Benchmarking Quantum Simulators Using Ergodic Quantum Dynamics

Daniel K. Mark,[1] Joonhee Choi,[2] Adam L. Shaw,[2] Manuel Endres,[2] and Soonwon Choi[1,*]
[1]Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[2]California Institute of Technology, Pasadena, California 91125, USA

(Received 7 July 2022; accepted 15 August 2023; published 12 September 2023)

We propose and analyze a sample-efficient protocol to estimate the fidelity between an experimentally prepared state and an ideal target state, applicable to a wide class of analog quantum simulators without advanced spatiotemporal control. Our protocol relies on universal fluctuations emerging from generic Hamiltonian dynamics, which we discover in the present work. It does not require fine-tuned control over state preparation, quantum evolution, or readout capability, while achieving near optimal sample complexity: a percent-level precision is obtained with $\sim 10^3$ measurements, independent of system size. Furthermore, the accuracy of our fidelity estimation improves exponentially with increasing system size. We numerically demonstrate our protocol in a variety of quantum simulator platforms, including quantum gas microscopes, trapped ions, and Rydberg atom arrays. We discuss applications of our method for tasks such as multiparameter estimation of quantum states and processes.

DOI: 10.1103/PhysRevLett.131.110601

Introduction.—Recent advances in quantum technology have opened new ways to probe quantum many-body physics, leading to the first observations of novel phases of matter [1–5], quantum thermalization [6–8], and nonequilibrium phenomena [9–14]. However, in order to advance to the stage where quantum devices produce highly accurate data, it is important to quantify the performance of said devices. One method to do so is quantum device benchmarking [15]—verifying that a device accurately produces a state ρ close to the desired state $|\Psi\rangle$, in the presence of imperfections and noise, measured by the fidelity $F = \langle\Psi|\rho|\Psi\rangle$. A high fidelity certifies that any property of the prepared state is close to that of the target state [16], hence is widely used in theory to quantify the goodness of state preparation. Experimentally measuring the fidelity is important for building, characterizing, and improving increasingly complex and precise systems.

Several methods to benchmark quantum devices have been proposed. A naïve approach is to perform quantum state tomography [16–19], in which an experimental state is fully characterized by measurements in many different bases. This approach, however, is impractical even for relatively small systems as it requires prohibitively many measurements. Alternatively, recent proposals pointed out that one can directly estimate the fidelity with a small number of measurements in randomly chosen bases [20–27]. These methods rely on implementing highly engineered quantum gates that satisfy certain statistical properties and are not readily applicable to quantum devices with limited controllability. Other existing benchmarking protocols require sophisticated controls and are challenging to implement [28–36]. In particular, we emphasize that analog quantum simulators are typically designed to realize specific forms of many-body Hamiltonians and lack the ability to implement arbitrary unitary operations. Thus, it remains an outstanding challenge to develop a general benchmarking method with minimal requirements on hardware capability.

In this Letter, we propose and analyze a benchmarking protocol that requires minimal experimental control: one prepares an initial state, time-evolves it under a natural Hamiltonian of the system, and performs measurements in a fixed basis (Fig. 1). We show that, with appropriate data processing (enabled by classical computation), this simple experiment gives an estimate for the fidelity F—encapsulating the combined effects of errors in state preparation, quench evolution, and readout [37]—with a small number of measurements. Most importantly, our method works for generic quench dynamics far from fine-tuned cases, including at finite effective temperatures, in the presence of symmetries, and in nonqubit based systems such as itinerant particles on optical lattices, making it suitable for a wide class of existing platforms.

The key behind our approach is our discovery of universal statistical fluctuations in the measurement outcome distributions $p(z)$ that arise from generic quantum dynamics (Fig. 2). Previously, such universal fluctuations in $p(z)$ were only known to occur in ideal, controlled dynamics such as random unitary circuits (RUCs), where $\{p(z)\}$ approximately follows the Porter-Thomas distribution [31,55]. Leveraging our discovery and classical computation, we design a novel statistic: a real number $f(z)$ associated to every measurement outcome z such that its average over experimentally obtained samples, $\hat{F}_d \equiv \langle f(z)\rangle_{\text{exp}}$, converges quickly to the many-body fidelity F. In other words, $\hat{F}_d$ is a

