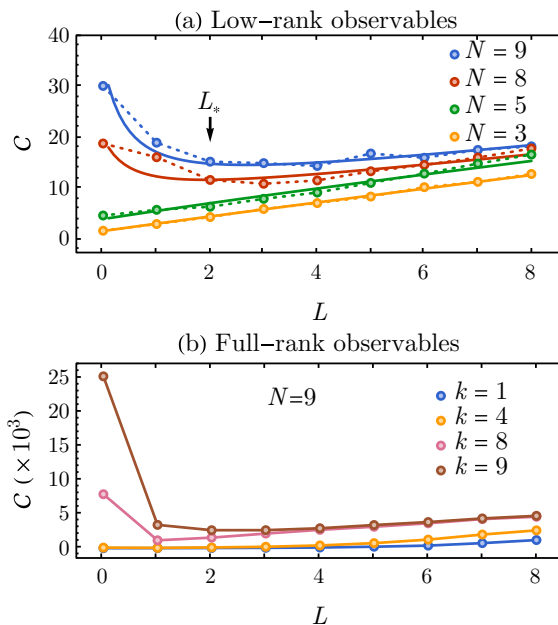

FIG. 7. (a) Tomography complexity $\mathcal{C} \propto (L+1)\,\text{Var}\,F$ as a function of circuit depth L for the fidelity (low-rank observable) estimation task. Dots are tomography complexities for GHZ states of qubit number N by our numerical simulation. Solid curves are best fits based on the empirical formula (37). (b) Tomography complexity $\mathcal{C} \propto (L+1)\,\text{Var}\,Z^{(k)}$ for the Pauli string (full-rank observable) estimation task. Dots are numerical simulation results. Solid curves are analytic calculations using the operator shadow norm formula (21).

scaling of variance with the circuit depth L, we can study the scaling of the sample complexity, as well as that of the tomography complexity $\mathcal{C}$. For fidelity estimation task, $\mathcal{C}$ scales as

$$\mathcal{C} = (L+1)M \propto (L+1)\,\text{Var}\,F$$

$$\propto (L+1)\exp\left(\frac{cN}{(L+1)^{\alpha}}\right). \qquad (37)$$

For sufficiently large systems, the complexity $\mathcal{C}$ can have a nontrivial minimum at a finite circuit depth $L_* \simeq (\alpha c N)^{1/\alpha} - 1$. Our simulation result in Fig. 7(a) verifies such behavior. For small systems ($N \lesssim 5$), random single-qubit measurements can efficiently benchmark the quantum state, so we do not need to use a finite-depth circuit for data acquisition. However, as the system size N gets larger, to maintain the prediction accuracy, single-qubit measurements will require more and more samples that have to grow exponentially with N. As shown in Fig. 6(a), applying a few layers of quantum circuits before the measurement can quickly bring down the single-shot variance (and hence reduce the sample complexity). However, we also do not want to go too far in the circuit depth because that would increase the circuit complexity. Therefore, we expect an optimal circuit depth L_* where the sample complexity and the circuit complexity reach a balance, and the total tomography complexity is minimized. This explains the advantage of shallow circuits in classical shadow tomography, as compared to the existing method that requires either onsite Haar random ($L \to 0$) or global Haar random ($L \to \infty$) unitaries.

We also study the tomography complexity $\mathcal{C} \propto (L + 1) \operatorname{Var} Z^{(k)}$ for the full-rank observables, such as Pauli strings $Z^{(k)}$, as shown in Fig. 6(b). In this case, what matters is the locality k of the full-rank operator (the length k of the Pauli string). For local operators (small k), onsite measurement will be most efficient. However, for nonlocal operators (large k), we observe that the tomography complexity is minimized at some finite circuit depth, again demonstrating the advantage of employing shallow circuits in classical shadow tomography. For different classes of physical observables, we can use the tomography complexity $\mathcal{C}$ as an objective function to guide the design of the optimal circuit structure. We will leave this promising direction for future investigation.

C. Classical shadow tomography with fixed quantum circuits or Hamiltonian dynamics

Compared to other classical shadow tomography protocols, our method can be applied to a large family of unitary ensembles that only requires the local scrambling condition, which is more appealing to near-term quantum devices. One of the biggest challenges in realizing the original proposal of global Clifford classical shadow tomography is that the realization of global Clifford unitary requires $\sim N^2$ many local Clifford gates (for a N-qubit system), which remains challenging for near-term quantum devices. Even though global Clifford shadow tomography is very efficient in predicting nonlocal properties, it has not been implemented even for few-qubit systems as far as we know.

As we have seen in Sec. III A, the *quantum entanglement* created by the unitary channel plays an important role in reducing the sample complexity. With the quantum entanglement generated by the unitary channel, the classical shadow tomography is essentially an *entanglement-assisted* nonlocal measurement protocol. To circumvent the difficulty of sampling (fully scrambled) deep random unitaries but still harness the power of entanglement, we can use the idea of locally scrambled unitaries to design randomized measurement protocols that have sandwich structures like Fig. 3(b), where random single-qubit Clifford gates (green boxes) are introduced at the beginning and the end of the unitary channel, and a fixed unitary circuit/quantum dynamics (the blue box) is sandwiched in-between to provide entanglement generation. This sandwiched protocol satisfies the local scrambling condition rigorously, therefore, the reconstruction map in Eq. (12) can be applied.

We will give two examples to demonstrate this sandwiched protocol. In the first example, as shown in Fig. 8(a), the fixed unitary is taken to be a fixed Clifford circuit consist of a sequence of controlled-NOT (CNOT) gates that generates entanglement. In the second example, as illustrated in Fig. 8(b), the fixed unitary is generated by the time evolution of a Rydberg atom Hamiltonian[3]:

$$H = \frac{\Omega}{2} \sum_i X_i - \Delta \sum_i Z_i + \Omega \sum_{i<j} \left(\frac{R_b}{a|i-j|} \right)^6 Z_i Z_j. \quad (38)$$

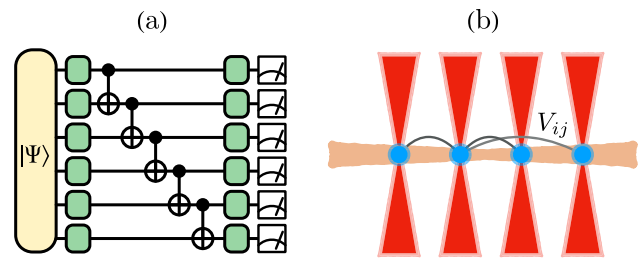[3]For simulation of Rydberg Hamiltonian, we choose parameters $\Omega = 2.75$, $\Delta = 1$, $R_b = 1$.

FIG. 8. Classical shadow tomography with a fixed unitary and onsite random Clifford gates. The fixed unitary can be generated with single quantum circuit, such as (a) CNOT gates, or fixed quantum dynamics, such as (b) Rydberg Hamiltonian dynamics.

Both cases are ready to be implemented with near-term quantum devices, such as trapped-ion-based quantum simulator or Rydberg-atom-based quantum simulator, given the fact that single-qubit Clifford gates can be efficiently implemented, and randomized Pauli measurements have been demonstrated [44].

For comparison, we use both our proposed sandwiched protocol and the standard randomized Pauli measurement to perform the classical shadow tomography on a GHZ state and to evaluate the fidelity of the reconstructed state. The results are shown in Fig. 9. As we can see, the variance of the fidelity estimation based on randomized Pauli measurements grows exponentially with increasing system size. As expected, the variance (or the sample complexity) reduces dramatically if one adds a fixed unitary generated by the CNOT circuit or the Rydberg Hamiltonian dynamics. More specifically, the variance of prediction is reduced by one order of magnitude even for a small system of $N = 9$ qubits. In both cases, the quantum entanglement generated by the locally scrambled quantum dynamics helps to improve the tomography efficiency. This
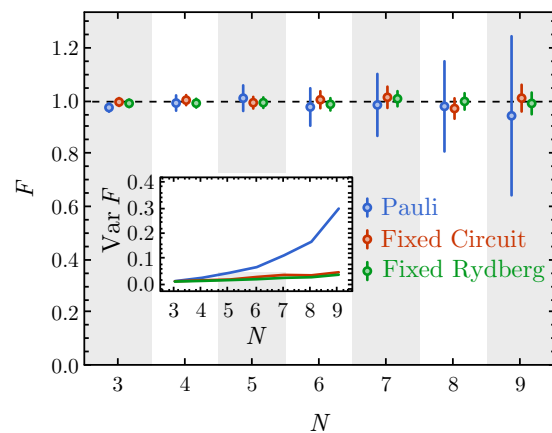

FIG. 9. Fidelity estimation of GHZ state using randomized Pauli measurements (denoted as Pauli), classical shadow tomography with fixed CNOT gates as in Fig. 8(a), and classical shadow tomography with fixed Rydberg Hamiltonian dynamics as in Fig. 8(b). The inset shows the variance $\operatorname{Var} F$ of the predicted fidelity as a function of system size N. The sample size is 10 000. Error bar indicates 3 standard deviations estimated by the bootstrap method. Points are split horizontally to avoid the overlap of markers.

result demonstrates the power of our protocol: it is both *very flexible* in terms of the design and *very efficient* in terms of the sample complexity.

