Enhanced Estimation of Quantum Properties with Common Randomized Measurements

Benoît Vermersch,[1,2,3,*] Aniket Rath,[1] Bharathan Sundar,[4] Cyril Branciard,[5] John Preskill,[4,6,7,8] and Andreas Elben[4,6,†]

[1] *Université Grenoble Alpes, CNRS, LPMMC, 38000 Grenoble, France*
[2] *Institute for Theoretical Physics, University of Innsbruck, 6020 Innsbruck, Austria*
[3] *Institute for Quantum Optics and Quantum Information of the Austrian Academy of Sciences, 6020 Innsbruck, Austria*
[4] *Institute for Quantum Information and Matter, Caltech, Pasadena, California 91125, USA*
[5] *Université Grenoble Alpes, CNRS, Grenoble INP, Institut Néel, 38000 Grenoble, France*
[6] *Walter Burke Institute for Theoretical Physics, Caltech, Pasadena, California 91125, USA*
[7] *Department of Computing and Mathematical Sciences, Caltech, Pasadena, California 91125, USA*
[8] *AWS Center for Quantum Computing, Pasadena, California 91125, USA*

(Received 3 May 2023; accepted 22 January 2024; published 28 March 2024)

We present a technique for enhancing the estimation of quantum state properties by incorporating approximate prior knowledge about the quantum state. This consists in performing randomized measurements on a quantum processor and comparing the results with those obtained from a classical computer that stores an approximation of the quantum state. We provide unbiased estimators for expectation values of multicopy observables and present performance guarantees in terms of variance bounds that depend on the prior knowledge accuracy. We demonstrate the effectiveness of our approach through experimental and numerical examples detecting mixed-state entanglement, and estimating polynomial approximations of the von Neumann entropy and state fidelities.

DOI: 10.1103/PRXQuantum.5.010352

I. INTRODUCTION

Classical shadows [1] are a key element in the randomized measurement (RM) toolbox [2–11]. Previous RM protocols [2,5,12,13] focused on estimating quantum state properties expressible as polynomial functions of a density matrix ρ. Classical shadows enable efficient access to the expectation values $\mathrm{Tr}(O\rho)$ of few-body observables O. This is particularly important for the variational quantum eigensolver algorithm, which typically requires the measurement of a local Hamiltonian [14,15]. More generally, classical shadows provide access to multicopy observables (MCOs) $\mathrm{Tr}(O\rho^{\otimes n})$ $(n \geq 1)$. Many physical properties, such as Rényi entropies and partial-transpose moments related to a mixed state and entanglement, can be represented as MCOs [16–19]. MCOs also yield bounds on the quantum Fisher information [20–23] and other entanglement detection quantities [24–26] and naturally connect to error mitigation [27,28]. With use of RMs, MCOs have been measured in various recent experiments [8,12,16,17,21,25,28–33] (see also Refs. [10,11]).

A central question for RMs concerns minimizing the number of measurements required to maintain statistical errors at a certain level. While numerous studies have addressed statistical error reduction in classical shadows for single-copy observables [34–39], optimized methods for reducing statistical errors are especially vital for general MCOs, where the required number of measurements typically scales exponentially with (sub)system size [10]. In this work, we propose a framework for enhancing estimations, i.e., reducing statistical errors, for general MCOs by incorporating approximate knowledge of the quantum state of interest. This is relevant for estimating linear $(n = 1)$ and nonlinear $(n > 1)$ observables with reduced statistical errors.

Our approach is based on *common random numbers* [40]. Suppose we aim to estimate the expectation value $\mathbb{E}[X]$ of a random variable X. If we estimate $\mathbb{E}[X]$ by averaging over multiple samples X_i, the statistical error is quantified by the variance $\mathbb{V}[X]$. Now assume that we

*benoit.vermersch@lpmmc.cnrs.fr
†aelben@caltech.edu

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

have access to a random variable Y, strongly correlated with X [41], whose average value $\mathbb{E}[Y]$ is known. We can estimate $\mathbb{E}[X]$ with reduced variance $\mathbb{V}[X-Y] < \mathbb{V}[X]$ by averaging the random variable $X - Y + \mathbb{E}[Y]$ over *commonly* sampled variables X_i and Y_i.

Here we introduce *common randomized measurements* (CRMs). Our starting point is (standard) RMs that have been experimentally performed on a quantum state ρ [10]. To enhance the estimation of (multicopy) observables, we use (approximate) knowledge of the experimental state ρ, provided in the form of a classically representable approximation σ, during the classical postprocessing stage. Here σ can be derived from approximate theoretical modeling of the experiment or from data obtained in companion experiments. CRMs are realized by simulating classically RMs on σ using the same random unitaries as applied in the experiment. If ρ and σ are sufficiently close, the results of experimentally realized (on ρ) and simulated (on σ) RMs will be strongly correlated. Then we can construct powerful CRM estimators for MCOs with reduced statistical error compared with the "standard" classical shadow approach. Our protocol is presented schematically in Fig. 1, together with an illustration for entanglement detection using the experimental data from Ref. [29]; see below for details.

To demonstrate the power of our approach, we present analytical variance bounds based on combining results on MCO [22,25] and multishot [28,42,43] shadow estimations, as well as two numerical examples. We note that the approximated state σ is typically available in noisy intermediate scale quantum devices via approximate classical simulations. As shown below, σ can, for instance, be obtained from tensor network simulations; see also, e.g., Ref. [44] in the context of simulating large quantum computers.

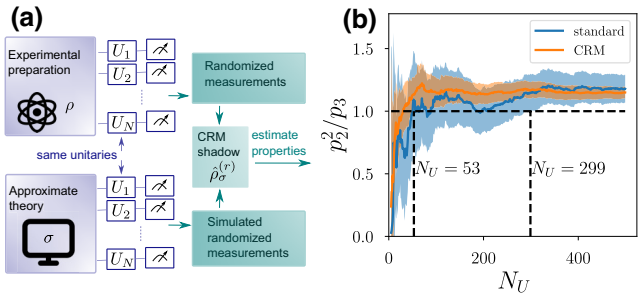

FIG. 1. Common randomized measurements. (a) Experimental data from RMs on a quantum state ρ are combined with data from simulated RMs on classical approximations σ to form CRM shadows $\hat{\rho}_\sigma^{(r)}$ for enhanced property prediction. (b) Enhanced entanglement detection via the p_3 positive partial transpose condition (p_3-PPT) condition [16] in a many-body quantum state using experimental data from Ref. [29] (with quench time $t = 1$ ms). A value greater than unity indicates bipartite entanglement between two partitions $A = [1, 2]$ and $B = [3, 4]$ in a total system with ten qubits. This can be achieved with a statistical accuracy of 1 standard deviation (shaded areas) with only $N_U \approx 50$ random unitaries with use of CRM shadows compared with $N_U \approx 300$ with use of standard classical shadows (vertical dashed lines).

II. THE COMMON RANDOMIZED MEASUREMENT FRAMEWORK

In this section, we present our formalism. We start by reviewing randomized measurements and classical shadows, and then move on to the idea of common randomized measurements. Finally, we discuss statistical errors, in comparison with statistical errors in previous methods.

A. Background on randomized measurements and classical shadows

Classical shadows [1] are classical snapshots of a quantum state that can be constructed efficiently from the experimental data acquired through RMs [10]. For concreteness, we consider a system of N qubits described by a density matrix ρ. RMs are generated by application of a random unitary U on ρ, sampled from a suitable ensemble (specified below). After application of the unitary U, a projective measurement on the rotated state $U\rho U^\dagger$ is performed in the computational basis $|\mathbf{s}\rangle = |s_1, \ldots, s_N\rangle$, with $s_i \in 0, 1$. We assume that a total of $N_U N_M$ such RMs are performed, with N_U denoting the number of sampled random unitaries $U^{(r)}$ and N_M representing the number of projective measurements per random unitary. The measurement data thus consist of $N_U N_M$ bitstrings, which we label as $\mathbf{s}^{(r,b)} = (s_1^{(r,b)}, \ldots, s_N^{(r,b)})$, for $r = 1, \ldots, N_U$ and $b = 1, \ldots, N_M$.

One can then construct N_U "standard" classical shadows

$$\hat{\rho}^{(r)} = \sum_{\mathbf{s}} \widehat{P}_\rho(\mathbf{s}|U^{(r)}) \mathcal{M}^{-1}\left(U^{(r)\dagger}|\mathbf{s}\rangle\langle\mathbf{s}|U^{(r)}\right), \quad (1)$$

with $r = 1, \ldots, N_U$ and $\widehat{P}_\rho(\mathbf{s}|U^{(r)}) = \sum_b \delta_{\mathbf{s},\mathbf{s}^{(r,b)}}/N_M$ denoting the *experimentally estimated outcome probabilities* of computational basis measurements performed on $U^{(r)}\rho U^{(r)\dagger}$. The inverse shadow channel $\mathcal{M}^{-1}$ is constructed such that, given the distribution of U, $\hat{\rho}^{(r)}$ is an unbiased estimator of ρ, i.e., $\mathbb{E}[\hat{\rho}^{(r)}] = \mathbb{E}_U \mathbb{E}_{\mathrm{QM}}[\hat{\rho}^{(r)}] = \rho$ [1]. Here $\mathbb{E}_U$ denotes the average over U and $\mathbb{E}_{QM}$ denotes the quantum mechanical expectation value (for a given U). While our construction of CRM shadows applies to any type of RM setting, we consider in the following examples Pauli measurements using $U = \bigotimes_{i=1}^N U_i$ where each U_i is uniformly sampled in $\left\{\mathbb{1}_2, 1/\sqrt{2}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, 1/\sqrt{2}\begin{pmatrix}1 & -i\\ 1 & +i\end{pmatrix}\right\}$, so that $U_i^\dagger Z U_i = Z, X, Y$, respectively (with Z, X, and Y being the Pauli matrices). The corresponding inverse channel $\mathcal{M}^{-1}(\bigotimes_i O_i) = \bigotimes_i (3O_i - \mathrm{Tr}(O_i)\mathbb{1}_2)$ [1].

B. Common randomized measurements

Our central idea is to construct classical shadows that incorporate (approximate) knowledge of the state ρ in the form of some classically representable approximation σ. We assume that σ is Hermitian but not necessarily positive semidefinite or trace one and call it a "pseudostate" for this reason. We propose building CRM shadows as

$$\hat{\rho}_\sigma^{(r)} = \hat{\rho}^{(r)} - \sigma^{(r)} + \sigma, \tag{2}$$

where $\sigma^{(r)}$ is constructed from σ as

$$\sigma^{(r)} = \sum_{\mathbf{s}} P_\sigma(\mathbf{s}|U^{(r)}) \mathcal{M}^{-1}\left(U^{(r)\dagger}|\mathbf{s}\rangle\langle\mathbf{s}|U^{(r)}\right), \tag{3}$$

with $P_\sigma(\mathbf{s}|U^{(r)}) = \langle\mathbf{s}|U^{(r)}\sigma U^{(r)\dagger}|\mathbf{s}\rangle$ being the *exact theoretical outcome probabilities* of (fictitious) computational basis measurements on the pseudostate $U^{(r)}\sigma U^{(r)\dagger}$—i.e., after σ is rotated by the same unitary $U^{(r)}$ that has been applied in the experiment. Using the definition of the inverse shadow channel [1], we find $\mathbb{E}[\sigma^{(r)}] = \mathbb{E}_U[\sigma^{(r)}] = \sigma$. Thus, $\hat{\rho}_\sigma^{(r)}$ is an unbiased estimator of ρ, as $\mathbb{E}[\hat{\rho}_\sigma^{(r)}] = \rho - \sigma + \sigma = \rho$, irrespective of the choice of σ. Crucially, the data acquisition is independent of σ, which enters only during postprocessing. In particular, an optimal σ can be chosen *after* the experiment, for instance, if a new or more accurate theoretical modeling of the experiment becomes available.

