

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,*}

¹Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²California Institute of Technology, Pasadena, CA 91125, USA

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, that 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 multi-parameter estimation of quantum states and processes.

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 non-equilibrium 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

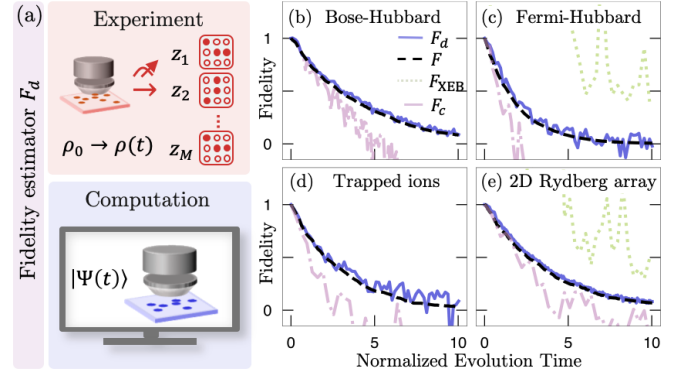


Figure 1. (a) Schematic of our benchmarking protocol. The fidelity estimator F_d is evaluated from experimental snapshots $\{z_1, \dots, z_M\}$ of a state $\rho(t)$ obtained after quench dynamics, compared against classical computation of the ideal state $|\Psi(t)\rangle$ in the absence of error (see Table I). (b-e) Numerical demonstrations. F_d closely tracks the fidelity decay over evolution time normalized in units of Rabi frequency or tunnelling 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 SM [37] 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.

requirements on hardware capability.

In this work, 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 [38]—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

* soonwon@mit.edu

the presence of symmetries, and in non-qubit 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, 39]. 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 computationally-assisted, efficiently measurable observable [40, 41] 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, 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 $\{z_1, \dots, z_M\}$, 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 $p_{\text{avg}}(z) \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T p(z, t) dt$. 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 $\tilde{p}(z, t) \equiv p(z, t)/p_{\text{avg}}(z)$ and the normalization factor $\mathcal{Z}(t) \equiv \sum_z p_{\text{avg}}(z) \tilde{p}(z, t)^2$.

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

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

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. $p_{\text{avg}}(z) \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T p(z, t) dt$,
 $\tilde{p}(z, t) \equiv p(z, t)/p_{\text{avg}}(z)$.
3. $\mathcal{Z}(t) \equiv \sum_z p_{\text{avg}}(z) \tilde{p}(z, t)^2$.

Data processing: Evaluate

$$\hat{F}_d(t) \equiv \frac{2}{M} \left[\sum_{i=1}^M \tilde{p}(z_i, t) \right] / \mathcal{Z}(t) - 1 \approx F_d(t),$$

which approximates the fidelity $F = \langle \Psi(t) | \rho(t) | \Psi(t) \rangle$.

This explicitly defines the statistic $f(z)$, that also depends on $|\Psi_0\rangle$, H , and t . In the limit $M \rightarrow \infty$, this converges to our benchmark $F_d(t) = 2 \left[\sum_z q(z, t) \tilde{p}(z, t) \right] / \mathcal{Z}(t) - 1$. 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 [37, 42–44], and rigorously prove our statement for isolated single errors and long evolution times.

Figure 1(b-e) numerically demonstrates 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 [37]. 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) 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 random unitary circuits (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 z 's, $p(z)$ 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 in-

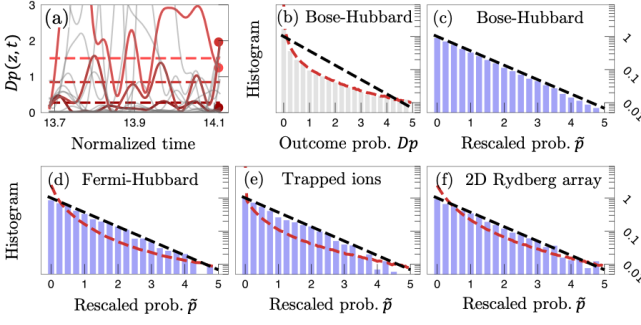


Figure 2. Emergent universal statistics. (a) During many-body Hamiltonian evolution of a pure state, the probability $p(z, t)$ of measuring an outcome z fluctuates around its average value $p_{\text{avg}}(z)$ (dashed, three distinct z 's highlighted). Here, we consider a Bose-Hubbard model and present $p(z, t)$ 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 non-universal (red dashed), far from the Porter-Thomas (PT) distribution (black dashed). (c) Rescaling each $p(z, t)$ by $p_{\text{avg}}(z)$ yields $\tilde{p}(z, t)$, whose histogram follows the PT distribution. (d,e,f) The histograms of $\tilde{p}(z, t)$ (blue bars) follow the universal PT distribution in all models considered in this work, whereas those of the bare $p(z, t)$ rescaled by D are non-universal (red dashed).

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 [Fig. 2(a,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 [37]. 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 $\tilde{p}(z) \equiv p(z)/p_{\text{avg}}(z)$ that follow the PT distribution with mean $\mu = 1$ [Fig. 2(c-f)].

Theorem. (informal) Consider an initial state $|\Psi_0\rangle$, evolved for time t under a Hamiltonian H satisfying the k -th 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 $D_\beta^{-1} \equiv \sum_{z,E} |\langle z|E\rangle|^4 |\langle E|\Psi_0\rangle|^4 / p_{\text{avg}}(z)$, where $\{|E\rangle\}$ are the eigenstates of H .

Rigorous statements of our Theorem and their proofs are presented in the SM [37]. The only assumption of this Theorem is the k -th 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) [45–49]. That is, the eigenvalues $\{E_j\}$ of H possess no resonant structures. This condition is expected to hold for generic ergodic Hamiltonians [45, 46], and we find that it even holds in some integrable systems such as the 1D Fermi-Hubbard model [Fig. 1(c)] [37, 50]. 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 $p(z, t) = p_{\text{avg}}(z) \times \tilde{p}(z, t)$: systematic values $p_{\text{avg}}(z)$ 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 [37]. 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 $\mathcal{Z} = \sum_z p_{\text{avg}}(z) \tilde{p}(z)^2 \approx 2$ 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 $\mathbb{E}_z[\cdot] \equiv \sum_z p_{\text{avg}}(z)[\cdot]$ and $\tilde{p}_\perp \equiv p_\perp/p_{\text{avg}}$. The second property relies on an assumption that the speckle pattern in $\tilde{p}$ significantly changes under any error, which has been rigorously proven for RUCs [42, 43].

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 $F_d \approx F$ [37]. 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 [37]. 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 [37]. Furthermore, we also verify that $F_d \approx F$ holds even for relatively

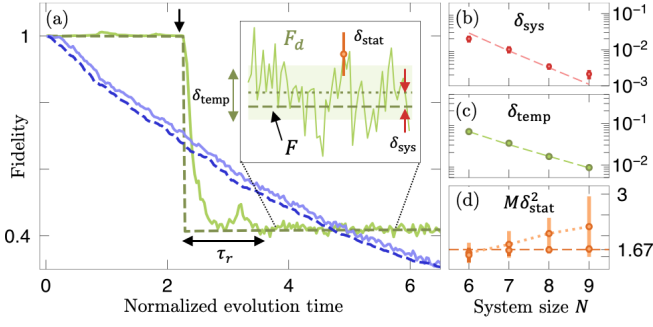


Figure 3. Performance analysis. (a) Numerical simulation of open system dynamics of a 1D Bose-Hubbard model with $N = 9$ particles on N sites ($D = 24310$). When a single error occurs (black arrow), F_d (solid line, green) closely approximates the fidelity F (dashed line, green) after a short delay time 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 $M\delta_{\text{stat}}^2$ 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 [37] for details.

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 Fig. 1(b-e) and the SM.

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 [37].

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 “unravelling” into an ensemble of stochastic pure state trajectories [51] each

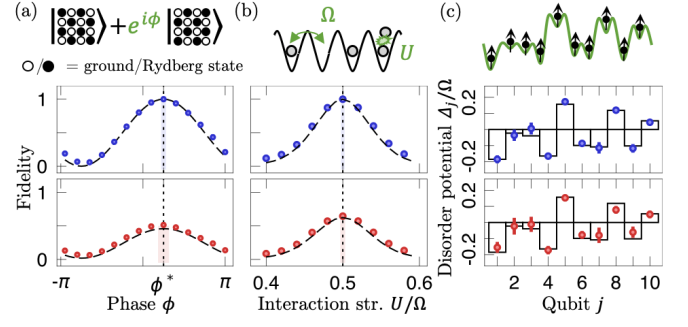


Figure 4. State and process benchmarking with F_d . We numerically simulate the estimation of: (a) the phase ϕ of a GHZ-like initial state in a 2D Rydberg system; (b) the ratio of the interaction and tunneling strengths U/Ω in a 1D Bose-Hubbard model; and (c) ten disordered on-site potentials in a trapped 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 [37] 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).

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 [42, 43].

We showcase the performance of F_d under realistic conditions by numerically simulating the 1D Bose-Hubbard model. See Figure 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 [37] 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 $O(D_\beta^{-1/2})$ [37]. Hence, both the accuracy and precision of our benchmark improve exponentially with increasing system size, explicitly confirmed in our numerical simulations [Fig. 3(b,c)]. Meanwhile, the sample complexity has optimal scaling. It is system size independent for long evolutions: $M\delta_{\text{stat}}^2 \approx 1 + 2F - F^2$ [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 [37, 42, 52].

Limitations.—While our protocol is applicable to generic quantum many-body systems, it may fail in spe-

cial cases in which our Theorem is not applicable. Examples include systems with weakly- or non-ergodic dynamics, or the presence of correlated nonlocal errors. We provide detailed analysis and potential resolutions of known failure cases in the SM [37].

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 Figure 4.

Acknowledgments.— 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 award 1753386 and NSF CAREER award 2237244. Support is also acknowledged from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. JC acknowledges support from the IQIM postdoctoral fellowship. ALS acknowledges support from the Eddleman Quantum graduate fellowship.

-
- [1] M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch, Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms, *Nature (London)* **415**, 39 (2002).
 - [2] M. Aidelsburger, M. Atala, M. Lohse, J. T. Barreiro, B. Paredes, and I. Bloch, Realization of the Hofstadter Hamiltonian with ultracold atoms in optical lattices, *Phys. Rev. Lett.* **111**, 185301 (2013).
 - [3] G. Jotzu, M. Messer, R. Desbuquois, M. Lebrat, T. Uehlinger, D. Greif, and T. Esslinger, Experimental realization of the topological Haldane model with ultracold fermions, *Nature (London)* **515**, 237 (2014).
 - [4] C. Gross and I. Bloch, Quantum simulations with ultracold atoms in optical lattices, *Science* **357**, 995 (2017).
 - [5] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, *et al.*, Quantum phases of matter on a 256-atom programmable quantum simulator, *Nature (London)* **595**, 227 (2021).
 - [6] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Quantum thermalization through entanglement in an isolated many-body system, *Science* **353**, 794 (2016).
 - [7] Y. Tang, W. Kao, K.-Y. Li, S. Seo, K. Mallayya, M. Rigol, S. Gopalakrishnan, and B. L. Lev, Thermalization near integrability in a dipolar quantum Newton's cradle, *Phys. Rev. X* **8**, 021030 (2018).
 - [8] F. Chen, Z.-H. Sun, M. Gong, Q. Zhu, Y.-R. Zhang, Y. Wu, Y. Ye, C. Zha, S. Li, S. Guo, H. Qian, H.-L. Huang, J. Yu, H. Deng, H. Rong, J. Lin, Y. Xu, L. Sun, C. Guo, N. Li, F. Liang, C.-Z. Peng, H. Fan, X. Zhu, and J.-W. Pan, Observation of strong and weak thermalization in a superconducting quantum processor, *Phys. Rev. Lett.* **127**, 020602 (2021).
 - [9] B. Neyenhuis, J. Zhang, P. W. Hess, J. Smith, A. C. Lee, P. Richerme, Z.-X. Gong, A. V. Gorshkov, and C. Monroe, Observation of prethermalization in long-range interacting spin chains, *Science Adv.* **3**, e1700672 (2017).
 - [10] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, *Nature (London)* **551**, 601 (2017).
 - [11] J. Choi, H. Zhou, S. Choi, R. Landig, W. W. Ho, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, D. A. Abanin, and M. D. Lukin, Probing quantum thermalization of a disordered dipolar spin ensemble with discrete time-crystalline order, *Phys. Rev. Lett.* **122**, 043603 (2019).
 - [12] A. Rubio-Abadal, M. Ippoliti, S. Hollerith, D. Wei, J. Rui, S. L. Sondhi, V. Khemani, C. Gross, and I. Bloch, Floquet prethermalization in a Bose-Hubbard system, *Phys. Rev. X* **10**, 021044 (2020).
 - [13] P. Peng, C. Yin, X. Huang, C. Ramanathan, and P. Cappellaro, Floquet prethermalization in dipolar spin chains, *Nature Physics* **17**, 444 (2021).
 - [14] X. Mi, M. Ippoliti, C. Quintana, A. Greene, Z. Chen, J. Gross, F. Arute, K. Arya, J. Atalaya, R. Babbush, *et al.*, Time-crystalline eigenstate order on a quantum processor, *Nature (London)* **601**, 531–536 (2021).
 - [15] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nature Rev. Phys.* **2**, 382 (2020).
 - [16] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
 - [17] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, *Nature Comm.* **1**, 149 (2010).
 - [18] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum state tomography via compressed sensing, *Phys. Rev. Lett.* **105**, 150401 (2010).
 - [19] M. Christandl and R. Renner, Reliable quantum state tomography, *Phys. Rev. Lett.* **109**, 120403 (2012).
 - [20] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, *Phys. Rev. Lett.* **106**, 230501 (2011).
 - [21] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomog-