D. Approximate classical shadow tomography with local Hamiltonian dynamics

Requiring an unitary ensemble to be strictly locally scrambled could be restrictive. To this end, we would like to explore a broader class of unitary ensembles that are only approximately locally scrambled. In particular, we study unitary evolutions $U = e^{-iHT}$ generated by a local Hamiltonian H for finite amount of time T, as depicted in Fig. 3(b). Two classes of Hamiltonians are of particular interest. In the first class, we consider a model of random local Hamiltonians

$$H = \sum_i H_{i,i+1}, \tag{39}$$

where each term $H_{i,i+1}$ is independently sampled as 2-local GUE random matrices. We dub this class the *GUE2 ensemble* to remind ourselves that the Hamiltonian is 2-local. The local Hamiltonian describes a disordered one-dimensional quantum system in general. Once every $H_{i,i+1}$ term is sampled, we will use the Hamiltonian H to drive the time evolution without changing H during the evolution. The unitary GUE2 ensemble is only invariant under $U \to V^\dagger U V$ (not $U \to U V$) for $V \in U(d)^N$, such that its corresponding prior snapshot ensemble $\mathcal{E}_\sigma$ will transform as $\hat{\sigma}_{U,b} = U^\dagger |b\rangle\langle b| U \to V^\dagger U^\dagger V |b\rangle\langle b| V^\dagger U V \neq V^\dagger \hat{\sigma}_{U,b} V$, which does not satisfy the locally scrambling condition at the state level (i.e., the invariance under $\hat{\sigma} \to V^\dagger \hat{\sigma} V$). However, we anticipate that under a sufficient amount of time evolution, the original local-basis choice (of $|b\rangle$) will be quickly randomized given the chaotic nature of the local Hamiltonian, such that the initial choice of $V|b\rangle\langle b|V^\dagger$ or $|b\rangle\langle b|$ will not make a substantial difference statistically, so the GUE2 ensemble will become approximately locally scrambled after some local thermalization (scrambling) time T_{Th}.

Another more realistic class of random Hamiltonians to be considered is based on the quantum Ising model with both disordered coupling in space and random fields in time:

$$H_t = \sum_{\langle ij \rangle} J_{ij} X_i X_j + h \sum_i (\cos\theta_t X_i + \sin\theta_t Y_i), \tag{40}$$

where the local coupling $J_{ij} \sim \text{Uni}[J - \frac{J}{2}, J + \frac{J}{2}]$ is drawn from a uniform distribution, and the angle of magnetic field $\theta_t \sim \text{Uni}[0, 2\pi]$ is also random. We use this Hamiltonian to drive the quantum dynamics in discrete time steps. In each period of time, the magnetic field h will be applied along a different random direction θ_t in the $x-y$ plane for all spins uniformly. However, J_{ij} will remain the same throughout the time evolution. The ensemble of unitary consists of

$$U = \prod_{t=1}^T e^{-iH_t}. \tag{41}$$

which we name as the *disordered quantum Ising model* or DQIM for short. The DQIM ensemble is friendly for quantum technology such as Rydberg-atom-based [28] or trapped-ion-based [29] quantum simulators. Similar construction of

approximate unitary designs by Hamiltonian evolution with random quenches in time was also proposed in Refs. [39,40]. We would like to investigate how well our framework applies to these two cases.

Each approximately locally scrambled unitary ensembles $\mathcal{E}_U$ leads to a prior snapshot ensemble $\mathcal{E}_\sigma = \{\hat{\sigma}_{U,b}|b \in \{0, 1\}^{\times N}, U \in \mathcal{E}_U\}$ that is also approximately locally scrambled. We propose to characterize how close the prior snapshot ensemble $\mathcal{E}_\sigma$ is towards its local-basis invariant limit by the following frame potential:

$$\mathcal{F}_{\mathcal{E}_\sigma}^{(k)} = \underset{\hat{\sigma}, \hat{\sigma}' \in \mathcal{E}_\sigma}{\mathbb{E}} (\text{Tr}\, \hat{\sigma} \hat{\sigma}')^k. \tag{42}$$

Recall that in deriving Eq. (6) from (3), we only require the second moment to match, i.e.,

$$\underset{\hat{\sigma} \in \mathcal{E}_\sigma}{\mathbb{E}} \hat{\sigma}^{\otimes 2} = \underset{V \in U(d)^N}{\mathbb{E}} \underset{\hat{\sigma} \in \mathcal{E}_\sigma}{\mathbb{E}} (V^\dagger \hat{\sigma} V)^{\otimes 2}, \tag{43}$$

therefore, we will be most interested in the second frame potential $\mathcal{F}_{\mathcal{E}_\sigma}^{(2)}$. The frame potential $\mathcal{F}_{\mathcal{E}_\sigma}^{(2)}$ for any ensemble $\mathcal{E}_\sigma$ is lower bounded by its locally scrambled [$U(d)^N$-twirled] limit $\mathcal{F}_{\mathcal{E}_\sigma^{\text{LS}}}^{(2)}$ as

$$\mathcal{F}_{\mathcal{E}_\sigma}^{(2)} \geqslant \mathcal{F}_{\mathcal{E}_\sigma^{\text{LS}}}^{(2)} = \sum_{A,B} W_{\mathcal{E}_\sigma, A}^{(2)} \text{Wg}_{A,B} W_{\mathcal{E}_\sigma, B}^{(2)}. \tag{44}$$

The fact that $\mathcal{F}_{\mathcal{E}_\sigma^{\text{LS}}}^{(2)}$ is expressed purely in terms of the entanglement feature of $\mathcal{E}_\sigma$ indicates that it is indeed free of any local-basis-dependent information. We can define the gap between the frame potential and its locally scrambled limit as

$$\begin{aligned} \Delta_{\mathcal{E}_\sigma}^{(2)} &= \mathcal{F}_{\mathcal{E}_\sigma}^{(2)} - \mathcal{F}_{\mathcal{E}_\sigma^{\text{LS}}}^{(2)} \\ &= \text{Tr} \left(\underset{\hat{\sigma} \in \mathcal{E}_\sigma}{\mathbb{E}} \left[\hat{\sigma}^{\otimes 2} - \underset{V \in U(d)^N}{\mathbb{E}} (V^\dagger \hat{\sigma} V)^{\otimes 2} \right] \right)^2, \end{aligned} \tag{45}$$

which turns out to match the trace-square-difference between the second moment $\mathbb{E}_{\hat{\sigma}} \hat{\sigma}^{\otimes 2}$ and its local twirling $\mathbb{E}_{V,\hat{\sigma}}(V^\dagger \hat{\sigma} V)^{\otimes 2}$. The frame potential gap $\Delta_{\mathcal{E}_\sigma}^{(2)}$ serves as an indicator of the validity of our approach, as it vanishes if $\mathcal{E}_\sigma$ is locally scrambled such that our construction becomes exact.

Different unitary ensembles can lead to different frame potential gaps of $\mathcal{E}_\sigma$, which can be used to evaluate the quality of the unitary ensemble in obeying the local scrambling condition. In Fig. 10(a), we first focus on the frame potential gap $\Delta^{(2)}$ for the GUE2 ensemble. We find the gap will first decay exponentially and then saturate to a plateau at a very low level. The quickly vanishing gap implies that the GUE2 ensemble quickly becomes approximately locally scrambled as time evolves. We define the characteristic time associated with the exponential decay as T_{Th}, i.e., $\Delta^{(2)}(T) \propto \exp(-T/T_{\text{Th}})$, which can be considered as the local scrambling (thermalization) time. Such an exponential decaying behavior in the early time regime is generally expected for noncritical quantum dynamics, which admit typical local energy scales (or timescales). In addition, the inset plot in Fig. 10(a) shows that T_{Th} is independent of the system size N, as the slope remains the same for different N within error bar. Unlike global scrambling (global thermalization) which requires a long time ($\sim N$) to achieve, achieving local scrambling only requires a fixed amount of time set by the ultraviolet energy scale that is independent of
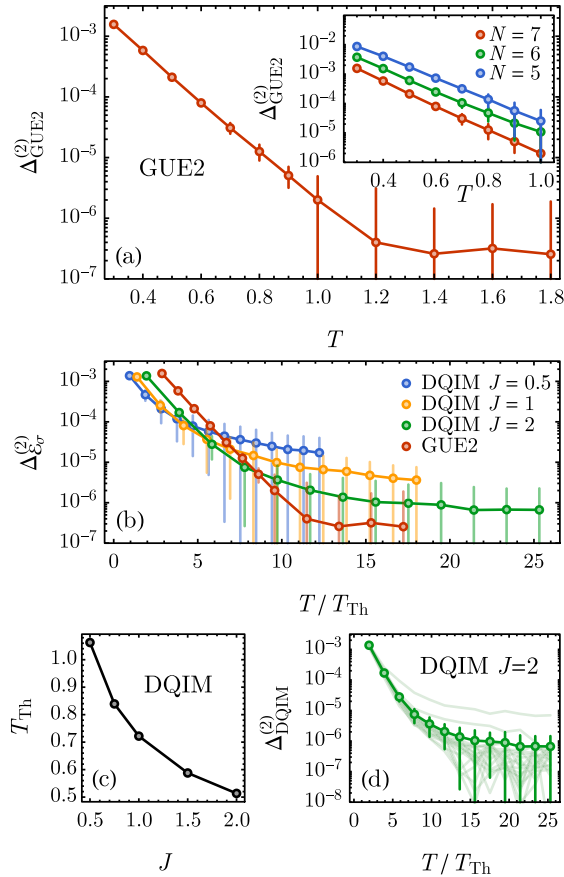
FIG. 10. (a) Frame potential gap $\Delta_{\mathcal{E}_\sigma}^{(2)}$ of the GUE2 ensemble as a function of evolution time T. The inset shows the decay behavior for different system sizes N. (b) Frame potential gap of DQIM ensemble at different coupling strength J, in comparison with that of the GUE2 ensemble. (c) The dependence of the local scrambling time T_{Th} on the coupling strength J. (d) Frame potential gap for single instances in the DQIM ensemble. Each instance corresponds to a light-green curve in the background.

the system size N. This is another advantage of using locally scrambled quantum dynamics for classical shadow tomography in practice.