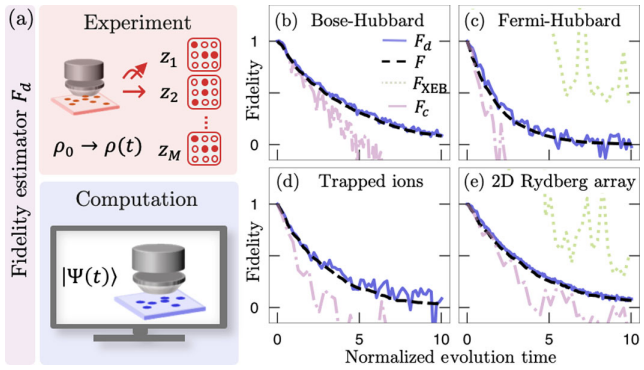

FIG. 1. (a) Schematic of our benchmarking protocol. The fidelity estimator F_d is evaluated from experimental snapshots $\{z_1, \ldots, z_M\}$ of a state $\rho(t)$ obtained after quench dynamics, compared against classical computation of the ideal state $|\Psi(t)\rangle$ in the absence of error (see Table I). (b)–(e) Numerical demonstrations. F_d closely tracks the fidelity decay over evolution time normalized in units of Rabi frequency or tunneling strength (black dashed) between noisy and ideal quench dynamics in a wide class of analog simulators, including 1D Bose-Hubbard, integrable 1D Fermi-Hubbard, 1D trapped-ion, and 2D Rydberg array models at finite effective temperature; see Supplemental Material (SM) [38] for details. Previously proposed benchmarks F_{XEB} [30] [green dotted, out of scale in (b),(d)] or F_c [36] (purple dot-dashed) fail to estimate F for these systems.

computationally assisted, efficiently measurable observable [56,57] that estimates the fidelity.

The ability to estimate fidelity serves as a foundation for two tasks: (i) *target state benchmarking*, where the overlap between an experimentally prepared state and a pure target state is measured via a high-fidelity quench time evolution,

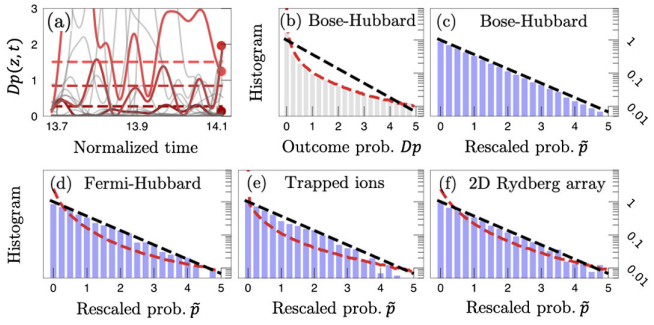

FIG. 2. Emergent universal statistics. (a) During many-body Hamiltonian evolution of a pure state, the probability $p(z,t)$ of measuring an outcome z fluctuates around its average value $p_{\mathrm{avg}}(z)$ (dashed, three distinct zs highlighted). Here, we consider a Bose-Hubbard model and present $p(z,t)$ rescaled by the Hilbert space dimension D. (b) The histogram of $p(z,t)$ over all z at a fixed t [red dots in (a)], is nonuniversal (red dashed), far from the Porter-Thomas (PT) distribution (black dashed). (c) Rescaling each $p(z,t)$ by $p_{\mathrm{avg}}(z)$ yields $\tilde{p}(z,t)$, whose histogram follows the PT distribution. (d)–(f) The histograms of $\tilde{p}(z,t)$ (blue bars) follow the universal PT distribution in all models considered in this Letter, whereas those of the bare $p(z,t)$ rescaled by D are nonuniversal (red dashed).

TABLE I. Proposed benchmarking protocol.

Experiment:
1. Prepare an initial state ρ_0, which approximates a pure state $|\Psi_0\rangle\langle\Psi_0|$.
2. Evolve the system under its natural Hamiltonian H for a time t.
3. Measure the evolved state $\rho(t)$ in a natural basis, obtaining configurations $\{z_j\}_{j=1}^{M}$.

Computation: Classically compute
1. $p(z,t) \equiv |\langle z|\Psi(t)\rangle|^2 = |\langle z|\exp(-iHt)|\Psi_0\rangle|^2$,
2. $p_{\mathrm{avg}}(z) \equiv \lim_{T\to\infty}(1/T)\int_0^T p(z,t)dt$, $\tilde{p}(z,t) \equiv p(z,t)/p_{\mathrm{avg}}(z)$.
3. $\mathcal{Z}(t) \equiv \sum_z p_{\mathrm{avg}}(z)\tilde{p}(z,t)^2$.

Data processing: Evaluate
$\hat{F}_d(t) \equiv (2/M)\left[\sum_{i=1}^{M}\tilde{p}(z_i,t)\right]/\mathcal{Z}(t) - 1 \approx F_d(t)$, which approximates the fidelity $F = \langle\Psi(t)|\rho(t)|\Psi(t)\rangle$.

and (ii) *quantum process benchmarking*, in which the fidelity decay of quench dynamics is monitored over the course of evolution.

Protocol.—We focus on describing and numerically demonstrating our protocol, before returning to why it works. Our benchmarking method consists of three steps: experiment, computation, and data processing [Table I and Fig. 1(a)]. The initial state of our protocol can either be an easy-to-prepare state or a more complex state that one wishes to benchmark. After quench evolution for a fixed time t, the experimental state $\rho(t)$ is measured in any fixed basis $\{|z\rangle\}$. Convenient choices of $\{|z\rangle\}$ include the set of bitstrings in two-level (qubit) systems or real-space particle number configurations in quantum gas microscopes. Repeating the state preparation and measurement M times, one obtains measured configurations $\{z_1, \ldots, z_M\}$, with each z_i sampled from the distribution $q(z,t) \equiv \langle z|\rho(t)|z\rangle$. Our protocol estimates the fidelity by using a small number of samples to compare the empirical distribution $q(z,t)$ against a theoretical, target distribution $p(z,t)$. By classical computation, we obtain $p(z,t)$ and its infinite-time average $p_{\mathrm{avg}}(z) \equiv \lim_{T\to\infty}(1/T)\int_0^T p(z,t)dt$. In practice, one may average over a finite duration T as an approximation, at the expense of slightly larger statistical errors. Then, we evaluate the rescaled outcome probabilities $\tilde{p}(z,t) \equiv p(z,t)/p_{\mathrm{avg}}(z)$ and the normalization factor $\mathcal{Z}(t) \equiv \sum_z p_{\mathrm{avg}}(z)\tilde{p}(z,t)^2$.