The power of CRM shadows can be intuitively understood in the limit $N_M \gg 1$. Then, if $\rho \approx \sigma$, $\hat{\rho}^{(r)} \approx \rho^{(r)}$ and $\sigma^{(r)}$ are strongly positively correlated since they share a common source of randomness (the matrix elements of the random unitary $U^{(r)}$). Consequently, the variances of the matrix elements of $\hat{\rho}^{(r)} - \sigma^{(r)}$ are smaller than those of $\hat{\rho}^{(r)}$. Below, we turn this intuition into rigorous performance guarantees.

We note that constructing $\sigma^{(r)}$ incurs overhead in terms of postprocessing compared with standard shadow estimations. However, as we show below, this step can be efficiently executed (in terms of both time and memory) using suitable representations, such as tensor networks [45]. Moreover, instead of using a theoretical state σ, one can build σ from classical shadows obtained from a companion experiment that produces a state σ close to ρ. This idea is presented in Appendix A, where we present expressions for CRM shadows that are built from the data associated with both ρ and σ and that allow unbiased MCO estimations for ρ.

C. Estimation of Pauli observables

We first consider estimators $\hat{O} = 1/N_U \sum_{r=1}^{N_U} \mathrm{Tr}\left(O\hat{\rho}_\sigma^{(r)}\right)$ of expectation values $\mathrm{Tr}(O\rho)$ of (single-copy) Pauli observables $O = \bigotimes_{i=1}^N O_i$ where each $O_i \in \{\mathbb{1}_2, X, Y, Z\}$ is a Pauli matrix. As shown in Appendix C, we find for the variance of $\hat{O}$,

$$\mathbb{V}[\hat{O}] \leq \frac{3^{N_A}}{N_U}\left(\mathrm{Tr}[O(\rho-\sigma)]^2 + \frac{1}{N_M}\right), \tag{4}$$

where N_A denotes the size of the support of O (i.e., of the set A of qubits i where $O_i \neq \mathbb{1}_2$). With standard shadows, the same expression applies after replacement of σ by 0, and our bound is consistent with Theorem 2 in Ref. [42]. This result demonstrates the power of CRMs: statistical errors in estimations with classical shadows originate from the finite values of N_U and N_M. With CRMs, we can significantly decrease the former such that, for any value of N_M, the variance given by CRM shadows is smaller than the variance given by standard shadows if $|\mathrm{Tr}[O(\rho-\sigma)]| \leq |\mathrm{Tr}(O\rho)|$. The fact that CRM shadows are useful to reduce the variance associated with finite N_U is highly relevant in experiments with significant calibration times such as experiments involving trapped ions [29] or superconducting qubits [32]. Here N_U is limited, while the value of N_M can typically be taken to be large.

D. Estimation of multicopy observables

Expectation values $\mathrm{Tr}(O\rho^{\otimes n})$ of MCOs can be estimated with (CRM) shadows with use of U statistics [1,16]. Here we use the method of "batch shadows" [25], which reduces the data processing time: For an integer $m \geq n$, m batch shadows $\hat{\rho}_\sigma^{[t]}$, $t = 1, \ldots, m$, are formed by averaging m distinct groups of N_U/m shadows $\hat{\rho}_\sigma^{(r)}$ (see Appendix B). We then define an estimator $\hat{O}$ of $\mathrm{Tr}(O\rho^{\otimes n})$ as

$$\hat{O} = \frac{(m-n)!}{m!} \sum_{t_1 \neq \cdots \neq t_n} \mathrm{Tr}\left[O\left(\hat{\rho}_\sigma^{[t_1]} \otimes \cdots \otimes \hat{\rho}_\sigma^{[t_n]}\right)\right]. \tag{5}$$

Since the batch shadows $\hat{\rho}_\sigma^{[t_i]}$ are statistically independent, and $\mathbb{E}[\hat{\rho}_\sigma^{[t_i]}] = \rho$, $\mathbb{E}[\hat{O}] = \mathrm{Tr}(O\rho^{\otimes n})$. As derived in Appendix C, the variance of $\hat{O}$ is bounded by

$$\mathbb{V}[\hat{O}] \leq \frac{n^2 ||O_A^{(1)}||_2^2}{N_U}\left(3^{N_A}||\rho_A - \sigma_A||_2^2 + \frac{2^{N_A}}{N_M}\right) + \mathcal{O}\left(\frac{1}{N_U^2}\right), \tag{6}$$

where $||\cdot||_2 = \sqrt{\mathrm{Tr}[(\cdot)^2]}$ is the Hilbert-Schmidt norm and the support $A = \mathrm{supp}(O)$ of O denotes a subset of N_A qubits on which the MCO O acts nontrivially in at least one of the copies. Also, $\rho_A = \mathrm{Tr}_{\bar{A}}(\rho)$ and $\sigma_A = \mathrm{Tr}_{\bar{A}}(\sigma)$, where $\bar{A}$ is the complementary subset to A, are reduced density matrices and $O_A^{(1)}$ is an operator that acts on (n copies of) A while depending on O and in general on ρ. This represents a key result of our work: provided that $||\rho_A - \sigma_A||_2^2 \ll ||\rho_A||_2^2$ and $N_M \gg (2/3)^{N_A}||\rho_A||_2^{-2}$, the required number of unitaries N_U is significantly reduced

compared with the number required with standard shadows [Eq. (6) with $\sigma \to 0$]. Finally, we note that Eq. (6) is independent of m and hence also applies to the case of the "original" multicopy estimators [1,16], obtained with $m = N_U$; see Appendixes B and C.

III. APPLICATIONS OF COMMON RANDOMIZED MEASUREMENTS

We now present various applications of CRMs that will illustrate their practical utility for accessing quantities with reduced measurement effort.

A. Enhanced experimental entanglement detection

We first illustrate the practical aspects of our method using existing experimental data [29]. In the related experiments, quantum quench dynamics were studied in a ten-qubit trapped-ion quantum simulator. Randomized measurements, implemented with $N_U = 500$ random unitaires and $N_M = 150$ measurements per unitary, were used to detect the generation of bipartite entanglement during the dynamics [16,29].

To model this experiment, unitary dynamics simulations providing some state $\sigma = |\psi\rangle\langle\psi|$ are available [29], and can be used here to construct CRM shadows. In the following we show that even though these simulations are approximate because the experimental state is necessarily affected by decoherence, the corresponding CRM shadows allow entanglement detection with far fewer randomized measurements compared with standard shadows. We focus on the "p_3 positive partial transpose condition (p_3-PPT)" bipartite entanglement condition, $p_2^2/p_3 > 1$ [16], where $p_n = \mathrm{Tr}([\rho_{AB}^{T_A}]^n)$ are moments of the partial transpose of the density matrix ρ_{AB} on two subsystems A and B. The moments p_n can be written as MCOs on n copies [46,47] and can be estimated with classical shadows [16]. To demonstrate the power of CRM shadows, we consider only a reduced dataset consisting of data generated in an experiment and classical simulation with the same N_U random unitaries ($N_M = 150$). In Fig. 1(b) we show the corresponding CRM estimates for p_2^2/p_3 as a function of N_U. Error bars are estimated via jackknife resampling [29] and are drawn at 1 standard deviation ε. Extracting the value of N_U such that $p_2^2/p_3 - \varepsilon > 1$, we see that CRM shadows allow us to certify the presence of entanglement within 1 standard deviation for $N_U = 53$, significantly fewer than with standard classical shadows, $N_U = 299$. In practical terms, this means that the required number of measurements to certify entanglement is divided in this case by approximately 6 (which represents saving hours in data acquisition).

B. Measuring the von Neumann entropy

To further illustrate the power of CRM shadows in the context of quantum simulation, we consider now the estimation of polynomial approximations of the von Neumann entropy $S = -\mathrm{Tr}(\rho_A \log \rho_A)$ of a subsystem A of N_A qubits, using trace moments $\mathcal{P}_n = \mathrm{Tr}[\rho_A^n]$. The von Neumann entropy is an entanglement measure [48] and can be used to distinguish quantum phases and transitions in many-body quantum systems [49]. To obtain a polynomial approximation of S, we rewrite $S = -\sum_\lambda \lambda \log \lambda$ expressed by the eigenvalues λ of ρ_A and perform a least-squares function approximation of $f(x) = -x\log(x)$ on in the interval $x \in (0,1)$ using polynomials of the type $f_{n_{\max}}(x) = \sum_{n=1}^{n_{\max}} a_n x^n$. For $n_{\max} = 3$, we obtain, for instance, $f_3(x) = 137x/60 - 4x^2 + 7x^3/4$. We then build

$$S_{n_{\max}} = \mathrm{Tr}\left[f_{n_{\max}}(\rho_A)\right] = \sum_{n=1}^{n_{\max}} a_n \mathcal{P}_n. \tag{7}$$

In Appendix D, we present the analytical expressions for $f_{n_{\max}}$ and present an upper bound for the convergence error $|S_{n_{\max}} - S|$. We note that for the quantum states considered below as an illustration, our fitting procedure provides more accurate approximations $S_{n_{\max}}$ compared with other polynomial interpolations of the same order [50].

To estimate $S_{n_{\max}}$, we rewrite each $\mathcal{P}_n$ as an expectation value of an MCO [51], namely, $\mathcal{P}_n = \mathrm{Tr}(\tau_A^{(n)} \rho_A^{\otimes n})$, with the n-copy MCO acting as $\tau_A^{(n)} |\mathbf{s}_A^{(1)}\rangle \cdots |\mathbf{s}_A^{(n)}\rangle = |\mathbf{s}_A^{(n)}\rangle |\mathbf{s}_A^{(1)}\rangle \cdots |\mathbf{s}_A^{(n-1)}\rangle$, and use the batch shadow estimator [Eq. (5)] with $m = n_{\max}$ batches. As shown in Appendix E, the variance bound Eq. (6) for estimating $\mathcal{P}_n$ is evaluated to $O_A^{(1)} = \rho_A^{n-1}$.

