Shallow Shadows: Expectation Estimation Using Low-Depth Random Clifford Circuits

Christian Bertoni,[1] Jonas Haferkamp,[1] Marcel Hinsche,[1] Marios Ioannou,[1] Jens Eisert,[1,2,3] and Hakop Pashayan[1]
[1]Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany
[2]Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany
[3]Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany

(Received 20 March 2023; accepted 17 May 2024; published 10 July 2024)

We provide practical and powerful schemes for learning properties of a quantum state using a small number of measurements. Specifically, we present a randomized measurement scheme modulated by the depth of a random quantum circuit in one spatial dimension. This scheme interpolates between two known classical shadows schemes based on random Pauli measurements and random Clifford measurements. We focus on the regime where depth scales logarithmically in the system size and provide evidence that this retains the desirable sample complexity properties of both extremal schemes while also being experimentally feasible. We present methods for two key tasks; estimating expectation values of certain observables from generated classical shadows and, computing upper bounds on the depth-modulated shadow norm, thus providing rigorous guarantees on the accuracy of the output estimates. We achieve our findings by bringing together tools from shadow estimation, random circuits, and tensor networks.

DOI: 10.1103/PhysRevLett.133.020602

It is a well known fact that learning a classical description of a quantum state from access to a limited number of samples (copies of the state) is a demanding task for quantum systems involving a large number of degrees of freedom. Quantum state tomography produces an accurate classical description of the unknown quantum states, but consumes an enormous number of samples [1–3]. Recently, the focus has shifted towards sample efficient approaches that satisfy weaker standards of state learning but are nevertheless very useful. Huang, Kueng and Preskill's (HKP) seminal work [4], building on ideas from Ref. [5] is a state learning protocol based on randomized measurements [6–9]. In particular, the protocol of classical shadows, suggested by HKP, allows one to estimate the expectation values of many observables from the outcomes of suitable randomized measurements of the unknown state. This protocol is oblivious to the observable, i.e., once the necessary classical data has been gathered from the state, it can be reused for many observables. The scheme comes with rigorous performance guarantees, permitting one to determine the trade-off between the number of samples and the accuracy of the estimated expectation value for any given observable. Classical shadows have found many applications including estimating: Expected molecular energies [10,11], the purity and ground state proximity of spin chains [12] and gate-set properties [13] for noise characterization. It has been used to detect entanglement [14–16] and chaos [17] in many body systems; classify quantum data [18], and develop improved variational search algorithms [19,20]. For example, by relying on the scheme's ability to obliviously estimate the expectation values of many observables, Ref. [19] has proposed a promising new method for training variational circuits to find Hamiltonian eigenstates and Ref. [21] improved existing computational methods and used them to estimate ground state atomization energies of molecules.

Classical shadows have seen some experimental implementation [14,22,23], inspired substantial further study [10,24–26] and many related proposals and generalizations including; noise-robustness analysis and noise-robust variants [27–29], variants for fermionic systems [30] and performance improvement in the setting of pure states [31]. Variants of the protocol have been considered where the randomized measurements are: derandomized [32], reused [33,34], or generalized to joint measurements [31]. Randomized measurements generated by low depth unitary ensembles [35,36], locally scrambled unitary ensembles [37,38], or certain Hamiltonian evolutions [39–41] have also been considered.

The classical shadows protocol as proposed by HKP has two variants; one randomizes the computational basis measurement by first applying a random global Clifford unitary to each copy of the unknown quantum state (also known as random Clifford measurements), the other by first applying a layer of random single qubit Clifford gates (also known as random Pauli measurements). Both prescriptions give rise to interesting and complementary schemes, each able to estimate expectation values for a large class of observables using a sparing number of copies of the unknown state. The global Clifford scheme is sample efficient for observables with low Frobenius norm including all quantum states (for fidelity estimation) and low-rank operators with bounded spectral norm. However, this

scheme performs poorly for high rank observables such as local observables and Paulis, even those with low weight. Additionally, the implementation of global Cliffords requires a linear depth circuit making the scheme infeasible even for moderate sized systems. In contrast, the single qubit Cliffords scheme is technologically far less demanding and has seen experimental implementation [22,23,42]. This scheme is on the one hand sample efficient for bounded observables that are linear combinations of low weight Paulis. On the other hand, the sample complexity can scale exponentially in the weight of the observable, rendering this scheme unsuitable for globally supported observables, e.g., for fidelity estimation.