- raphy, *Phys. Rev. Lett.* **107**, 210404 (2011).
- [22] S. Aaronson, Shadow tomography of quantum states, *SIAM Journal on Computing* **49**, STOC18 (2019).
- [23] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nature Physics* **16**, 1050 (2020).
- [24] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states, *Phys. Rev. A* **99**, 052323 (2019).
- [25] 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**, 010504 (2020).
- [26] Y. Liu, M. Otten, R. Bassirianjahromi, L. Jiang, and B. Fefferman, Benchmarking near-term quantum computers via random circuit sampling, *arXiv:2105.05232* (2021).
- [27] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, *Nature Reviews Physics* , 1 (2022).
- [28] M. Ohliger, V. Nesme, and J. Eisert, Efficient and feasible state tomography of quantum many-body systems, *N. J. Phys.* **15**, 015024 (2013).
- [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] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, *Nature Physics* **14**, 595 (2018).
- [31] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandão, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [32] Z. Li, L. Zou, and T. H. Hsieh, Hamiltonian tomography via quantum quench, *Phys. Rev. Lett.* **124**, 160502 (2020).
- [33] C. Kokail, R. van Bijnen, A. Elben, B. Vermersch, and P. Zoller, Entanglement hamiltonian tomography in quantum simulation, *Nature Physics* **17**, 936 (2021).
- [34] M. Gluza and J. Eisert, Recovering quantum correlations in optical lattices from interaction quenches, *Phys. Rev. Lett.* **127**, 090503 (2021).
- [35] H.-Y. Hu, S. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, *arXiv:2107.04817* (2021).
- [36] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, R. Finkelstein, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale, *et al.*, Preparing random states and benchmarking with many-body quantum chaos, *Nature (London)* **613**, 468 (2023).
- [37] See Supplemental material online for the proof of our theorem, detailed analysis of the performance, numerical demonstration, and limitations of our protocol, which includes Refs. [53–67].
- [38] 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.
- [39] C. E. Porter and R. G. Thomas, Fluctuations of nuclear reaction widths, *Phys. Rev.* **104**, 483 (1956).
- [40] S. J. Garratt, Z. Weinstein, and E. Altman, Measurements conspire nonlocally to restructure critical quantum states, *arXiv:2207.09476* (2022).
- [41] J. Y. Lee, W. Ji, Z. Bi, and M. P. A. Fisher, Decoding measurement-prepared quantum phases and transitions: from ising model to gauge theory, and beyond, *arXiv:2208.11699* (2022).
- [42] X. Gao, M. Kalinowski, C.-N. Chou, M. D. Lukin, B. Barak, and S. Choi, Limitations of linear cross-entropy as a measure for quantum advantage, *arXiv:2112.01657* (2021).
- [43] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits transform local noise into global white noise, *arXiv:2111.14907* (2021).
- [44] K. Noh, L. Jiang, and B. Fefferman, Efficient classical simulation of noisy random quantum circuits in one dimension, *Quantum* **4**, 318 (2020).
- [45] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghi, On the distribution of the wave function for systems in thermal equilibrium, *J. Stat. Phys.* **125**, 1193 (2006).
- [46] P. Reimann, Foundation of statistical mechanics under experimentally realistic conditions, *Phys. Rev. Lett.* **101**, 190403 (2008).
- [47] N. Linden, S. Popescu, A. J. Short, and A. Winter, Quantum mechanical evolution towards thermal equilibrium, *Phys. Rev. E* **79**, 061103 (2009).
- [48] K. Kaneko, E. Iyoda, and T. Sagawa, Characterizing complexity of many-body quantum dynamics by higher-order eigenstate thermalization, *Phys. Rev. A* **101**, 042126 (2020).
- [49] Y. Huang, Extensive entropy from unitary evolution, *arXiv:2104.02053* (2021).
- [50] F. H. Essler, H. Frahm, F. Göhmann, A. Klümper, and V. E. Korepin, *The one-dimensional Hubbard model* (Cambridge University Press, 2005).
- [51] A. J. Daley, Quantum trajectories and open many-body quantum systems, *Adv. Phys.* **63**, 77 (2014).
- [52] X. Mi, P. Roushan, C. Quintana, S. Mandrà, J. Marshall, C. Neill, F. Arute, K. Arya, J. Atalaya, R. Babbush, *et al.*, Information scrambling in quantum circuits, *Science* **374**, 1479 (2021).
- [53] A. J. Short, Equilibration of quantum systems and subsystems, *N. J. Phys.* **13**, 053009 (2011).
- [54] J. S. Cotler, D. K. Mark, H.-Y. Huang, F. Hernández, J. Choi, A. L. Shaw, M. Endres, and S. Choi, Emergent quantum state designs from individual many-body wave functions, *PRX Quantum* **4**, 010311 (2023).
- [55] 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).
- [56] 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).
- [57] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, *et al.*, Probing many-body dynamics on a 51-atom quantum simulator, *Nature (London)* **551**, 579 (2017).
- [58] R. Vasseur and J. E. Moore, Nonequilibrium quantum

- dynamics and transport: from integrability to many-body localization, *J. Stat. Mech.* **2016**, 064010 (2016).
- [59] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80**, 012304 (2009).
- [60] J. M. Deutsch, Eigenstate thermalization hypothesis, *Rep. Prog. Phys.* **81**, 082001 (2018).
- [61] B. Collins and I. Nechita, Random matrix techniques in quantum information theory, *Journal of Mathematical Physics* **57**, 015215 (2016).
- [62] A. Nahum, S. Vijay, and J. Haah, Operator spreading in random unitary circuits, *Phys. Rev. X* **8**, 021014 (2018).
- [63] V. Khemani, A. Vishwanath, and D. A. Huse, Operator spreading and the emergence of dissipative hydrodynamics under unitary evolution with conservation laws, *Phys. Rev. X* **8**, 031057 (2018).
- [64] D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller, Cold bosonic atoms in optical lattices, *Phys. Rev. Lett.* **81**, 3108 (1998).
- [65] 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).
- [66] D. Bluvstein, A. Omran, H. Levine, A. Keesling, G. Semeghini, S. Ebadi, T. T. Wang, A. A. Michailidis, N. Maskara, W. W. Ho, *et al.*, Controlling quantum many-body dynamics in driven rydberg atom arrays, *Science* **371**, 1355 (2021).
- [67] T. D. Kühner and H. Monien, Phases of the one-dimensional Bose-Hubbard model, *Phys. Rev. B* **58**, R14741 (1998).
- [68] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, *Journal of mathematical physics* **48**, 052104 (2007).
-

Supplemental Material for: Benchmarking quantum simulators using ergodic quantum dynamics

CONTENTS

A. Previous fidelity benchmarks	9
B. Emergence of universal statistical properties from Hamiltonian dynamics	9
1. k -th no-resonance condition	9
2. k -th no resonance with degeneracies	9
3. k -th Hamiltonian twirling identity	10
4. Porter-Thomas distribution over time	11
5. Porter-Thomas distribution over outcomes	14
C. Detailed performance analysis for long quench dynamics	15
1. Ansatz for how speckle patterns change under errors	16
2. Systematic error	17
a. Argument based on Eigenstate Thermalization Hypothesis	17
b. Extensions beyond local errors	19
3. Variance over time	19
4. Sample complexity	20
5. Moderate error rate: Effects of delay time τ_d	20
6. Large error rate: Critical error rate and operator scrambling time	22
7. Symmetry Resolution	23
8. Alternative fidelity estimator F_e	23
D. Detailed performance analysis for short quench dynamics	24
1. Overview: random matrix product states	24
2. Remarks	25
3. Derivation of result	25
E. Details of numerical simulations	26
1. Bose-Hubbard model	26
a. Disorder Averaging	27
2. Fermi-Hubbard model	27
3. Trapped ion model	27
4. Rydberg blockaded model	27
F. Limitations of benchmarking protocol: Failure cases	28
1. Violating the no-resonance condition: Particle-hole symmetry	28
a. Real Porter-Thomas distribution	29
2. Violating the no-resonance condition: Non-interacting models	29
3. Small effective dimension: Low temperature quench	29
4. Slow equilibration: Quantum many-body scar states	30
5. Slow equilibration: Inhomogeneous initial state	30
6. Non-local error on low entanglement states	31
7. Non-ergodicity due to symmetry and readout in symmetry-resolved basis	32
G. Greedy algorithm for multiparameter estimation	32
1. Algorithm	32
H. Target state benchmarking	33
1. Coherent local errors	33
2. Comparing non-locally deformed states	33

Appendix A: Previous fidelity benchmarks

The cross-entropy benchmark was introduced in Ref. [30] and was prominently used in Ref. [31] to demonstrate quantum computational advantage. It is defined as

$$F_{\text{XEB}} \equiv D \sum_z p(z)q(z) - 1, \quad (\text{S1})$$

where as in the main text, $p(z)$ and $q(z)$ are the probabilities of obtaining the outcome z from the ideal simulated state and experimental states respectively, and D is the Hilbert space dimension, here $D = 2^N$ where N is the number of qubits. F_{XEB} has been argued to estimate the fidelity in deep random unitary circuits (RUCs) [30, 31, 42, 43].

The estimator F_c was introduced in Ref. [36], defined as

$$F_c \equiv 2 \frac{\sum_z p(z)q(z)}{\sum_z p(z)^2} - 1. \quad (\text{S2})$$

F_c is a slight modification of F_{XEB} to reliably estimate the fidelity in shallow circuits. This is because of the onset of “local scrambling” even in shallow circuits, discussed in Ref. [54]. In Ref. [36], F_c was experimentally demonstrated to estimate the fidelity in a Rydberg atom array quantum simulator, at infinite effective temperature and short quench evolution times. F_d can be taken as a further improvement of F_c , that works for arbitrary quench initial states and Hamiltonians.

Appendix B: Emergence of universal statistical properties from Hamiltonian dynamics

In this section, we present precise statements of our Theorem in the main text and prove them. The significance of our Theorems 1 and 2 is that a Porter-Thomas distribution emerges from measurement outcomes $p(z, t) = |\langle z | \Psi(t) \rangle|^2$ after generic Hamiltonian dynamics, upon appropriate data processing. To prove our theorems, we first establish a new identity, the *k-th Hamiltonian twirling identity*, which only requires one assumption: the eigenvalues of the Hamiltonian H satisfy the *k-th no-resonance condition*. We utilize this identity to establish theorems on the emergence of a Porter-Thomas distribution in the rescaled measurement outcomes $\tilde{p}(z, t)$.

1. k-th no-resonance condition

Definition (k-th no-resonance): Let E_j be the energy eigenvalues of a Hamiltonian H . H satisfies the *k-th no-resonance condition* [46–49] if

$$E_{\alpha_1} + E_{\alpha_2} + \cdots + E_{\alpha_k} = E_{\beta_1} + E_{\beta_2} + \cdots + E_{\beta_k} \quad (\text{S3})$$

implies that the indices $(\alpha_1, \alpha_2, \dots, \alpha_k)$ and $(\beta_1, \beta_2, \dots, \beta_k)$ are equal up to reordering. In other words, $\beta_j = \alpha_{\sigma(j)}$ for some permutation $\sigma \in S_k$, where S_k is the group of permutations of k elements (the symmetric group). It is easy to see that if the *k-th no-resonance condition* is satisfied, then so are the *k'-th no-resonance conditions*, for $k' < k$. This assumption is considered generically satisfied: it is believed that if the condition does not hold in a given ergodic system, any small perturbation will generically break any *k-th resonances* [48].

2. k-th no resonance with degeneracies

The *k-th no-resonance conditions* are, strictly speaking, violated if the Hamiltonian contains degeneracies: $E_i = E_j, i \neq j$. This violates the first no-resonance condition and so all higher *k-th no-resonance conditions*. This seems to preclude systems with symmetry-protected degeneracies, which can still be ergodic. For our purposes, however, we will consider *pure states* $\rho = |\Psi_0\rangle\langle\Psi_0|$. In this case, one can replace any set of degenerate states $\mathcal{E} = \{|\Psi_E\rangle\}$ with a *single* representative state, defined as the projection of $|\Psi_0\rangle$ onto the degenerate subspace: $\mathcal{P}_{\mathcal{E}} |\Psi_0\rangle$. In this way, all degeneracies are effectively eliminated. Therefore, while Lemma 1 (below) requires the full *k-th no-resonance condition*, one can modify this condition by demanding the *k-th no-resonance condition* only after eliminating all degenerates as prescribed above. Our main Theorems only require this weakened version of the *k-th no-resonance condition*. Hence, our Theorems holds generically even in the presence of degeneracies protected by symmetries.

The k -th no-resonance condition is also notably violated for $k \geq 2$ if there is a *particle-hole symmetry*, in which the energy spectrum is symmetric about $E = 0$. This case is discussed in Section F 1.

We finally remark that there are other, indirect ways in which degeneracies can affect the performance of our benchmark. For example, if there is a large degeneracy of eigenstates, the quench dynamics will be less ergodic and our protocol will have larger uncertainties. However, this is entirely captured by our effective dimension parameter D_β , and no further modification is required.

3. k -th Hamiltonian twirling identity

We first present the k -th Hamiltonian twirling identity. This identity is a rewriting of a k -fold quantum channel: we take a state $\rho^{(k)}$ in k -copies of the Hilbert space $\mathcal{H}^{\otimes k}$, and average it over all possible times of time-evolution $U_t^{\otimes k}$ over k copies, with $U_t \equiv \exp(-iHt)$. That is, the channel $C^{(k)}$ is $C^{(k)}[\rho^{(k)}] = \mathbb{E}_t [U_t^{\otimes k} \rho^{(k)} U_{-t}^{\otimes k}]$ where $\mathbb{E}_t[f(t)] \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(t) dt$ here and below. We dub this the k -th Hamiltonian twirling channel, in analogy to twirling channels, in which a state is averaged over all possible unitaries in a group, most commonly the ensemble of Haar unitaries or the Clifford or Pauli groups [59].

We find that the k -th Hamiltonian twirling channel can be written as sum of (i) a dephasing operator in the energy eigenbasis, followed by a sum of permutations of k elements and (ii) additional terms associated with the double counting of certain permutations. We associate these terms with additional correlations associated with energy conservation, hence we dub these “energy-conservation terms”.

Such a rewriting is motivated by *Haar-unitary twirling channels*, which can be written as a sum of permutations [68]. In our Theorem below, we consider a specific quantity: $\mathbb{E}_t[p^k(z, t)] = \langle z |^{\otimes k} C^{(k)}[|\Psi_0\rangle\langle\Psi_0|^{\otimes k}] |z\rangle^{\otimes k}$. In this case, the terms in (i) are the dominant contributions and we are able to bound the magnitude of the energy-conservation terms in (ii).

We explicitly state our identity for $k = 2$.

Lemma (Second Hamiltonian twirling identity). *If a Hamiltonian H satisfies the second no-resonance condition, the following identity holds:*

$$\begin{aligned} C^{(2)}[\rho^{(2)}] &\equiv \mathbb{E}_t [U_t^{\otimes 2} \rho^{(2)} U_{-t}^{\otimes 2}] \\ &= \sum_{E, E'} \left[|E, E'\rangle\langle E, E'| \rho^{(2)} |E, E'\rangle\langle E, E'| + |E, E'\rangle\langle E, E'| \rho^{(2)} |E', E\rangle\langle E', E| \right] - \sum_E |E, E\rangle\langle E, E| \rho^{(2)} |E, E\rangle\langle E, E| \\ &= \mathcal{D}^{\otimes 2}[\rho^{(2)}] \mathbb{I} + \mathcal{D}^{\otimes 2}[\rho^{(2)} \mathcal{S}] \mathcal{S} - \sum_E |E, E\rangle\langle E, E| \rho^{(2)} |E, E\rangle\langle E, E|, \end{aligned} \quad (\text{S4})$$

where $U_t \equiv \exp(-iHt)$, $\mathbb{E}_t[f(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(t) dt$ is the average of $f(t)$ over infinite time. $\rho^{(2)}$ is any state on two copies of the Hilbert space, and $|E\rangle$ are energy eigenstates of the Hamiltonian. $\mathbb{I}$ and $\mathcal{S}$ are the identity and swap operators on $\mathcal{H}^{\otimes 2}$, defined as $\mathbb{I}(|\psi_1\rangle \otimes |\psi_2\rangle) = |\psi_1\rangle \otimes |\psi_2\rangle$ and $\mathcal{S}(|\psi_1\rangle \otimes |\psi_2\rangle) = |\psi_2\rangle \otimes |\psi_1\rangle$ for any $|\psi_1\rangle, |\psi_2\rangle$. Lastly, $\mathcal{D}^{\otimes 2}$ is the dephasing channel in the energy eigenbasis: $\mathcal{D}^{\otimes 2}[\rho^{(2)}] \equiv \sum_{E, E'} |E, E'\rangle\langle E, E'| \rho^{(2)} |E, E'\rangle\langle E, E'|$. Finally, we remark that the state $\rho^{(2)}$ on $\mathcal{H}^{\otimes 2}$ can be generalized to any operator A acting on $\mathcal{H}^{\otimes 2}$.

Proof.

$$\mathbb{E}_t [U_t^{\otimes 2} \rho^{(2)} U_{-t}^{\otimes 2}] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt U_t^{\otimes 2} \rho^{(2)} U_{-t}^{\otimes 2} \quad (\text{S5})$$

$$= \sum_{E_1, E_2, E_3, E_4} \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt e^{-i(E_1 - E_2 + E_3 - E_4)t} |E_1, E_3\rangle\langle E_1, E_3| \rho^{(2)} |E_2, E_4\rangle\langle E_2, E_4| \right] \quad (\text{S6})$$

$$= \sum_{E_1, E_2, E_3, E_4} \delta(E_1 - E_2 + E_3 - E_4) |E_1, E_3\rangle\langle E_1, E_3| \rho^{(2)} |E_2, E_4\rangle\langle E_2, E_4| \quad (\text{S7})$$

$$\stackrel{2\text{-NR}}{=} \sum_{E, E'} \left[|E, E'\rangle\langle E, E'| \rho^{(2)} |E, E'\rangle\langle E, E'| + |E, E'\rangle\langle E, E'| \rho^{(2)} |E', E\rangle\langle E', E| \right] - \sum_E |E, E\rangle\langle E, E| \rho^{(2)} |E, E\rangle\langle E, E|, \quad (\text{S8})$$

where “(2-NR)” indicates the use of the second no-resonance condition. The second no-resonance condition states that the delta function $\delta(E_1 - E_2 + E_3 - E_4)$ is non-zero in only two cases: $E_1 = E_2, E_3 = E_4$ or $E_1 = E_4, E_3 = E_2$, which

give the two terms above. The final, energy-conservation term in Eq. (S8) corresponds to $E_1 = E_2 = E_3 = E_4$. This special case belongs to the above two terms simultaneously and is double counted. Therefore, the energy-conservation term is subtracted to correct for this double counting. $\square$

A diagrammatic representation of this Lemma is shown in Fig. S1(a,b). The energy-conservation terms are illustrated in Fig. S1(c). As with the second Hamiltonian twirling identity, one can define the k -th twirling identity.