As an illustration, we consider the ground state $|G\rangle$ of the critical Ising chain $H = -\sum_{i=1}^N Z_i Z_{i+1} + X_i$ of length N (Z_i and X_i are Pauli matrices at sites $i = 1, \ldots, N$, and $Z_{N+1} = 0$). Since we consider the model at a critical point, the entanglement entropy S of the reduced density matrix $\rho_A = \mathrm{Tr}_{N/2+1,\ldots,N}(|G\rangle\langle G|)$ of the half partition (with $N_A = N/2$ qubits) grows as $S = c/6 \log(N_A) + \mathrm{const}$, where the central charge $c = 1/2$ characterizes the transition's universality class [52]. In Fig. 2(a), we represent $S_{n_{\max}}$ as a function of $N_A = N/2$ for different values of $n_{\max}$. Here $|G\rangle$ is calculated from the density matrix renormalization group algorithm [53]. Already for $n_{\max} = 3$, we observe the characteristic logarithmic scaling with N_A [52], while $n_{\max} = 5$ and $n_{\max} = 7$ provide quantitative agreements with S.

We now numerically simulate a measurement of $S_{n_{\max}}$ with "standard" classical and CRM shadows. In our simulations, the N-qubit ground state $|G\rangle$ is expressed with a matrix product state (MPS) [53] of large bond dimension $\chi_G \approx 40$. We then obtain MPS approximations $|\psi_\chi\rangle$ by truncating $|G\rangle$ to much smaller bond dimensions $\chi = 1, 2, 3$. The corresponding reduced state σ of the first N qubits is a matrix product operator of bond dimension χ_G^2 [53]. As χ increases, σ converges to ρ, where we expect the optimal performances for CRM shadows.

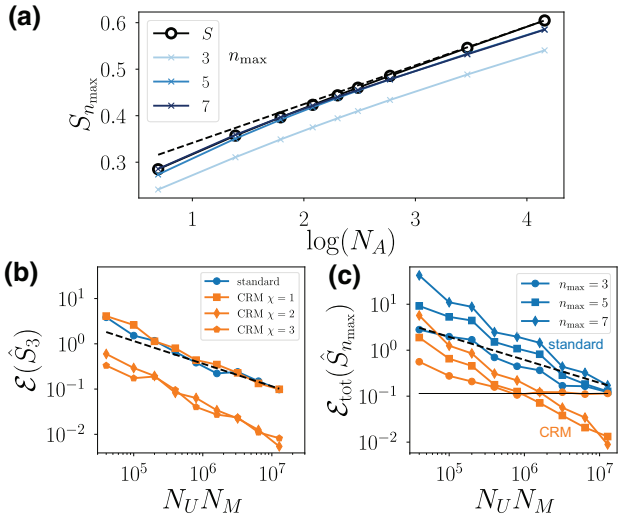

FIG. 2. Estimation of the von Neumann entropy in the critical Ising chain. (a) Approximations $S_{n_{\max}}$ as a function of subsystem size $N_A = N/2$ for $n_{\max} = 3, 5, 7$ compared with S. The dashed line is a guide for the eye proportional to $c/6 \log(N_A)$, with $c = 1/2$, the central charge of the Ising universality class [52]. (b) Statistical relative error $\mathcal{E}(\hat{S}_3) = \mathbb{E}[|\hat{S}_3 - S_3|]/S_3$ of estimations of S_3, obtained with use of the standard batch shadow estimation [1,25] (blue) and the CRM method (orange) with σ obtained from MPS approximations $|\psi_\chi\rangle$ of bond dimensions $\chi = 1, 2, 3$. (c) Total relative error $\mathcal{E}_{\text{tot}}(\hat{S}_{n_{\max}}) = \mathbb{E}[|\hat{S}_{n_{\max}} - S|]/S$ for $n_{\max} = 3, 5, 7$, and $\chi = 2$. The horizontal line denotes $|S_3 - S|/S$. In (b),(c), we use $N_A = N/2 = 8$, $N_M = 1000$ is fixed, and we vary N_U. The dashed black lines are guides for the eye proportional $1/\sqrt{N_U N_M}$. The average errors are obtained by our averaging over 20 simulations for (b) and 50 simulations for (c).

In Figs. 2(b) and 2(c) we show the relative statistical error $\mathcal{E}(\hat{S}_{n_{\max}}) = \mathbb{E}[|\hat{S}_{n_{\max}} - S_{n_{\max}}|]/S_{n_{\max}}$ as a function of $N_U N_M$ for $n_{\max} = 3, 5, 7$. We choose $N_A = N/2 = 8$, use $N_M = 1000$, and vary N_U. In Fig. 2(b), we study the behavior of $\mathcal{E}(\hat{S}_3)$ for $\chi = 1, 2, 3$. For $\chi = 1$, the approximation σ corresponds to a product state, which is too inaccurate to obtain any improvement with CRM shadows over standard shadows. For $\chi = 2, 3$, the approximation σ_χ is sufficiently accurate to significantly decrease the statistical errors. In Fig. 2(c), we study the *total* relative error $\mathcal{E}_{\text{tot}}(\hat{S}_{n_{\max}}) = \mathbb{E}[|\hat{S}_{n_{\max}} - S|]/S$. This error includes statistical errors in estimating $S_{n_{\max}}$, but also the systematic error $|S_{n_{\max}} - S|$. For small values of $N_U N_M$, where statistical errors dominate, the error increases with increasing $n_{\max}$ [which we attribute to the prefactor n^2 in the variance bound Eq. (6)]. For large numbers of measurements, the error saturates to the systematic error $|S_{n_{\max}} - S|/S$ (visualized here only for $n_{\max} = 3$, black line), and it becomes more advantageous to use larger values of $n_{\max}$.

C. Fidelity estimation

Finally, we discuss a third example that focuses on single-copy observables ($n = 1$); specifically, we consider direct fidelity estimation [54–56]. We aim to estimate the fidelity $\mathcal{F}_\psi = \langle\psi|\rho|\psi\rangle$ between the prepared quantum state ρ and a pure theoretical state $|\psi\rangle$, i.e., O is the projector $O = |\psi\rangle\langle\psi|$.

Our motivation for enhanced CRM fidelity estimates is twofold: Firstly, fidelity estimation allows us to certify the preparation of a quantum state within a quantum device. However, while $\mathcal{F}_\psi$ can be efficiently estimated with (standard) classical shadows constructed from *global* Clifford measurements [1], fidelity estimation can be challenging with (standard) *local* RMs due to a potential exponential scaling of the required number of measurements [1]. Secondly, fidelity estimation can also be used to identify suitable CRM priors σ for estimating other MCOs: direct inspection of Eq. (6) reveals that CRM shadows provide lower variance compared with standard shadows when $\mathcal{F}_\phi \geq 1/2$ (considering for simplicity an MCO O with full support ($N_A = N$) and a pure state prior $\sigma = |\phi\rangle\langle\phi|$).

We propose an iterative procedure to find useful priors for CRM shadows as follows: (i) Starting with a prior $\sigma = |\phi\rangle\langle\phi|$, we estimate $\mathcal{F}_\phi$ using either CRM shadows $\hat{\rho}_\sigma^{(r)}$ or standard shadows $\hat{\rho}^{(r)}$. The choice can be made during postprocessing by comparing empirical variances, as illustrated in the numerical example below. (ii) If $\mathcal{F}_\phi \leq \mathsf{F}$ falls below a specific threshold $\mathsf{F} \geq 1/2$, we define a new prior, which may involve more classical computation. We then repeat step (i). Once we have found a prior $\sigma = |\phi\rangle\langle\phi|$ characterized by a sufficiently high fidelity $\mathcal{F}_\phi$, we can perform enhanced estimations on arbitrary MCOs O. This includes fidelities $\mathcal{F}_\psi$ to any other quantum state $|\psi\rangle$. Performance guarantees are provided by Eq. (6) with the measured value of $\mathcal{F}_\phi$. Importantly, the entire iterative procedure can be conducted on a single RM dataset, as the choice of the prior σ is incorporated only during postprocessing. This is in contrast with importance sampling methods [54,55], where the choice of measurement settings for data acquisition depends on the prior.

As a numerical highlight, we consider a state ρ that is prepared with a unidimensional random circuit composed of d alternating layers of single and neighboring two-qubit Haar-random gates. Each gate is subject to local depolarization noise with probability p [57]. We then numerically simulate the randomized measurements. We consider CRM priors $\sigma = |\phi\rangle\langle\phi|$ as MPSs $|\phi\rangle$ with bond dimensions χ obtained by our truncating the exact output state of the noiseless quantum circuit. Note that with these priors, CRM estimations can be computed in poly(χ) time in the MPS formalism [53]. As χ increases, the fidelity $\mathcal{F}_\phi$ increases, but the computational cost of estimating $\widehat{\mathcal{F}}_\phi$ with CRM shadows also grows.

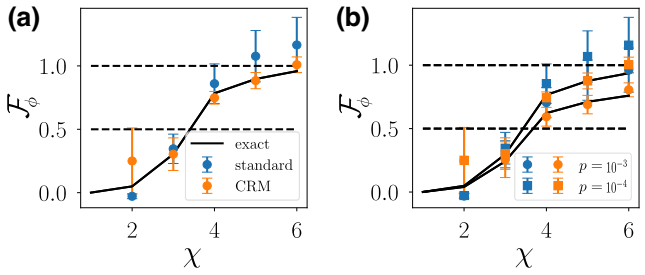

FIG. 3. Fidelity estimation in (noisy) random quantum circuits. (a) Estimated fidelities $\widehat{\mathcal{F}}_\phi$ of the prepared state ρ and the theoretical prior states $\sigma = |\phi\rangle\langle\phi|$ as a function of their bond dimension χ. Here ρ is a 30-qubit pure state generated from an ideal noiseless ($p = 0$) random quantum circuit of depth $d = 6$ and $|\phi\rangle$ are obtained by our truncating ρ to bond dimension χ. (b) Each gate in the circuit is perturbed by local depolarization noise with strength p resulting in a mixed state ρ. The prior state σ is the same as in (a). For (a),(b), we compare CRM estimation (orange dots) with standard shadow estimation (blue dots). We fix $N_U = 15$ and $N_M = 10^5$. The error bars are evaluated as standard errors of the mean over random unitaries. The solid black lines denote the exact fidelity $\mathcal{F}_\phi$. The dashed black lines are guides for the eye for 0.5 and 1 respectively.

Figure 3 shows $\widehat{\mathcal{F}}_\phi$ for a 30-qubit noiseless state [$p = 0$, Fig. 3(a)] and noisy state [$p = 10^{-3}$ and 10^{-4}, Fig. 3(b)], with error bars calculated as the standard error of the mean over random unitaries. When χ increases, the estimated fidelity $\mathcal{F}_\phi$ increases, and the error bars of the CRM estimations decrease as the CRM shadows become more accurate. For small χ, the CRM shadows fail to provide improved estimations and have larger error bars compared with (standard) classical shadows as seen in Fig. 3(a). These features are similarly observed in the case of the noisy experimental state in Fig. 3(b).