The classical computation determines our statistic $f(z)$, while the experimental samples determine which outcomes zs to use when evaluating the statistic. More specifically, we estimate the fidelity with the empirical average

$$\hat{F}_d(t) = \langle f(z)\rangle_{\mathrm{exp}} = \frac{1}{M}\sum_{i=1}^{M} 2\tilde{p}(z_i,t)/\mathcal{Z}(t) - 1. \quad (1)$$

This explicitly defines the statistic $f(z)$, which also depends on $|\Psi_0\rangle$, H, and t. In the limit $M \to \infty$, this

converges to our benchmark $F_d(t) = 2[\sum_z q(z,t)\tilde{p}(z,t)]/\mathcal{Z}(t) - 1$. This benchmark can be understood as a weighted covariance between the empirical and ideal distributions. We show that $F_d(t)$ approximates the fidelity F for a wide class of quantum systems, both for uncorrelated, infrequent incoherent errors, or local and weak global coherent errors [38,58–60], and rigorously prove our statement for isolated single errors and long evolution times.

Figures 1(b)–1(e) numerically demonstrate the use of our estimator for process benchmarking: tracking the decay of fidelity over time in four different quantum simulation platforms. For each platform, we simulate an initial product state undergoing natural Hamiltonian dynamics in the presence of experimentally relevant errors [38]. We confirm that F_d successfully traces the fidelity decay in regimes where previously proposed fidelity estimators F_{XEB} [30] and F_c [36] do not. This is because F_c and F_{XEB} (reviewed in the SM [38]) assume that $p(z,t)$ satisfy statistical properties (discussed below) that are in general not satisfied by natural Hamiltonian dynamics, e.g., F_c requires the system to evolve at infinite effective temperature. We now turn to the underlying principles of our protocol: *emergent universal statistics*, *speckle-based benchmarking*, and *measurement-basis independence*.

Emergent universal statistics.—The statistical properties of $p(z)$ have been extensively studied in deep RUCs. When the output state of a typical deep RUC is measured, $p(z)$ is not perfectly uniform, but exhibits a *speckle pattern*: over different zs, $p(z)$ fluctuates about $1/D$ due to random interference in coherent quantum dynamics, with D the Hilbert space dimension. While the details of the fluctuations—which $p(z)$s are larger—sensitively depends on the particular choice of RUC, the statistical properties of $p(z)$ are universal. Specifically, the fraction of $p(z)$s in a given interval $p(z) \in [x, x+dx]$ is given by the Porter-Thomas (PT) distribution: $P[p(z) = x]dx = \mu^{-1}\exp(-x/\mu)dx$ with mean $\mu = 1/D$. This enables the existing benchmarks F_{XEB} and F_c. Specifically, they utilize the fact that the PT distribution has a second moment equal to two [30,31,36]. Previously, it was unclear under what conditions the PT distribution can arise, other than from RUCs and fine-tuned Hamiltonian dynamics.

In fact, for generic time-independent Hamiltonian dynamics, the raw distribution $p(z)$ does not follow the PT distribution [Figs. 2(a) and 2(b)]. This is due to the presence of energy conservation or symmetries, which causes systematic trends in $p(z)$ and distorts its distribution away from PT. For example, in any state with positive effective temperature, low-energy configurations are measured more frequently than high-energy ones [38]. While previous work discovered PT distributions in certain *local observables* [36], this work concerns global observables under general conditions such as finite effective temperature.

Our key insight is that the systematic trends in $p(z)$ can be removed simply by rescaling $p(z)$ with its time-averaged value, leaving only random relative fluctuations $\tilde{p}(z) \equiv p(z)/p_{\mathrm{avg}}(z)$ that follow the PT distribution with mean $\mu = 1$ [Figs. 2(c)–2(f)].

Theorem (informal).—Consider an initial state $|\Psi_0\rangle$, evolved for time t under a Hamiltonian H satisfying the kth no-resonance condition for a large integer k, and measured in a complete basis $\{|z\rangle\}$. For sufficiently late times t, the rescaled probabilities $\tilde{p}(z,t)$ follow the Porter-Thomas distribution, up to a correction bounded by the inverse effective Hilbert space dimension $D_\beta^{-1} \equiv \sum_{z,E} |\langle z|E\rangle|^4 |\langle E|\Psi_0\rangle|^4 / p_{\mathrm{avg}}(z)$, where $\{|E\rangle\}$ are the eigenstates of H.