Lemma 1 (k -th Hamiltonian twirling identity). *If the Hamiltonian satisfies the k -th no-resonance condition, its k -th twirling channel is:*

$$\begin{aligned} C^{(k)}[\rho^{(k)}] &\equiv \mathbb{E}_t \left[U_t^{\otimes k} \rho^{(k)} U_{-t}^{\otimes k} \right] \\ &= \sum_{M=(E_1, \dots, E_k)} \left[\sum_{\sigma \in \pi(M)} |E_1, \dots, E_k\rangle\langle E_1, \dots, E_k| \rho^{(k)} |E_{\sigma(1)}, \dots, E_{\sigma(k)}\rangle\langle E_{\sigma(1)}, \dots, E_{\sigma(k)}| \right], \end{aligned} \quad (\text{S9})$$

$$= \mathcal{D}^{\otimes k}[\rho^{(k)} \hat{\sigma}^{-1}] \sum_{\sigma \in S_k} \hat{\sigma} + \text{E.C.} \quad (\text{S10})$$

where $\mathbb{E}_t[f(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(t) dt$ is averaging over infinite time, $\rho^{(k)}$ is any state on k copies of the Hilbert space $\mathcal{H}^{\otimes k}$, and $\hat{\sigma}$ —with $\sigma \in S_k$ —is a permutation operator acting on the k -copy Hilbert space:

$$\hat{\sigma} = \sum_{\{E_i\}} |E_1, E_2, \dots, E_k\rangle\langle E_{\sigma(1)}, E_{\sigma(2)}, \dots, E_{\sigma(k)}|. \quad (\text{S11})$$

In (S9), the sum is taken over all possible lists of k energies $M = (E_1, \dots, E_k)$. If M has r unique energies with multiplicities $n_1, \dots, n_r$ (such that $\sum_{i=1}^r n_i = k$), its group of permutations $\pi(M)$ has $k! / \prod_{i=1}^r n_i!$ elements and is a subgroup of the symmetric group S_k . In (S10), this identity is re-written in terms of permutations, the dephasing channel $\mathcal{D}^{\otimes k}[\rho^{(k)}] \equiv \sum_{E_1, \dots, E_k} |E_1, \dots, E_k\rangle\langle E_1, \dots, E_k| \rho^{(k)} |E_1, \dots, E_k\rangle\langle E_1, \dots, E_k|$, and energy-conservation terms E.C. associated with double, or multiple, counting (similar to the $k = 2$ case). All terms can be explicitly derived from (S9).

The proof of the k -th twirling identity is analogous to that of the second twirling identity: in this case, there are $k!$ permutations relevant for the k -th no resonance theorem, along with more energy-conservation terms than the simplest one in the $k = 2$ case. As an example, the third twirling identity is illustrated diagrammatically in Fig. S1(d). In addition to the energy-conservation terms associated with cases where two energies are the same $E_i = E_j$, there is a higher order energy-conservation term when all three energies are the same: $E_1 = E_2 = E_3$.

Lastly, we note that this identity also holds in Floquet systems, as long the k -th no resonance condition is tightened to $\sum_i E_{\alpha_i} \equiv \sum_i E_{\beta_i} \pmod{2\pi} \iff \{\alpha_i\} = \{\beta_i\}$, and the integral over all time is replaced by a discrete sum over integer Floquet periods nT .

4. Porter-Thomas distribution over time

We will use the k -th Hamiltonian twirling identity, with $\rho^{\otimes k} = |\Psi_0\rangle\langle\Psi_0|^{\otimes k}$, to prove that the quantity $\tilde{p}(z, t) \equiv p(z, t)/p_{\text{avg}}(z)$ asymptotically follows the *Porter-Thomas*, or exponential distribution [39].

The Porter-Thomas (PT) distribution naturally arises from the distribution of overlaps p of the Haar ensemble — the unique distribution of states on the Hilbert space that is invariant under any unitary — to a fixed vector $|\Phi\rangle$: $p = |\langle\Psi|\Phi\rangle|^2$, with $|\Psi\rangle$ drawn from the Haar ensemble. A given $|\Psi\rangle$ and $|\Phi\rangle$ have overlap $|\langle\Psi|\Phi\rangle|^2$. The probability density of obtaining an overlap with value p is $P(p) = \int d\Psi \delta(|\langle\Psi|\Phi\rangle|^2 - p)$. In deep random circuits, the wavefunction $|\Psi\rangle$ is approximately Haar-random and we can evaluate $P(p)$ for any Hilbert space dimension D , which converges to the Porter-Thomas distribution in the limit $D \rightarrow \infty$.

$$P(p)dp = (D-1)(1-p)^{D-2}dp \xrightarrow{D \rightarrow \infty} \exp(-Dp)d(Dp), \quad (\text{S12})$$

It is natural to rescale the quantity p by D . In this case, the exponential distribution has a k -th moment $\int_0^\infty (Dp)^k P(Dp)d(Dp) = k!$.

We formally state the Theorem from the main text in Theorems 1 and 2 below. Informally, our theorems show that the rescaled distribution $\tilde{p}(z, t) \equiv p(z, t)/p_{\text{avg}}(z)$ asymptotically follows the PT distribution. As a reminder, $p(z, t) = |\langle z|\Psi(t)\rangle|^2$ is the probability of measuring outcome z , and $p_{\text{avg}}(z) = \mathbb{E}_t[p(z, t)]$ is its time-averaged value.

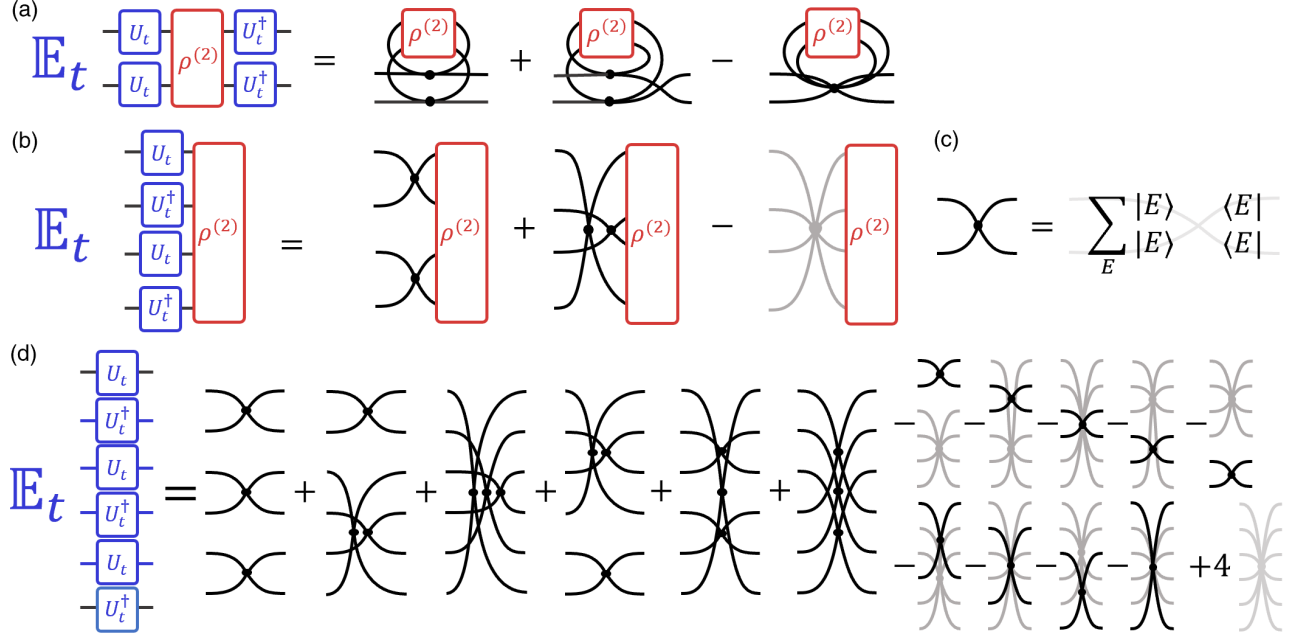


Figure S1. Second and third Hamiltonian twirling identities. (a) The second Hamiltonian twirling identity comprises identity and swap permutations terms, along with an *energy-conservation* correction term that in our context will have an exponentially small contribution. (b) The same as (a), but open legs representing the channel input/outputs are rearranged in order to simplify the diagram. We keep this convention for the rest of diagrams in this document. (c) Diagrammatic notation for the basic building block in our diagrams: the dephasing operator in the energy eigenbasis: $D[\rho] = \sum_E |E\rangle\langle E| \rho |E\rangle\langle E|$. (d) The third Hamiltonian twirling identity comprises of six permutation and several energy-conservation terms. For our purposes, these additional terms will have small contributions.

By showing that the k -th moments of $\tilde{p}(z, t)$ are close to $k!$, we show that $\tilde{p}(z, t)$ asymptotically approaches the Porter-Thomas distribution. We can choose to fix the bitstring z , and consider the ensemble over times t [Theorem 1, Fig. S2(d,e)], or fix a large but arbitrary time t , and consider the ensemble of $\tilde{p}(z, t)$ over outcomes z [Theorem 2, Fig. S2(d,f)]. For brevity, we will occasionally denote these ensembles as $\{\tilde{p}(z, t)\}_t$ and $\{\tilde{p}(z, t)\}_z$ respectively, with the subscript denoting the variable over which the ensemble is formed (the other variable is kept fixed). Formally stated, our first Theorem is the following:

Theorem 1 (Temporal Porter-Thomas distribution). *For an initial state $|\Psi_0\rangle$ and a Hamiltonian H satisfying the k -th no-resonance condition, the temporal k -th moment of $\tilde{p}(z, t)$ (with fixed z), approaches a universal value $\mathbb{E}_t [\tilde{p}(z, t)^k] = k! + O(D_\beta^{-1}(z))$. $D_\beta(z)$ is the effective Hilbert space dimension of the outcome z defined as $D_\beta^{-1}(z) \equiv (\sum_E |\langle z|E\rangle|^4 |\langle E|\Psi_0\rangle|^4) / p_{\text{avg}}(z)^2$.*

Proof. Our proof is a direct application of our k -th Hamiltonian twirling identity on the temporal k -th moment $\mathbb{E}_t [\tilde{p}(z, t)^k] = k! + O(D_\beta^{-1})$: To compute $\mathbb{E}_t [\tilde{p}(z, t)^k]$, we apply Lemma 1 on $\rho^{(k)} = |\Psi_0\rangle\langle\Psi_0|^{\otimes k}$:

$$\mathbb{E}_t [\tilde{p}(z, t)^k] = \frac{\langle z|^{\otimes k} \mathbb{E}_t [U_t^{\otimes k} |\Psi_0\rangle\langle\Psi_0|^{\otimes k} U_{-t}^{\otimes k}] |z\rangle^{\otimes k}}{p_{\text{avg}}(z)^k} \quad (\text{S13})$$

$$= \frac{\sum_{M=(E_1, \dots, E_k)} \sum_{\sigma \in \pi(M)} \left[\prod_{i=1}^k |\langle z|E_i\rangle \langle E_{\sigma_i}|\Psi_0\rangle|^2 \right]}{p_{\text{avg}}(z)^k} = \frac{\sum_{M=(E_1, \dots, E_k)} |\pi(M)| \left[\prod_{i=1}^k |\langle z|E_i\rangle \langle E_i|\Psi_0\rangle|^2 \right]}{p_{\text{avg}}(z)^k} \quad (\text{S14})$$

where we have used Eq. (S9). Now using Eq. (S10), where the contribution from the first term is equivalent to replacing $|\pi(M)|$ with $k!$. This gives:

$$\frac{\langle z|^{\otimes k} \mathcal{D}^{\otimes k} [|\Psi_0\rangle\langle\Psi_0|^{\otimes k}] [\sum_{\sigma \in S_k} \hat{\sigma}] |z\rangle^{\otimes k}}{p_{\text{avg}}(z)^k} = k! \frac{\sum_{\{E_j\}} \prod_{j=1}^k |\langle z|E_j\rangle \langle E_j|\Psi_0\rangle|^2}{p_{\text{avg}}(z)^k} = k! , \quad (\text{S15})$$

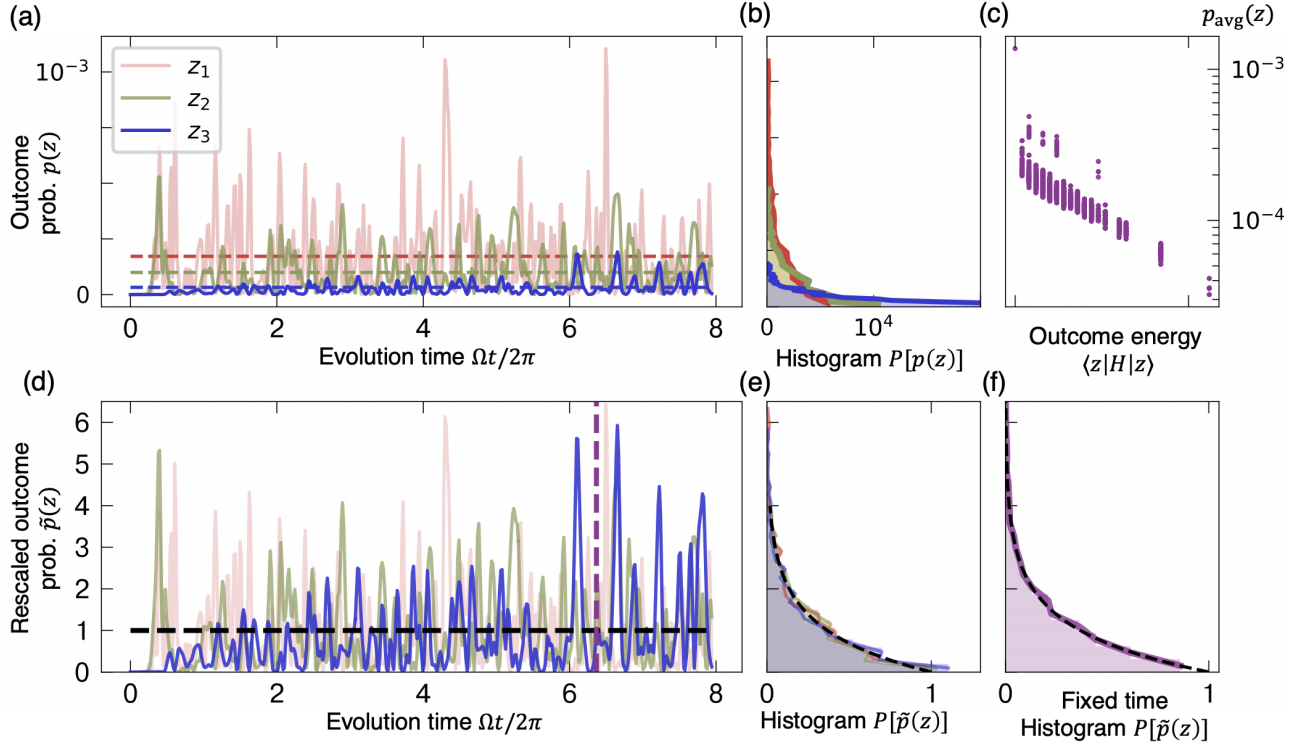


Figure S2. (a) Outcome probabilities $p(z, t)$ over time, for three outcomes z_1, z_2, z_3 in a Bose-Hubbard model with 8 particles on 8 sites in red, green and blue respectively. For each z , $p(z, t)$ fluctuates about its mean value $p_{\text{avg}}(z)$ (dashed lines). (b) Plotting the histogram of $p(z, t)$ over time, each histogram follows an exponential distribution with differing means $p_{\text{avg}}(z)$. (c) $p_{\text{avg}}(z)$ is approximately exponentially dependent on the outcome energy $\langle z|H|z \rangle$, analogous to a Boltzmann distribution. (d) The rescaled outcome probabilities $\tilde{p}(z, t)$ follow (e) a universal exponential (Porter Thomas) distribution, which follows from Theorem 1. (f) Taking a histogram at fixed time [purple, (d)] over bitstrings z gives the same distribution (Theorem 2).

where in the first equality we used the fact that $\hat{\sigma}$ operator acts trivially on the product state $|z\rangle^{\otimes k}$ and in the second equality, we used the relation

$$p_{\text{avg}}(z) \equiv \mathbb{E}_t [p(z, t)] = \sum_E |\langle z|E\rangle \langle E|\Psi_0\rangle|^2. \quad (\text{S16})$$

This relation is used several times, motivating our definition of the quantity

$$q_z(E) \equiv \frac{|\langle z|E\rangle \langle E|\Psi_0\rangle|^2}{p_{\text{avg}}(z)} \quad (\text{S17})$$

In fact, $q_z(E)$ has the properties of a probability distribution over the energies E : $0 \leq q_z(E) \leq 1$ and $\sum_E q_z(E) = 1$.