As for the DQIM ensemble, we fix the strength of the magnetic field at $h = \pi/4$ since this value produces the fastest onsite scrambling of a single qubit. According to the definition (40), the only tuning parameter will be the mean value J of Ising couplings (which also sets their disorder strength). We calculate the frame potential gap $\Delta^{(2)}$ for DQIM ensemble with different J. We observe that the frame potential gap always decays exponentially in the early time regime, in correspondence to the local thermalization process. Then it will typically crossover to a plateau (i.e., saturate to a finite constant) in the late time. The early-time exponential decay region is larger for larger J, and we use the exponential decay regime to define the local scrambling time T_{Th}. The result is shown in Fig. 10(b). The DQIM ensemble also approaches local scrambling as time evolves, although the final saturation plateau is not as low as the GUE2 ensemble. Larger Ising coupling J will lower the saturation plateau and shorter the local scrambling time T_{Th}, as shown in Fig. 10(c). In
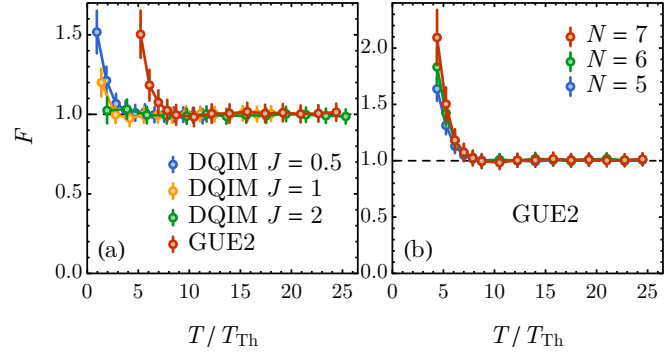

FIG. 11. Fidelity prediction by (a) different approximated locally scrambled ensembles, and (b) the GUE ensemble at different system sizes N. Sample number is 10 000 and error bar indicates 3 standard deviations.

addition, as shown in Fig. 10(d), we find the frame potential gap for a single realization quenched-disorder Hamiltonian does not deviate significantly from the ensemble mean value. This indicates that a single fixed disordered Ising chain under a randomly rotating uniform magnetic field is already good to generate an approximately locally scrambled ensemble that can be used for classical shadow tomography.

In practice, we use the two proposed approximated ensembles, (i) the GUE2 ensemble and (ii) a single instance of the DQIM ensemble, to perform the tomography task and predict the fidelity of a seven-qubit GHZ state. In Fig. 11(a), we see the predicted fidelity will be biased in the beginning (the biased fidelity can be greater than one, see Appendix E for more discussions), due to the fact that the quantum dynamics is still on its way to establish local scrambling. After around $T \sim 10T_{\mathrm{Th}}$, the local scrambling condition is approximately established, then the entanglement-feature-based reconstruction map $\mathcal{M}_{\mathrm{EF}}^{-1}$ can provide a good reconstruction of the quantum state, as indicated by the convergence of the quantum fidelity to identity. In Appendix E, we further investigate the quantum fidelity of $\tilde{\rho}$ projected to the physical space (to tame the unphysical $F > 1$ behavior) and show that the reconstruction is nearly perfect after around $T \sim 10T_{\mathrm{Th}}$. In addition, Fig. 11(b) also shows the local scrambling time for GUE2 is independent of system size, which is consistent with the same behavior in Fig. 10(a). The results in Fig. 11 suggest that the entanglement-feature-based approach could be applicable for approximately locally scrambled unitary ensembles. The reconstruction bias vanishes as the frame potential gap decays. As long as the frame potential gap is low enough, the bias is also expected to be vanishingly small for all predictions. This significantly broadens the application of classical shadow tomography to a large class of quantum dynamics that can be achieved on NISQ devices.

IV. SUMMARY AND DISCUSSIONS

Our result can be further extended to more general measurement channels, which can involve ancilla qubits and partial measurements. The unitary channel can be noisy and the measurements can be weak. Under generalized measurements, the state ρ collapses to $\rho \to K_a \rho K_a^\dagger/(\mathrm{Tr}\, K_a \rho K_a^\dagger)$,

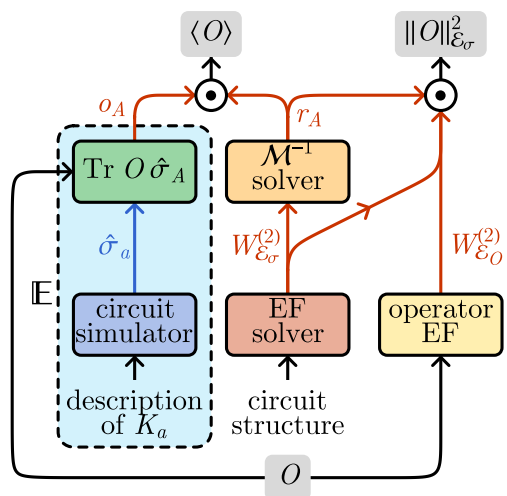FIG. 12. Classical postprocessing protocol to estimate the operator expectation value and shadow norm.

FIG. 13. Illustration of holographic classical shadow tomography scheme, where the quantum circuit is arranged in a hierarchical structure (forming the hyperbolic bulk space).

where K_a is the Kraus operator [45] associated with the measurement outcome a. We can define the measurement operator $\hat{\sigma}_a = K_a^\dagger K_a$ (with the standard normalization $\sum_a \hat{\sigma}_a = \mathbb{1}$), which forms the prior snapshot ensemble $\mathcal{E}_\sigma = \{\hat{\sigma}_a | P(\hat{\sigma}_a) = d^{-N}\}$, and the posterior snapshot ensemble will be $\mathcal{E}_{\sigma|\rho} = \{\hat{\sigma}_a | P(\hat{\sigma}_a|\rho) = \text{Tr}\,\hat{\sigma}_a\rho\}$ correspondingly. As long as the generalized prior snapshot ensemble $\mathcal{E}_\sigma$ is locally scrambled, i.e., $\forall V \in \text{U}(d)^N : P(\hat{\sigma}) = P(V^\dagger \hat{\sigma} V)$, our theoretical framework automatically applies, and all formulations in this work remain valid in the same form. This enables us to consider classical shadow tomography with very general data acquisition protocols.

The entanglement feature formalism plays a central role in our approach. Figure 12 summarizes the proposed classical postprocessing protocol to predict the expectation value $\langle O \rangle$ of a physical observable O, together with its estimated variance (given by the shadow norm $\|O\|^2_{\mathcal{E}_\sigma}$ divided by the sample size M). Given the circuit structure, the entanglement feature (EF) solver calculates the entanglement feature $W^{(2)}_{\mathcal{E}_\sigma}$ of the prior snapshot ensemble as defined in Eq. (8) (the algorithm is developed in previous works [14,20,38]). The result is passed to the inverse channel solver to calculate the reconstruction coefficients r_A by solving Eq. (12). With r_A, we can predict any physical observable O by $\langle O \rangle = d^N \sum_A r_A o_A$ where $o_A = \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \text{Tr}\, O\hat{\sigma}_A$ (the median-of-means trick [8] can be used here if multiple observables are to be predicted). For every sample of classical description of the Kraus operator K, a quantum circuit simulator (running on a classical computer) is needed to construct the (efficient representation of) measurement operator $\hat{\sigma} = K^\dagger K$. The classical simulation could be efficient if the circuit is Clifford [46] (our formalism applies to random Clifford circuits with no problem). The part of computation in the dashed box of Fig. 12 should be repeated for every sample to evaluate the ensemble average. Finally, given the reconstruction coefficient r and the entanglement features $W^{(2)}_{\mathcal{E}_\sigma}$ and $W^{(2)}_{\mathcal{E}_O}$, the shadow norm $\|O\|_{\mathcal{E}_\sigma}$ can be calculated, which provides an estimation for variance of the predicted observable. Although it takes some effort to process the entanglement feature data and to calculate the
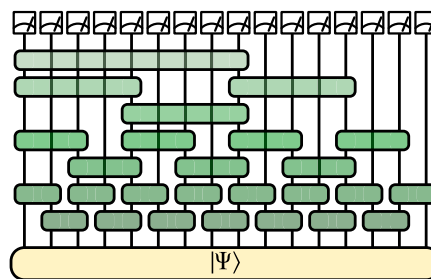
reconstruction coefficients, such computation (everything outside the dashed box in Fig. 12) only occurs once for a given circuit structure, therefore this computational effort is usually affordable (especially when efficient tensor-network approaches are developed and employed) [47].

The theoretical framework established in this work extends the classical shadow tomography to general quantum circuits, which opens up many possible applications. As one interesting example, we consider performing the classical shadow tomography in the "holographic bulk" by transforming the original state by a random Clifford circuit arranged in a hierarchical structure (see Fig. 13), similar to the multiscale entanglement renormalization ansatz (MERA) network [48,49] or the holographic quantum error-correcting code [50]. Following the idea of holographic duality, local measurements in the holographic bulk translate to measurements at all different scales on the holographic boundary. Therefore, it is conceivable that the holographic classical shadow tomography could achieve high sample efficiency for operators of all scales, potentially evading the dichotomy between sample complexity and circuit complexity.