Rigorous statements of our theorem and their proofs are presented in the SM [38]. The only assumption of this Theorem is the kth no-resonance condition, stating that $\sum_{i=1}^k E_{\alpha_i} = \sum_{i=1}^k E_{\beta_i}$ if and only if the k indices (α_i) are a permutation of (β_i) [61–65]. That is, the eigenvalues $\{E_j\}$ of H possess no resonant structures. This condition is expected to hold for generic ergodic Hamiltonians [61,62], and we find that it even holds in some integrable systems such as the 1D Fermi-Hubbard model [Fig. 1(c)] [38,66]. The effective dimension D_β quantifies the size of the Hilbert space explored during quench evolution (that can be probed in the $\{|z\rangle\}$ basis). D_β is similar to a participation ratio of $|\Psi_0\rangle$ and $|z\rangle$, when they are decomposed in the energy eigenbasis, which generically increases exponentially with increasing system size, leading to a better agreement with the PT distribution and enabling our protocol to be increasingly accurate.

Our theorem states that the outcome distribution factorizes into two parts $p(z,t) = p_{\mathrm{avg}}(z) \times \tilde{p}(z,t)$: systematic values $p_{\mathrm{avg}}(z)$ and random Porter-Thomas fluctuations $\tilde{p}(z,t)$. The systematic value is related to thermalization and does not distinguish between pure and mixed states, usually set by coarse-grained information such as the total energy $\langle\Psi_0|H|\Psi_0\rangle$. Meanwhile, the fluctuations originate from random interference and average away in a mixed state [38]. These fluctuations are highly sensitive to details of the initial state and evolution, serving as a "fingerprint" that enables their benchmarking.

Speckle-based benchmarking.—We provide an intuitive explanation of our benchmark F_d, based on two properties: (i) the second moment of the rescaled $\tilde{p}(z)$ is $\mathcal{Z} = \sum_z p_{\mathrm{avg}}(z)\tilde{p}(z)^2 \approx 2$ for $p(z)$ arising from ideal unitary evolution, (ii) the experimental distribution $q(z)$ in the presence of errors can be expressed as a linear combination $q(z) = Fp(z) + (1-F)p_\perp(z)$, where $p_\perp(z)$ is uncorrelated with the ideal distribution $p(z)$ in the following sense: $\mathbb{E}_z[\tilde{p}(z)\tilde{p}_\perp(z)] \approx \mathbb{E}_z[\tilde{p}(z)]\mathbb{E}_z[\tilde{p}_\perp(z)] = 1$, with $\mathbb{E}_z[\cdot] \equiv \sum_z p_{\mathrm{avg}}(z)[\cdot]$ and $\tilde{p}_\perp \equiv p_\perp/p_{\mathrm{avg}}$. The second property relies on an assumption that the speckle pattern in $\tilde{p}$

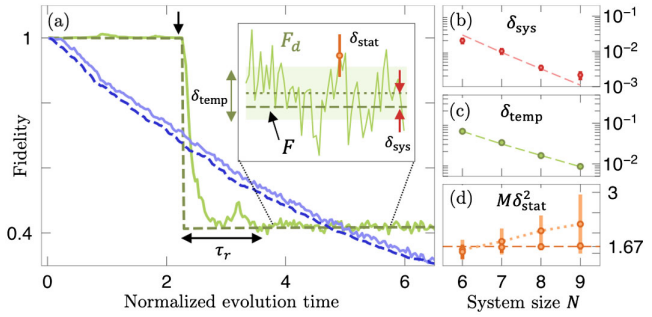

FIG. 3. Performance analysis. (a) Numerical simulation of open system dynamics of a 1D Bose-Hubbard model with $N = 9$ particles on N sites ($D = 24310$). When a single error occurs (black arrow), F_d (solid line, green) closely approximates the fidelity F (dashed line, green) after a short delay time τ_r. Averaging F_d and F over (potentially many) stochastic errors at different times gives their values for the mixed state ρ (blue). The inset shows uncertainties and errors in our method. In particular, F_d slightly deviates from F, quantified by the difference δ_{sys} between F and the time-averaged F_d (red arrow) and by the fluctuations δ_{temp} of F_d over time (green arrow). Furthermore, a finite number of samples M results in a statistical uncertainty δ_{stat} in the unbiased estimator $\hat{F}_d$ (orange error bar). (b),(c) Both δ_{sys} and δ_{temp} decrease exponentially with system size, quantitatively agreeing with analytic predictions (dashed lines). (d) The sample complexity $M\delta^2_{\text{stat}}$ increases weakly with N at fixed, early times (dotted line). At late times, it approaches the N-independent value $1 + 2F - F^2$ (dashed line). Error bars in (b)–(d) indicate variations over an ensemble of disordered Hamiltonians. See SM [38] for details.

significantly changes under any error, which has been rigorously proven for RUCs [58,59].

Using these properties, our estimator F_d is designed to isolate the desired "fingerprint," taking value 1 when $q(z) = p(z)$ and 0 when $q(z) = p_\perp(z)$, which in turn implies $F_d \approx F$ [38]. We emphasize that it is essential to use the rescaled $\tilde{p}(z)$ to estimate the fidelity; otherwise $p(z)$ and $p_\perp(z)$ exhibit large correlations due to their shared systematic values $p_{\text{avg}}(z)$.