IV. CONCLUSION AND OUTLOOK

CRM shadows provide a readily applicable tool to significantly enhance the estimation of linear and multicopy observables. We envision a wide range of applications, from gradient estimation in variational quantum algorithms [58] to the probing of quantum phases of matter [59–61]. In addition, CRM shadows can be straightforwardly extended towards the estimation of properties of quantum processes [62–64].

For future work, it will be interesting to study the potential benefits of using importance sampling or adaptive techniques such as the one developed here to access the purity p_2 with RM [65], or methods using auxiliary systems [66], in addition to our method. In a broader context, it was recently shown that correlations between quantum experiments and classical simulations can be used to probe measurement-induced entanglement phase transitions [67]. It will be intriguing to combine those ideas with CRMs and investigate more generally how classical simulations can be used to enhance the learning of model-independent quantum properties.

ACKNOWLEDGMENTS

We thank Daniel K. Mark and Steve Flammia for helpful discussions and Richard Kueng and Hsin-Yuan (Robert) Huang for their valuable comments on the manuscript. Work in Grenoble is funded by the French National Research Agency via the JCJC project QRand (Grant No. ANR-20-CE47-0005) and via the research programs EPIQ (Gran No. ANR-22-PETQ-0007, Plan France 2030), and QUBITAF (Grant No. ANR-22-PETQ-0004, Plan France 2030). B.V. acknowledges funding from the Austrian Science Fund (Grant No. P 32597 N). A.R. acknowledges support from the Laboratoire d'excellence LANEF in Grenoble (Grant No. ANR-10-LABX-51-01) and from the Grenoble Nanoscience Foundation. B.S. is supported by the Caltech Summer Undergraduate Research Fellowship. J.P. acknowledges funding from the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (Grants No. DE-NA0003525 and No DE-SC0020290), the U.S. Department of Energy Quantum Systems Accelerator, and the National Science Foundation (Grant No. PHY-1733907). The Institute for Quantum Information and Matter is a National Science Foundation Physics Frontiers Center. A.E. acknowledges funding by the German National Academy of Sciences Leopoldina under Grant No. LPDS 2021-02 and by the Walter Burke Institute for Theoretical Physics at Caltech. Our JULIA code uses ITensor [68] and PastaQ [57], and is available from Ref. [69].

APPENDIX A: BUILDING CRM SHADOWS WITH A COMPANION EXPERIMENT

In the main text, we assumed that we have *a priori* access to a theoretical state σ. We now expand our method to accommodate a more state-agnostic scenario where no such theoretical description is available. Instead, we use data gathered in a companion experiment realizing a quantum state σ to measure an MCO associated with ρ. Our only requirement is that the state σ represents an approximation of the state ρ, which we can then use to construct CRM shadows for ρ. We note that if we could guarantee $\sigma = \rho$, it would likely be more efficient to simply create independent standard shadows with all available data to estimate MCOs. However, the scenario we envision is one where the companion experiment has not precisely implemented ρ. Instead, it has the ability to perform RMs on σ with a greater number of random unitaries ($N'_U \gg N_U$) than in the original experiment implementing ρ. This could be due to a faster setup or the ability to aggregate data from multiple prior experiments.

To construct CRM shadows for ρ, we proceed as follows: (i) We construct N_U classical shadows $\hat{\rho}^{(r)}$ from the original experiment with $r = 1, \ldots, N_U$. (ii) We construct N_U classical shadows $\hat{\sigma}^{(r)}$ from the companion experiment, using the same unitaries $U^{(r)}$ as in the previous step. (iii) We construct N'_U additional, independently sampled, classical shadows $\hat{\sigma}^{(r)}$, $r = N_U + 1, \ldots, N_U + N'_U$, from the companion experiment. Note that the time ordering is not important here. From the data gathered in steps (i), (ii), and (iii), respectively, we form three sets of $t = 1, \ldots, m$ batch shadows, $\hat{\rho}^{[t]}$, $\hat{\sigma}^{[t]}$, and $(\hat{\sigma}')^{[t]}$, which we combine as

$$\hat{\rho}^{[t]}_{\sigma} = \hat{\rho}^{[t]} - \hat{\sigma}^{[t]} + (\hat{\sigma}')^{[t]} \tag{A1}$$

and which satisfy the desired property $\mathbb{E}[\hat{\rho}^{[t]}_{\sigma}] = \rho$. Here $\mathbb{E}$ includes the averaging over the N'_U additional unitaries of the companion experiments for the last term in Eq. (A1). The finite value of N'_U introduces statistical errors (see the numerical examples below). In the limit $N'_U \gg 1$, the CRM shadows of Eq. (A1) become equivalent to the ones defined in Eq. (2) [because $(\hat{\sigma}')^{[t]}$ converges to σ], and the variance bound Eq. (6) applies.

In Fig. 4, we show the relative error $\mathcal{E}(\hat{S}_3)$ using CRM shadows built from the companion experiment, considering again the example of the critical Ising chain (as in Fig. 2). We use here $m = n_{\max} = 3$ batches. We consider the scenario in which the ground state $|G'\rangle$ of a Hamiltonian H' implemented in the companion experiment slightly differs from $|G\rangle$ by choosing $H' = H + \sum_i \epsilon_i Z_i$, with ϵ_i sampled independently in $[0, 0.02]$. For $N'_U = 4000$ (orange circles), we obtain significant error reduction, but we also observe a plateau effect that comes from the finite value of N'_U. With increasing N'_U (orange squares), the plateau's height is reduced, and we obtain excellent CRM

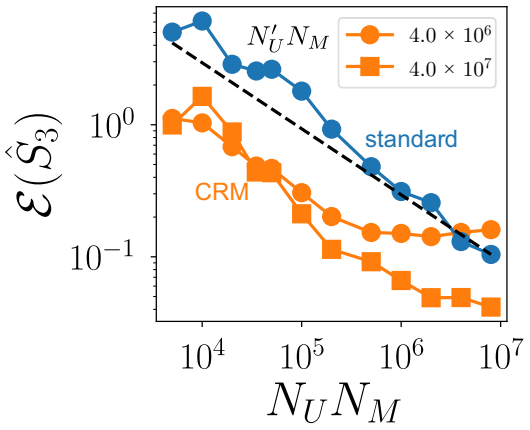

FIG. 4. Estimation of the von Neumann entropy in the critical Ising chain, (as in Fig. 2), but using CRM shadows built from a companion experiment. Statistical relative error $\mathcal{E}(\hat{S}_3)$, where the CRM shadows are formed from a companion experiment with use of N'_U additional unitaries ($N_M = 1000$), as in Eq. (A1). Here we use $N_A = 6$ ($N = 12$).

estimations compared with standard shadow estimations for all presented values of $N_U N_M$.

APPENDIX B: STATISTICAL ANALYSIS OF CRM SHADOWS

In this appendix, we provide a general variance bound for estimating (multicopy) observables with CRM shadows.

1. General variance formula

Here we recapitulate the variance of the batch shadow estimator $\hat{O}$, defined in Eq. (5), as derived in Ref. [25] and adapt it to the case of CRM shadows. The batch shadows that appear in Eq. (5) are defined as [25]

$$\hat{\rho}^{[t]}_{\sigma} = \frac{N_U}{m} \sum_{1+(t-1)(N_U/m)}^{t(N_U/m)} \hat{\rho}^{(r)}_{\sigma}, \tag{B1}$$

where the CRM shadows $\hat{\rho}^{(r)}_{\sigma}$ are defined in Eq. (2). Here $t = 1, \ldots, m$, with $m \geq n$, and we assume for simplicity that m divides N_U. An unbiased estimator of $\mathrm{Tr}[O\rho^{\otimes n}]$ is then given by the U statistic of batch shadows [70] (see also Refs. [1,16]):

$$\hat{O} = \frac{(m-n)!}{m!} \sum_{\substack{t_1 \neq t_2 \neq \cdots \neq t_n \\ t_i \in \{1,\ldots,m\}}} \mathrm{Tr}\left[O\left(\hat{\rho}^{[t_1]}_{\sigma} \otimes \hat{\rho}^{[t_2]}_{\sigma} \otimes \cdots \otimes \hat{\rho}^{[t_n]}_{\sigma}\right)\right]. \tag{B2}$$

As stated in the main text, since the batch shadows $\hat{\rho}^{[t_i]}_{\sigma}$ are statistically independent, and $\mathbb{E}[\hat{\rho}^{[t_i]}_{\sigma}] = \rho$, it follows that $\mathbb{E}[\hat{O}] = \mathrm{Tr}(O\rho^{\otimes n})$. We mention three relevant limiting cases for our analysis: For $\sigma = 0$, we recover estimations with standard (batch) shadows [25]. For $\sigma = 0$ and $m = N_U$, $\hat{O}$ coincides with the standard shadow estimator for the MCO presented in Refs. [1,16]. Finally, for linear observables $n = 1$, the notion of batch shadows becomes meaningless as the two averages in Eqs. (B1) and (B2) can be combined, i.e., the estimation does not depend on m anymore.

The variance of the batch shadow estimator $\hat{O}$ was calculated in Ref. [25] for the case of "standard classical shadows." As this derivation relies only on the condition $\mathbb{E}[\hat{\rho}^{(r)}] = \rho$, the same result applies for CRM shadows [see Eq. (C27) in Ref. [25]]:

$$\begin{aligned}\mathbb{V}[\hat{O}] &= \sum_{\ell=1}^{n} \frac{\binom{n}{\ell}^2}{\binom{m}{\ell}} \left(\frac{m}{N_U}\right)^{\ell} \left[\sum_{k=1}^{\ell} \binom{\ell}{k} (-1)^{\ell-k} \mathbb{V}_k\right] \\ &= \frac{n^2}{N_U}\mathbb{V}_1 + \frac{n^2(n-1)^2 \frac{m}{m-1}}{2N_U^2}(\mathbb{V}_2 - 2\mathbb{V}_1) + \mathcal{O}\left(\frac{1}{N_U^3}\right),\end{aligned} \tag{B3}$$

with

$$\mathbb{V}_k = \mathbb{V}\left[\mathrm{Tr}\left[O_{\mathrm{sym}}\left(\bigotimes_{r=1}^{k}\hat{\rho}_\sigma^{(r)}\otimes\rho^{\otimes(n-k)}\right)\right]\right] = \mathbb{V}\left[\mathrm{Tr}\left(O^{(k)}\bigotimes_{r=1}^{k}\hat{\rho}_\sigma^{(r)}\right)\right] \quad \text{for} \quad k=1,\dots,n, \tag{B4}$$

depending on the CRM shadows $\hat{\rho}_\sigma^{(r)}$. Here we defined a symmetrized n-copy operator $O_{\mathrm{sym}} = 1/n!\sum_{\pi\in\mathcal{S}_n} W_\pi^\dagger O W_\pi$ and its ρ-dependent partial traces $O^{(k)} = \mathrm{Tr}_{k+1,\dots,n}\left[O_{\mathrm{sym}}\left(\mathbb{1}_{2^N}^{\otimes k}\otimes\rho^{\otimes(n-k)}\right)\right]$. The operators W_π are n-copy permutation operators, with $\pi=(\pi(1),\dots,\pi(n))$, acting as $W_\pi(\otimes_i|\mathbf{s}_i\rangle) = \otimes_i|\mathbf{s}_{\pi(i)}\rangle$, and $\mathcal{S}_n$ denotes the symmetric group.