In this work, we provide evidence that much of the power of the global Clifford scheme can be retained when resorting to shallow (logarithmic depth) circuits. This significantly reduces the experimental cost. Specifically, we study the family of shadow schemes between these two extremes and identify an intermediate regime that is experimentally feasible and inherits much of the favorable sample complexity scaling of both extremal schemes. Our schemes are modulated by the depth of the randomly sampled two-local Clifford circuit used to randomize measurements (Fig. 1). Our zero depth and infinite depth schemes correspond to the random Pauli measurements and random Clifford measurements schemes of HKP, respectively. To be practical, such a protocol needs to be both sample efficient and computationally efficient. That is, only a polynomial number of copies of the state should be required to estimate expectation values to the desired accuracy while the classical pre- and postprocessing should take at most polynomial time. Some aspects of this last requirement in particular are straightforward in the HKP protocol but challenging in our intermediate regime, as intermediate depth Clifford unitary circuits do not form a group. However, even for the HKP protocols, the computational efficiency is highly dependent on the choice of observables: Indeed, one can come up with observables for which the sample complexity is polynomially bounded, but

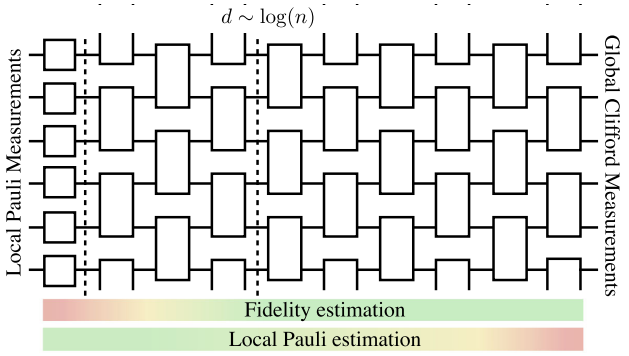

FIG. 1. Brickwork circuit and illustration of the performance of the protocol in various regimes. At depths logarithmic in the system size, both Pauli estimation and fidelity estimation become viable.

computing the expectation values requires exponential time. We therefore identify two classes of observables for which our protocol is computationally efficient for $d = \mathcal{O}(\log(n))$ under reasonable assumptions. We also introduce computationally efficient methods for computing upper bounds on the estimation error. We study the sample complexity of our schemes at depth $d = \Theta(\log n)$ for large classes of observables. For observables that have polynomial sample complexity under the zero depth scheme (Pauli measurements) we prove that our depth $d = \Theta(\log n)$ scheme also achieves polynomial sample complexity. For observables that have polynomial sample complexity under the infinite depth scheme (global Cliffords) we give theoretical and numerical evidence that our scheme achieves polynomial sample complexity. We thus put forward that our $d = \Theta(\log n)$ scheme constitutes the sweet spot between the two extremes.

Depth modulated classical shadows protocol.—Given many identical samples of an unknown n-qubit quantum state ρ, our goal is to compute a classical description $\hat{\rho}$ such that for any observable O in some large class, $\mathrm{tr}(O\rho) \approx \mathrm{tr}(O\hat{\rho})$. The protocol has three distinct components, namely, (a) the data acquisition step which implements a randomized measurement and outputs measurement data; (b) the estimate computation step that takes the measurement data and a description of O as input and outputs an estimate $\hat{e}$ of $e = \mathrm{tr}(O\rho)$; (c) the accuracy guarantee step that computes an upper bound on the estimation error $\epsilon = |\hat{e} - e|$ using the number of samples taken in the data acquisition step and a description of O. In the following, we describe these three steps.

Step 1: Data acquisition: We now describe the collection of randomized measurement data for a given circuit depth $d \in \{\infty, 0, 1, \ldots\}$. We define $\mathcal{U}_d$ as the ensemble of Clifford unitaries constructed by first applying a uniformly random single qubit Clifford gate to each qubit (the 0th layer). If $d = 0$ then we are done, alternatively, we apply d layers of independent and uniformly sampled 2-qubit Clifford gates in an alternating brickwork pattern (cf. Fig. 1). The resulting n-qubit, depth d Clifford circuit is a single sample from the ensemble $\mathcal{U}_d$. For each copy of ρ, we independently apply a Clifford unitary U sampled from $\mathcal{U}_d$, then measure in the computational basis, with the resulting measurement outcome $b \in \{0, 1\}^n$ and unitary U labeling the measurement data produced with respect to each sample of ρ. For concreteness, we assume a circular brickwork architecture where the first and last qubit are identified as nearest neighbors. While this makes the analysis easier our results can be extended to other local circuit architectures.