Another interesting application is to consider random circuits hybrid with random measurements inserted into the circuit at a fixed rate [51–55]. Conditioned on the intermediate measurement outcomes, the hybrid quantum circuit forms a quantum channel that transmits quantum information from end to end. Driven by the measurement rate, the final state can undergo an entanglement transition [19,56,57] (or the quantum channel can undergo a purification transition [58] equivalently). When the measurement rate is high, the quantum information in the initial state can be efficiently extracted by intermediate measurements (eavesdroppers), such that the channel has zero transmission capacity. When the measurement rate is lower than a critical threshold, the channel will have a finite capacity and can transmit quantum information in an error-correcting manner [18,20,59]. However, it is unclear how to take advantage of the self-organized quantum error correction in these hybrid quantum circuits. We anticipate that the classical shadow tomography with a flexible measurement scheme can help decoding the measurement-induced quantum error-correcting code. We will leave these interesting applications for future explorations.

Finally, the classical shadow tomography provides an efficient interface that converts quantum states to classical shadow data, which enables us to exploit the power of

classical computation, especially data-driven and machine-learning approaches, to advance our understanding of complex quantum systems and to solve challenging quantum many-body problems. As shown in Ref. [60], classical algorithms that learn from the classical shadow data have provable performance advantages over conventional numerical approaches that do not learn from data. Our work further adds to this promising direction by providing a more flexible classical shadow tomography scheme that works with very general measurement protocols (beyond onsite Pauli measurements), which could lead to potentially more efficient classical-shadow-based learning algorithms.

APPENDIX A: ENTANGLEMENT FEATURE AND THE RECONSTRUCTION CHANNEL

In this Appendix, we discuss the details about the derivation on measurement channel $\sigma = \mathcal{M}[\rho]$ and reconstruction channel $\rho = \mathcal{M}^{-1}[\sigma]$. With the notation developed in Sec. II A, the expected classical snapshot σ is expressed as

$$\sigma = \mathcal{M}[\rho] = \mathop{\mathbb{E}}_{\hat{\sigma} \in \mathcal{E}_\sigma} \hat{\sigma} \, \text{Tr}(\hat{\sigma}\rho)d^N. \tag{A1}$$

By utilizing the assumption that prior ensemble $\mathcal{E}_\sigma$ is locally scrambled, i.e., $P(\hat{\sigma}) = P(V^\dagger \hat{\sigma} V), \forall V \in \text{U}(d)^N$, we are free to insert local-basis transformations V, and average it out. By doing so, we have

$$\sigma = \mathbb{E}_{V \in \text{U}(d)^N} \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_\sigma} V^\dagger \hat{\sigma} V \, \text{Tr}(V^\dagger \hat{\sigma} V \rho) d^N$$

$$= \sum_{B,C \in 2^{\Omega_N}} d^N \text{Wg}_{B,C} \left(\cdots \right), \tag{A2}$$

where each B_i and C_i have two choices: swap operator (1) or identity operator (0), and $\text{Wg}_{B,C} = (d^2 - 1)^{-N}(-1/d)^{|B \ominus C|}$ is the Weingarten function of regions B and C, where $B \ominus C = (B \setminus C) \cup (C \setminus B)$ denotes their symmetric difference. In the above tensor diagram, short parallel lines indicate the periodic boundary condition, and the summation of B and C is over all possible subregions of the N qudit system. As we can see, if we choose a subregion B to be the swap operators, then ρ will be traced out on the counter part $\bar{B}$. In addition, the identity operators (red lines) on $\bar{B}$ are inserted. So the first tensor diagram in Eq. (A2) is the reduced density matrix embedded back into the total Hilbert space. We spoil the notation and use $\rho_B d^{\bar{B}} = (\text{Tr}_{\bar{B}} \rho) \otimes \mathbb{1}_{\bar{B}}$ to denote the first tensor diagram, but one should remember the identity operators are supported in region $\bar{B}$. The tensor product $\otimes$ notation indicates that $(\text{Tr}_{\bar{B}} \rho)$

and $\mathbb{1}_{\bar{B}}$ act separately in regions B and $\bar{B}$, which does *not* imply that B should be a consecutive region "in front of" $\bar{B}$ (as in the conventional notation). The second tensor diagram in Eq. (A2) is the second *entanglement feature* of the prior positive-operator-valued mechanism (POVM) $\mathcal{E}_\sigma$,

$$W^{(2)}_{\mathcal{E}_\sigma, C} \equiv \mathop{\mathbb{E}}_{\hat{\sigma} \in \mathcal{E}_\sigma} \text{Tr}_C(\text{Tr}_{\bar{C}} \hat{\sigma})^2 = \mathop{\mathbb{E}}_{\hat{\sigma} \in \mathcal{E}_\sigma} e^{-S_C^{(2)}(\hat{\sigma})}, \tag{A3}$$

where $S_C^{(2)}(\hat{\sigma})$ denotes the second Rényi entanglement entropy of the state $\hat{\sigma}$ in region C. The above tensor diagram representation is equivalent to Eq. (7) in the main text.

APPENDIX B: VARIANCE ESTIMATION AND SAMPLE COMPLEXITY

In the main text, we relate the sample complexity M with the ρ-dependent shadow norm $\|O\|^2_{\mathcal{E}_{\sigma|\rho}}$, by

$$M \geqslant \|O\|^2_{\mathcal{E}_{\sigma|\rho}} / \epsilon^2 \delta. \tag{B1}$$

However, the ρ-dependent shadow norm $\|O\|^2_{\mathcal{E}_{\sigma|\rho}}$ is generally complicated to evaluate. If we are not interested in the shadow norm for a specific state ρ, but rather the expectation of the shadow norm over an ensemble of states $\{V\rho V^\dagger\}$ that are similar to ρ by local-basis transformations $V \in \text{U}(d)^N$, we can actually define a ρ-independent shadow norm by averaging over V. The result is similar to Eq. (15),

$$\|O\|^2_{\mathcal{E}_\sigma} \equiv \mathop{\mathbb{E}}_{V \in \text{U}(d)^N} \|O\|^2_{\mathcal{E}_{\sigma|V\rho V^\dagger}}$$
$$= \sum_{g,h \in S_3^N} \|O\|^2_g \text{Wg}_{g,h} W^{(3)}_{\mathcal{E}_\sigma, h}, \tag{B2}$$

where $\|O\|^2_g$ is inherited from Eq. (16),

$$\|O\|^2_g \equiv \mathop{\mathbb{E}}_{V \in \text{U}(d)^N} \|O\|^2_{V\rho V^\dagger, g}$$
$$= \text{Tr}((\mathcal{M}^{-1}[O]^{\otimes 2} \otimes \mathbb{1})\chi_g). \tag{B3}$$

Compared with Eq. (16), we can see that the ensemble average $\mathbb{E}_{V \in \text{U}(d)^N}$ in Eq. (B3) removes the ρ dependence by effectively replacing ρ with $d^{-N}\mathbb{1}$ (the prior density matrix that defines the prior POVM $\mathcal{E}_\sigma$). This explains the consistency in our notation that $\|O\|^2_{\mathcal{E}_\sigma} = \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_\sigma} \hat{\sigma}(\hat{\sigma})^2$ follows from essentially the same definition as in Eq. (14).

Note that the reconstruction map $\mathcal{M}^{-1}$ always commutes with the local-basis transformation $V = \prod_i V_i$, i.e., $\mathcal{M}^{-1}[V^\dagger O V] = V^\dagger \mathcal{M}^{-1}[O]V$, because V_i acts on each qudit separately and hence does not interfere with the partial trace operation. This indicates that the norm $\|O\|^2_g = \|V^\dagger O V\|^2_g$ is invariant under the transformation V. This suggests us to define a locally scrambled ensemble $\mathcal{E}_O$ [or known as $U(d)^N$ twirling] associated with any given observable O,

$$\mathcal{E}_O \equiv \{V^\dagger O V | V \in \text{U}(d)^N\}, \tag{B4}$$

such that $\|O\|_g^2$ in Eq. (B3) can be redefined as its ensemble average

$$
\begin{aligned}
\|O\|_g^2 &= \mathbb{E}_{O \in \mathcal{E}_O} \|O\|_g^2 \\
&= \mathbb{E}_{V \in \mathrm{U}(d)^N} \mathrm{Tr}((\mathcal{M}^{-1}[V^\dagger O V]^{\otimes 2} \otimes \mathbb{1})\chi_g) \\
&= \sum_{A,B,C,D \in 2^{\Omega_N}} d^{2N} r_A r_B \mathsf{Wg}_{C,D} W_{\mathcal{E}_o,D}^{(2)} \mathrm{Tr}([(\chi_C)_{A,B} \otimes \mathbb{1}]\chi_g).
\end{aligned}
$$

(B5)

Here χ_C denotes the swap operator supported in region C that acts between the first two copies of the Hilbert space, and $(\chi_C)_{A,B}$ denotes the reduction of χ_C in regions A and B, respectively, in the first and the second copies of the Hilbert space, which results in $(\chi_C)_{A,B} = \chi_{A \cap B \cap C} d^{|A \cap B \cap C| - |C|}$. The operator entanglement feature $W_{\mathcal{E}_o,D}^{(2)} = \mathbb{E}_{O \in \mathcal{E}_O} \mathrm{Tr}_D(\mathrm{Tr}_{\bar{D}} O)^2$ follows from the same definition given in Eq. (8). r_A, r_B are the reconstruction coefficients given by the solution of Eq. (12). Substituting Eq. (B5) to (B2), we can evaluate the summation of g, h given that $\sum_{g,h \in S_3^N} \mathrm{Tr}(\chi_{A \cap B \cap C} \chi_g) \mathsf{Wg}_{g,h} W_{\mathcal{E}_\sigma,h}^{(3)} = W_{\mathcal{E}_\sigma,A \cap B \cap C}^{(2)}$. The reduction of the third entanglement feature to the second entanglement feature is a consequence of the fact that ρ drops out from the tensor product in Eq. (B3), such that only twofold Hilbert space is required to define $\|O\|_{\mathcal{E}_\sigma}^2$.