In fact, the second condition can be relaxed. Under local coherent or incoherent errors, the relation $F_d \approx F$ can be shown at late times without any assumption on $q(z)$, solely based on the eigenstate thermalization hypothesis and no-resonance conditions [38]. We also argue that this result extends to multiple stochastic errors, and small coherent errors in the quench Hamiltonian, as verified by various numerical simulations [38]. Furthermore, we also verify that $F_d \approx F$ holds even for relatively short quenches, well before the PT distribution emerges in $\tilde{p}(z)$, owing to the time-dependent adjustment factor $\mathcal{Z}(t)$ in F_d. This factor is inspired by F_c [36] and its effect is illustrated in Figs. 1(b)–1(e) and the SM [38].

Measurement-basis independence.—A surprising feature of our approach is that the fidelity is estimated from measurements in a fixed basis, despite the fact that the fidelity also depends on phase information not accessed from such measurements. Nevertheless, our protocol works because quench evolution transforms the effects of physically relevant errors, including phase errors, into a form detectable by generic local measurements, after a short delay time [Fig. 3(a)].

In our examples above, the quench dynamics plays two roles simultaneously: it enables our protocol, but also generates imperfect quantum evolution (due to errors), whose fidelity decay is measured. If the quench evolution were perfect, the measured fidelity would only reflect the state preparation error of the (potentially interesting) target initial state. We present numerical demonstrations of such target state benchmarking in the SM [38].

Performance analysis.—Our theorem enables us to predict how well F_d estimates the fidelity. First, we point out that it suffices to study the effect of a single error when the error rate is sufficiently small. This is because incoherent noisy dynamics can be "unraveled" into an ensemble of stochastic pure state trajectories [67] each corresponding to a fixed occurrence of errors [Fig. 3(a)]. As long as they are sufficiently infrequent, the effect of multiple errors can be understood from that of a single error [58,59].

We showcase the performance of F_d under realistic conditions by numerically simulating the 1D Bose-Hubbard model; see Fig. 3(a). We quantify the performance of F_d along several axes [Fig. 3(a), inset]: the systematic error δ_{sys} refers to the difference between the true fidelity F and the time-averaged F_d, while the δ_{temp} quantifies how F_d fluctuates over time. Finally, the statistical fluctuation (or, *sample complexity*) δ_{stat} measures the uncertainty of the estimated $\hat{F}_d$ associated with a finite number of samples M, and hence the number of samples required to determine F_d up to a desired precision. See SM [38] for further details.

Our theorem allows analytical estimation of these quantities in the limit of long evolution: δ_{sys} and δ_{temp} are, respectively, $O(D_\beta^{-1})$ and $O(D_\beta^{-1/2})$ [38]. Hence, both the accuracy and precision of our benchmark improve exponentially with increasing system size, explicitly confirmed in our numerical simulations [Figs. 3(b) and 3(c)]. Meanwhile, the sample complexity has optimal scaling. It is system size independent for long evolutions: $M\delta^2_{\text{stat}} \approx 1 + 2F - F^2$ [Fig. 3(d), dashed line]. For short quench evolution, the sample complexity grows weakly with system size [Fig. 3(d), dotted line]. Finally, in the presence of incoherent errors, the finite response time τ_r leads to a slight delay between F_d and the continuously decaying fidelity [38,58,68].

Limitations.—While our protocol is applicable to generic quantum many-body systems, it may fail in special cases in which our theorem is not applicable. Examples include systems with weakly or nonergodic dynamics, or the

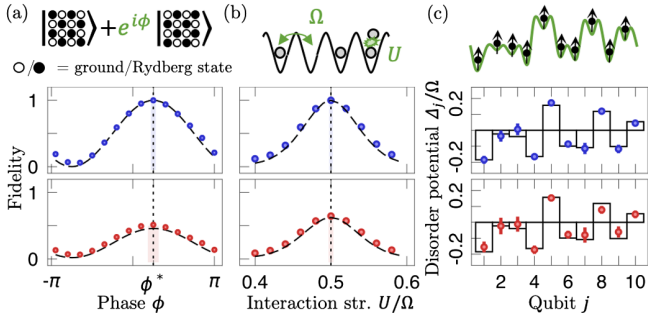

FIG. 4. State and process benchmarking with F_d. We numerically simulate the estimation of (a) the phase ϕ of a GHZ-like initial state in a 2D Rydberg system, (b) the ratio of the interaction and tunneling strengths U/Ω in a 1D Bose-Hubbard model, and (c) ten disordered on-site potentials in a trapped ion model. Parameters are estimated by maximizing $\hat{F}_d$ over simulated parameter values, using measurements after error-free (blue) or noisy (red) quench evolution. The error bars and shaded regions indicate the statistical uncertainties in $\hat{F}_d$ and the parameter values with 1000 samples. See SM [38] for details. (a), (b) Both F (black lines) and $\hat{F}_d$ (markers) are consistent and simultaneously maximized at the true parameter value (dotted line). (c) Reconstructed disorder potential values (markers) are consistent with their true values (lines).

presence of correlated nonlocal errors. We provide detailed analysis and potential resolutions of known failure cases in the SM [38].