Step 2: Estimate computation: We now describe how the measurement data $\{(U_i, b_i)\}_i$ from step 1 can be used to produce an estimate of an expectation value for a given observable O. On average over the randomness of (U, b), the data acquisition step defines a quantum channel

$$\mathcal{M}_d(\rho) \coloneqq \mathop{\mathbb{E}}_{U\sim\mathcal{U}_d}\left[\sum_{b\in\{0,1\}^n}\langle b|U\rho U^\dagger|b\rangle U^\dagger|b\rangle\langle b|U\right], \quad (1)$$

henceforth referred to as the measurement channel. That is, it produces states $\tau_i \coloneqq U_i^\dagger|b_i\rangle\langle b_i|U_i$ with a probability depending on the state ρ and the ensemble of unitaries. Since τ_i is the output of a depth d quantum circuit, it can be represented as a bond dimension $\mathcal{O}(2^d)$ matrix product state (MPS). As a linear map, $\mathcal{M}_d$ is positive definite and hence invertible (cf. Theorem 1 and Sec. II of the Supplemental Material [43]). This map is central to our study due to the observation that $\mathcal{M}_d^{-1}(\tau)$ is an unbiased estimator of ρ. Thus, each measurement outcome (U_i, b_i) can be converted into an independent instance $\hat{o}_i \coloneqq \mathrm{tr}[O\mathcal{M}_d^{-1}(\tau_i)] = \mathrm{tr}[\mathcal{M}_d^{-1}(O)\tau_i]$ of a random variable $\hat{o}$ which has mean $e = \mathrm{tr}(O\rho)$. Just as in the standard HKP protocols, the final estimate $\hat{e}$ can be computed from these single-shot estimates as follows: first, we partition all of the single-shot estimates $\hat{o}_i$ into equal blocks of appropriate size and average the estimates in each block, then, the final estimate $\hat{e}$ is the median of the averages for each block. We have now specified our estimators; these are natural generalizations of the HKP estimators to the intermediate depth regime. The remainder of this section focuses on how to efficiently compute a single-shot estimate of $\hat{o}$ given τ and a target observable O. We consider two efficiently describable classes of observables: those given as a linear combination of at most $\mathrm{poly}(n)$ Paulis (we call these sparse observables) and those given as a matrix product operator (MPO) with a $\mathrm{poly}(n)$ bounded bond dimension (we call these shallow observables). For both classes of observables, to compute an estimate, we need a way of computing $\mathcal{M}^{-1}(O)$ for our observable of choice, O. From now on, we label Pauli operators with bit strings of length $2n$. Given $\lambda\in\{0,1\}^{2n}$, the corresponding Pauli is denoted P^λ. We rely on the following key results, proven in the Supplemental Material [43], Sec. II:

Theorem 1.—(Action of measurement map) For $\lambda\in\{0,1\}^{2n}$, let $t_{\lambda,d}\in[0,1]$ be defined as

$$t_{\lambda,d} \coloneqq \Pr_{U\sim\mathcal{U}_d}(UP^\lambda U^\dagger \in \pm\mathcal{Z}), \quad (2)$$

where $\pm\mathcal{Z}$ is the set of all strings of Pauli Z and I with phase ±1, then

$$\mathcal{M}_d(P) = t_{\lambda,d}P^\lambda. \quad (3)$$

One can verify that the expression in Eq. (2) corresponds to the expected limiting values in the extreme cases $d=\infty$ and $d=0$. In these cases, the measurement channel is given, respectively, by global and local depolarizing channels, at $d=\infty$ we have then for any A, $\mathcal{M}_\infty(A) = [1/(2^n+1)][A+\mathrm{tr}(A)I]$ while at $d=0$, the channel acts like $\mathcal{M}_\infty$ for $n=1$ on each qubit, such that for a Pauli P with support on k qubits $\mathcal{M}_0(P) = 3^{-k}P$. Indeed, the probability that a randomly drawn Pauli operator is in $\pm\mathcal{Z}$ is $(2^n-1)/(4^n-1) = 1/(2^n+1)$, while if we draw a random Pauli on each of k sites, this probability is 3^{-k}.