Now we wish to bound the deviation $|\mathbb{E}_t [\tilde{p}(z, t)^k] - k!|$, that is the contribution from all energy-conservation (E.C.) terms. To do so, we note the E.C. terms must come from lists of energies $M = (E_1, \dots, E_k)$ with at least one repeated energy. That is:

$$\begin{aligned} |\mathbb{E}_t [\tilde{p}(z, t)^k] - k!| &\equiv \left| \frac{\langle z|^{\otimes k} (\text{E.C.}) |z\rangle^{\otimes k}}{p_{\text{avg}}(z)^k} \right| = \sum_{\substack{M=(E_1, \dots, E_k) \\ \text{not all } E_j \text{ distinct}}} |\pi(M) - k!| \prod_{j=1}^k q_z(E_j) \\ &\leq k! \left[\sum_{E_1=E_2} \sum_{E_3, \dots, E_k} \prod_{j=1}^k q_z(E_j) + \sum_{E_1=E_3} \sum_{E_2, E_4, \dots, E_k} \prod_{j=1}^k q_z(E_j) + \dots \right] \end{aligned} \quad (\text{S18})$$

In the last line, we have used two separate inequalities: we bound the difference $|\pi(M) - k!| \leq k!$. This enables us to contain all possible lists M with at least one repeated energy in the sums above, where we have only explicitly demanded that one pair of energies coincide: $E_a = E_b$. There are $\binom{k}{2}$ choices of a and b . Each choice has the same

value. Therefore without loss of generality, we set $a = 1$ and $b = 2$. Then we obtain:

$$|\mathbb{E}_t [\tilde{p}(z, t)^k] - k!| \leq k! \binom{k}{2} \sum_{E_1} q_z(E_1)^2 \prod_{j=3}^k \sum_{E_j} q_z(E_j) = k! \binom{k}{2} \sum_{E_1} q_z(E_1)^2 \equiv k! \binom{k}{2} D_\beta^{-1}(z). \quad (\text{S19})$$

Therefore, we obtain the desired bound

$$|\mathbb{E}_t [\tilde{p}(z, t)^k] - k!| \leq k! \binom{k}{2} D_\beta^{-1}(z) = O(D_\beta^{-1}(z)). \quad (\text{S20})$$

□

In fact, we expect this bound to be close to tight: the contribution from the simplest energy-conservation terms — those with exactly one repeated energy — is close to our bound. Other energy-conservation terms are expected to be exponentially smaller than these simplest terms. Here, the factor $\pi(M) = k!/2$, and we obtain:

$$\mathbb{E}_t [\tilde{p}(z, t)^k] \approx k! \left(1 - \frac{k(k-1)}{2} \frac{D_\beta^{-1}(z)}{2} \right). \quad (\text{S21})$$

Eq. (S21) agrees with the leading order (in D^{-1}) correction to the k -th moment of the Porter-Thomas distribution:

$$\int_0^1 dp p^k (D-1)(1-p)^{D-2} = k! \frac{D^k}{D \cdots (D+k-1)} = k! \left[1 - \frac{k(k-1)}{2} D^{-1} + O(D^{-2}) \right]. \quad (\text{S22})$$

Therefore, it is apt to call $D_\beta(z)$ as the effective dimension: the deviation of the k -th moment from $k!$ for the temporal distribution is the same as that of the Porter-Thomas distribution for a system with finite dimension $2D_\beta$. $\{\tilde{p}(z, t)\}_t$ approximately follows a Porter-Thomas distribution over a vector space of dimension $2D_\beta(z)$. Note, however, that this agreement does not extend to corrections of higher orders in D^{-1} .

For a fixed z , the inverse effective dimension $D_\beta^{-1}(z)$ can be understood as the *collision probability* associated with $q_z(E)$. More explicitly, we can define $D_\beta^{-1}(z) = \sum_E q_z(E)^2$ such that $D_\beta(z)$ describes the participation ratio (or equivalently the size of effective Hilbert space dimension) corresponding to a particular bitstring z . Then the overall inverse participation ratio (inverse effective dimension) D_β^{-1} is the weighted average: $D_\beta^{-1} = \sum_z p_{\text{avg}}(z) D_\beta^{-1}(z)$.

5. Porter-Thomas distribution over outcomes

Theorem 2 concerns the distribution of $\tilde{p}(z, t)$ for fixed t , treated as samples labeled by z .

Theorem 2 (Porter-Thomas distribution over outcomes). *Consider an initial state $|\Psi_0\rangle$ and a Hamiltonian H satisfying the $2k$ -th no-resonance condition. At a typical time $t \in [0, \infty)$, the weighted k -th moment of $\tilde{p}(z, t)$ approaches a universal value: $P^{(k)}(t) \equiv \sum_z p_{\text{avg}}(z) \tilde{p}(z, t)^k = k! + O(D_\beta^{-1/2})$, where $D_\beta^{-1} \equiv \sum_{z, E} |\langle z|E\rangle|^4 |\langle E|\Psi_0\rangle|^4 / p_{\text{avg}}(z)$ is the inverse effective Hilbert space dimension associated with the quench evolution, and $\{|E\rangle\}$ are the eigenstates of H .*

Proof. Our proof strategy is the following. We first compute the *time-averaged* value $\mathbb{E}_t [P^{(k)}(t)]$. From Theorem B, this quantity is $k! + O(D_\beta^{-1})$. We next compute the fluctuations of $P^{(k)}(t)$ over time, and we will find that $\text{std}_t [P^{(k)}(t)] = O(D_\beta^{-1/2})$. Combining these two results leads to our desired conclusion: $P^{(k)}(t) = k! + O(D_\beta^{-1/2})$ with high probability.

In Fig. S3(a), we illustrate this computation of $\mathbb{E}_t [P^{(k)}(t)]$ for the case $k = 2$. In general, we have:

$$|\mathbb{E}_t [P^{(k)}(t)] - k!| \leq \sum_z p_{\text{avg}}(z) |\mathbb{E}_t [\tilde{p}(z, t)^k] - k!| \leq \binom{k}{2} k! \sum_z p_{\text{avg}}(z) D_\beta^{-1}(z) = O(D_\beta^{-1}) \quad (\text{S23})$$

Up to this point, we have only used the k -th Hamiltonian twirling identity. To estimate the temporal fluctuations of $P^{(k)}(t)$, we need the $2k$ -th twirling identity.

$$\text{var}_t [P^{(k)}(t)] = \mathbb{E}_t [P^{(k)}(t)^2] - \mathbb{E}_t [P^{(k)}(t)]^2 = \sum_{z, z'} p_{\text{avg}}(z) p_{\text{avg}}(z') \frac{\mathbb{E}_t [p(z, t)^k p(z', t)^k] - \mathbb{E}_t [p(z, t)^k] \mathbb{E}_t [p(z', t)^k]}{p_{\text{avg}}(z)^k p_{\text{avg}}(z')^k} \quad (\text{S24})$$

By using the Hamiltonian k -th and $2k$ -th twirling identities, any term in $\mathbb{E}_t[P^{(k)}(t)^2]$ that does not cancel with a term in $\mathbb{E}_t[P^{(k)}(t)]^2$ must involve a permutation between z and z' [illustrated for $k = 2$ in Fig. S3(b)]. As in our proof of Theorem 1, without loss of generality, we let the energies corresponding to this permutation be E_1 and E_2 , and multiply our overall expression by a factor of $2\binom{2k}{2}$ (here, E_1 and E_2 are inequivalent). Then their sum is of the form

$$\text{var}_t[P^{(k)}(t)] = \sum_{z,z'} p_{\text{avg}}(z)p_{\text{avg}}(z') 2\binom{2k}{2} \frac{\sum_{E_1,E_2} |\langle E_1|\Psi_0\rangle|^2 \langle z|E_1\rangle \langle E_1|z'\rangle |\langle E_2|\Psi_0\rangle|^2 \langle z'|E_2\rangle \langle E_2|z\rangle}{p_{\text{avg}}(z)p_{\text{avg}}(z')} (\dots) \quad (\text{S25})$$

$$\equiv \sum_{z,z'} p_{\text{avg}}(z)p_{\text{avg}}(z') 2\binom{2k}{2} \frac{\langle z|\rho_d|z'\rangle \langle z'|\rho_d|z\rangle}{p_{\text{avg}}(z)p_{\text{avg}}(z')} (\dots), \quad (\text{S26})$$

$$(\dots) = \sum_{\text{perms. } j=3} \prod_{j=3}^{2k} \frac{\sum_{E_j} |\langle E_j|\Psi_0\rangle|^2 \langle z_j^L|E_j\rangle \langle E_j|z_j^R\rangle}{\sqrt{p_{\text{avg}}(z_j^L)p_{\text{avg}}(z_j^R)}} \equiv \sum_{\text{perms. } j=3} \prod_{j=3}^{2k} \frac{\langle z_j^L|\rho_d|z_j^R\rangle}{\sqrt{p_{\text{avg}}(z_j^L)p_{\text{avg}}(z_j^R)}} \leq \sum_{\text{perms.}} 1 \leq (2k-2)! \quad (\text{S27})$$

where z_j^L and z_j^R can either be z or z' , such that there exactly $k-1$ of the $\{z_j^L\}$ are z and the other $k-1$ are z' , and likewise for the $\{z_j^R\}$. For brevity, we have also rewritten our expressions with $\rho_d \equiv \sum_E |\langle E|\Psi_0\rangle|^2 |E\rangle\langle E|$, the commonly studied *diagonal ensemble* [47, 53]. Note that $p_{\text{avg}}(z) = \langle z|\rho_d|z\rangle$.

The inequalities in (S27) are obtained as follows. The term involving ρ_d is bounded by a simple application of the Cauchy-Schwarz identity, gives $|\langle z|\rho_d|z'\rangle|^2 \leq p_{\text{avg}}(z)p_{\text{avg}}(z')$. Since $|\langle z|\rho_d|z\rangle|^2 \equiv p_{\text{avg}}(z)p_{\text{avg}}(z')$, it follows that $|\langle z_j^L|\rho_d|z_j^R\rangle|/\sqrt{p_{\text{avg}}(z_j^L)p_{\text{avg}}(z_j^R)} \leq 1$ regardless of the values of z_j^L and z_j^R . Finally, the sum $\sum_{\text{perms.}}$ is a sum over allowed permutations of a given list of energies $(E_1, \dots, E_{2k})$; we bound this by $(2k-2)!$. (Formally, $\sum_{\text{perms.}}$ should be *inside* the summation over energies, with different permutations giving rise to different realizations of $\{z_j^L, z_j^R\}$. However, we use the notation above because this sum only gives a constant factor.)

Combining (S26) and (S27), we obtain

$$\text{var}_t[P^{(k)}(t)] = (2k)! \sum_{z,z'} p_{\text{avg}}(z)p_{\text{avg}}(z') \frac{\langle z|\rho_d|z'\rangle \langle z'|\rho_d|z\rangle}{p_{\text{avg}}(z)p_{\text{avg}}(z')} = (2k)! \text{tr}(\rho_d^2) \leq (2k)! D_\beta^{-1} \quad (\text{S28})$$

In our last inequality, we used $\text{tr}(\rho_d^2) \leq D_\beta^{-1}$. This is derived from Sedrakyan's inequality, which states that for positive u_i and v_i , $(\sum_i u_i)^2 / \sum_i v_i \leq \sum_i (u_i^2/v_i)$. In our context,

$$\text{tr}(\rho_d^2) \equiv \sum_E |\langle \Psi_0|E\rangle|^4 = \sum_E |\langle \Psi_0|E\rangle|^4 \frac{\left(\sum_z |\langle z|E\rangle|^2\right)^2}{\sum_z p_{\text{avg}}(z)} \leq \sum_{z,E} \frac{|\langle z|E\rangle \langle E|\Psi_0\rangle|^4}{p_{\text{avg}}(z)} \equiv D_\beta^{-1}. \quad (\text{S29})$$

This completes our proof that for a typical t , $P^{(k)}(t) = k! + O(D_\beta^{-1/2})$. $\square$

Our theorem states that when D_β is sufficiently large, the k -th moment of $\{p(z,t)\}$ is close to $k!$. We can then conclude that $\{p(z,t)\}$ follows the exponential distribution when D_β is sufficiently large. Our numerical results support this: fixing a late enough time t (we numerically find that a time scaling linearly in system size is sufficient for our purposes) and plotting the histogram of the rescaled probabilities $\{\tilde{p}(z,t)\}_z$ (weighted by $p_{\text{avg}}(z)$) shows a good agreement with the exponential distribution, as seen in Fig. S2(f).

Theorems 1 and 2 together suggest an analogue to the ergodic theorem in classical dynamical systems. Consider an ensemble of particles in a classical dynamical system, such as a billiard stadium. The ergodic theorem states that the spatial distribution of this ensemble at a fixed time is the same as the distribution of a fixed particle over time. In our context, the distributions of $\tilde{p}(z,t)$ are asymptotically the same in two settings: with fixed t over configurations z and with fixed z over t . Furthermore, both distributions limit to the universal Porter-Thomas distribution.

Appendix C: Detailed performance analysis for long quench dynamics

In this section we apply the above results to characterize the accuracy of our fidelity estimator F_d . Our analysis is valid when the state has reached global equilibrium, quantified by the quantity $P^{(2)}(t) \equiv \sum_z p_{\text{avg}}(z)\tilde{p}(z,t)^2$ reaching its equilibrium value of 2. In Theorem 2, we have bounded the deviation of the k -th moments $P^{(k)}(t)$ from its ideal value of $k!$. Under an approximation [Eq. (S6)], our bound for $P^{(2)}(t)$ is sufficient to claim that the F_d formula

approximates F with accuracy that is exponentially good in system size. By keeping the simplest energy-conservation terms that are the leading-order corrections in Theorem 2, we make quantitative predictions for the systematic error δ_{sys} and temporal fluctuations δ_{temp} . We explicitly confirm our theoretical analysis via numerical simulations presented in Fig. 3 of our main text and Fig. S4, which shows excellent agreement between theory prediction and numerical results. This suggests that our formalism presented here is able to capture the performance of the F_d formula to first order in D_β^{-1} , at least for the system sizes investigated.

We define the following relevant quantities:

$$F_{\text{id,d}}(t) \equiv \sum_z p_{\text{avg}}(z) \tilde{p}(z, t)^2 - 1 = P^{(2)}(t) - 1, \quad (\text{S1})$$

$$F_{\text{XEB,d}}(t) \equiv \sum_z q(z, t) \tilde{p}(z, t) - 1. \quad (\text{S2})$$

F_d is given by

$$F_d = 2 \frac{\sum_z q(z) \tilde{p}(z)}{\sum_z p(z) \tilde{p}(z)} - 1 = 2 \frac{F_{\text{XEB,d}} + 1}{F_{\text{id,d}} + 1} - 1 \quad (\text{S3})$$

$F_{\text{id,d}}$ is a measure of whether the state has reached global equilibrium: after equilibration, $F_{\text{id,d}} \approx 1$, while at very short times, $F_{\text{id,d}} \sim O(D)$ is exponentially large when the initial state is a product state or a lowly-entangled state. $F_{\text{XEB,d}}$ is analogous to the linear cross-entropy benchmark F_{XEB} .