For later use, we define the support $\mathrm{supp}(O)$ of a multicopy operator O as the subpartition of the quantum system $\mathcal{S}$ on which O acts nontrivially in at least one of the copies. Then, by the definition of $O^{(k)}$, $\mathrm{supp}(O^{(k)})\subseteq\mathrm{supp}(O)$. Denoting $A\equiv\mathrm{supp}(O)$, we can factorize $O^{(k)} = O_A^{(k)}\otimes\mathbb{1}_{\bar{A}}^{\otimes k}$, with $O_A^{(k)} = \mathrm{Tr}_{k+1,\dots,n}\left[(O_A)_{\mathrm{sym}}\left(\mathbb{1}_A^{\otimes k}\otimes\rho_A^{\otimes(n-k)}\right)\right]$ and $(O_A)_{\mathrm{sym}}$ the symmetrization of the restriction O_A of O to A.

2. Leading-order term $(n^2/N_U)\mathbb{V}_1$

We now evaluate the leading-order term $(n^2/N_U)\mathbb{V}_1$ in Eq. (B3). It is dominant in the limit $N_U\to\infty$, and solely determines the variance of the estimation of standard single-copy observables [$n=1$ in Eq. (B3)] as higher-order terms in Eq. (B3) vanish. Interestingly, this term does not depend on the number of batches m, so we typically choose a minimal number $m=n$ of batch shadows to evaluate Eq. (B2), as it leads to minimal postprocessing effort: $\mathcal{O}(m^n)$ terms have to be evaluated in Eq. (B2).

Here, and in the following, we use the fact that the inverse shadow channel $\mathcal{M}^{-1}$ used to define $\hat{\rho}_\sigma^{(r)}$ is Hermitian-preserving and self-adjoint (see Ref. [1] to see that this is necessarily the case for shadows built from randomized measurements). Using this, we first evaluate

$$\begin{aligned}\mathrm{Tr}(O^{(1)}\hat{\rho}_\sigma^{(r)}) &= \sum_{\mathbf{s}}\left(\widehat{P}_\rho(\mathbf{s}|U)-P_\sigma(\mathbf{s}|U)\right)\mathrm{Tr}\left[O^{(1)}\mathcal{M}^{-1}\left(U^\dagger|\mathbf{s}\rangle\langle\mathbf{s}|U\right)\right]+\mathrm{Tr}(O^{(1)}\sigma)\\ &= \sum_{\mathbf{s}}\left(\widehat{P}_\rho(\mathbf{s}|U)-P_\sigma(\mathbf{s}|U)\right)[\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})+\mathrm{Tr}(O^{(1)}\sigma),\end{aligned} \tag{B5}$$

where we dropped the label (r) and introduced the short-hand notation $\mathrm{Tr}\left[\mathcal{M}^{-1}\left(O^{(1)}\right)U^\dagger|\mathbf{s}\rangle\langle\mathbf{s}|U\right] = [\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})$. With this, we obtain

$$\begin{aligned}\mathbb{V}_1 &= \mathbb{V}[\mathrm{Tr}(O^{(1)}\hat{\rho}_\sigma^{(r)})]\\ &= \mathbb{V}\left[\sum_{\mathbf{s}}\left(\widehat{P}_\rho(\mathbf{s}|U)-P_\sigma(\mathbf{s}|U)\right)[\mathcal{M}^{-1}(O^{(1)})](U,s)\right]\\ &= \mathbb{E}_U\left[\sum_{\mathbf{s},\mathbf{s}'}\mathbb{E}_{\mathrm{QM}}\left[\left(\widehat{P}_\rho(\mathbf{s}|U)-P_\sigma(\mathbf{s}|U)\right)\left(\widehat{P}_\rho(\mathbf{s}'|U)-P_\sigma(\mathbf{s}'|U)\right)\right][\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})[\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s}')\right]\\ &\quad-\mathrm{Tr}\left[O^{(1)}(\rho-\sigma)\right]^2.\end{aligned} \tag{B6}$$

Now we use (see, e.g., Refs. [6,71])

$$\mathbb{E}_{\mathrm{QM}}[\widehat{P}_\rho(\mathbf{s}|U)\widehat{P}_\rho(\mathbf{s}'|U)] = P_\rho(\mathbf{s}|U)P_\rho(\mathbf{s}'|U)+\frac{\delta_{\mathbf{s},\mathbf{s}'}P_\rho(\mathbf{s}|U)-P_\rho(\mathbf{s}|U)P_\rho(\mathbf{s}'|U)}{N_M}. \tag{B7}$$

Inserting this into Eq. (B6), we find

$$\mathbb{V}_1 = \mathbb{V}_U[f_{\rho,\sigma}(U)]+\frac{\mathbb{E}_U[g_\rho(U)]}{N_M}, \tag{B8}$$

with $\mathbb{V}_U$ and $\mathbb{E}_U$ denoting variance and expectation with respect to only the distribution of the unitaries U (as we

have performed the quantum mechanical average explicitly). The functions can be written explicitly as

$$f_{\rho,\sigma}(U) = \sum_{\mathbf{s}} \big(P_\rho(\mathbf{s}|U) - P_\sigma(\mathbf{s}|U)\big)[\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s}) \tag{B9}$$

and

$$g_\rho(U) = \sum_{\mathbf{s}} P_\rho(\mathbf{s}|U)\Big([\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})\Big)^2 - \Big(\sum_{\mathbf{s}} P_\rho(\mathbf{s}|U)[\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})\Big)^2. \tag{B10}$$

The N_M-independent first term in Eq. (B8) depends on both ρ and σ and vanishes for $\rho = \sigma$ ($f_{\rho,\rho} = 0$ for any ρ). It accounts for the variance of $\hat{O}$ due to a finite number N_U of random unitaries U and is present even for $N_M \to \infty$. In contrast, the N_M-dependent second term in Eq. (B8) quantifies the average quantum shot noise arising from a finite number N_M of computational basis measurements per random unitary and depends only on the experimental state ρ.

Combining these results and inserting them into Eq. (B3), we obtain

$$\mathbb{V}(\hat{O}) = \frac{n^2}{N_U}\left(\mathbb{V}_U[f_{\rho,\sigma}(U)] + \frac{\mathbb{E}_U[g_\rho(U)]}{N_M}\right) + \mathcal{O}\left(\frac{1}{N_U^2}\right). \tag{B11}$$

Although it is difficult to bound the higher-order term $\mathcal{O}(1/N_U^2)$ explicitly, we have the guarantee that this term decays to zero when $N_M \to \infty$ and $\sigma \to \rho$. In this case, the CRM shadows involved in the estimator $\hat{O}$ satisfy $\hat{\rho}^{(r)} \to \sigma^{(r)}$, and therefore $\hat{\rho}_\sigma^{(r)} \to \sigma$ becomes constant.

APPENDIX C: VARIANCE BOUNDS FOR LOCAL PAULI MEASUREMENTS

The analysis in Appendix B applies to all unitary ensembles that can be used to define classical shadows (i.e., that give rise to a tomographically complete set of measurements). Here we consider Pauli measurements realized by local random unitaries of the form $U = U_1 \otimes \cdots \otimes U_N$, where each U_i is sampled independently from the set $\mathcal{U} = \{\mathbb{1}_2, 1/\sqrt{2}\left(\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right), 1/\sqrt{2}\left(\begin{smallmatrix}1 & -i\\ 1 & +i\end{smallmatrix}\right)\}$, so that $U_i^\dagger Z U_i = Z, X, Y$, respectively. In this case, the inverse shadow channel used to define $\hat{\rho}_\sigma$ [Eq. (2)] reads as

$$\mathcal{M}^{-1}\Big(\bigotimes_i O_i\Big) = \bigotimes_i \mathcal{M}_i^{-1}(O_i), \quad \text{with}$$
$$\mathcal{M}_i^{-1}(O_i) = 3O_i - \mathbb{1}_2\mathrm{Tr}(O_i), \tag{C1}$$

for product observables $O = \bigotimes_i O_i$, and one can extend this definition to nonproduct operators by linearity [1].

1. Estimating single-copy Pauli observables

We first consider single-copy ($n = 1$) observables O. In this case, the operator $O^{(1)}$ defined below Eq. (B4) is evaluated to $O^{(1)} = O$. In addition, we specify O to be a Pauli string of the form $O = \gamma = \bigotimes_{i=1}^N \gamma_i$, with $\gamma_i \in \{\mathbb{1}_2, X, Y, Z\}$ and X, Y, Z. We denote by $A = \mathrm{supp}(O)$ the subset of N_A qubits where O acts nontrivially, i.e., $\gamma_i \neq \mathbb{1}_2$ for $i \in A$ and $\gamma_i = \mathbb{1}_2$ for $i \in \bar{A}$. We define $U_A = \bigotimes_{i\in A} U_i$.

We bound $\mathbb{V}(\hat{O})$ starting from Eq. (B11). First, it follows directly from Eq. (C1) that $\mathcal{M}^{-1}(\gamma) = 3^{N_A}\gamma$. Introducing the unitary $V_\gamma = \bigotimes_{i\in A} V_i$, with $V_i \in \mathcal{U}$ such that $\gamma_i = V_i^\dagger Z V_i$ for all $i \in A$, we can write

$$\gamma(U,\mathbf{s}) \equiv \mathrm{Tr}\left[\gamma U^\dagger |\mathbf{s}\rangle\langle\mathbf{s}| U\right] = \delta_{U_A,V_\gamma}\langle\mathbf{s}|Z_A|\mathbf{s}\rangle, \tag{C2}$$

with $Z_A = \bigotimes_i (Z\delta_{i\in A} + \mathbb{1}_2\delta_{i\notin A}) = \sum_{\mathbf{s}}(-1)^{\sum_{i\in A} s_i}|\mathbf{s}\rangle\langle\mathbf{s}|$. With these definitions, we rewrite the first term in Eq. (B11) as

$$\begin{aligned} f_{\rho,\sigma}(U) &= 3^{N_A}\sum_{\mathbf{s}}\big(P_\rho(\mathbf{s}|U) - P_\sigma(\mathbf{s}|U)\big)\gamma(U,\mathbf{s}) \\ &= 3^{N_A}\sum_{\mathbf{s}}\big(P_\rho(\mathbf{s}|U) - P_\sigma(\mathbf{s}|U)\big)\delta_{U_A,V_\gamma}\langle\mathbf{s}|Z_A|\mathbf{s}\rangle \\ &= 3^{N_A}\delta_{U_A,V_\gamma}\mathrm{Tr}\left((\rho-\sigma)U^\dagger Z_A U\right) \\ &= 3^{N_A}\delta_{U_A,V_\gamma}\mathrm{Tr}((\rho-\sigma)\gamma), \end{aligned} \tag{C3}$$