Thus, we finally arrive at the expression for the operator shadow norm purely in terms of the entanglement features of $\mathcal{E}_\sigma$ and $\mathcal{E}_O$,

$$
\|O\|_{\mathcal{E}_\sigma}^2 = \sum_{A,B,C,D \in 2^{\Omega_N}} v_{A,B,C,D} W_{\mathcal{E}_\sigma,A \cap B \cap C}^{(2)} W_{\mathcal{E}_o,D}^{(2)},
$$

(B6)

where the coefficient $v_{A,B,C,D}$ is given by

$$
v_{A,B,C,D} = r_A r_B \left(\frac{d^2}{d^2 - 1}\right)^N d^{|A \cap B \cap C| - |C|} \left(-\frac{1}{d}\right)^{|C \ominus D|}.
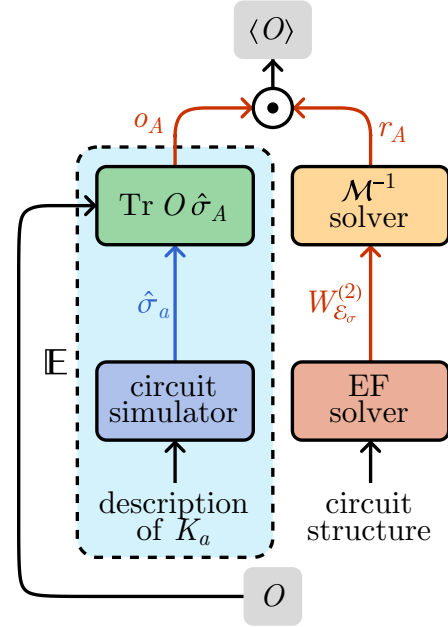$$

(B7)

FIG. 14. Classical postprocessing protocol to estimate the operator expectation value and shadow norm.

On the left of the workflow, we do experiments on quantum devices and calculate classical shadows $\hat{\sigma}$. And one should notice that our formulation is general enough to include Clifford circuits that do not have group structure. And the classical shadows of those circuits can be calculated efficiently. Then we can combine the classical shadows $\hat{\sigma}$ and tensor network representation of r_A to predict operator expectations $\langle O \rangle$.

A detailed discussion on how to model both $\mathrm{Tr}(O\hat{\sigma}_A)$ and r_A with tensor network is discussed in the following two subsections.

APPENDIX C: EFFICIENT CLASSICAL POSTPROCESSING ALGORITHM WITH TENSOR NETWORK METHOD

1. Overview of tensor network based classical postprocessing

In the main text, we have derived the following protocol for predicting quantities of states with locally scrambled quantum circuits: It seems that the number of coefficients r_A scales exponentially with system size, therefore not efficient. Surprisingly, with clever design and the help of tensor network, there indeed exists efficient tensor network method that can achieve efficient classical postprocessing. Figure 14 summarizes the classical postprocessing workflow of predicting operator expectations with tensor network methods.

On the right side of workflow, given the circuit structure, the entanglement feature can be efficiently encoded as a tensor network using "EF solver." Then the tensor network representation of the entanglement feature is inputted into the $\mathcal{M}^{-1}$ solver, whose output is a tensor network representation of the reconstruction coefficient r_A. The nice thing is that one only need to solve the tensor network representation for r_A once. And this representation can be stored for future usage.

2. Efficient matrix product state representation of $\vec{o} = \{o_A \mid \mathrm{Tr}(O\sigma_A)\}$

In the main text, we have shown the reconstruction channel under local scrambling assumption can be written as

$$
\rho = \mathcal{M}^{-1}[\sigma] = d^N \sum_{A \in 2^{\Omega_N}} r_A \sigma_A.
$$

(C1)

At first sight, the exponential summation of subregion A seems to be troublesome. However, it can be circumvented by tensor network methods. Here, we will introduce a concrete algorithm. First of all, if the unitaries used in the classical shadow experiment are Clifford gates, then classical shadows $\hat{\sigma}$ are stabilizer states, and each of them can be efficiently stored with $O(N^2)$ memory on a classical computer, where N is the system size.

Proposition I. Given O is a Pauli observable and σ is a stabilizer state, the vector $\vec{o} = \{o_A \mid o_A = \mathrm{Tr}(O\sigma_A)\}$ has an efficient matrix product state (MPS) representation with internal bond dimension equal to one, where $\sigma_A = \mathcal{D}_{\bar{A}}[\sigma]$. And $\mathcal{D}_A[\circ]$ is the depolarizing channel acting on region A.

Proof. If the circuit is composed of Clifford gates, then classical shadow σ is a stabilizer state with stabilizer group generated by

$$\mathcal{S} = \langle (-1)^{b_1} U^\dagger Z_1 U, \ldots, (-1)^{b_n} U^\dagger Z_N U \rangle = \langle \widetilde{Z}_1, \ldots, \widetilde{Z}_N \rangle,$$
(C2)

$$\sigma = \prod_{i=1}^{n} \frac{\mathbb{1} + \widetilde{Z}_i}{2} = \frac{1}{2^N} \sum_{g \in \mathcal{S}} g.$$
(C3)

The reduced state $\sigma_A = \mathcal{D}_{\bar{A}}[\sigma]$ restricted to region A is also a stabilizer state with stabilizer group $\mathcal{S}_A \subseteq \mathcal{S}$ defined by taking the elements of $\mathcal{S}$ which have zero support on $\bar{A}$. This is obviously a subgroup of $\mathcal{S}$ since it is closed under multiplication and inversion. Without loss of generality, we can write

$$\sigma_A = \left(\frac{1}{2^{|A|}} \sum_{g \in \mathcal{S}_A} g \right) \otimes \left(\frac{\mathbb{1}_2}{2} \right)^{\otimes (N-|A|)}.$$
(C4)

It is obvious that the expectation $\mathrm{Tr}(O\sigma_A) = 0$ when $\mathrm{supp}(O) \nsubseteq A$. Moreover, the only scenario when $\mathrm{Tr}(O\sigma_A)$ is nonzero is $\pm O \in \mathcal{S}_A$. Therefore, we have

$$\mathrm{Tr}(O\sigma_A) = \begin{cases} 0, & \pm O \notin \mathcal{S}_A \\ \mathrm{Tr}(O\sigma_A) = \mathrm{Tr}(O\sigma), & \pm O \in \mathcal{S}_A. \end{cases}$$
(C5)

From the above equation, it is clear that for any Pauli observable O, $o_A = \mathrm{Tr}(O\sigma_A)$ can be represented as a trivial MPS with bond dimension $D = 1$:

$$o_A = \mathrm{Tr}(\sigma O) \mathrm{Tr}\left(o_1^{(a_1)} o_2^{(a_2)} \ldots o_N^{(a_N)} \right)$$
$$= \mathrm{Tr}(\sigma O) o_1^{(a_1)} o_2^{(a_2)} \ldots o_N^{(a_N)},$$
(C6)

where we drop the second trace since the internal bond dimension is one, and each tensor $o_i^{(a_i)}$ on site i with binary physical index ($a_i = 0$ or 1) is

$$o_i = \begin{cases} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, & i \notin \mathrm{supp}(O) \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, & i \in \mathrm{supp}(O). \end{cases}$$
(C7)

This concludes that even the vector $\vec{o} = \{o_A | o_A = \mathrm{Tr}(O\sigma_A)\}$ contains exponentially many elements, it has an efficient MPS representation with bond dimension $D = 1$. This MPS representation can be easily constructed: given Pauli observable O, first calculate $\mathrm{Tr}(O\sigma)$, then construct the MPS using Eq. (C7). $\mathrm{Tr}(O\sigma)$ can computed in $O(N^2)$ time because O is a Pauli observable and σ is a stabilizer state. The remaining MPS tensors can be constructed in $O(N)$ time by traversing through the Pauli string.

3. Encoding reconstruction coefficient r_A with variational MPS method

In the main text, we argued that the vector r_A can be represented as a MPS. Here, we illustrate how to find such a MPS using variational method. First of all, we have shown that the reconstruction coefficient r_A satisfies the linear equation

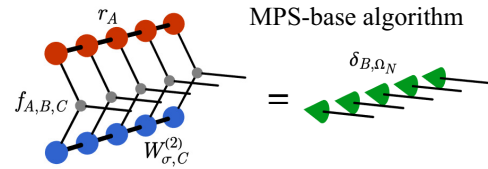$$\sum_{A,C \in 2^{\Omega_N}} r_A f_{A,B,C} W_C^{(2)} = \delta_{B,\Omega_N},$$
(C8)

FIG. 15. A cartoon illustration of variational solving MPS representation of r_A.

where $W_C^{(2)}$ is the second entanglement feature vector created by the unitary ensemble. In Ref. [47], the authors show entanglement feature vector $W_C^{(2)}$ can be efficiently encoded using MPS representation. The physical intuition behind this efficient representation is that if one views $W_C^{(2)}$ as a weight of a quantum state, i.e., $|\psi\rangle = \sum_{C \in 2^{\Omega_N}} W_C^{(2)} |C\rangle$, then this state will possess low entanglement Ref. [61]. Therefore, it can be represented as a MPS with low bond dimension. In Fig. 15, the blue nodes indicate the MPS representation of $W_C^{(2)}$. For translation-invariant circuit structure (such as the brick-wall circuit), the time complexity to construct the MPS representation for $W_C^{(2)}$ is $O(1)$ (independent of the system size). For general circuit structure, the time complexity is at most $O(N)$.