Applications.—The ability to measure the fidelity enables further applications. As examples, we show that one can simultaneously estimate multiple parameters of quantum states or Hamiltonians. The key observation is that, given the ability to measure the fidelity between a theoretical model and experiment, one can vary classical simulation model parameters to maximize the estimated fidelity [31,36]. This optimization requires classical computation but no further data acquisition. We numerically demonstrate this idea in three different examples and verify that the extracted parameter values are close to the actual ones, even in the presence of noise; see Fig. 4.

We would like to thank Wen Wei Ho, Anant Kale, Eun Jong Kim, Sherry Zhang, and Peter Zoller for useful discussions. In particular, we thank Sherry Zhang for a careful reading of this manuscript and Anant Kale for insightful discussion at the early stage of the project. We acknowledge funding provided by the NSF Physics Frontiers Centers IQIM (NSF PHY-1733907) and CUA (NSF PHY-1734011), the AFOSR YIP (FA9550-19-1-0044), the DARPA ONISQ program (W911NF2010021), the Army Research Office MURI program (W911NF2010136), the NSF QLCI program (2016245), the DOE (DE-SC0021951), NSF CAREER Grant No. 1753386 and NSF CAREER Grant No. 2237244, and the DOE Quantum Systems Accelerator Center (7571809). J. C. acknowledges support from the IQIM postdoctoral fellowship. A. L. S. acknowledges support from the Eddleman Quantum graduate fellowship.

*Corresponding author: soonwon@mit.edu

[1] Markus Greiner, Olaf Mandel, Tilman Esslinger, Theodor W. Hänsch, and Immanuel Bloch, Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms, Nature (London) **415**, 39 (2002).
[2] Monika Aidelsburger, Marcos Atala, Michael Lohse, Julio T. Barreiro, B. Paredes, and Immanuel Bloch, Realization of the Hofstadter Hamiltonian with Ultracold Atoms in Optical Lattices, Phys. Rev. Lett. **111**, 185301 (2013).
[3] Gregor Jotzu, Michael Messer, Rémi Desbuquois, Martin Lebrat, Thomas Uehlinger, Daniel Greif, and Tilman Esslinger, Experimental realization of the topological Haldane model with ultracold fermions, Nature (London) **515**, 237 (2014).
[4] Christian Gross and Immanuel Bloch, Quantum simulations with ultracold atoms in optical lattices, Science **357**, 995 (2017).
[5] Sepehr Ebadi, Tout T Wang, Harry Levine, Alexander Keesling, Giulia Semeghini, Ahmed Omran, Dolev Bluvstein, Rhine Samajdar, Hannes Pichler, Wen Wei Ho *et al.*, Quantum phases of matter on a 256-atom programmable quantum simulator, Nature (London) **595**, 227 (2021).
[6] Adam M. Kaufman, M Eric Tai, Alexander Lukin, Matthew Rispoli, Robert Schittko, Philipp M. Preiss, and Markus Greiner, Quantum thermalization through entanglement in an isolated many-body system, Science **353**, 794 (2016).
[7] Yijun Tang, Wil Kao, Kuan-Yu Li, Sangwon Seo, Krishnanand Mallayya, Marcos Rigol, Sarang Gopalakrishnan, and Benjamin L. Lev, Thermalization Near Integrability in a Dipolar Quantum Newton's Cradle, Phys. Rev. X **8**, 021030 (2018).
[8] Fusheng Chen *et al.*, Observation of Strong and Weak Thermalization in a Superconducting Quantum Processor, Phys. Rev. Lett. **127**, 020602 (2021).
[9] Brian Neyenhuis, Jiehang Zhang, Paul W. Hess, Jacob Smith, Aaron C. Lee, Phil Richerme, Zhe-Xuan Gong, Alexey V. Gorshkov, and Christopher Monroe, Observation of prethermalization in long-range interacting spin chains, Sci. Adv. **3**, e1700672 (2017).
[10] Jiehang Zhang, Guido Pagano, Paul W. Hess, Antonis Kyprianidis, Patrick Becker, Harvey Kaplan, Alexey V. Gorshkov, Z.-X. Gong, and Christopher Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, Nature (London) **551**, 601 (2017).
[11] Joonhee Choi, Hengyun Zhou, Soonwon Choi, Renate Landig, Wen Wei Ho, Junichi Isoya, Fedor Jelezko, Shinobu Onoda, Hitoshi Sumiya, Dmitry A. Abanin, and Mikhail D. Lukin, Probing Quantum Thermalization of a Disordered Dipolar Spin Ensemble with Discrete Time-Crystalline Order, Phys. Rev. Lett. **122**, 043603 (2019).
[12] Antonio Rubio-Abadal, Matteo Ippoliti, Simon Hollerith, David Wei, Jun Rui, S. L. Sondhi, Vedika Khemani, Christian Gross, and Immanuel Bloch, Floquet