where we used the fact that Z_A commutes with each $|\mathbf{s}\rangle\langle\mathbf{s}|$. We obtain

$$\begin{aligned} \mathbb{V}_U[f_{\rho,\sigma}(U)] &= 9^{N_A}\mathrm{Tr}((\rho-\sigma)\gamma)^2\,\mathbb{V}_U[\delta_{U_A,V_\gamma}] \\ &= (3^{N_A}-1)\,\mathrm{Tr}((\rho-\sigma)\gamma)^2 \end{aligned} \tag{C4}$$

using the fact that $\mathbb{V}_U[\delta_{U_A,V_\gamma}] = \mathbb{E}_U[\delta^2_{U_A,V_\gamma}] - \mathbb{E}_U[\delta_{U_A,V_\gamma}]^2 = 1/3^{N_A} - 1/9^{N_A}$ irrespective of γ. We can proceed similarly with the second term in Eq. (B11):

$$\begin{aligned} g_\rho(U) &= \sum_{\mathbf{s}} P_\rho(\mathbf{s}|U)\mathcal{M}^{-1}(\gamma)(U,\mathbf{s})^2 \\ &\quad - \left(\sum_{\mathbf{s}} P_\rho(\mathbf{s}|U)\mathcal{M}^{-1}(\gamma)(U,\mathbf{s})\right)^2 \\ &= 9^{N_A}\left(\sum_{\mathbf{s}} P_\rho(\mathbf{s}|U)\gamma(U,\mathbf{s})^2\right) \\ &\quad - 9^{N_A}\delta_{U_A,V_\gamma}\mathrm{Tr}(\rho\gamma)^2 = 9^{N_A}\delta_{U_A,V_\gamma}(1-\mathrm{Tr}(\rho\gamma)^2), \end{aligned} \tag{C5}$$

and thus

$$E_U[g_\rho(U)] = 3^{N_A}(1-\mathrm{Tr}(\rho\gamma)^2). \tag{C6}$$

Inserting these results into Eq. (B3) and recalling that higher-orders terms $\mathcal{O}(1/N_U^2)$ are absent for linear observables, we find

$$\begin{aligned}\mathbb{V}(\hat{O}) &= \frac{1}{N_U}\left((3^{N_A}-1)\mathrm{Tr}(O(\rho-\sigma))^2 \right.\\ &\quad\left. + \frac{3^{N_A}(1-\mathrm{Tr}(\rho\gamma)^2)}{N_M}\right)\\ &\leq \frac{3^{N_A}}{N_U}\left(\mathrm{Tr}[O(\rho-\sigma)]^2 + \frac{1}{N_M}\right)\end{aligned} \tag{C7}$$

as stated in the main text.

2. Estimating general MCOs

We now bound the variance equation (B11) for a general multicopy observable O with corresponding Hermitian operator $O^{(1)}$ defined below Eq. (B4). We denote the support of O by $A = \mathrm{supp}(O) \supseteq \mathrm{supp}(O^{(1)})$, such that, up to relabeling of the qubits, we can write $O^{(1)} = O_A^{(1)} \otimes \mathbb{1}_{\bar{A}}$. We write $O_A^{(1)}$ in the basis of the 4^{N_A} Pauli strings γ_A, $O_A^{(1)} = 1/2^{N_A}\sum_{\gamma_A}\mathrm{Tr}(O_A^{(1)}\gamma_A)\gamma_A$ [where the Pauli strings are orthogonal: $\mathrm{Tr}(\gamma_A\gamma_A') = 2^{N_A}\delta_{\gamma_A,\gamma_A'}$], and first note that

$$\begin{aligned}[\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s}) &= \mathrm{Tr}\left[\mathcal{M}^{-1}(O^{(1)})U^\dagger\,|\mathbf{s}\rangle\,\langle\mathbf{s}|\,U\right]\\ &= \mathrm{Tr}\left[\mathcal{M}_A^{-1}(O_A^{(1)})U_A^\dagger\,|\mathbf{s}_A\rangle\,\langle\mathbf{s}_A|\,U_A\right]\\ &= \frac{1}{2^{N_A}}\sum_{\gamma_A}3^{N_\Gamma}\mathrm{Tr}(O_A^{(1)}\gamma_A)\gamma_A(U_A,\mathbf{s}_A),\end{aligned} \tag{C8}$$

with $\Gamma = \mathrm{supp}(\gamma_A) \subseteq A$ denoting the support of γ_A consisting of N_Γ qubits. This implies that $f_{\rho,\sigma}(U)$ depends only on reduced quantities acting on A only:

$$\begin{aligned}f_{\rho,\sigma}(U) &= \sum_{\mathbf{s}}\left(P_\rho(\mathbf{s}|U) - P_\sigma(\mathbf{s}|U)\right)[\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})\\ &= \sum_{\mathbf{s}_A}\left(P_{\rho_A}(\mathbf{s}_A|U_A) - P_{\sigma_A}(\mathbf{s}_A|U_A)\right)\\ &\quad\times\left(\frac{1}{2^{N_A}}\sum_{\gamma_A}3^{N_\Gamma}\mathrm{Tr}(O_A^{(1)}\gamma_A)\gamma_A(U_A,\mathbf{s}_A),\right)\end{aligned} \tag{C9}$$

with the reduced density matrices $\rho_A = \mathrm{Tr}_{\bar{A}}(\rho), \sigma_A = \mathrm{Tr}_{\bar{A}}(\sigma)$. We now use the Cauchy-Schwartz inequality

$$\begin{aligned}f_{\rho,\sigma}(U)^2 &\leq \left(\sum_{\mathbf{s}_A}\left[P_{\rho_A}(\mathbf{s}_A|U_A) - P_{\sigma_A}(\mathbf{s}_A|U_A)\right]^2\right)\\ &\quad\times\left(\sum_{\mathbf{s}_A}\left[\frac{1}{2^{N_A}}\sum_{\gamma_A}3^{N_\Gamma}\mathrm{Tr}(O_A^{(1)}\gamma_A)\gamma_A(U_A,\mathbf{s}_A)\right]^2\right).\end{aligned} \tag{C10}$$

The first factor can be bounded as

$$\begin{aligned}&\sum_{\mathbf{s}_A}\left(P_{\rho_A}(\mathbf{s}_A|U_A) - P_{\sigma_A}(\mathbf{s}_A|U_A)\right)^2\\ &\quad= \sum_{\mathbf{s}_A}\langle\mathbf{s}_A|U_A(\rho_A-\sigma_A)U_A^\dagger|\mathbf{s}_A\rangle^2\\ &\quad\leq \sum_{\mathbf{s}_A}\langle\mathbf{s}_A|[U_A(\rho_A-\sigma_A)U_A^\dagger]^2|\mathbf{s}_A\rangle = ||\rho_A-\sigma_A||_2^2,\end{aligned} \tag{C11}$$

where we used the fact that $\rho_A - \sigma_A$ is Hermitian. As before, we denote by $V_\gamma = \bigotimes_{i\in\Gamma}V_i$, $V_i \in \mathcal{U}$, the unitary that maps γ_i to Z_i for all $i \in \Gamma$. Further, we define $Z_\Gamma = \bigotimes_{i\in A}(Z\delta_{i\in\Gamma} + \mathbb{1}_2\delta_{i\notin\Gamma})$, such that $\gamma(U_A,\mathbf{s}_A) = \langle\mathbf{s}_A|Z_\Gamma|\mathbf{s}_A\rangle\,\delta_{U_\Gamma,V_\gamma}$, and analogously for $(\gamma_A, V_\gamma,\ Z_\Gamma) \to (\gamma_A', V_{\gamma'},\ Z_{\Gamma'})$. We then have

$$\begin{aligned}&\sum_{\mathbf{s}_A}\gamma_A(U_A,\mathbf{s}_A)\gamma_A'(U_A,\mathbf{s}_A)\\ &\quad= \sum_{\mathbf{s}_A}\langle\mathbf{s}_A|Z_\Gamma|\mathbf{s}_A\rangle\,\langle\mathbf{s}_A|Z_{\Gamma'}|\mathbf{s}_A\rangle\,\delta_{U_\Gamma,V_\gamma}\delta_{U_{\Gamma'},V_{\gamma'}}\\ &\quad= \prod_i\left(1+(-1)^{\delta_{i\in\Gamma}+\delta_{i\in\Gamma'}}\right)\delta_{U_\Gamma,V_\gamma}\delta_{U_{\Gamma'},V_{\gamma'}}\\ &\quad= \prod_i\left(2[\delta_{i\in\Gamma}\delta_{i\in\Gamma'} + \delta_{i\notin\Gamma}\delta_{i\notin\Gamma'}]\right)\delta_{U_\Gamma,V_\gamma}\delta_{U_{\Gamma'},V_{\gamma'}}\\ &\quad= 2^{N_A}\delta_{\Gamma,\Gamma'}\delta_{U_\Gamma,V_\gamma}\delta_{U_{\Gamma'},V_{\gamma'}}\\ &\quad= 2^{N_A}\delta_{\gamma_A,\gamma_A'}\delta_{U_\Gamma,V_\gamma},\end{aligned} \tag{C12}$$

where we used in the last equality the fact that two Pauli strings that have the same support $\Gamma = \Gamma'$ and that are mapped to $Z_\Gamma = Z_{\Gamma'}$ via the same transformation U_A are necessarily equal. Hence, we get

$$f_{\rho,\sigma}(U)^2 \leq \frac{1}{2^{N_A}}||\rho_A-\sigma_A||_2^2\sum_{\gamma_A}\delta_{U_\Gamma,V_\gamma}9^{N_\Gamma}\mathrm{Tr}(O_A^{(1)}\gamma_A)^2. \tag{C13}$$

With $\mathbb{E}_U[\delta_{U_\Gamma,V_\gamma}] = 1/3^{N_\Gamma}$, we obtain

$$\begin{aligned}\mathbb{V}_U[f_{\rho,\sigma}(U)^2] &\leq \mathbb{E}_U[f_{\rho,\sigma}(U)^2]\\ &\leq \frac{||\rho_A-\sigma_A||_2^2}{2^{N_A}}\sum_{\gamma_A}3^{N_\Gamma}\mathrm{Tr}(O_A^{(1)}\gamma_A)^2\\ &\leq 3^{N_A}\,\frac{||\rho_A-\sigma_A||_2^2}{2^{N_A}}\sum_{\gamma_A}\mathrm{Tr}(O_A^{(1)}\gamma_A)^2\\ &= 3^{N_A}\,||O_A^{(1)}||_2^2\,||\rho_A-\sigma_A||_2^2,\end{aligned} \tag{C14}$$

where we used the fact that $||O_A^{(1)}||_2^2 = \mathrm{Tr}([O_A^{(1)}]^2) = \sum_{\gamma_A} \mathrm{Tr}(O_A^{(1)}\gamma_A)^2/2^{N_A}$ [which can be easily proven with use of $\mathrm{Tr}(\gamma_A\gamma_A') = 2^{N_A}\delta_{\gamma_A,\gamma_A'}$].