We can bound $F_{\text{id,d}} \geq 0$ and $F_{\text{XEB,d}} \geq -1$, and we obtain (weak) bounds on F_d :

$$-1 \leq F_d \leq 2 \sqrt{\frac{\sum_z p_{\text{avg}}(z) \tilde{q}(z)^2}{\sum_z p_{\text{avg}}(z) \tilde{p}(z)^2}} - 1, \quad (\text{S4})$$

We compute the average values over time of F_d as well as its variance over time. We find that the typical variance is exponentially small in system size, therefore sampling at a single late time will give values close to its mean.

We use “exponentially small in system size” as a shorthand to mean that a quantity is proportional to a measure of the effective dimension: D_β^{-1} or $\text{tr}(\rho_d^2)$. These are related by the following inequalities:

$$D^{-1} \leq \sum_z p_{\text{avg}}(z)^2 \leq \text{tr}(\rho_d^2) \leq D_\beta^{-1} \leq 1. \quad (\text{S5})$$

The first inequality is a straightforward inequality between the arithmetic and quadratic means. The second inequality is obtained immediately by noting that $\text{tr}(\rho_d^2) = \sum_z p_{\text{avg}}(z)^2 + \sum_{z \neq z'} |\langle z | \rho_d | z' \rangle|^2 \geq \sum_z p_{\text{avg}}(z)^2$, and the final inequality was established in Eq. (S29). D_β^{-1} is in general only upper bounded by 1, since choosing $|z\rangle = |E\rangle = |\Psi_0\rangle$, gives $D_\beta^{-1} = 1$. However, for generic initial states, we expect D_β^{-1} to be exponentially small in system size.

1. Ansatz for how speckle patterns change under errors

In the long-time limit, it is fruitful to approximate $q(z)$ as:

$$q(z) \approx F p(z) + (1 - F) p_\perp(z), \quad (\text{S6})$$

where $p_\perp(z)$ is a distribution uncorrelated with $p(z)$. In general, in the case of incoherent errors, $\tilde{p}_\perp(z) \equiv p_\perp(z)/p_{\text{avg}}(z)$ does not follow a PT distribution.

We assume that $p(z)$ and $p_\perp(z)$ are uncorrelated. Specifically, we assume that:

$$\sum_z p_\perp(z) \tilde{p}(z)^k \approx \mathbb{E}_z[p_\perp(z)] \sum_{z'} \tilde{p}(z')^k \approx k!, \quad (\text{S7})$$

This illustrates that we do not require $\tilde{p}_\perp(z)$ to follow the Porter-Thomas distribution — the energy density of the perturbed state can be quite different from the target state, as is the case when many errors occur. As an example, applying this on Eq. (S2) gives the desired $F_{\text{XEB,d}} \approx F$.

A related approximation applicable to higher order correction terms, is to assume a coherent error V , and write the experimental state as $|\psi'(t)\rangle = V(t) |\psi_0(t)\rangle$. $V(t)$ is unitary, and therefore $V^\dagger(t)V(t) = \mathbb{I}$. However, we replace instances of $V(t)$ or $V^\dagger(t)$ that do not cancel with their Hermitian conjugates with

$$V(t), V^\dagger(t) \rightarrow \sqrt{F}. \quad (\text{S8})$$

While crude, this approximation is useful to make quick estimates. Furthermore, its predictions quantitatively agree with our numerical results, indicating that the dominant sources of error are due to the structure of the F_d formula, which we discuss below.

2. Systematic error

Here, we characterize the systematic error $\delta_{\text{sys.}} \equiv \mathbb{E}_t[F_d(t)] - F$: the difference between the time-averaged F_d and the true fidelity F after an isolated error. To do so, we first study the time average of the simpler quantities $F_{\text{id,d}}(t) \equiv \sum_z p_{\text{avg}}(z) \tilde{p}(z, t)^2 - 1$ and $F_{\text{XEB,d}}(t) \equiv \sum_z p_{\text{avg}}(z) \tilde{p}(z, t) \tilde{q}(z, t) - 1$.

The k -th moment of the PT distribution (Theorems 1,2) immediately allows us to estimate $F_{\text{id,d}}$:

$$\mathbb{E}_t[F_{\text{id,d}}(t)] = \mathbb{E}_t \left[\sum_z p_{\text{avg}}(z) \tilde{p}^2(z, t) - 1 \right] = 1 - D_\beta^{-1}. \quad (\text{S9})$$

Using (S8) gives

$$F_{\text{XEB,d}} \approx F(1 - D_\beta^{-1}). \quad (\text{S10})$$

We would like to use these results to obtain the time-averaged value of F_d . The time average of the numerators and denominators $F_{\text{XEB,d}}$ and $F_{\text{id,d}}$ should be done together. We can estimate its effects with the following Taylor expansion:

$$\mathbb{E} \left[\frac{X}{Y} \right] \approx \frac{\mathbb{E}[X]}{\mathbb{E}[Y]} - \frac{\text{cov}[X, Y]}{\mathbb{E}[Y]^2} + \frac{\text{var}[Y]\mathbb{E}[X]}{\mathbb{E}[Y]^3}, \quad (\text{S11})$$

which gives the estimate for the systematic error in the presence of a single error:

$$\delta_{\text{sys}} = \mathbb{E}_t[F_d] - F \approx (1 - F) \left[\frac{D_\beta^{-1}}{2} + \frac{\text{var}_t[F_{\text{id,d}}]}{8} \right]. \quad (\text{S12})$$

This prediction agrees well with numerical simulations in Fig. S4(a) and Fig. 3(b) in the main text.

a. Argument based on Eigenstate Thermalization Hypothesis

In the limit of low error rates and long-time evolution, we can show that $F_d \approx F$, independent of our approximation for $\tilde{q}$ (S6). To do so, we consider the effect of a single error V that occurs at a variable time t . In the limit of long evolution, $F_d \approx F_{\text{XEB,d}} \equiv \sum_z p_{\text{avg}}(z) \tilde{q}(z) \tilde{p}(z) - 1$ since the denominator in F_d approaches 2. Therefore, we will show that $F_{\text{XEB,d}} \approx F$ for a single error averaged over the error occurrence time. In the presence of multiple errors, the above relation will hold, as long as the errors are sufficiently far apart in time. We average over two quantities: (i) the time t at which the error V occurs, and (ii) the time τ after the error happens when the measurement is made. At this time, the ideal state is $|\Psi(t + \tau)\rangle$, while the perturbed state is $V(\tau) |\Psi(t + \tau)\rangle$, with $V(\tau) \equiv \exp(-iH\tau)V\exp(iH\tau)$ the Heisenberg evolution of V . Here, the fidelity $F(t) = |\langle \Psi(t + \tau) | V(\tau) |\Psi(t + \tau)\rangle|^2 = |\langle \Psi(t) | V |\Psi(t)\rangle|^2$ is independent of τ and only depends on t (the time when the error is applied), while the XEB depends on both t and τ . We will average both t and τ . We first average over t to compute the *average* value of $F_{\text{XEB,d}}$ at a time τ after an error is applied:

$$\mathbb{E}_t[F_{\text{XEB,d}}(t, \tau)] = \sum_z \frac{|\langle z | V(\tau) \rho_d | z \rangle|^2}{p_{\text{avg}}(z)} - \sum_z \frac{\langle z, z | (V^\dagger(\tau) \otimes \mathbb{I}) \rho_d^{(2)} (V(\tau) \otimes \mathbb{I}) | z, z \rangle}{p_{\text{avg}}(z)}, \quad (\text{S13})$$

where we have defined the *diagonal ensemble* ρ_d and its analog $\rho_d^{(2)}$ over $\mathcal{H}^{\otimes 2}$ as follows:

$$\rho_d \equiv \sum_E |\langle E | \Psi_0 \rangle|^2 |E\rangle\langle E| = \mathbb{E}_t[|\Psi(t)\rangle\langle\Psi(t)|] \quad (\text{S14})$$

$$\rho_d^{(2)} \equiv \sum_E |\langle E | \Psi_0 \rangle|^4 |E, E\rangle\langle E, E| \quad (\text{S15})$$

$\rho_d^{(2)}$ is an analog of ρ_d defined over two-copies of the Hilbert space. The first term in Eq. (S13) is the dominant contribution, while the second term comes from the energy-conservation term in Eq. (S4).

We next average over the time τ after the error, in order to claim that $\mathbb{E}_t [F_{\text{XEB,d}}(t, \tau)] \approx \mathbb{E}_t [F(t)]$. The averaging over τ gives:

$$\mathbb{E}_\tau \mathbb{E}_t [F_{\text{XEB,d}}(t, \tau)] = \sum_z \frac{\langle z | \rho_d V_d | z \rangle \langle z | \rho_d V_d^\dagger | z \rangle}{p_{\text{avg}}(z)} - \sum_z \frac{\langle z, z | \rho_d^{(2)} (V_d \otimes V_d^\dagger) | z, z \rangle}{p_{\text{avg}}(z)}, \quad (\text{S16})$$

where $V_d \equiv \sum_E |E\rangle\langle E| V |E\rangle\langle E|$ is the error operator *dephased* in the energy basis, or the time-averaged operator $\mathbb{E}_\tau V(\tau)$. We interpret the first term — the dominant contribution to (S22) — as the average fidelity of states $|\tilde{z}\rangle \propto \sqrt{\rho_d} |z\rangle$ after the error operator V_d :

$$\sum_z \frac{\langle z | \rho_d V_d | z \rangle \langle z | \rho_d V_d^\dagger | z \rangle}{p_{\text{avg}}(z)} = \sum_z \frac{\langle z | \sqrt{\rho_d} V_d \sqrt{\rho_d} | z \rangle \langle z | \sqrt{\rho_d} V_d^\dagger \sqrt{\rho_d} | z \rangle}{p_{\text{avg}}(z)} = \sum_z p_{\text{avg}}(z) |\langle \tilde{z} | V_d | \tilde{z} \rangle|^2, \quad (\text{S17})$$

$|\tilde{z}\rangle \equiv \sqrt{\rho_d} |z\rangle / \sqrt{p_{\text{avg}}(z)}$ is a normalized wavefunction, after a transformation of a computational basis state. We interpret this transformation as an imaginary time evolution generated by the modular Hamiltonian of ρ . We now interpret F_d as an average fidelity $|\langle \tilde{z} | V_d | \tilde{z} \rangle|^2$ of an error operator V_d is applied to $|\tilde{z}\rangle$.

We want to relate $\mathbb{E}_\tau \mathbb{E}_t [F_{\text{XEB,d}}(t, \tau)]$ to the fidelity $\mathbb{E}_t [F(t)]$ averaged over the time t that the error occurred. This is given by:

$$\mathbb{E}_t [F(t)] = \mathbb{E}_t [|\langle \Psi(t) | V | \Psi(t) \rangle|^2] = |\text{tr}(\rho_d V)|^2 + \text{tr}(\rho_d V \rho_d V^\dagger) - \text{tr}[\rho_d^{(2)} (V \otimes V^\dagger)], \quad (\text{S18})$$

with the terms ordered by (expected) magnitude.

In order to relate Eq. (S13) and (S18), we make two observations: (i) for a product initial state $|\Psi_0\rangle$ and local Hamiltonian H , ρ_d approaches a microcanonical state since its energy variance vanishes relative to the size of the spectrum. This is because H is local, implying that $\text{tr}(H^2) \propto N^2$. Therefore, the eigenvalues of H are spread over an energies scaling with total system size N . Meanwhile, (ii) by the locality of H the uncertainty in energy of a product state is of order $\sqrt{N}$: $\Delta E^2 \equiv \langle \Psi_0 | H^2 | \Psi_0 \rangle - \langle \Psi_0 | H | \Psi_0 \rangle^2 \propto N$. We model the distribution of energies as a Gaussian distribution with mean $\bar{E} \equiv \langle \Psi_0 | H | \Psi_0 \rangle$ and variance ΔE^2 :

$$\sum_{E=\bar{E}-\Delta E/2}^{\bar{E}+\Delta E/2} |\langle E | \Psi_0 \rangle|^2 \approx \rho(\bar{E}) \overline{|\langle E | \Psi_0 \rangle|^2} dE \approx \frac{1}{\Delta E \sqrt{2\pi}} \exp\left(-\frac{(E-\bar{E})^2}{2(\Delta E)^2}\right) dE, \quad (\text{S19})$$

where $\rho(\bar{E})$ is the density of states around $\bar{E}$. We define this such that $\int \rho(E) dE = D$, such that $\rho(\bar{E}) = \exp(cN)$ for some $c > 0$. Eq. (S19) also gives the estimate $\sum_{E=\bar{E}-\Delta E/2}^{\bar{E}+\Delta E/2} |\langle E | \Psi_0 \rangle|^4 \propto 1/[\rho(\bar{E})\Delta E] \sim \exp(-cN)$ (omitting polynomial factors of N).

(iii) Lastly, the Eigenstate Thermalization Hypothesis (ETH) states that the diagonal elements $\langle E | V | E \rangle$ of a local operator V are well described by a smooth function of E , $\langle E | V | E \rangle \approx v(E)$. The off-diagonal elements are exponentially smaller: $|\langle E | V | E' \rangle|^2 \approx f(E - E') \langle V^2 \rangle_{\text{micro}} \exp(-S(\bar{E})/2)$, for some function f which is $O(1)$ when the separation $E - E'$ is small, and $\bar{E} = (E + E')/2$. $\langle V^2 \rangle_{\text{micro}}$ is the microcanonical value of V^2 near $\bar{E}$, and $S(\bar{E}) = -\ln(\rho(\bar{E}))$ is the thermodynamic entropy [60]. Then the expression (S18) for the fidelity can be approximated as follows:

$$\mathbb{E}_t [F] = |\text{tr}(\rho_d V)|^2 + \text{tr}(\rho_d V \rho_d V^\dagger) - \text{tr}(\rho_d^{(2)} (V \otimes V^\dagger)) \quad (\text{S20})$$

$$\begin{aligned} &= \left| \sum_E \langle E | V | E \rangle |\langle E | \Psi_0 \rangle|^2 \right|^2 + \sum_{E, E'} \langle E | V | E' \rangle \langle E' | V^\dagger | E \rangle |\langle E | \Psi_0 \rangle|^2 |\langle E' | \Psi_0 \rangle|^2 - \sum_E |\langle E | V | E \rangle|^2 |\langle E | \Psi_0 \rangle|^4 \\ &\stackrel{\text{ETH}}{\approx} |v(\bar{E})|^2 + \langle V^2 \rangle_{\text{micro}} \exp(-c'N), \end{aligned} \quad (\text{S21})$$

for some constant $c' > 0$.

Meanwhile, to analyze $\mathbb{E}_\tau \mathbb{E}_t [F_{\text{XEB,d}}(t, \tau)]$, we approximate $|\langle z | \rho_d V_d | z \rangle| = \sum_E |\langle z | E \rangle \langle E | \Psi_0 \rangle|^2 |\langle E | V | E \rangle| \stackrel{\text{ETH}}{\approx}$

$\overline{\langle E|V|E\rangle}_E \sum_E |\langle z|E\rangle \langle E|\Psi_0\rangle|^2 = |v(\bar{E})| p_{\text{avg}}(z)$. Then

$$\mathbb{E}_\tau \mathbb{E}_t [F_{\text{XEB,d}}(t, \tau)] = \sum_z \frac{\langle z|\rho_d V_d |z\rangle \langle z|\rho_d V_d^\dagger |z\rangle}{p_{\text{avg}}(z)} - \sum_z \frac{\langle z, z|\rho_d^{(2)} (V_d \otimes V_d^\dagger) |z, z\rangle}{p_{\text{avg}}(z)}, \quad (\text{S22})$$

$$= \sum_z \frac{1}{p_{\text{avg}}(z)} \left[\left| \sum_E |\langle z|E\rangle \langle E|\Psi_0\rangle|^2 \langle E|V|E\rangle|^2 - \sum_E |\langle z|E\rangle \langle E|\Psi_0\rangle|^4 |\langle E|V|E\rangle|^2 \right] \quad (\text{S23})$$

$$\stackrel{\text{ETH}}{\approx} \sum_z \frac{|v(\bar{E})|^2}{p_{\text{avg}}(z)} \left[\left(\sum_E |\langle z|E\rangle \langle E|\Psi_0\rangle|^2 \right)^2 - \sum_E |\langle z|E\rangle \langle E|\Psi_0\rangle|^4 \right] \quad (\text{S24})$$

$$= |v(\bar{E})|^2 (1 - D_\beta^{-1}) \quad (\text{S25})$$

Since D_β^{-1} is also typically exponentially small in system size, we conclude that $|\mathbb{E}_\tau \mathbb{E}_t [F_{\text{XEB,d}}(t, \tau)] - \mathbb{E}_t [F(t)]| = O(\exp[-dN])$, for some $d > 0$. This is our desired result, obtained here through the ETH.

b. Extensions beyond local errors

The above argument shows that in the presence of local coherent errors, F_d faithfully estimates F in the long time limit. This can be immediately extended to incoherent errors by averaging over coherent errors. By unraveling an error channel into quantum trajectories, these results also apply to stochastic, local error channels, as long as the error rate is sufficiently low (numerically investigated in Fig. S4).