This expression permits the construction of an efficient tensor network representation for $t_{\lambda,d}$: for two qubits, a four-legged tensor can be constructed representing the probability that a Pauli P is mapped to another Pauli Q by a random Clifford gate. By arranging copies of this tensor in a brickwork fashion, one obtains the probabilities that any possible Pauli string is produced by the circuit, given a certain input string. Finally, summing over the outputs corresponding to elements of $\pm\mathcal{Z}$ we get the following.

Theorem 2.—(Tensor network formulation) $t_{\lambda,d}$ can be written as a one-dimensional tensor network with bond dimension at most 2^{d-1}. For depth $d=\mathcal{O}(\log n)$, it can then be exactly computed in run-time $n^{\mathcal{O}(1)}$.

The construction is described in detail in Sec. II of the Supplemental Material [43]. We now have two possible choices: if we are interested in estimating a sparse observable, we can efficiently compute an estimate $\hat{o}$ by evaluating term by term

$$\hat{o} = \mathrm{tr}(\mathcal{M}_d^{-1}(O)\tau) = \sum_\lambda \frac{1}{t_{\lambda,d}}\beta_\lambda \mathrm{tr}(P^\lambda\tau). \quad (4)$$

The polynomially many terms can each be efficiently computed since both $t_{\lambda,d}$ and the expectation value of a Pauli with respect to the MPS τ are efficiently computable. If instead O is a shallow observable, there will generally be exponentially many nonzero Pauli coefficients in the expansion of O, so the term by term computation of Eq. (4) is no longer efficient. Instead, we first construct an approximate MPS representation of the eigenvalues $1/t_{\lambda,d}$ of the inverse measurement map. Given such an MPS, and O in the form of a shallow MPO, we can represent $\mathcal{M}_d^{-1}(O)$ as an MPO with controlled bond-dimension. Since τ is also given as a low bond dimension MPS, for shallow observables, $\hat{o}$ can be efficiently approximated by a tensor contraction.

We briefly discuss the construction of an approximate MPS representation of the inverse measurement map (see Supplemental Material [43], Sec. III for details). Let us first note that it is nontrivial to construct a low bond-dimenson MPS that evaluates $1/t_{\lambda,d}$ given a low bond-dimenson MPS that evaluates $t_{\lambda,d}$. However, given a candidate for the former we can efficiently compute how much it deviates from the exact MPS representation of the inverse measurement map. Using this property, we are able to employ a DMRG style heuristic that repeatedly sweeps through all components of a candidate MPS locally optimizing each component (while all others remain fixed). An accurate approximate MPS representation of the inverse measurment channel can be precomputed, since it only depends on

n and d. We do not guarantee that this can be done efficiently but note that in practice, our method produces high precision results, e.g., $n = 10$, $d = 3$ can be computed in less than a minute on a laptop. For reasonable parameter choices, the error induced by this approximation is much smaller than the statistical error (cf. Supplemental Material [43], Sec. IV, see also Ref. [52]).

Step 3: Accuracy guarantee: We now discuss how to upper bound the error in the estimate from step 2, given a description of a target observable O and the number of samples of ρ used in step 2. We bound the variance of the estimators $\hat{o}_i$, since this implies a bound on the errors for a given number of samples (see Supplemental Material [43], Sec. V).