In Eq. (C8), the fusion coefficient $f_{A,B,C}$ is

$$f_{A,B,C} = \left(\frac{d^3}{d^2-1} \right)^N \sum_{D \in 2^{\Omega_N}} \delta_{B,A \cap D} d^{-|D|} \left(-\frac{1}{d} \right)^{|C \ominus D|}.$$
(C9)

Note that this fusion factor $f_{A,B,C}$ can be factorized to each site as $f_{A,B,C} = \prod_i f_{a_i,b_i,c_i}$ where

$$f_{a_i,b_i,c_i} = \begin{pmatrix} \binom{d}{0} & \binom{0}{0} \\ \frac{d^2}{d^2-1}\binom{d}{-1} & \frac{d}{d^2-1}\binom{-1}{d} \end{pmatrix},$$
(C10)

as the tensor subscripts $a_i, b_i, c_i = 0, 1$ enumerate over Boolean variables. Therefore, the fusion factor $f_{A,B,C}$ can be represented as the gray tensors in Fig. 15.

To find the MPS representation of vector r_A, we use the variational method. We can write an MPS ansatz for r_A and try to find the best parameters in the MPS by doing variational optimization. The same idea has been explored in machine-learning tensor network optimization [62–64], and differential programming of tensor networks [65–67]. With differential programming, we can find the best parameters in
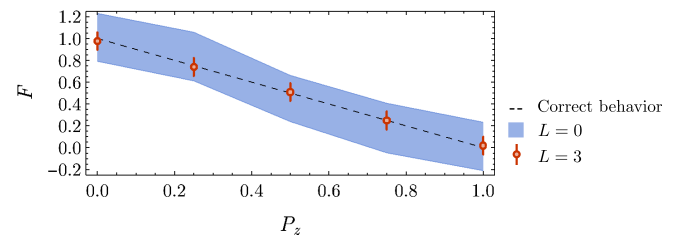

FIG. 16. Fidelity estimation between mix state and target state. 5000 experimental classical snapshots are prepared for both random Pauli measurement ($L = 0$) and shallow random unitary circuit ($L = 3$). Error bar indicates 3 standard deviation.
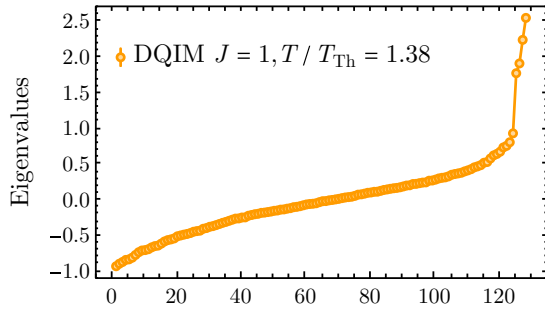
FIG. 17. Eigenvalues of reconstructed density matrix $\tilde{\rho}$ of seven-qubit GHZ state using mismatched channels. The unitary ensemble is single instance of DQIM ensemble with $J = 1$, and $T/T_{\mathrm{Th}} = 1.38$. Under this condition, the unitary ensemble is not locally scrambled.

the MPS ansatz for r_A by minimizing the L1 or L2 loss of the left-hand side tensor and right-hand side tensor of Fig. 15. With a fixed bond dimension, the algorithm complexity is $O(N)$. In addition, we can utilize the symmetry of the unitary ensemble to minimize the training parameters in the MPS

ansatz. In practice, we find that r_A can be represented as a MPS with a low bond dimension using the variational method. A detailed discussion of this new computational method will be in another paper.

In addition, we would like to point out that our proposal has caught much attention from both theoretical and experimental sides. Especially, the formal solution of Eq. (C8) can be solved [68]

$$r_A = \frac{(-1)^{-|A|}}{2^N} \sum_{A \subseteq S} \frac{3^{|S|}}{\sum_{B \subseteq S}(-2)^{|B|} W_B^{(2)}}. \tag{C11}$$

It would be also interesting to directly encode Eq. (C11) with a MPS without the help of variational optimization. And we leave this to a future study.

APPENDIX D: FIDELITY ESTIMATION FOR MIXED STATE

Our method is not restricted to pure state. In variational quantum state preparation, even the target state is some pure
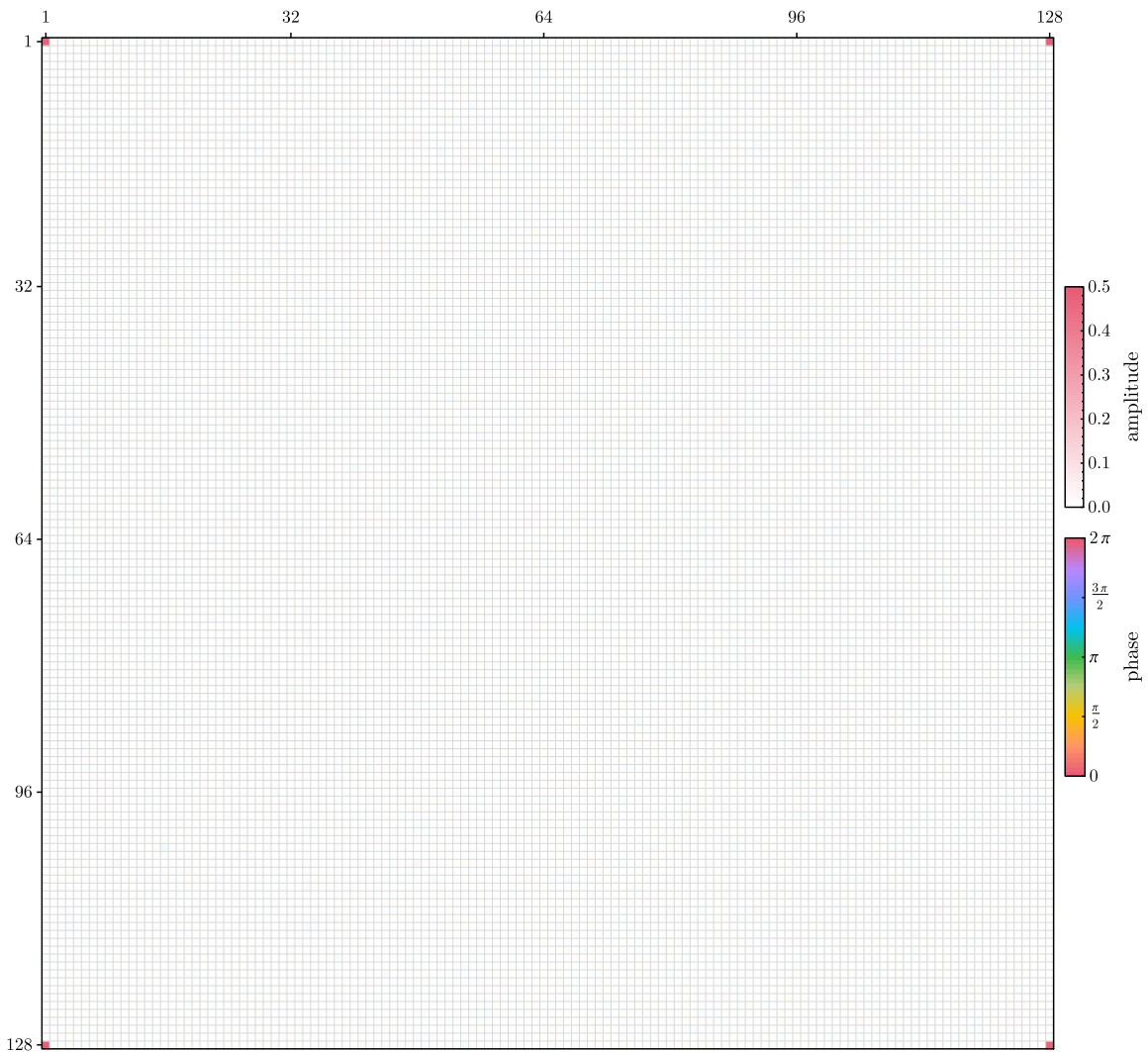