Finally, we argue that these results also apply to coherent Hamiltonian errors, i.e. quench evolution by different Hamiltonian parameter, utilized in our process benchmarking examples. Time evolution under a slightly perturbed local Hamiltonian can be expanded into a coherent sum of paths representing evolution by the original Hamiltonian, weakly interspersed by the perturbation term. This is similar to a stochastic error channel, except we can have coherent interference between different paths. We argue that this interference effect from off-diagonal terms is negligible: different paths have vanishing overlap and essentially uncorrelated speckle patterns, and their sum is further suppressed by their random phases. An analogous argument applies to target state benchmarking within a ground state phase diagram [Fig. S9], which can be understood as applying imaginary-time evolution (cooling) between states. This effect is perturbative and can also be decomposed into paths of local excitations, as long as the two states are sufficiently nearby.

3. Variance over time

The variance over time utilizes the fourth Hamiltonian twirling identity. The proof of Theorem 2 estimates this as $O(\text{tr}[\rho_d^2])$. We explicitly compute the prefactors of the expected leading order contribution arising from the simplest energy-conserving terms, illustrated in Fig. S3(b):

$$\text{var}_t [F_{\text{id,d}}] = \mathbb{E}_t [F_{\text{id,d}}(t)^2] - \mathbb{E}_t [F_{\text{id,d}}(t)]^2 = 16 \left[\cancel{\text{tr}(\rho_d^2)} - \cancel{\text{tr}(\rho_d^{(2)})} \right]^0 + 4 \sum_{z,z'} \frac{|\langle z|\rho_d|z'\rangle|^4}{p_{\text{avg}}(z)p_{\text{avg}}(z')} + \dots \quad (\text{S26})$$

While this expression allows us to theoretically understand the expected the magnitude of $\text{var}_t [F_{\text{id,d}}]$, in practice we numerically evaluate $\text{var}_t [F_{\text{id,d}}]$ and investigate its relation to $\text{var}_t [F_d]$. In the presence of errors, the variance of $F_{\text{XEB,d}}$ can be calculated using Eq. (S7). This gives $\text{var}_t [F_{\text{XEB,d}}] \approx F^2 \text{var}_t [F_{\text{id,d}}]$ for incoherent errors. In the case of a *single, coherent error*, we must include the coherence of $\tilde{p}_\perp^2$: $\langle \tilde{p}_\perp^2 \rangle \approx 2$ and we obtain $\text{var}_t [F_{\text{XEB,d}}] \approx [F^2 + (1-F)^2] \text{var}_t [F_{\text{id,d}}]$ instead. In both cases, we also have $\text{cov}_t [F_{\text{XEB,d}}, F_{\text{id,d}}] \approx F \text{var}_t [F_{\text{id,d}}]$.

Performing the same Taylor expansion as in Eq. (S11),

$$\text{var} \left[\frac{X}{Y} \right] \approx \frac{\mathbb{E}[X]^2}{\mathbb{E}[Y]^2} \left[\frac{\text{var}[X]}{\mathbb{E}[X]^2} + \frac{\text{var}[Y]}{\mathbb{E}[Y]^2} - 2 \frac{\text{cov}[X, Y]}{\mathbb{E}[X]\mathbb{E}[Y]} \right] \quad (\text{S27})$$

gives the estimate

$$\delta_{\text{temp}}^2 = \text{var}_t [F_d] \approx \frac{1}{4} (1-F)^2 \text{var}_t [F_{\text{id,d}}] \quad (\text{open system dynamics}) \quad (\text{S28})$$

$$\approx \frac{5}{4} (1-F)^2 \text{var}_t [F_{\text{id,d}}] \quad (\text{coherent errors}) \quad (\text{S29})$$

$$\begin{aligned}
\text{(a)} \quad \mathbb{E}_t \left[\sum_z \frac{1}{p_{\text{avg}}(z)} \begin{array}{c} \langle z| \boxed{U_t} |\Psi_0\rangle \\ \langle z| \boxed{U_t^\dagger} |\Psi_0\rangle \\ \langle z| \boxed{U_t} |\Psi_0\rangle \\ \langle z| \boxed{U_t^\dagger} |\Psi_0\rangle \end{array} \right] &= \sum_z \frac{1}{p_{\text{avg}}(z)} \left(\begin{array}{c} \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \end{array} + \begin{array}{c} \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \end{array} - \begin{array}{c} \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \end{array} \right) = 2 - D_\beta^{-1}
\end{aligned}$$

$$\begin{aligned}
\text{(b)} \quad \text{var}_t \left[\sum_z \frac{1}{p_{\text{avg}}(z)} \begin{array}{c} \langle z| \boxed{U_t} |\Psi_0\rangle \\ \langle z| \boxed{U_t^\dagger} |\Psi_0\rangle \\ \langle z| \boxed{U_t} |\Psi_0\rangle \\ \langle z| \boxed{U_t^\dagger} |\Psi_0\rangle \end{array} \right] &= \sum_{z,z'} \frac{1}{p_{\text{avg}}(z)} \left(\begin{array}{c} \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \\ \langle z| \text{---} |\Psi_0\rangle \end{array} \right. \\
&\quad \left. \frac{1}{p_{\text{avg}}(z')} \left(\begin{array}{c} \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \end{array} \right) \right) \\
&\quad + \frac{1}{p_{\text{avg}}(z')} \left(\begin{array}{c} \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \end{array} \right) - \frac{1}{p_{\text{avg}}(z')} \left(\begin{array}{c} \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \\ \langle z'| \text{---} |\Psi_0\rangle \end{array} \right) + \dots
\end{aligned}$$

Figure S3. Diagrammatic derivation of terms that contribute to (a) $\mathbb{E}_t [F_{\text{id,d}}(t)]$ and (b) $\text{var}_t [F_{\text{id,d}}(t)]$. In (a), the two permutations and collision term give $2 - D_\beta^{-1}$. In (b), we list the (expected) leading order contributions. 16 out of the 24 possible permutations each give $\text{tr}(\rho_d^2)$, these are cancelled out by corresponding collision terms: representative permutations and collision terms are illustrated here. Another four permutations give the leading order contribution to $\text{var}_t [F_{\text{id,d}}(t)] = 4 \sum_{z,z'} |\langle z | \rho_d | z' \rangle|^4 / (p_{\text{avg}}(z)p_{\text{avg}}(z')) \leq 4 \text{tr}(\rho_d^2)$.

Here, the larger fluctuations for a single coherent error are associated with the fact that it is pure. We numerically evaluate $\text{var}_t [F_{\text{id,d}}]^{-1}$, and compare our predictions for $\text{var}_t [F_d]$ in terms of $\text{var}_t [F_{\text{id,d}}]$ and F in Fig. 3(c) of the main text and Fig. S4(b,d) (respectively single error and open system dynamics).

4. Sample complexity

The sample complexity of F_d is related to the variance of the estimator $\hat{F}_d$ with a finite number of samples M . The variance δ_{stat}^2 is given by

$$\delta_{\text{stat}}^2 \equiv \text{var}[\hat{F}_d] = \frac{4 \left(\sum_z q(z) \tilde{p}(z)^2 - [\sum_z q(z) \tilde{p}(z)]^2 \right)}{M [\sum_z p_{\text{avg}}(z) \tilde{p}(z)^2]^2} \stackrel{\text{large } t}{\approx} \frac{1}{M} \left(\sum_z q(z) \tilde{p}(z)^2 - \left[\sum_z q(z) \tilde{p}(z) \right]^2 \right) \quad (\text{S30})$$

We estimate this by computing its time-averaged value, with the third Hamiltonian twirling identity, which gives $\mathbb{E}_t [\sum_z p_{\text{avg}}(z) \tilde{p}(z, t)^3] = 3! - (2!)^2 + \dots \approx 2$. Our ansatz for $\tilde{q}(z, t)$ (S7) gives:

$$\mathbb{E}_t [M \delta_{\text{stat}}^2] = \mathbb{E}_t \left[\sum_z q(z, t) \tilde{p}(z, t)^2 - \left(\sum_z q(z, t) \tilde{p}(z, t) \right)^2 \right] \approx 1 + 2F - F^2. \quad (\text{S31})$$

This non-trivial prediction is verified in Fig. S4(c) and Fig. 3(d) of the main text.

5. Moderate error rate: Effects of delay time τ_d

Here we argue that under open system dynamics with a moderate error rate, the ratio F_d/F approaches a constant $F_d/F \approx \exp(\gamma\tau_d)$, where γ is the *total* error-rate and τ_d is a response time independent of system size. When the fidelity F shows a simple exponential decay, the delay time leads to F_d lagging behind F by a constant time τ_d . This is in contrast to the constant difference $F_d - F$ seen under a single error, which is exponentially small in system size. We quantify this delay time effect by introducing the notion of a response function of F_d after a single error, $F_d^{(1)}$, and making the assumption that the effects of multiple errors on F_d are independent and factorize.

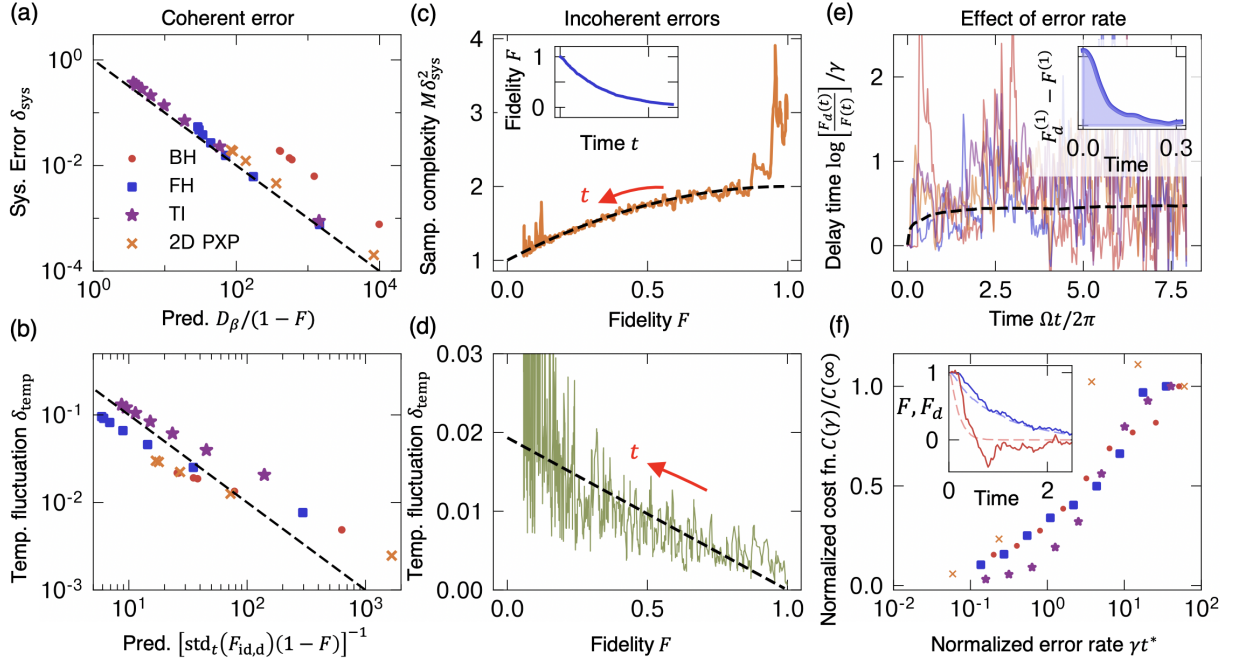


Figure S4. Performance metrics for F_d under single and continuous errors. In all cases, our numerical results quantitatively agree with our theoretical predictions (black dashed lines), with all parameters numerically extracted from independent numerical data. (a,b) After a single error, the (a) systematic error $\delta_{\text{sys.}}$ and (b) temporal standard deviation $\delta_{\text{temp.}}$ quantitatively agrees with the ansatz $\delta_{\text{sys.}} = (1-F)D_\beta^{-1}$ and $\delta_{\text{temp.}} = \sqrt{5/4}(1-F)\text{std}_t[F_{\text{id,d}}]$, across all four systems studied: a Bose-Hubbard model (red), Fermi-Hubbard model (blue), trapped-ion spin chain Hamiltonian (purple), and 2D Rydberg model (orange). (c,d) In the presence of continuous errors, F decays with time from 1 to 0 (inset). Therefore, we can test our predictions for the dependence of our quantities of interest—the sample complexity (c), and the temporal fluctuations (d)—on the fidelity F in a single time series of open system dynamics. We present such a time series for a Bose-Hubbard model with 8 bosons on 8 sites, where we plot the (c) sample complexity at each point in time and the (d) temporal fluctuations in a narrow time window against the fidelity F at that point in time. After a short thermalization time, the sample complexity follows our universal prediction in Eq. (S31): $M\delta_{\text{stat.}}^2 = (1 + 2F - F^2)$, and the temporal fluctuations satisfy Eq. (S28): $\delta_{\text{temp.}} = (1-F)\text{std}_t[F_{\text{id,d}}]/2$. (e) Under continuous errors, $F_d(t)$ will appear to lag behind $F(t)$ with a delay time. In particular, the rescaled curves of F_d/F for a range of error rates $\gamma = 0.0005$ to $\gamma = 0.005$ collapse, and are close to our prediction in Eq. (S34), obtained from the single-error response curve (inset). (f) At low error rates, F_d estimates F well, but they deviate significantly at high error rates. Inset: Across all platforms, a normalized cost function $C(\gamma)/C(\infty)$ (S37) shows a crossover when the average time between errors is less than an error autocorrelation time t^* (S36).

We assume that an error operator X can occur at any time during the open system dynamics. Without loss of generality, we normalize X so it does not change the norm of a typical state: $\langle \psi(t) | X^\dagger X | \psi(t) \rangle \approx 1$. The fidelity F as a function of time is:

$$\begin{aligned} F(t) &\approx p_{ne}(t) + \int_0^t dt_1 p_e(t_1) |\langle X(t_1) \rangle|^2 + \int_0^t dt_1 \int_{t_1}^t dt_2 p_e(t_1, t_2) |\langle X(t_2) X(t_1) \rangle|^2 + \dots \\ &\approx p_{ne} + \sum_{n=1}^{\infty} \frac{1}{n!} \left[\int_0^t dt_i p_e(t_i) |\langle X(t_i) \rangle|^2 \right]^n = \exp[(F_\beta - 1) \gamma t] , \end{aligned} \quad (\text{S32})$$

where we denote $p_{ne}(t) \propto \exp(-\gamma t)$ as the probability of not having an error from time 0 to t and $p_e(t_1, \dots, t_n) \propto \gamma^n \exp(-\gamma t)$ as the probability density of errors occurring at times $t_1, \dots, t_n$. If n errors occur during a trajectory, the pure state associated with this trajectory will have a fidelity $|\langle \psi(t_n) | X \dots X e^{-iH(t_2-t_1)} X | \psi(t_1) \rangle|^2$. As long as the error rate is low enough that errors are typically far apart in time, we assume that this expression factorizes: $F = |\langle X(t_n) \dots X(t_1) \rangle|^2 \approx |\langle X \rangle|_\beta^{2n} \equiv F_\beta^n$. Here the quantity $F_\beta \equiv |\langle X \rangle|_\beta^2$ should be interpreted as the fidelity after a *single error* X on a typical thermal state.