We upper bound the variance of $\hat{o}$ by its second moment, $\mathbb{E}[\hat{o}^2]$, and refer to this quantity as the state-dependent shadow norm denoted by $\|O\|_{\mathrm{s}(d),\rho}$. To get a worst case upper bound we maximize this over all density states ρ; this defines the shadow norm $\|O\|_{\mathrm{s}(d)}$ [4]. We will also consider the locally scrambled shadow norm [37,38], $\|O\|_{\mathrm{s}(d),\mathrm{LS}}$. This is the average of $\|O\|_{\mathrm{s}(d),\rho}$ over ρ sampled from a state 1-design, i.e., any ensemble of pure states $\mathcal{E}$ such that $\mathbb{E}_{\sigma\sim\mathcal{E}}[\sigma] = \mathbb{I}/2^n$. In the Supplemental Material [43], Sec. V, we give explicit expressions for all of the above quantities for a general observable O in terms of a quantity $\tau_{(\lambda,\lambda'),d}$ which is a second order analog of the eigenvalues $t_{\lambda,d}$. The explicit expression for the locally scrambled shadow norm is efficiently computable for both shallow and sparse observables. The expression for the shadow norm is efficient only in the sparse case. In the shallow case, given low bond dimension approximations to $1/t_{\lambda,d}$, we give efficiently computable upper bounds to the shadow norm. Furthermore, in the Supplemental Material [43], Sec. VI, we consider the shadow norm of stabilizer projectors, and, in particular, we explicitly identify the states that maximize the state-dependent shadow norm for these observables, that is, the states yielding the worst possible performance.

On sample complexity at logarithmic depth.—Finally, we consider the sample complexity of our $d = \mathcal{O}(\log n)$ scheme with respect to two important classes of observables. First, we consider the subset of sparse observables that can be written as a linear combination of $\mathcal{O}(\log n)$ weight Pauli operators with coefficients of size at most $\mathrm{poly}(n)$, e.g., one-dimensional local Hamiltonians. These observables are known to have $\mathrm{poly}(n)$ sample complexity with respect to the $d = 0$ scheme of HKP. By bounding the shadow norm, we show that for this class, $\mathrm{poly}(n)$ sample complexity is also achieved by our $d = \mathcal{O}(\log n)$ scheme. Then, we consider the subset of observables with $\mathrm{poly}(n)$ bounded Frobenius norm. These observables are known to have $\mathrm{poly}(n)$ sample complexity with respect to the $d = \infty$ scheme of HKP. We give evidence that this also holds for our $d = \mathcal{O}(\log n)$ scheme: by bounding the locally scrambled shadow norm, we show that for most states, the sample complexity is polynomial. Specifically, we show that if the unknown ρ were sampled from any state 1-design, then for at least a $1 - 1/\mathrm{poly}(n)$ fraction of states, $\mathrm{poly}(n)$ copies of the state would suffice. We show this by using a mapping to a path counting problem in a statistical mechanics model [44,53,54]. All of this can be summarized in the following Theorem, which we prove in Sec. VII of the Supplemental Material [43].

Theorem 3.—(Performance guarantee) Let $d = \Theta(\log(n))$. If $O = \sum_{k=1}^{r} \beta_{\lambda_k} P^{\lambda_k}$ is a linear combination of Pauli operators such that each Pauli is supported on a region of length upper bounded by $\mathcal{O}(\log(n))$, then

$$\|O\|_{\mathrm{s}(d)} \leq n^{\mathcal{O}(1)} \sum_{k=1}^{r} |\beta_{\lambda_k}|. \tag{5}$$

Furthermore, for any traceless O, we have

$$\|O\|^2_{s(d),\mathrm{LS}} \leq 2\|O\|_F^2 \left(1 + \frac{1}{n^{\mathcal{O}(1)}}\right). \tag{6}$$

Numerics.—We now present numerical experiments performed by classically simulating step 1 of our scheme (the component that is to be implemented on a quantum computer) followed by the implementation of our protocol as described in steps 2 and 3 [55].

In Fig. 2, we choose ρ to be the GHZ state on $n = 8$ qubits. For each depth, we obtain 100 independent