FIG. 18. Unbiased reconstruction of a seven-qubit GHZ density matrix, using a single instance of Hamiltonian in the DQIM ensemble at $T/T_{\mathrm{Th}} = 25.3$ (after the local scrambling condition is achieved).
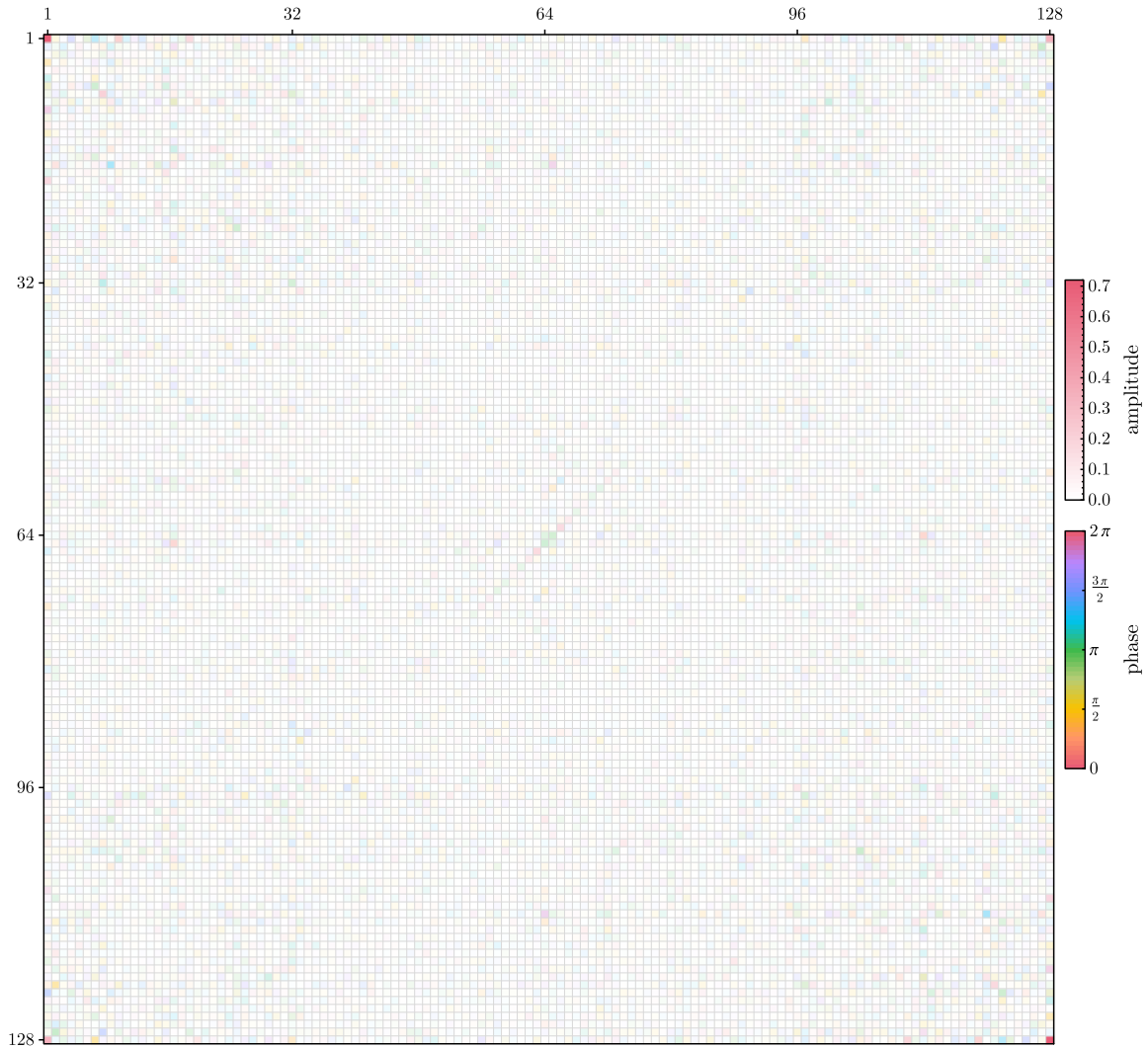
FIG. 19. Biased reconstruction of a seven-qubit GHZ density matrix, using a single instance of Hamiltonian in the DQIM ensemble at $T/T_{\mathrm{Th}} = 1.95$ (before the local scrambling condition is achieved).

state, noise in the preparation circuit could make the final state in experiments a mixed state. We can use the shallow circuit classical shadow tomography to efficiently estimate the quantum fidelity between final prepared quantum state ρ_P and the target quantum state ρ_T. Fast access to this quantity is crucial for variational quantum state preparation, error mitigation, etc. As an example, we consider the noisy preparation of a perfect GHZ state with Z errors occurs at probability p. The prepared state can be expressed as

$$\rho_P = (1 - p)|\psi_{\mathrm{GHZ}}^+\rangle\langle\psi_{\mathrm{GHZ}}^+| + p|\psi_{\mathrm{GHZ}}^-\rangle\langle\psi_{\mathrm{GHZ}}^-|, \qquad \text{(D1)}$$

where $|\psi_{\mathrm{GHZ}}^\pm\rangle = \frac{1}{\sqrt{2}}(|0^{\otimes N}\rangle \pm |1^{\otimes N}\rangle)$. We compare the performance between random Pauli measurement and shallow circuit shadow tomography with three layers of local random unitaries. Experiments are performed on a nine-qubit system and 5000 classical snapshots are collected for both random Pauli shadow tomography and shallow circuits shadow tomography. The result is shown in Fig. 16. As we can see, for 5000 experiments, the quantum fidelity estimated using random Pauli measurement has huge error bar, indicated by

the blue shaded region. However, the same amount of data collected after shallow circuit evolution can give accurate estimation of quantum fidelity, and the error bar is almost four times smaller. Practically, this makes the usage of shallow circuits more appealing.

APPENDIX E: APPROXIMATED UNITARY ENSEMBLE AND PURIFICATION

In the main text, we have seen when the measurement channel $\mathcal{M}$ in data acquisition and the reconstruction channel $\mathcal{M}^{-1}$ in classical postprocessing mismatch, the reconstructed density matrix $\frac{1}{M}\sum_{\hat{\sigma} \in \mathcal{E}_{\sigma|\rho}} \mathcal{M}^{-1}[\hat{\sigma}]$ may not be positive definite. And it results in biased prediction of physical quantities. In Figs. 5(a) and 11, we have seen the biased prediction of fidelity that is larger than one. In Fig. 17, we plot the eigenvalues of reconstructed density matrix of seven-qubit GHZ state using DQIM ensemble with $T/T_{\mathrm{Th}} = 1.38$. In the main text, we have seen the DQIM ensemble with one period of evolutional time or $T/T_{\mathrm{Th}} = 1.38$ is not sufficient to achieve the local scrambling assumption, such that there is mismatch
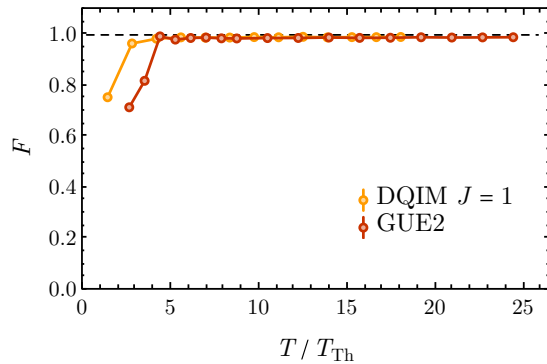
FIG. 20. Fidelity estimation of approximated unitary ensemble after purification. After around $T \sim 10 T_{\text{Th}}$, the fidelity is around 0.99. Same data are used as Fig. 11.

between data acquisition channel $\mathcal{M}$ and reconstruction channel $\mathcal{M}^{-1}$. We see the spectrum of density matrix contains some negative eigenvalues.

In addition, the approximate shadow tomography based on locally scrambling Hamiltonian evolution, such as DQIM ensemble or GUE2 ensemble, is approximately unbiased when local scrambling is approximately satisfied or frame potential gap is vanishingly small. In Fig. 11, we have seen they all can give unbiased prediction of quantum fidelity when $T \geqslant 10 T_{\text{Th}}$. We directly visualize the reconstructed density matrix using approximated DQIM ensemble in Figs. 18 and 19. As we see in Fig. 19, at $T/T_{\text{Th}} = 1.95$, the locally scrambling

assumption is not satisfied, and reconstructed density matrix is biased. In contrast, at $T/T_{\text{Th}} = 25.3$ (Fig. 18), the reconstructed density matrix using a single instance of DQIM Hamiltonian is perfect, justifying the validity of our approach when the locally scrambling assumption is approximated satisfied.

Furthermore, for biased reconstruction, in order to make it positive definite, we can nonlinear project the reconstructed ρ to the convex set of physical states $C = \{\rho | \rho \succeq 1, \text{Tr}(\rho) = 1\}$ by minimizing

$$\Pi_C(\sigma) = \arg \min_{\rho \in C} \text{Tr}[(\rho - \sigma)^2], \tag{E1}$$

which is the method mentioned in Ref. [30]. If we have more prior knowledge about the quantum state, such as it is a pure state, then we can further impose those assumptions into the projection. Here, as an illustration, we utilize the knowledge that the target quantum state is pure, and we project the reconstructed ρ to a pure state $\tilde{\rho}$ in C by choosing the eigenstate of ρ with the largest eigenvalue. As shown in Fig. 20, for the approximated ensembles, the GUE2 and DQIM are biased in the short-time region, and projected state $\tilde{\rho}$ has a fidelity less than one. And when locally scrambling assumption is approximately satisfied, the projected $\tilde{\rho}$ will have fidelity that is approximately 0.99. With these checks,

(i) unbiased prediction of physical quantities (see Fig. 11),

(ii) high fidelity of reconstructed density matrix projected back to physical space (see Fig. 20)

we confirm the approximated shadow tomography can perform unbiased reconstruction.

---

[1] K. Vogel and H. Risken, Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase, Phys. Rev. A **40**, 2847 (1989).

[2] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[3] C. M. Caves, C. A. Fuchs, and R. Schack, Unknown quantum states: The quantum de Finetti representation, J. Math. Phys. **43**, 4537 (2002).

[4] R. O'Donnell and J. Wright, Efficient quantum tomography, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing, STOC '16* (Association for Computing Machinery, New York, 2016), pp. 899–912.

[5] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, IEEE Trans. Inf. Theory **63**, 5628 (2017).

[6] S. Aaronson, Shadow tomography of quantum states, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018* (Association for Computing Machinery, New York, 2018), pp. 325–338.

[7] S. Aaronson and G. N. Rothblum, Gentle measurement of quantum states and differential privacy, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019* (Association for Computing Machinery, New York, 2019), pp. 322–333.