To bound the second term in Eq. (B11), we use the previously established bounds for standard shadows (Proposition 3 in Ref. [1]):

$$\begin{aligned}\mathbb{E}_U[g_\rho(u)] &\le \mathbb{E}_U\left[\sum_{\mathbf{s}} P_\rho(\mathbf{s}|U)\Big([\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})\Big)^2\right] \\ &\le 2^{N_A}||O_A^{(1)}||_2^2, \end{aligned} \tag{C15}$$

where we used

$$\begin{aligned}&\mathbb{E}_U\left[\sum_{\mathbf{s}} P_\rho(\mathbf{s}|U)\Big([\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})\Big)^2\right] \\ &\le \max_{\sigma \text{ state}} \mathbb{E}_U\left[\sum_{\mathbf{s}} P_\sigma(\mathbf{s}|U)\Big([\mathcal{M}^{-1}(O^{(1)})](U,\mathbf{s})\Big)^2\right] \\ &= \|O^{(1)}\|_{\text{shadow}}^2 \le 2^{N_A}||O_A^{(1)}||_2^2\end{aligned}$$

with the shadow norm defined in Ref. [1] and used in Eq. (S57) in Ref. [1] to obtain a bound of the shadow norm in terms of the Hilbert-Schmidt norm.

Summarizing and inserting these results into Eq. (B3), we find

$$\mathbb{V}(\hat{O}) \le \frac{n^2||O_A^{(1)}||^2}{N_U}\left(3^{N_A}\,\|\rho_A - \sigma_A\|_2^2 + \frac{2^{N_A}}{N_M}\right) + \mathcal{O}\left(\frac{1}{N_U^2}\right). \tag{C16}$$

Note that the term $||O_A^{(1)}||_2^2$ is state dependent. In particular, for specific states and MCOs with $n > 1$, $\mathrm{supp}(O^{(1)})$ could be smaller than $A = \mathrm{supp}(O)$, i.e., $\mathrm{supp}(O^{(1)}) \subsetneq A$. In this case, we can obtain a tighter bound by replacing A (N_A) in Eqs. (C16) and (6) with $\mathrm{supp}(O^{(1)})$ [$|\mathrm{supp}(O^{(1)})|$]. Lastly, we note that for trace moments $\mathrm{Tr}(\rho_A^n)$ for which $O = \tau_A^{(n)} \otimes \mathbb{1}_{\bar{A}}^{\otimes n}$, we have, as shown in Appendix E, $O_A^{(1)} = \rho_A^{n-1}$.

APPENDIX D: POLYNOMIAL APPROXIMATIONS OF THE VON NEUMANN ENTROPY VIA LEAST-SQUARES MINIMIZATION

In this appendix, we explain how to construct the polynomial approximations $S_{n_{\max}}$ introduced in the main text. Our aim is to derive the coefficients a_n, $n = 1, \ldots, n_{\max}$, that minimize the least-squares error

$$I_{n_{\max}} = \int_0^1 [f(x) - f_{n_{\max}}(x)]^2 dx, \tag{D1}$$

where $f(x) = -x\log(x)$ and $f_{n_{\max}}(x) = \sum_{n=1}^{n_{\max}} a_n x^n$ a polynomial of degree $n_{\max}$. We find

$$I_{n_{\max}} = -\sum_{n=1}^{n_{\max}} \frac{2a_n}{(2+n)^2} + \sum_{n,n'=1}^{n_{\max}} \frac{a_n a_{n'}}{1+n+n'} + \text{const}, \tag{D2}$$

which we can differentiate as

$$\frac{\partial I_{n_{\max}}}{\partial a_n} = -\frac{2}{(2+n)^2} + 2\sum_{n'=1}^{n_{\max}} \frac{a_{n'}}{1+n+n'} = 0. \tag{D3}$$

It follows immediately that $\partial^2 I_{n_{\max}}/\partial a_n^2 > 0$. Equation (D3) corresponds to a matrix inversion problem $A.a = Y$, with $A_{n,n'} = (1+n+n')^{-1}$ and $Y_n = 1/(2+n)^2$. The inverse of a Cauchy matrix $A_{n,n'} = 1/(x_n + y_{n'})$ appears naturally in polynomial approximation problems and is found to be [72]

$$A^{-1}_{n,n'} = \frac{\prod_{k=1}^n (x_{n'} + y_k)(x_k + y_n)}{(x_{n'} + y_n)\prod_{k\neq n'}(x_{n'} - x_k)\prod_{k\neq n}(y_n - y_k)}. \tag{D4}$$

In our case, $x_n = n+1$ and $y_{n'} = n'$. We obtain

$$\begin{aligned} a_n &= \sum_{n'} A^{-1}_{n,n'} Y_{n'} \\ &= \sum_{n'}\left(\frac{\prod_{k=1}^n (n'+1+k)(k+1+n)}{(n+n'+1)\prod_{k\neq n'}(n'-k)\prod_{k\neq n}(n-k)}\right) \\ &\quad \times \frac{1}{(2+n')^2}. \end{aligned} \tag{D5}$$

Having derived explicit expressions for the coefficients a_n that determine $f_{n_{\max}}(x)$, we can quantify convergence aspects via the least-squares error $I_{n_{\max}}$, which we plot in Fig. 5.

Once we have built $f_{n_{\max}}$, we can bound the error $|S - S_{n_{\max}}|$ as follows. First, we can find numerically an upper

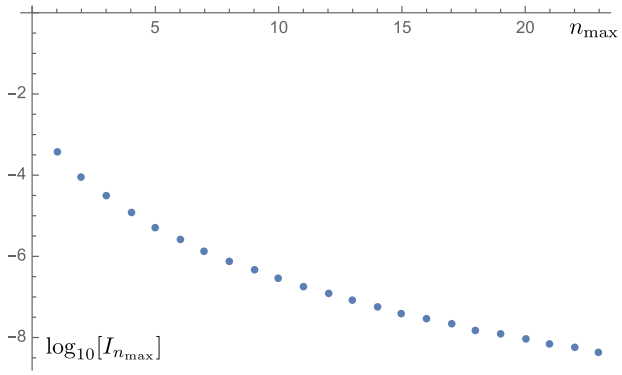

FIG. 5. Least-squares error $I_{n_{\max}}$ as a function of $n_{\max}$.

bound $\alpha_{n_{\max}}$ for the function $|f(x) - f_{n_{\max}}(x)|$ in the interval $[0, 1]$. For instance, we find $\alpha_3 \approx 0.046$, $\alpha_4 \approx 0.028$, and $\alpha_5 \approx 0.019$. Then we find

$$\begin{aligned} |S - S_{n_{\max}}| &= \left| \sum_{\lambda \in \mathrm{Spec}(\rho)} f(\lambda) - f_{n_{\max}}(\lambda) \right| \\ &\leq \sum_{\lambda \in \mathrm{Spec}(\rho)} |f(\lambda) - f_{n_{\max}}(\lambda)| \\ &= \sum_{\lambda \in \mathrm{Spec}(\rho), \lambda \neq 0} |f(\lambda) - f_{n_{\max}}(\lambda)| \leq \alpha_{n_{\max}} \mathrm{rank}(\rho), \end{aligned} \tag{D6}$$

where we used in the second line the fact that $f(0) = f_{n_{\max}}(0)$.

APPENDIX E: COMPUTING $O_A^{(1)}$ FOR A SHIFT OPERATOR $\tau_A^{(n)}$

Our aim is to calculate the linear operator $O_A^{(1)}$ for $O = \tau_A^{(n)} \otimes \mathbb{1}_{\bar{A}}^{\otimes n}$ for $\tau_A^{(n)}$ being the shift operator on n copies of A as defined in the main text. As argued in the statements following Eq. (B4), we can restrict ourselves entirely to the subsystem A. We thus drop the subscript A in the remainder of this appendix.

By the definition of $O^{(1)}$ [see below Eq. (B4)],

$$\begin{aligned} O^{(1)} &= \mathrm{Tr}_{\{2,\dots,n\}} \left(O_{\mathrm{sym}} \left[\mathbb{1}_{2^N} \otimes \rho^{\otimes(n-1)} \right] \right) \\ &= \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \mathrm{Tr}_{\{2,\dots,n\}} \left(W_\pi^\dagger \tau_n W_\pi \left[\mathbb{1}_{2^N} \otimes \rho^{\otimes(n-1)} \right] \right). \end{aligned} \tag{E1}$$

Writing now $\tau_n = \sum_{\mathbf{s}_1,\dots,\mathbf{s}_n} |\mathbf{s}_n, \mathbf{s}_1, \dots, \mathbf{s}_{n-1}\rangle\langle \mathbf{s}_1, \dots, \mathbf{s}_n|$ and $W_\pi = \sum_{\mathbf{s}'_1,\dots,\mathbf{s}'_n} |\mathbf{s}'_1, \dots, \mathbf{s}'_n\rangle\langle \mathbf{s}'_{\pi^{-1}(1)}, \dots, \mathbf{s}'_{\pi^{-1}(n)}|$, we get

$$\begin{aligned} &W_\pi^\dagger \tau_n W_\pi \\ &\quad = \sum_{\mathbf{s}_1,\dots,\mathbf{s}_n} |\mathbf{s}_{\pi^{-1}(1)-1}, \dots, \mathbf{s}_{\pi^{-1}(n)-1}\rangle\langle \mathbf{s}_{\pi^{-1}(1)}, \dots, \mathbf{s}_{\pi^{-1}(n)}|, \end{aligned} \tag{E2}$$

with $\mathbf{s}_0 \equiv \mathbf{s}_n$. Defining $j = \pi^{-1}(1)$ and noting that $\pi^{-1}(2), \dots, \pi^{-1}(n)$ give all other values $i \neq j$ from 1 to n, we get

$$\begin{aligned} &\mathrm{Tr}_{\{2,\dots,n\}} \left(W_\pi^\dagger \tau_n W_\pi [\mathbb{1}_{2^N} \otimes \rho^{\otimes(n-1)}] \right) \\ &\quad = \sum_{\mathbf{s}_1,\dots,\mathbf{s}_n} |\mathbf{s}_{j-1}\rangle\langle \mathbf{s}_j| \left(\prod_{i \neq j} \langle \mathbf{s}_i| \rho |\mathbf{s}_{i-1}\rangle \right). \end{aligned} \tag{E3}$$

Reordering all the terms (with the index i going down from $j-1$ to 1, and then from n to $j+1$), we get

$$\begin{aligned} &\mathrm{Tr}_{\{2,\dots,n\}} \left(W_\pi^\dagger \tau_n W_\pi [\mathbb{1}_{2^N} \otimes \rho^{\otimes(n-1)}] \right) \\ &\quad = \sum_{\mathbf{s}_1,\dots,\mathbf{s}_n} |\mathbf{s}_{j-1}\rangle \langle \mathbf{s}_{j-1}| \rho |\mathbf{s}_{j-2}\rangle \langle \mathbf{s}_{j-2}| \rho |\mathbf{s}_{j-3}\rangle \cdots \\ &\qquad \times \langle \mathbf{s}_1| \rho |\mathbf{s}_0\rangle \langle \mathbf{s}_n| \rho |\mathbf{s}_{n-1}\rangle \cdots \langle \mathbf{s}_{j+2}| \rho |\mathbf{s}_{j+1}\rangle \\ &\qquad \times \langle \mathbf{s}_{j+1}| \rho |\mathbf{s}_j\rangle \langle \mathbf{s}_j| = \rho^{n-1}, \end{aligned} \tag{E4}$$

where we used the sum rules $\sum_{\mathbf{s}_i} |\mathbf{s}_i\rangle\langle \mathbf{s}_i| = \mathbb{1}_{2^N}$ (recalling that $\mathbf{s}_0 \equiv \mathbf{s}_n$). Hence, after (trivially) averaging over π, we get

$$O^{(1)} = \rho^{n-1}. \tag{E5}$$

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

[2] S. J. van Enk and C. W. J. Beenakker, Measuring Tr ρ^n on single copies of ρ using random measurements, Phys. Rev. Lett. **108**, 110503 (2012).