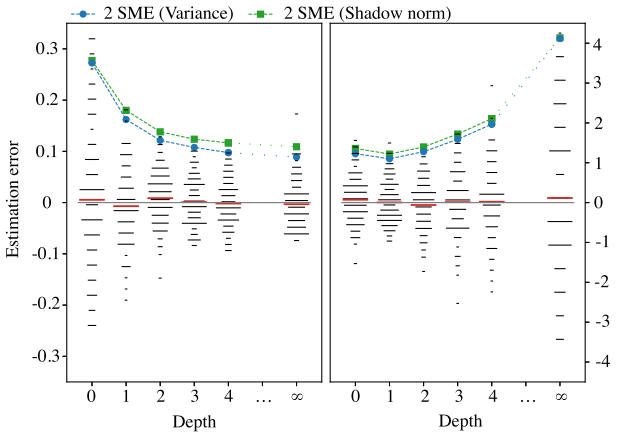

FIG. 2. Performance of the protocol for $O = \rho = |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$ (left) and $O = H$ (right) for various depths. For each depth we compute 100 independent estimates each based on 1000 samples and plot the estimation error (estimate minus the true value). We show the spread of estimation errors as well as a bound for twice the standard mean error (SME) corresponding to ∼95% confidence interval assuming Gaussian distributed errors. Each black line has a length proportional to the fraction of estimates with similar estimation error. The SMEs are computed using the empirical variance and the shadow norm. The red lines corresponds to the estimates obtained using the whole set of 100 000 samples.

estimates of $\mathrm{tr}(O\rho)$ using $N = 1000$ samples for each. On the left side of Fig. 2 we estimate the fidelity of the state with itself, i.e., pick $O = |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$. Since O has full support and $\|O\|_{\mathrm{F}} = 1$, the $d = 0$ scheme is expected to perform poorly and the $d = \infty$ scheme is expected to perform well. We see that the spread in estimated values depends on the depth and is consistent with both the empirically calculated variance and the theoretically computed shadow norms. It is apparent that the performance becomes dramatically better from $d = 0$ to $d = 2$ (decreasing shadow norm and variability in estimates). Thereafter, the improvement in performance for every additional layer is much smaller, indicating that the global Cliffords scheme's performance is already matched at short depth.

On the right side of Fig. 2, we consider expectations of a sparse observable, namely, the expected energy $\langle H\rangle$ for the Hamiltonian $H = \sum_i (Z_{i-1}Z_iZ_{i+1} + X_i)$, with periodic boundary conditions. As expected, low depth is better suited for this observable, and it is evident that the global $d = \infty$ protocol is ill suited. The shadow norm of a k-local observable (or a linear combination thereof) is minimized at some k dependent optimal depth, which appears to be $d = 1$ in this example. While this minimum does not necessarily occur at $d \sim \log(n)$, the protocol is still efficient, if not optimal, at log depth. This can be seen by the fact that the error only increases slowly, up to moderately high depths, compared to the global minimum. In the Supplemental Material [43], Sec. VIII, we present additional numerical results where we consider a randomly drawn MPS for ρ and an XXZ model for H. As expected the results are similar to those presented here, showcasing the generality of our method.

Discussion.—We provide a protocol for estimating properties of an unknown state by considering measurements in a basis generated by random quantum circuits of 2-local Clifford gates of arbitrary depth d. To go beyond the extremal depths of $d \in \{0, \infty\}$, we overcome new challenges by bringing together methods from random quantum circuits and tensor networks. This allows us to estimate expectation values of a large class of experimentally relevant observables. we expect the $d \sim \log(n)$ regime to be a "sweet spot" that inherits the benefits of the high depth scheme while requiring much fewer resources, being efficiently implementable, and retaining the desirable properties of the low depth scheme. Several questions remain open. At extremal depths, the shadows procedure is shown to be robust to noisy measurements [27], we expect that a similar robustness can be obtained in our protocol for local Pauli noise by correcting the MPS tensors of the measurement channel. We also expect that this protocol can be directly applied to improve applications of shadow tomography such as learning quantum processes [56,57]. Based on numerical evidence, low bond dimension MPS representations of the inverse measurement map appear to exist, however, we are not aware of theoretical bounds on their bond dimension. Furthermore, while our bounds on the locally scrambled shadow norm provide compelling evidence that our log-depth protocol is useful for low Frobenius norm observables, a rigorous proof of sample efficiency for this class of observables is missing. Finally, we provide ways of estimating the sample complexity for individual observables. We are confident that the computations involved in these procedures can be made more efficient by bond dimension reduction techniques.

Note added.—Recently, we became aware of the recent results of Ref. [52] which studies the MPS inversion problem in the more general setting of MPOs and arrive at very similar results as our MPS inversion algorithm.