Meanwhile, we can estimate how F_d behaves under open system dynamics in terms of its behavior in response to a single error. We denote $F_d^{(1)}(\tau)$ as the average F_d value at time τ after a single error.

$$\begin{aligned} F_d(t) &\approx p_{ne}(t) + \int_0^t dt_1 p_e(t_1) F_d(t - t_1) + \int_0^t dt_1 \int_{t_1}^t dt_2 p_e(t_1, t_2) F_d(t - t_2, t - t_1) + \dots \\ &\approx p_{ne} + \sum_{n=1}^{\infty} \frac{1}{n!} \left[\int_0^t dt_i p_e(t_i) F_d^{(1)}(t - t_i) \right]^n = \exp \left[\gamma \int_0^t d\tau \left(F_d^{(1)}(\tau) - 1 \right) \right] \end{aligned} \quad (\text{S33})$$

where we again assume that the value of F_d after n errors factorizes as: $F_d(t_1, \dots, t_n) \approx \prod_i F_d^{(1)}(t - t_i)$.

Combining Eqs. (S32), (S33), we have a simple expression for the ratios:

$$\frac{F_d(t)}{F(t)} \approx \exp \left[\gamma \int_0^t d\tau \left(F_d^{(1)}(\tau) - F_\beta \right) \right] \approx \exp[\gamma \tau_d] , \quad (\text{S34})$$

where the quantity $\tau_d \equiv \int_0^\infty d\tau \left(F_d^{(1)}(\tau) - F_\beta \right)$ has a natural interpretation as the *total area under the response curve* $F_d^{(1)}(\tau) - F_\beta$ (Fig. S4(e) inset) — if we ignore the (exponentially small) systematic error $\bar{F}_d^{(1)} - F$, this total area is a constant. This prediction is supported in Fig. S4(e), where we observe the ratio $F_d(t)/F$ saturating at our predicted value for $\exp[\gamma \tau_d]$, with τ_d extracted from the corresponding single error curve $F_d^{(1)}(t)$.

6. Large error rate: Critical error rate and operator scrambling time

As the error rate increases, there is a critical rate at which F_d shows significant deviations from F [Fig. S4(f) inset]. Naively, we would expect that this error rate is solely determined by the delay time discussed above. However, this is not necessarily the case. The delay time is a property of the relaxation dynamics of a local perturbation occurring after the state has reached thermal equilibrium. When the error rate is substantially larger than the local relaxation time, the fidelity becomes very small even before the state has a chance to reach thermal equilibrium. Therefore, the delay time might not be the appropriate timescale for the critical error rate. Instead, we seek to characterize a different timescale relevant in this non-equilibrium dynamics. We hypothesize that a suitable timescale is the error operator *autocorrelation time* — intuitively an operator scrambling time — t^* with respect to the nonequilibrium initial state. Our numerical results support our hypothesis: we find that the critical error rate is related to t^* .

We define the autocorrelation function of the error operator X_j , for a given initial state $|\Psi_0\rangle$ and quench Hamiltonian H , as:

$$A_j(t) = \frac{|\langle \Psi_0 | e^{iHt} X_j e^{-iHt} X_j | \Psi_0 \rangle|^2}{\langle \Psi_0 | e^{iHt} X_j^2 e^{-iHt} | \Psi_0 \rangle \langle \Psi_0 | X_j^2 | \Psi_0 \rangle} , \quad (\text{S35})$$

where we have chosen this normalization such that $0 \leq A(t) \leq 1$. The states $X_j |\Psi(t)\rangle$ are generally not normalized because the error operators X_j may not be unitary (for example, in the Bose-Hubbard model, the errors are associated with photon scattering, and we take X_j to be n_j). We define the autocorrelation time as:

$$t_j^* \equiv \frac{\int_0^\infty dt |A_j(t) - A_j(\infty)|}{1 - A_j(\infty)} , \quad (\text{S36})$$

where $A_j(\infty)$ is the infinite time value $A_j(\infty) = \lim_{t \rightarrow \infty} A_j(t)$. This can be analytically evaluated from the second Hamiltonian twirling identity [Eq. (S4)]. We take the overall operator scrambling time as the average of t_j^* over error operators X_j . This timescale is independent of system size, unlike other timescales such as the time for the entanglement entropy to saturate, which scales with system size.

Importantly, we find that the critical error rate is associated with the non-equilibrium dynamics of the initial state, as opposed to the dynamics after the state has reached thermal equilibrium. To numerically demonstrate this, we initialize the noisy dynamics on a time-evolved state $e^{-iHt} |\Psi_0\rangle$ for sufficiently large t , after the state has thermalized. We find that in the regime of error rates studied, there is no critical change in the accuracy of F_d , merely a constant delay time between F_d and F , as predicted in Section C5, which translates to a smooth increase in the relative error F_d/F . This is analogous to Ref. [42, 43], in which the error rate in random unitary circuits was found to have the largest effect at the start and the end of the evolution.

We define a cost function

$$C(\gamma) \equiv \int_0^T dt |F_d(t) - F(t)|, \quad (\text{S37})$$

that captures the total deviation between F_d and F over time. In Fig. S4(f), we find that the rescaled cost function $C(\gamma)/C(\infty)$ shows a crossover from 0 to 1 at $\gamma t^* \sim 1$, indicating that t^* is the relevant timescale for our setting. We choose such a normalization such that the curves from different simulator platforms can be plotted on the same graph: they demonstrate a cross over in a common regime $\gamma t^* \sim 1$.

7. Symmetry Resolution

In the main text, we described how F_d naturally incorporates symmetries of the quench Hamiltonian. Here, we discuss this in more detail and provide an explicit example. If the initial state belongs to a single symmetry sector s , the quenched state will remain in that sector. Assume for simplicity that the state is Haar random in s . The simpler formulae F_{XEB} and F_c would be appropriate if our measurements z were also states in s . However, this is generically not the case — if the outcome z does not have a well-defined symmetry quantum number, it should be weighted by its overlap with s . The factor $p_{\text{avg}}(z)$ automatically weights z appropriately, without explicit knowledge of the symmetry.

As an illustration, consider spatial inversion symmetry $\mathcal{I}$ on a lattice. The even parity sector has basis states $|e_z\rangle = \mathcal{N}(|z\rangle + |\bar{z}\rangle)$, where $\bar{z}$ is the mirrored configuration of z . If $\bar{z} = z$, $\mathcal{N} = 1/2$, otherwise $\mathcal{N} = 1/\sqrt{2}$. Consider the optimal case where $|\Psi_0\rangle$ is a random even-parity state and the dynamics is a random unitary that conserves parity. The universal statistics would be apparent if we could measure in the $\{|e_z\rangle\}$ basis — a challenging task. Instead, outcome configurations z have unequal average probabilities: $p_{\text{avg}}(z) = 1/D_s$ if $z = \bar{z}$ and $p_{\text{avg}}(z) = 1/(2D_s)$ otherwise, where D_s is the dimension of the symmetry sector. While $p(z)$ does not follow the Porter-Thomas distribution, $\tilde{p}(z)$ does. Therefore, F_d correctly estimates F , while F_c unequally weights some basis states. This argument generalizes to other symmetries such as translational invariance. Since $p_{\text{avg}}(z)$ automatically accounts for symmetries, our conclusions hold even in the presence of extensively many symmetries, as in the case of integrable systems. Finally, in a special case where the readout basis states have well-defined symmetry quantum number, F_d does not estimate F . This is discussed in Section F7.

8. Alternative fidelity estimator F_e

In the following section, we use the random MPS model to derive an alternative fidelity estimator

$$F_e \equiv \frac{\sum_z q(z) \tilde{p}(z) - 1}{\sum_z p(z) \tilde{p}(z) - 1} = \frac{F_{\text{XEB,d}}}{F_{\text{id,d}}}. \quad (\text{S38})$$

Here, we summarize some of its properties. While the temporal fluctuations and sample complexity of F_e are similar to those of F_d , the systematic error (after a long time) is smaller.

(i) Systematic error. The analysis in Section C2 indicates that the systematic error after a single error $\delta_{\text{sys}} = \mathbb{E}_t F_d(t) - F = O(D_\beta^{-1})$. Repeating the same analysis for F_e indicates that the systematic error vanishes for F_e , to first order. This indicates that F_e is less sensitive to small values of D_β , and hence may perform better in quench dynamics with low effective temperature. This is supported in Fig. S6(e,f). More fundamentally, F_e is more robust against deviations of the denominator $\sum_z p_{\text{avg}}(z) \tilde{p}(z)^2$ from 2. For example, in the presence of particle hole symmetry, this denominator can be 3. F_d can be appropriately modified to take this into account, but F_e estimates the fidelity without any modification [Fig. S6(a,b)].

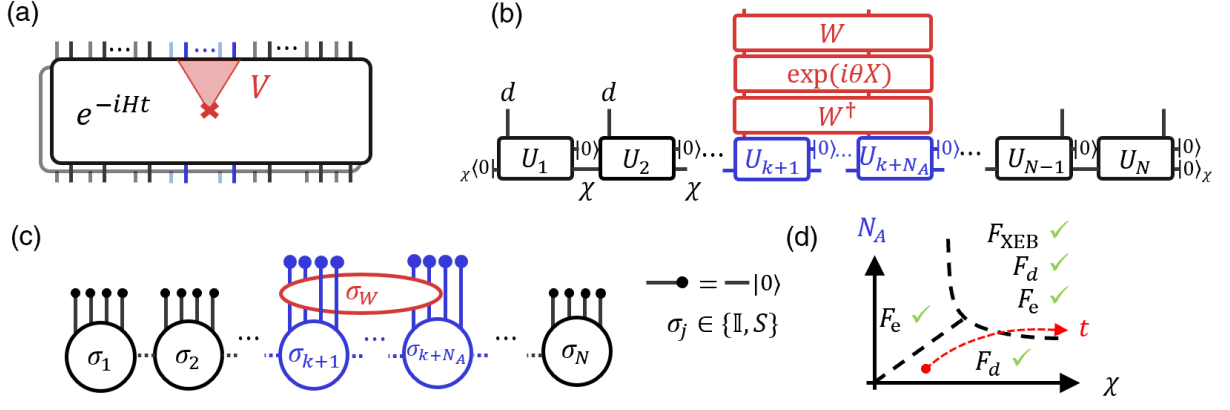


Figure S5. Random matrix product state model. (a) We consider two states evolved by a short quench. In one copy, an error has occurred, which spreads to a small subsystem A , and no error has occurred in the other copy. (b) A toy model for the non-equilibrium state is the *random MPS* — a matrix product state with a random $\chi d \times \chi d$ matrix at every site, with χ the MPS bond dimension, and d the local Hilbert space. The error V is also modeled with a random unitary W and an operator $\exp(i\theta X)$. (c) The expression for F_d , averaged over random unitaries is mapped to the partition function of an effective ferromagnetic Ising spin chain, with Ising degrees of freedom $\sigma_j, \sigma_W \in \{\mathbb{I}, S\}$. This partition function is given in Eq. (S2). (d) We predict regimes of validity for our benchmarking formulae F_d, F_e , and F_{XEB} , depending on the error operator size N_A and MPS bond dimension χ . A typical quench is always in the regime of validity of F_e (red arrow).

(ii) Non-local errors. Our analysis based on random matrix product states indicates that F_e will estimate F for highly non-local errors. This is numerically supported in Fig. S7(f).

(iii) Short-time response. A tradeoff of F_e is that it typically has a longer response time than F_d to learn the fidelity F . This is demonstrated in Fig. S7(e) and Fig. S8(b).

Appendix D: Detailed performance analysis for short quench dynamics

1. Overview: random matrix product states

The infinite-time averaging in Section C tells us that our fidelity estimators work at late times. Our formulae are also designed to work at early times, before the state has reached equilibrium. To capture this early time physics, we introduce a toy model, the *random matrix product state*.

We model the state $|\Psi_0(t)\rangle$ as a matrix product state (MPS) with bond dimension χ and local Hilbert space dimension d . The MPS is constructed by choosing a $\chi d \times \chi d$ random unitary U_j on each site j , as illustrated in Fig. S5(b). At every site, a d degree of freedom is eliminated by contracting to the state $|0\rangle$. At the boundary sites, an additional χ degree of freedom is removed by contracting to a state $|0\rangle_\chi$ on the bond-dimension space. This provides a toy model for a state with entanglement $S \sim \log(\chi)$.

We model the error operator V as $V = W \exp(i\theta X) W^\dagger$, where V is taken to be supported on some region A , and W is a random unitary on $\mathcal{H}_A$. $\exp(i\theta X)$ is an operator whose precise form will be irrelevant. However, the parameter θ measures the strength of the coherent error, and can be used to tune the fidelity between the perturbed and initial states. The random unitary W models the scrambling nature of $V(\tau)$ while making the calculation tractable using techniques of averaging over random unitaries [61].

This toy model captures what we believe to be the essential physics: entanglement and operator growth. In particular, it is well known that both entanglement entropy and operator size grow linearly with time. This corresponds to:

$$\chi \sim \exp(v_S t), \quad N_A \sim vt, \quad (\text{S1})$$

where v_S is an entanglement velocity, and v is an operator spreading velocity [62, 63].

Our calculations give us expressions for how $F_{\text{XEB}} \equiv D^{-1} \sum_z q(z)p(z) - 1$ and $F_{\text{id}} \equiv D^{-1} \sum_z p(z)^2 - 1$, and the fidelity F depend on N_A and χ . Because of the random unitary nature, there are no systematic patterns in $p(z, t)$: $p_{\text{avg}}(z) = 1/D$. Therefore, in this case $F_d = F_c$, $F_{\text{XEB},d} = F_{\text{XEB}}$ and $F_{\text{id},d} = F_{\text{id}}$. Furthermore, we will also consider $F_e = F_{\text{XEB}}/F_{\text{id}}$. Our main result is:

$$\mathbb{E}_{W,U} [F_{\text{XEB}} + 1] \approx \mathbb{E}_U [F_{\text{id}} + 1] \left(F + \exp \left[-\frac{d-1}{d} \frac{N_A}{\chi} \right] \frac{1-F}{2} \right), \quad (\text{S2})$$

where $\mathbb{E}_{W,\mathbf{U}}$ denotes averaging over the random unitaries W and $\mathbf{U} = \{U_j\}$, and $\mathbb{E}_{\mathbf{U}}$ denotes averaging over $\mathbf{U}$.

The second term describes a crossover between two regimes: $\chi \ll N_A$ and $\chi \gg N_A$. In the former regime, $\exp\left[-\frac{d-1}{d} \frac{N_A}{\chi}\right] \approx 0$, and $F_{\text{XEB}} + 1 \approx (F_{\text{id}} + 1) F$. Furthermore since χ is small, both F_{id} and F_{XEB} are exponentially large and

$$F \approx \frac{F_{\text{XEB}} + 1}{F_{\text{id}} + 1} \approx \frac{F_{\text{XEB}}}{F_{\text{id}}} = F_e. \quad (\text{S3})$$

In the latter regime, $\exp\left[-\frac{d-1}{d} \frac{N_A}{\chi}\right] \approx 1$, and $F_{\text{XEB}} + 1 \approx (F_{\text{id}} + 1) \frac{1+F}{2}$, and

$$F \approx 2 \frac{F_{\text{XEB}} + 1}{F_{\text{id}} + 1} - 1 = F_d. \quad (\text{S4})$$

Both F_d and $F_e \equiv [\sum_z p_{\text{avg}}(z) \tilde{q}(z) \tilde{p}(z) - 1] / [\sum_z p_{\text{avg}}(z) \tilde{p}(z)^2 - 1]$ accurately estimate F at late times. F_d is valid since $\chi \approx D \gg N_A$. However F_e is also valid since $F_{\text{id}} \approx 1$ and so $F_{\text{XEB}} \approx F$. These results give the regime of validity of F_d and F_e sketched in Fig. S5(d). Further properties of F_e are discussed in Sec. C8.