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

[4] M. C. Tran, B. Dakić, F. Arnault, W. Laskowski, and T. Paterek, Quantum entanglement from random measurements, Phys. Rev. A **92**, 050301 (2015).

[5] 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**, 50406 (2018).

[6] 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**, 1 (2019).

[7] A. Ketterer, N. Wyderka, and O. Gühne, Characterizing multipartite entanglement with moments of random correlations, Phys. Rev. Lett. **122**, 120505 (2019).

[8] L. Knips, J. Dziewior, W. Kłobus, W. Laskowski, T. Paterek, P. J. Shadbolt, H. Weinfurter, and J. D. A. Meinecke, Multipartite entanglement analysis from random correlations, Npj Quantum Inf. **6**, 51 (2020).

[9] S. Imai, N. Wyderka, A. Ketterer, and O. Gühne, Bound entanglement from randomized measurements, Phys. Rev. Lett. **126**, 150501 (2021).

[10] 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).

[11] P. Cieśliński, S. Imai, J. Dziewior, O. Gühne, L. Knips, W. Laskowski, J. Meinecke, T. Paterek, and T. Vértesi, Analysing quantum systems with randomised measurements, arXiv:2307.01251 (2023).

[12] A. Elben, B. Vermersch, R. Van Bijnen, C. Kokail, T. Brydges, C. Maier, M. K. Joshi, R. Blatt, C. F. Roos, and

P. Zoller, Cross-platform verification of intermediate scale quantum devices, Phys. Rev. Lett. **124**, 10504 (2020).
[13] A. Elben, J. Yu, G. Zhu, M. Hafezi, F. Pollmann, P. Zoller, and B. Vermersch, Many-body topological invariants from randomized measurements in synthetic quantum matter, Sci. Adv. **6** (2020).
[14] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).
[15] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).
[16] A. Elben, R. Kueng, H.-Y. R. 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).
[17] 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).
[18] X.-D. Yu, S. Imai, and O. Gühne, Optimal entanglement certification from moments of the partial transpose, Phys. Rev. Lett. **127**, 060504 (2021).
[19] J. Carrasco, M. Votto, V. Vitale, C. Kokail, A. Neven, P. Zoller, B. Vermersch, and B. Kraus, Entanglement phase diagrams from partial transpose moments, arXiv:2212.10181 (2022).
[20] M. Cerezo, A. Sone, J. L. Beckey, and P. J. Coles, Subquantum Fisher information, Quantum Sci. Technol. **6**, 035008 (2021).
[21] M. Yu, D. Li, J. Wang, Y. Chu, P. Yang, M. Gong, N. Goldman, and J. Cai, Experimental estimation of the quantum Fisher information from randomized measurements, Phys. Rev. Res. **3**, 043122 (2021).
[22] A. Rath, C. Branciard, A. Minguzzi, and B. Vermersch, Quantum Fisher information from randomized measurements, Phys. Rev. Lett. **127**, 260501 (2021).
[23] V. Vitale, A. Rath, P. Jurcevic, A. Elben, C. Branciard, and B. Vermersch, Estimation of the quantum Fisher information on a quantum processor, arXiv:2307.16882 [quant-ph] (2023).
[24] Z. Liu, Y. Tang, H. Dai, P. Liu, S. Chen, and X. Ma, Detecting entanglement in quantum many-body systems via permutation moments, Phys. Rev. Lett. **129**, 260501 (2022).
[25] A. Rath, V. Vitale, S. Murciano, M. Votto, J. Dubail, R. Kueng, C. Branciard, P. Calabrese, and B. Vermersch, Entanglement barrier and its symmetry resolution: Theory and experimental observation, PRX Quantum **4**, 010318 (2023).
[26] A. Rico and F. Huber, Entanglement detection with trace polynomials, arXiv:2303.07761 [quant-ph] (2023).
[27] J. Cotler, S. Choi, A. Lukin, H. Gharibyan, T. Grover, M. E. Tai, M. Rispoli, R. Schittko, P. M. Preiss, A. M. Kaufman, M. Greiner, H. Pichler, and P. Hayden, Quantum virtual cooling, Phys. Rev. X **9**, 031013 (2019).
[28] A. Seif, Z.-P. Cian, S. Zhou, S. Chen, and L. Jiang, Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors, PRX Quantum **4**, 010303 (2023).
[29] 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).
[30] J. Vovrosh and J. Knolle, Confinement and entanglement dynamics on a digital quantum computer, Sci. Rep. **11**, 11577 (2021).
[31] 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).
[32] K. J. Satzinger, Y.-J. Liu, A. Smith, C. Knapp, M. Newman, C. Jones, Z. Chen, C. Quintana, X. Mi, and A. Dunsworth *et al.*, Realizing topologically ordered states on a quantum processor, Science **374**, 1237 (2021).
[33] D. Zhu, Z. P. Cian, C. Noel, A. Risinger, D. Biswas, L. Egan, Y. Zhu, A. M. Green, C. H. Alderete, N. H. Nguyen, Q. Wang, A. Maksymov, Y. Nam, M. Cetina, N. M. Linke, M. Hafezi, and C. Monroe, Cross-platform comparison of arbitrary quantum states, Nat. Commun. **13**, 6620 (2022).
[34] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, arXiv:2006.15788 (2020).
[35] C. Hadfield, Adaptive Pauli shadows for energy estimation, arXiv:2105.12207 (2021).
[36] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, Phys. Rev. Lett. **127**, 030503 (2021).
[37] T.-C. Yen, A. Ganeshram, and A. F. Izmaylov, Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates, arXiv:2201.01471 (2022).
[38] K. Van Kirk, J. Cotler, H.-Y. Huang, and M. D. Lukin, Hardware-efficient learning of quantum many-body states, arXiv:2212.06084 (2022).
[39] B. Wu, J. Sun, Q. Huang, and X. Yuan, Overlapped grouping measurement: A unified framework for measuring quantum states, Quantum **7**, 896 (2023).
[40] P. Glasserman and D. D. Yao, Some guidelines and guarantees for common random numbers, Manage. Sci. **38**, 884 (1992).
[41] Specifically, we require that $\mathrm{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] > \mathbb{V}[Y]/2$, so that $\mathbb{V}[X - Y] < \mathbb{V}[X]$.
[42] Y. Zhou and Q. Liu, Performance analysis of multi-shot shadow estimation, arXiv:2212.11068 (2022).
[43] J. Helsen and M. Walter, Thrifty shadow estimation: re-using quantum circuits and bounding tails, arXiv:2212.06240 (2022).
[44] J. Tindall, M. Fishman, M. Stoudenmire, and D. Sels, Efficient tensor network simulation of IBM's Eagle kicked Ising experiment, arXiv:2306.14887 [quant-ph] (2023).
[45] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, theorems, Rev. Mod. Phys. **93**, 045003 (2021).

[46] P. Horodecki, Measuring quantum entanglement without prior state reconstruction, Phys. Rev. Lett. **90**, 167901 (2003).
[47] H. A. Carteret, Noiseless quantum circuits for the Peres separability criterion, Phys. Rev. Lett. **94**, 040502 (2005).
[48] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).
[49] J. Eisert, M. Cramer, and M. B. Plenio, Area laws for the entanglement entropy, Rev. Mod. Phys. **82**, 277 (2010).
[50] E.-M. Kontopoulou, G.-P. Dexter, W. Szpankowski, A. Grama, and P. Drineas, Randomized linear algebra approaches to estimate the von Neumann entropy of density matrices, arXiv:1801.01072 (2018).
[51] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Direct estimations of linear and nonlinear functionals of a quantum state, Phys. Rev. Lett. **88**, 217901 (2002).
[52] P. Calabrese and J. Cardy, Entanglement entropy and quantum field theory, J. Stat. Mech.: Theory Exp. **2004**, P06002.
[53] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. **326**, 96 (2011).
[54] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, Phys. Rev. Lett. **106**, 230501 (2011).
[55] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomography, Phys. Rev. Lett. **107**, 210404 (2011).
[56] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, Variational quantum fidelity estimation, Quantum **4**, 248 (2020).
[57] M. Fishman and G. Torlai, PastaQ: A package for simulation, tomography and analysis of quantum computers (2020).
[58] 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).
[59] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, Science **377** (2022).
[60] L. Lewis, H.-Y. Huang, V. T. Tran, S. Lehner, R. Kueng, and J. Preskill, Improved machine learning algorithm for predicting ground state properties, Nat. Commun. **15**, 895 (2024).
[61] E. Onorati, C. Rouze, D. S. Franca, and J. D. Watson, Efficient learning of ground & thermal states within phases of matter, arXiv:2301.12946 (2023).
[62] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, Shadow process tomography of quantum channels, Phys. Rev. A **107**, 042403 (2023).
[63] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, arXiv:2110.02965 (2021).
[64] D. S. França, L. A. Markovich, V. V. Dobrovitski, A. H. Werner, and J. Borregaard, Efficient and robust estimation of many-qubit hamiltonians, arXiv:2205.09567 (2022).
[65] A. Rath, R. van Bijnen, A. Elben, P. Zoller, and B. Vermersch, Importance sampling of randomized measurements for probing entanglement, Phys. Rev. Lett. **127**, 200503 (2021).
[66] M. C. Tran, D. K. Mark, W. W. Ho, and S. Choi, Measuring arbitrary physical properties in analog quantum simulation, arXiv:2212.02517 (2022).
[67] S. J. Garratt and E. Altman, Probing post-measurement entanglement without post-selection, arXiv:2305.20092 (2023).
[68] M. Fishman, S. White, and E. Stoudenmire, The ITensor software library for tensor network calculations, SciPost Physics Codebases (2022).
[69] https://github.com/bvermersch/RandomMeas.jl.
[70] W. Hoeffding, in *Breakthroughs in Statistics* (Springer, 1992), p. 308.
[71] 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).
[72] D. E. Knuth, *The Art of Computer Programming, Volume I: Fundamental Algorithms, 3rd Edition* (Addison-Wesley, 1997).