We thank Ingo Roth, David Wierichs, and Afrad Basheer for fruitful discussions. We also acknowledge a discussion with two groups of researchers: On the one hand, Ahmed A. Akhtar, Hong-Ye Hu, Yi-Zhuang You, and on the other hand Mirko Arienzo, Markus Heinrich, and Martin Kliesch that both independently and concurrently investigated intermediate classical shadows schemes. H. P. acknowledges the Centre for Quantum Software and Information at the University of Technology Sydney and Michael Bremner for hosting him as a visiting scholar. We thank the BMBF (Hybrid, MuniQC-Atoms, DAQC) the MATH+ Cluster of Excellence, the Einstein Foundation (Einstein Unit on Quantum Devices), the QuantERA (HQCC), the Munich Quantum Valley (K8), the Quantum Flagship (Millenion, PasQuans2), the ERC (DebuQC), and the DFG (CRC 183, EI 519 20-1, EI 519/21-1) for support.

[1] 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, NY, USA, 2016), pp. 899–912.

[2] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing, STOC '16* (Association for Computing Machinery, New York, NY, USA, 2016), pp. 913–925.

[3] M. G. D'Ariano, M. G. Paris, M. Sacchi, and F. Massimiliano, Quantum tomography, Adv. Imaging Electron Phys. **128**, 205 (2003).

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

[5] 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, NY, USA, 2018), pp. 325–338.

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

[7] 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.
[8] 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).
[9] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).
[10] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, Commun. Math. Phys. **391**, 951 (2022).
[11] C. Hadfield, Adaptive Pauli shadows for energy estimation, arXiv:2105.12207.
[12] S. Notarnicola, A. Elben, T. Lahaye, A. Browaeys, S. Montangero, and B. Vermersch, A randomized measurement toolbox for an interacting Rydberg-atom quantum simulator, New J. Phys. **25**, 103006 (2023).
[13] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A. H. Werner, J. Eisert, and I. Roth, Shadow estimation of gate-set properties from random sequences, Nat. Commun. **14**, 5039 (2023).
[14] A. Elben, R. Kueng, H.-Y. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermersch, Mixed-state entanglement from local randomized measurements, Phys. Rev. Lett. **125**, 200501 (2020).
[15] A. Neven, J. Carrasco, V. Vitale, C. Kokail, A. Elben, M. Dalmonte, P. Calabrese, P. Zoller, B. Vermersch, R. Kueng, and B. Kraus, Symmetry-resolved entanglement detection using partial transpose moments, npj Quantum Inf. **7**, 152 (2021).
[16] A. Rath, C. Branciard, A. Minguzzi, and B. Vermersch, Quantum Fisher information from randomized measurements, Phys. Rev. Lett. **127**, 260501 (2021).
[17] R. J. Garcia, Y. Zhou, and A. Jaffe, Quantum scrambling with classical shadows, Phys. Rev. Res. **3**, 033155 (2021).
[18] G. Li, Z. Song, and X. Wang, VSQL: Variational shadow quantum learning for classification, in *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI Press, Washington, DC, 2021), Vol. 35, pp. 8357–8365.
[19] G. Boyd and B. Koczor, Training variational quantum circuits with CoVaR: Covariance root finding with classical shadows, Phys. Rev. X **12**, 041022 (2022).
[20] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn, Avoiding barren plateaus using classical shadows, PRX Quantum **3**, 020365 (2022).
[21] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, Nature (London) **603**, 416 (2022).
[22] G. I. Struchalin, Y. A. Zagorovskii, E. V. Kovlakov, S. S. Straupe, and S. P. Kulik, Experimental estimation of quantum state properties from classical shadows, PRX Quantum **2**, 010307 (2021).
[23] T. Zhang, J. Sun, X.-X. Fang, X.-M. Zhang, X. Yuan, and H. Lu, Experimental quantum state measurement with classical shadows, Phys. Rev. Lett. **127**, 200501 (2021).
[24] J. M. Lukens, K. J. H. Law, and R. S. Bennink, A Bayesian analysis of classical shadows, npj Quantum Inf. **7**, 113 (2021).
[25] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, Exponential separations between learning with and without quantum memory, *Proceedings of the 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Computer Society, Los Alamitos, CA, 2022), p. 574.
[26] A. Acharya, S. Saha, and A. M. Sengupta, Informationally complete POVM-based shadow tomography, arXiv:2105.05992.
[27] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, PRX Quantum **2**, 030348 (2021).
[28] D. E. Koh and S. Grewal, Classical shadows with noise, Quantum **6**, 776 (2022).
[29] S. T. Flammia, Averaged circuit eigenvalue sampling, Proc. TQC 2022 **232**, 4:1 (2022).
[30] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic partial tomography via classical shadows, Phys. Rev. Lett. **127**, 110504 (2021).
[31] D. Grier, H. Pashayan, and L. Schaeffer, Sample-optimal classical shadows for pure states, Quantum **8**, 1373 (2024).
[32] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, Phys. Rev. Lett. **127**, 030503 (2021).
[33] J. Helsen and M. Walter, Thrifty shadow estimation: Reusing quantum circuits and bounding tails, Phys. Rev. Lett. **131**, 240602 (2023).
[34] Y. Zhou and Q. Liu, Performance analysis of multi-shot shadow estimation, Quantum **7**, 1044 (2023).
[35] A. A Akhtar, H.-Y. Hu, and Y.-Z. You, Scalable and flexible classical shadow tomography with tensor networks, Quantum **7**, 1026 (2023).
[36] M. Arienzo, M. Heinrich, I. Roth, and M. Kliesch, Closed-form analytic expressions for shadow estimation with brickwork circuits, QIC **23**, 961 (2023).
[37] K. Bu, D. E. Koh, R. J. Garcia, and A. Jaffe, Classical shadows with Pauli-invariant unitary ensembles, npj Quantum Inf. **10**, 6 (2024).
[38] H.-Y. Hu, S. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, arXiv:2107.04817.
[39] H.-Y. Hu and Y.-Z. You, Hamiltonian-driven shadow tomography of quantum states, Phys. Rev. Res. **4**, 013054 (2022).
[40] M. C. Tran, D. K. Mark, W. W. Ho, and S. Choi, Measuring arbitrary physical properties in analog quantum simulation, Phys. Rev. X **13**, 011049 (2023).
[41] M. McGinley and M. Fava, Shadow tomography from emergent state designs in analog quantum simulators, Phys. Rev. Lett. **131**, 160601 (2023).
[42] A. Elben, R. Kueng, H.-Y. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermersch, Mixed-state entanglement from local randomized measurements, Phys. Rev. Lett. **125**, 200501 (2020).
[43] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.133.020602, which also contains Refs. [44–51] for proofs of the theorems stated here as well as additional details.