Prethermalization in a Bose-Hubbard System, Phys. Rev. X **10**, 021044 (2020).
[13] Pai Peng, Chao Yin, Xiaoyang Huang, Chandrasekhar Ramanathan, and Paola Cappellaro, Floquet prethermalization in dipolar spin chains, Nat. Phys. **17**, 444 (2021).
[14] Xiao Mi, Matteo Ippoliti, Chris Quintana, Ami Greene, Zijun Chen, Jonathan Gross, Frank Arute, Kunal Arya, Juan Atalaya, Ryan Babbush *et al.*, Time-crystalline eigenstate order on a quantum processor, Nature (London) **601**, 531 (2022).
[15] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi, Quantum certification and benchmarking, Nat. Rev. Phys. **2**, 382 (2020).
[16] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, England, 2010).
[17] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu, Efficient quantum state tomography, Nat. Commun. **1**, 149 (2010).
[18] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert, Quantum State Tomography via Compressed Sensing, Phys. Rev. Lett. **105**, 150401 (2010).
[19] Matthias Christandl and Renato Renner, Reliable Quantum State Tomography, Phys. Rev. Lett. **109**, 120403 (2012).
[20] Steven T. Flammia and Yi-Kai Liu, Direct Fidelity Estimation from Few Pauli Measurements, Phys. Rev. Lett. **106**, 230501 (2011).
[21] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin, Practical Characterization of Quantum Devices without Tomography, Phys. Rev. Lett. **107**, 210404 (2011).
[22] Scott Aaronson, Shadow tomography of quantum states, SIAM J. Comput. **49**, STOC18–368 (2019).
[23] Hsin-Yuan Huang, Richard Kueng, and John Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).
[24] Andreas Elben, Benoît Vermersch, Christian F. Roos, and Peter Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states, Phys. Rev. A **99**, 052323 (2019).
[25] Andreas Elben, Benoît Vermersch, Rick van Bijnen, Christian Kokail, Tiff Brydges, Christine Maier, Manoj K. Joshi, Rainer Blatt, Christian F. Roos, and Peter Zoller, Cross-Platform Verification of Intermediate Scale Quantum Devices, Phys. Rev. Lett. **124**, 010504 (2020).
[26] Yunchao Liu, Matthew Otten, Roozbeh Bassirianjahromi, Liang Jiang, and Bill Fefferman, Benchmarking near-term quantum computers via random circuit sampling, arXiv:2105.05232.
[27] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).
[28] M. Ohliger, V. Nesme, and J. Eisert, Efficient and feasible state tomography of quantum many-body systems, New J. Phys. **15**, 015024 (2013).
[29] Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P. Lanyon, Peter Zoller, Rainer Blatt, and Christian F. Roos, Probing Rényi entanglement entropy via randomized measurements, Science **364**, 260 (2019).
[30] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven, Characterizing quantum supremacy in near-term devices, Nat. Phys. **14**, 595 (2018).
[31] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandão, David A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).
[32] Zhi Li, Liujun Zou, and Timothy H. Hsieh, Hamiltonian Tomography via Quantum Quench, Phys. Rev. Lett. **124**, 160502 (2020).
[33] Christian Kokail, Rick van Bijnen, Andreas Elben, Benoît Vermersch, and Peter Zoller, Entanglement Hamiltonian tomography in quantum simulation, Nat. Phys. **17**, 936 (2021).
[34] Marek Gluza and Jens Eisert, Recovering Quantum Correlations in Optical Lattices from Interaction Quenches, Phys. Rev. Lett. **127**, 090503 (2021).
[35] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You, Classical shadow tomography with locally scrambled quantum dynamics, arXiv:2107.04817.
[36] Joonhee Choi, Adam L. Shaw, Ivaylo S. Madjarov, Xin Xie, Ran Finkelstein, Jacob P. Covey, Jordan S. Cotler, Daniel K. Mark, Hsin-Yuan Huang, Anant Kale *et al.*, Preparing random states and benchmarking with many-body quantum chaos, Nature (London) **613**, 468 (2023).
[37] More accurately, the fidelity is a quantity defined between two quantum states and does not include readout errors. Our benchmark estimates the state fidelity when readout errors are negligible.
[38] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.131.110601 for the proof of our theorem, detailed analysis of the performance, numerical demonstration, and limitations of our protocol, which includes Refs. [39–54].
[39] Anthony J. Short, Equilibration of quantum systems and subsystems, New J. Phys. **13**, 053009 (2011).
[40] Jordan S. Cotler, Daniel K. Mark, Hsin-Yuan Huang, Felipe Hernández, Joonhee Choi, Adam L. Shaw, Manuel Endres, and Soonwon Choi, Emergent quantum state designs from individual many-body wave functions, PRX Quantum **4**, 010311 (2023).
[41] S. Sarkar, S. Langer, J. Schachenmayer, and A. J. Daley, Light scattering and dissipative dynamics of many fermionic atoms in an optical lattice, Phys. Rev. A **90**, 023618 (2014).
[42] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić, Quantum scarred eigenstates in a rydberg atom chain: Entanglement, breakdown of thermalization, and stability to perturbations, Phys. Rev. B **98**, 155134 (2018).
[43] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S. Zibrov, Manuel Endres, Markus Greiner