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

[9] M. Ohliger, V. Nesme, and J. Eisert, Efficient and feasible state tomography of quantum many-body systems, New J. Phys. **15**, 015024 (2013).

[10] H.-Y. Hu and Y.-Z. You, Hamiltonian-driven shadow tomography of quantum states, Phys. Rev. Res. **4**, 013054 (2022).

[11] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M. D. Lukin, Quantum phases of matter on a 256-atom programmable quantum simulator, Nature (London) **595**, 227 (2021).

[12] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, T. Lahaye, A. M. Läuchli, and A. Browaeys, Quantum simulation of 2D antiferromagnets with hundreds of rydberg atoms, Nature (London) **595**, 233 (2021).

[13] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z. X. Gong, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, Nature (London) **551**, 601 (2017).

[14] W.-T. Kuo, A. A. Akhtar, D. P. Arovas, and Y.-Z. You, Markovian entanglement dynamics under locally scrambled quantum evolution, Phys. Rev. B **101**, 224202 (2020).

[15] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, Quantum Entanglement Growth under Random Unitary Dynamics, Phys. Rev. X **7**, 031016 (2017).

[16] T. Zhou and A. Nahum, Emergent statistical mechanics of entanglement in random unitary circuits, Phys. Rev. B **99**, 174205 (2019).

[17] A. Nahum, S. Vijay, and J. Haah, Operator Spreading in Random Unitary Circuits, Phys. Rev. X **8**, 021014 (2018).

[18] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, Quantum Error Correction in Scrambling Dynamics and Measurement-Induced Phase Transition, Phys. Rev. Lett. **125**, 030505 (2020).

[19] Y. Bao, S. Choi, and E. Altman, Theory of the phase transition in random unitary circuits with measurements, Phys. Rev. B **101**, 104301 (2020).

[20] R. Fan, S. Vijay, A. Vishwanath, and Y.-Z. You, Self-organized error correction in random unitary circuits with measurement, Phys. Rev. B **103**, 174309 (2021).

[21] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, Towards the fast scrambling conjecture, J. High Energy Phys. 04 (2013) 022.

[22] S. Xu and B. Swingle, Locality, Quantum Fluctuations, and Scrambling, Phys. Rev. X **9**, 031048 (2019).

[23] H. Gharibyan, M. Hanada, S. H. Shenker, and M. Tezuka, Onset of random matrix behavior in scrambling systems, J. High Energy Phys. 07 (2018) 124.

[24] T. Zhou and X. Chen, Operator dynamics in a Brownian quantum circuit, Phys. Rev. E **99**, 052212 (2019).

[25] X. Chen and T. Zhou, Quantum chaos dynamics in long-range power law interaction systems, Phys. Rev. B **100**, 064305 (2019).

[26] Y.-Z. You, Z. Yang, and X.-L. Qi, Machine learning spatial geometry from entanglement features, Phys. Rev. B **97**, 045153 (2018).

[27] Y.-Z. You and Y. Gu, Entanglement features of random Hamiltonian dynamics, Phys. Rev. B **98**, 014309 (2018).

[28] M. Saffman, Quantum computing with atomic qubits and rydberg interactions: progress and challenges, J. Phys. B: At., Mol. Opt. Phys. **49**, 202001 (2016).

[29] C. Monroe, W. C. Campbell, L.-M. Duan, Z.-X. Gong, A. V. Gorshkov, P. W. Hess, R. Islam, K. Kim, N. M. Linke, G. Pagano, P. Richerme, C. Senko, and N. Y. Yao, Programmable quantum simulations of spin systems with trapped ions, Rev. Mod. Phys. **93**, 025001 (2021).

[30] A. Acharya, S. Saha, and A. M. Sengupta, Informationally complete POVM-based shadow tomography, arXiv:2105.05992.

[31] M. Guţă, J. Kahn, R. Kueng, and J. A. Tropp, Fast state tomography with optimal error bounds, J. Phys. A: Math. Theor. **53**, 204001 (2020).

[32] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states, Phys. Rev. A **99**, 052323 (2019).

[33] D. E. Koh and S. Grewal, Classical Shadows With Noise, Quantum **6**, 776 (2022).

[34] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, PRX Quantum **2**, 030348 (2021).

[35] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic Partial Tomography via Classical Shadows, Phys. Rev. Lett. **127**, 110504 (2021).

[36] D. Weingarten, Asymptotic behavior of group integrals in the limit of infinite rank, J. Math. Phys. **19**, 999 (1978).

[37] B. Collins and P. Śniady, Integration with respect to the haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys. **264**, 773 (2006).

[38] A. A. Akhtar and Y.-Z. You, Multiregion entanglement in locally scrambled quantum dynamics, Phys. Rev. B **102**, 134203 (2020).

[39] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller, Rényi Entropies from Random Quenches in Atomic Hubbard and Spin Models, Phys. Rev. Lett. **120**, 050406 (2018).

[40] B. Vermersch, A. Elben, M. Dalmonte, J. I. Cirac, and P. Zoller, Unitary n-designs via random quenches in atomic hubbard and spin models: Application to the measurement of rényi entropies, Phys. Rev. A **97**, 023604 (2018).

[41] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, Science **364**, 260 (2019).

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

[43] D. Aharonov, J. Cotler, and X.-L. Qi, Quantum algorithmic measurement, Nat. Commun. **13**, 887 (2022).

[44] X. Ma, T. Jackson, H. Zhou, J. Chen, D. Lu, M. D. Mazurek, K. A. G. Fisher, X. Peng, D. Kribs, K. J. Resch, Z. Ji, B. Zeng, and R. Laflamme, Pure-state tomography with the expectation value of pauli operators, Phys. Rev. A **93**, 032140 (2016).

[45] K. Kraus, General state changes in quantum theory, Ann. Phys. **64**, 311 (1971).

[46] D. Gottesman, The heisenberg representation of quantum computers, arXiv:quant-ph/9807006.

[47] A. A. Akhtar, H.-Y. Hu, and Y.-Z. You, Scalable and flexible classical shadow tomography with tensor networks, arXiv:2209.02093.

[48] G. Vidal, Entanglement Renormalization, Phys. Rev. Lett. **99**, 220405 (2007).

[49] G. Vidal, Class of Quantum Many-Body States That Can Be Efficiently Simulated, Phys. Rev. Lett. **101**, 110501 (2008).

[50] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence, J. High Energy Phys. 06 (2015) 149.

[51] Y. Li, X. Chen, and M. P. A. Fisher, Quantum Zeno effect and the many-body entanglement transition, Phys. Rev. B **98**, 205136 (2018).

[52] Y. Li, X. Chen, and M. P. A. Fisher, Measurement-driven entanglement transition in hybrid quantum circuits, Phys. Rev. B **100**, 134306 (2019).

[53] M. Szyniszewski, A. Romito, and H. Schomerus, Entanglement transition from variable-strength weak measurements, Phys. Rev. B **100**, 064204 (2019).

[54] A. Chan, R. M. Nandkishore, M. Pretko, and G. Smith, Unitary-projective entanglement dynamics, Phys. Rev. B **99**, 224307 (2019).

[55] B. Skinner, J. Ruhman, and A. Nahum, Measurement-Induced Phase Transitions in the Dynamics of Entanglement, Phys. Rev. X **9**, 031009 (2019).

[56] R. Vasseur, A. C. Potter, Y.-Z. You, and A. W. W. Ludwig, Entanglement transitions from holographic random tensor networks, Phys. Rev. B **100**, 134203 (2019).

[57] C.-M. Jian, Y.-Z. You, R. Vasseur, and A. W. W. Ludwig, Measurement-induced criticality in random quantum circuits, Phys. Rev. B **101**, 104302 (2020).

[58] M. J. Gullans and D. A. Huse, Dynamical Purification Phase Transition Induced by Quantum Measurements, Phys. Rev. X **10**, 041020 (2020).

[59] M. J. Gullans, S. Krastanov, D. A. Huse, L. Jiang, and S. T. Flammia, Quantum Coding with Low-Depth Random Circuits, Phys. Rev. X **11**, 031066 (2021).

[60] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, Science **377**, eabk3333 (2022).

[61] T. Grover and M. P. A. Fisher, Entanglement and the sign structure of quantum states, Phys. Rev. A **92**, 042308 (2015).

[62] E. Stoudenmire and D. J. Schwab, "Supervised learning with tensor networks," in *Advances in Neural Information Processing Systems*, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon,

and R. Garnett (Curran Associates, Inc., Red Hook, NY, 2016), Vol. 29.

[63] S. Lu, M. Kanász-Nagy, I. Kukuljan, and J. I. Cirac, Tensor networks and efficient descriptions of classical data, arXiv:2103.06872.

[64] J. A. Reyes and E. M. Stoudenmire, Multi-scale tensor network architecture for machine learning, Mach. Learn.: Sci. Technol. **2**, 035036 (2021).

[65] M. Hauru, M. V. Damme, and J. Haegeman, Riemannian optimization of isometric tensor networks, SciPost Phys. **10**, 040 (2021).

[66] C. Geng, H.-Y. Hu, and Y. Zou, Differentiable programming of isometric tensor networks, Mach. Learn.: Sci. Technol. **3**, 015020 (2022).

[67] R. Haghshenas, Optimization schemes for unitary tensor-network circuit, Phys. Rev. Res. **3**, 023148 (2021).

[68] K. Bu, D. Enshan Koh, R. J. Garcia, and A. Jaffe, Classical shadows with Pauli-invariant unitary ensembles, arXiv:2202.03272.