[44] N. Hunter-Jones, Unitary designs from statistical mechanics in random quantum circuits, arXiv:1905.12053.
[45] M. Sanz, Tensor networks in condensed matter, Ph.D. thesis, Technische Universität München, 2011.
[46] L. Grippo and M. Sciandrone, On the convergence of the block nonlinear Gauss-Seidel method under convex constraints, Operations Research Letters **26**, 127 (2000).
[47] A. Nahum, S. Vijay, and J. Haah, Operator spreading in random unitary circuits, Phys. Rev. X **8**, 021014 (2018).
[48] T. Zhou and A. Nahum, Emergent statistical mechanics of entanglement in random unitary circuits, Phys. Rev. B **99**, 174205 (2019).
[49] H. Zhu, Multiqubit Clifford groups are unitary 3-designs, Phys. Rev. A **96**, 062336 (2017).
[50] P. W. Brouwer and C. W. J. Beenakker, Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems, J. Math. Phys. (N.Y.) **37**, 4904 (1996).
[51] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys. **264**, 773 (2006).
[52] Y. Guo and S. Yang, Quantum error mitigation via matrix product operators, PRX Quantum **3**, 040313 (2022).
[53] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits anticoncentrate in log depth, PRX Quantum **3**, 010333 (2022).
[54] B. Barak, C.-N. Chou, and X. Gao, Spoofing linear cross-entropy benchmarking in shallow quantum circuits, arXiv:2005.02421.
[55] C. Bertoni, J. Haferkamp, M. Hinsche, M. Ioannou, J. Eisert, and H. Pashayan, Shallow-shadows, GitHub repository, 2023, https://github.com/christianbertoni/shallow-shadows.
[56] Jonathan Kunjummen, Minh C. Tran, Daniel Carney, and Jacob M. Taylor, Shadow process tomography of quantum channels, Phys. Rev. A **107**, 042403 (2023).
[57] Ryan Levy, Di Luo, and Bryan K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, Phys. Rev. Res. **6**, 013029 (2024).