et al., Probing many-body dynamics on a 51-atom quantum simulator, Nature (London) **551**, 579 (2017).
[44] Romain Vasseur and Joel E. Moore, Nonequilibrium quantum dynamics and transport: From integrability to many-body localization, J. Stat. Mech. (2016) 064010.
[45] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).
[46] David Gross, Koenraad Audenaert, and Jens Eisert, Evenly distributed unitaries: On the structure of unitary designs, J. Math. Phys. (N.Y.) **48**, 052104 (2007).
[47] Joshua M. Deutsch, Eigenstate thermalization hypothesis, Rep. Prog. Phys. **81**, 082001 (2018).
[48] Benoît Collins and Ion Nechita, Random matrix techniques in quantum information theory, J. Math. Phys. (N.Y.) **57**, 015215 (2016).
[49] Adam Nahum, Sagar Vijay, and Jeongwan Haah, Operator Spreading in Random Unitary Circuits, Phys. Rev. X **8**, 021014 (2018).
[50] Vedika Khemani, Ashvin Vishwanath, and David A. Huse, Operator Spreading and the Emergence of Dissipative Hydrodynamics under Unitary Evolution with Conservation Laws, Phys. Rev. X **8**, 031057 (2018).
[51] Dieter Jaksch, Christoph Bruder, Juan Ignacio Cirac, Crispin W. Gardiner, and Peter Zoller, Cold Bosonic Atoms in Optical Lattices, Phys. Rev. Lett. **81**, 3108 (1998).
[52] H. Pichler, A. J. Daley, and P. Zoller, Nonequilibrium dynamics of bosonic atoms in optical lattices: Decoherence of many-body states due to spontaneous emission, Phys. Rev. A **82**, 063605 (2010).
[53] Dolev Bluvstein, Ahmed Omran, Harry Levine, Alexander Keesling, Giulia Semeghini, Sepehr Ebadi, Tout T. Wang, Alexios A. Michailidis, Nishad Maskara, Wen Wei Ho *et al.*, Controlling quantum many-body dynamics in driven rydberg atom arrays, Science **371**, 1355 (2021).
[54] T. D. Kühner and H. Monien, Phases of the one-dimensional Bose-Hubbard model, Phys. Rev. B **58**, R14741 (1998).
[55] Charles E. Porter and Robert G. Thomas, Fluctuations of nuclear reaction widths, Phys. Rev. **104**, 483 (1956).
[56] Samuel J. Garratt, Zack Weinstein, and Ehud Altman, Measurements Conspire Nonlocally to Restructure Critical Quantum States, Phys. Rev. X **13**, 021026 (2023).
[57] Jong Yeon Lee, Wenjie Ji, Zhen Bi, and Matthew P. A. Fisher, Decoding measurement-prepared quantum phases and transitions: From Ising model to gauge theory, and beyond, arXiv:2208.11699.
[58] Xun Gao, Marcin Kalinowski, Chi-Ning Chou, Mikhail D. Lukin, Boaz Barak, and Soonwon Choi, Limitations of linear cross-entropy as a measure for quantum advantage, arXiv:2112.01657.
[59] Alexander M. Dalzell, Nicholas Hunter-Jones, and Fernando G. S. L. Brandão, Random quantum circuits transform local noise into global white noise, arXiv:2111.14907.
[60] Kyungjoo Noh, Liang Jiang, and Bill Fefferman, Efficient classical simulation of noisy random quantum circuits in one dimension, Quantum **4**, 318 (2020).
[61] Sheldon Goldstein, Joel L. Lebowitz, Roderich Tumulka, and Nino Zanghi, On the distribution of the wave function for systems in thermal equilibrium, J. Stat. Phys. **125**, 1193 (2006).
[62] Peter Reimann, Foundation of Statistical Mechanics under Experimentally Realistic Conditions, Phys. Rev. Lett. **101**, 190403 (2008).
[63] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter, Quantum mechanical evolution towards thermal equilibrium, Phys. Rev. E **79**, 061103 (2009).
[64] Kazuya Kaneko, Eiki Iyoda, and Takahiro Sagawa, Characterizing complexity of many-body quantum dynamics by higher-order eigenstate thermalization, Phys. Rev. A **101**, 042126 (2020).
[65] Yichen Huang, Extensive entropy from unitary evolution, arXiv:2104.02053.
[66] Fabian H. L. Essler, Holger Frahm, Frank Göhmann, Andreas Klümper, and Vladimir E. Korepin, *The One-Dimensional Hubbard Model* (Cambridge University Press, Cambridge, England, 2005).
[67] Andrew J. Daley, Quantum trajectories and open many-body quantum systems, Adv. Phys. **63**, 77 (2014).
[68] Xiao Mi, Pedram Roushan, Chris Quintana, Salvatore Mandrà, Jeffrey Marshall, Charles Neill, Frank Arute, Kunal Arya, Juan Atalaya, Ryan Babbush *et al.*, Information scrambling in quantum circuits, Science **374**, 1479 (2021).