2. Remarks

The random MPS model qualitatively captures the short time physics. However, it is not quantitatively accurate, at least to first order in D^{-1} . The simplest example of this is that the state norm has non-zero variance: $\text{var}_{\mathbf{U}} [|\langle \Psi_{\mathbf{U}} | \Psi_{\mathbf{U}} \rangle|^2] \neq 0$. As another example, F_{id} is, on average, bigger than 1, while $\mathbb{E}_t [F_{\text{id},d}] \leq 1$ for Hamiltonian dynamics [Eq. (S9)]. Furthermore, the randomness in the MPS does not capture the systematic trends in finite temperature states that lead to non-trivial $p_{\text{avg}}(z)$ factors. Therefore, the random MPS model cannot be a qualitative model for non-equilibrium quench-evolved states at either finite or infinite temperature. Nevertheless, the qualitative conclusions of the random MPS model — in particular the regimes of validity of F_d and F_e — are consistent with numerical results at finite temperature.

3. Derivation of result

We model the error operator $V(\tau) = W e^{i\theta X} W^\dagger$, where W is a Haar-random unitary supported on A [Fig. S5(b)]. This models the effect of scrambling and makes the calculation tractable. We first integrate over W (a unitary two-design calculation [61, 68]), and obtain two terms corresponding to the permutations of two elements: $\sigma_W = \mathbb{I}$ and $\sigma_W = \mathcal{S}$. We then integrate over U to obtain effective Ising chains for each value of σ_W : the integration over unitaries U_j leaves a degree of freedom on each sites with two possible states, again corresponding to (local) $\mathbb{I}$ and $\mathcal{S}$ permutations [Fig. S5(c)]. These degrees of freedom “interact” with their nearest neighbors ferromagnetically: the lowest energy configurations are those with all sites in the $\mathbb{I}$ state or all sites in the $\mathcal{S}$ state. This ferromagnetic interaction is encoded in the transfer matrix (of the partition function) T_0 (S8).

The presence of the error operator $V(\tau)$ modifies this effective Ising chain when $\sigma_W = \mathcal{S}$: it introduces an effective magnetic field on the region A supporting the error.

Our quantities of interest can be expressed in terms of the partition functions of these two Ising chains. We call these partition functions $f_{X,\mathbb{I}}$ and $f_{X,\mathcal{S}}$, corresponding to the value of σ_W . For example, the average value of $F_{\text{id}} + 1$ over all possible choices of U_j is exactly the partition function $f_{X,\mathbb{I}}$. The presence of an error changes the value of $F_{\text{XEB}} + 1$. $F_{\text{XEB}} + 1$, again averaged over unitaries U_j and W , can be expressed in terms of $f_{X,\mathbb{I}}$, $f_{X,\mathcal{S}}$, and the (averaged) fidelity F .

Our calculation gives the expressions:

$$\mathbb{E}_{W,\mathbf{U}} [F] = \frac{|\text{tr } V|^2 - 1}{D_A^2 - 1} \mathbb{E}_{\mathbf{U}} [|\langle \Psi_{\mathbf{U}} | \Psi_{\mathbf{U}} \rangle|^2] + \frac{|D_A^2 - \text{tr } V|^2}{D_A (D_A^2 - 1)} \mathbb{E}_{\mathbf{U}} [\text{tr}(\rho_A^2)] \approx \frac{|\text{tr } V|^2 - 1}{D_A^2 - 1} \equiv \bar{F}, \quad (\text{S5})$$

$$\mathbb{E}_{W,\mathbf{U}} [F_{\text{XEB}} + 1] = f_{X,\mathbb{I}} \bar{F} + \frac{f_{X,\mathcal{S}}}{D_A} (1 - \bar{F}), \quad (\text{S6})$$

$$\mathbb{E}_{\mathbf{U}} [F_{\text{id}} + 1] = f_{X,\mathbb{I}}, \quad (\text{S7})$$

where the coefficients $f_{X,\mathbb{I}}$ and $f_{X,\mathcal{S}}$ are the partition functions of the Ising spin models, defined by the transfer matrices:

$$T_0 = \frac{1}{\chi^2 d^2 - 1} \begin{pmatrix} \chi^2 d^2 - d & \chi d(d-1) \\ \chi d(d-1) & \chi^2 d^2 - d \end{pmatrix}, \quad T_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{d} \end{pmatrix} T_0 \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{d} \end{pmatrix}. \quad (\text{S8})$$

These transfer matrices are obtained from the Weingarten coefficients that arise from the integration over random unitaries [61]. The partition functions are

$$f_{X,\mathbb{I}} \equiv \frac{\chi}{\chi+1} \frac{\chi d + d}{\chi d + 1} \begin{pmatrix} 1 & 1 \end{pmatrix} T_0^{N-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2 \frac{\chi}{\chi+1} \left(\frac{\chi d + d}{\chi d + 1} \right)^N, \quad (\text{S9})$$

$$f_{X,\mathcal{S}} \equiv \frac{\chi}{\chi+1} \frac{\chi d + d}{\chi d + 1} \begin{pmatrix} 1 & 1 \end{pmatrix} T_0^{k-1} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{d} \end{pmatrix} T_1^{N_A} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{d} \end{pmatrix} T_0^{N-N_A-k} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (\text{S10})$$

As a reminder, $f_{X,\mathbb{I}}$ is the partition function of the ferromagnetic effective Ising model, while $f_{X,\mathcal{S}}$ is the partition function of the same Ising spin chain with an additional (very strong) effective magnetic field applied on region A , that effectively locks all the Ising spins in A to the swap state $\mathcal{S}$.

We can simplify $f_{X,\mathbb{I}}$ and $f_{X,\mathcal{S}}$ by noting that the dominant contributions to $f_{X,\mathbb{I}}$ are the configurations with all sites in the $\mathbb{I}$ state, and all sites in the $\mathcal{S}$ state. Meanwhile, the dominant contribution to $f_{X,\mathcal{S}}$ is the configuration with all sites in the $\mathcal{S}$ state. Furthermore, T_0 has eigenvalues $\frac{\chi d - d}{\chi d - 1}$ and $\frac{\chi d + d}{\chi d + 1}$, while T_1 has eigenvalues $\frac{(\chi d)^2 - d^2}{(\chi d)^2 - 1}$ and d . Therefore, for large N_A , $\frac{f_{X,\mathcal{S}}}{D_A f_{X,\mathbb{I}}}$ is proportional to the ratio of their largest eigenvalues:

$$\frac{1}{D_A} \frac{f_{X,\mathcal{S}}}{f_{X,\mathbb{I}}} \approx \frac{1}{2} \left(\frac{\chi d + 1}{\chi d + d} \right)^{N_A} \approx \frac{1}{2} \exp \left(-\frac{d-1}{d} \frac{N_A}{\chi} \right). \quad (\text{S11})$$

where the factor of a half comes from the breaking of the global Ising $\mathbb{I} \leftrightarrow \mathcal{S}$ symmetry.

This gives our main result Eq. (S2), which predicts the relationship between the cross-entropy benchmark F_{XEB} , fidelity, and the parameters in our model: the operator size N_A and MPS bond dimension χ .

Appendix E: Details of numerical simulations

Below we summarize the details of the models simulated: their Hamiltonians, initial states, error models, and Hamiltonian parameter values.

1. Bose-Hubbard model

We numerically simulate quench evolution under the one-dimensional (1D) Bose-Hubbard Hamiltonian [64] in the presence of noise, which is often dominated by the photon scattering from optical lattices. In the quantum trajectory formalism [51], such noisy dynamics is described by stochastically applying the particle number operator $\{n_j\}$ at a rate γ for each site. An ensemble of such pure state evolution represents a mixed state $\rho(t)$ in open dynamics. The 1D Bose-Hubbard Hamiltonian is:

$$H_{\text{BH}} = -\Omega \sum_{j=1}^{N-1} \left(b_j^\dagger b_{j+1} + \text{h.c.} \right) + \frac{U}{2} \sum_{j=1}^N n_j (n_j - 1), \quad (\text{S1})$$

where $b_j^\dagger$ is the bosonic creation operator at site j , and $n_j = b_j^\dagger b_j$ is the bosonic number operator at site j . In Fig. 1 and 2, we use the parameters $(\Omega, U) = (1, 2.87)$. To increase the ergodicity of our quench, in Fig. 3,4, we use the parameters $(\Omega, U) = (1, 0.5)$, whose ground state is in the superfluid phase. We take as initial state the “unity filling state” $|1, 1, \dots, 1\rangle$, which is the ground state deep in the Mott insulating phase $U \gg \Omega$.

Our error operators are $\{n_j\}$, which model the effects of scattering from optical lattice photons [65]. In Fig. 1(b), we use an overall error rate of $\gamma = 0.05$.

In Fig. 2(a), we show data for a system of 6 bosons on 6 sites, and in Fig. 1(b) and Fig. 2(b,c), we show data for a system with 10 bosons on 10 sites (with total Hilbert space dimension 92378).

a. Disorder Averaging

In Fig. 3(b-d), we present disorder averaged results for the response curve and the sample complexity of F_d against time. Disorder averaging is done to remove the large fluctuations present at short times in order to clearly observe systematic behavior.

For simplicity, we choose the following ensemble of Bose-Hubbard models with disordered hopping strengths:

$$H_{\text{BH}} = - \sum_{j=1}^{N-1} \Omega_j \left(b_j^\dagger b_{j+1} + \text{h.c.} \right) + \frac{U}{2} \sum_{j=1}^N n_j (n_j - 1) , \quad (\text{S2})$$

with $\Omega_j = 0.2m_j$, for m_j a random integer from 1 to 10. As in Eq. (S2), we choose $U = 0.5$. We chose to disorder the hopping strengths in order to observe different dynamics at short terms. If, for example, we had chosen a disordered potential $\Delta_j n_j$ instead, due to the homegeneous initial state $|1, \dots, 1\rangle$, the resultant states $|\psi(t)\rangle$ would not be significantly different from each other across disorder realizations. Lastly, an error at $n_{\lfloor L/2 \rfloor}$ is applied at a fixed, early time $t = 5$. This is because the error operator n_j acts trivially on the initial state $|1, \dots, 1\rangle$.

2. Fermi-Hubbard model

The 1D Fermi-Hubbard model Hamiltonian is:

$$H_{\text{FH}} = \sum_{j=1}^N \left[-\Omega \sum_{\sigma=\uparrow,\downarrow} \left(c_{j,\sigma}^\dagger c_{j+1,\sigma} + \text{h.c.} \right) + U n_{j,\uparrow} n_{j,\downarrow} \right] , \quad (\text{S3})$$

where $c_{j,\sigma}^\dagger$ is the creation operator of a fermion at site j , with spin $\sigma \in \{\uparrow, \downarrow\}$, and $n_{j,\sigma} = c_{j,\sigma}^\dagger c_{j,\sigma} \in \{0, 1\}$ is the fermion occupation number. In Fig. 2 we use the parameter value $(\Omega, U) = (1, 3)$, while in other figures we use $(\Omega, U) = (1, 1)$. For any value of U/Ω , this model is integrable by Bethe Ansatz [50], but we numerically observe that it still passes the 2nd no-resonance condition, after disregarding degeneracies [Eq. (S3)]. As discussed in Section B, degeneracies can be disregarded for the purposes of Theorem 2.

We take as initial state the anti-ferromagnetic state at half filling, $|\uparrow, \downarrow, \dots\rangle$, and as error operators the total fermion occupation $\{n_{j,\uparrow} + n_{j,\downarrow}\}$ at each site. As with the Bose-Hubbard model, this models scattering from optical lattice photons [55]. In Fig. 1(c), we use a total error rate of $\gamma = 0.2$, in a system with 5 up and 5 down fermions on 10 sites (with total Hilbert space dimension 63504).

3. Trapped ion model

Our trapped ion model has Hamiltonian:

$$H_{\text{TI}} = \sum_{j=1}^{N-1} \sum_{i>j} (\Omega/|i-j|^\alpha) \sigma_i^x \sigma_j^x + h_z \sum_{j=1}^N \sigma_j^z , \quad (\text{S4})$$

with σ_j^x and σ_j^z the Pauli x - and z -operators on site j . H_{TI} is adapted from the experiment of Ref. [10]. Due to the long-ranged interactions, following Ref. [10] we fix the “rescaled” parameter value $\tilde{h}_z = h_z / (N^{-1} \sum_{j=1}^{N-1} \sum_{i>j} 1/|i-j|^\alpha)$ to ensure consistent behavior across system sizes, where α is the exponent of the power-law interactions. As in Ref. [10], our initial state is the all-downs computational basis state: $|\downarrow, \dots, \downarrow\rangle$. We use the rescaled parameters $(\Omega, \tilde{h}_z, \alpha) = (1, 0.7, 1)$. In Fig. 1(d), we use on-site spin flip and dephasing errors: $\{\sigma_j^x, \sigma_j^z\}$, with a total error rate of $\gamma = 0.05$ in a system of 16 qubits (with a total Hilbert space dimension of $2^{16} = 65536$).

In Fig. 4(c) we learn the disorder strengths Δ_j of a disorder term $H_{\text{dis}} = \sum_j \Delta_j \sigma_j^z$.

4. Rydberg blockaded model

We simulate a Rydberg blockaded model in two dimensions (2D), as recently experimentally realized in Ref. [66]. Here each site can be either in the ground $|\downarrow\rangle$ or Rydberg $|\uparrow\rangle$ state, subject to the constraint that there are no two

adjacent Rydberg excitations $|\uparrow\rangle$. Its Hamiltonian is:

$$H_{\text{PXP}} = \Omega \sum_{j=1}^N P X_j + \delta \sum_j n_j, \quad (\text{S5})$$

with $PX_j \equiv \left[\bigotimes_{\langle i,j \rangle} |\downarrow\rangle\langle\downarrow|_i \right] \sigma_j^x$ the projected spin flip operator — a spin flip on site j is allowed only if all the nearest neighbors i of site j are in state $|\downarrow\rangle$.

We use the all-downs state $|\downarrow, \dots, \downarrow\rangle$ as our initial state. At zero detuning $\delta = 0$, this would be an infinite temperature state ($\langle H \rangle = \text{tr}(H)/D$). In Fig. 2 we use the parameter values $(\Omega, \delta) = (1, 1.5)$ and in all other figures we use the parameters $(\Omega, \delta) = (1, 0.7)$ to tune $|\downarrow, \dots, \downarrow\rangle$ away from infinite temperature. Furthermore, when $\delta = 0$, the Hamiltonian is particle-hole symmetric, leading to a different kind of universal behavior, detailed in Section F 1. In Fig. 1(e), we use dephasing error $\{\sigma_j^z\}$ at a total error rate of $\gamma = 0.05$, in a 5×5 square grid of Rydberg atoms (with total Hilbert space dimension of 55447).

Appendix F: Limitations of benchmarking protocol: Failure cases

As discussed in the main text, our benchmarking formula can fail when: 1) the k -th no resonance condition is not satisfied, 2) the effective dimension D_β is small, 3) the approach to equilibrium is slow, 4) the error is highly non-local, or 5) the readout basis states have well-defined symmetry quantum numbers. Below, we detail our findings for each case, and discuss alternate benchmarking formulae.

1. Violating the no-resonance condition: Particle-hole symmetry

The second no-resonance condition (S3) is extensively violated in the presence of a particle-hole symmetry (PHS). By this we mean a reflection symmetry in the spectrum: for every eigenvalue E , there is one at $-E$. For any given E, E' , the equation $E_1 - E_2 + E_3 - E_4 = 0$ not only has solutions $(E_1, E_2, E_3, E_4) = (E, E, E', E')$ and (E, E', E', E) , but also has solutions: $(E_1, E_2, E_3, E_4) = (E, -E, E', -E')$, $(E, -E, -E', E')$. This will give rise to additional terms in the second Hamiltonian twirling identity [Eq. (S4)]:

$$\begin{aligned} \mathbb{E}_t [U_t^{\otimes 2} A U_{-t}^{\otimes 2}] \\ \approx \sum_{E, E'} \left[|E, E'\rangle\langle E, E'| A |E, E'\rangle\langle E, E'| + |E, E'\rangle\langle E, E'| A |E', E'\rangle\langle E', E| + |E, -E'\rangle\langle E, -E'| A |E', -E'\rangle\langle E', -E'| \right], \end{aligned} \quad (\text{S1})$$

suppressing the subleading collision terms. This can arise from, an *anti-unitary* charge conjugation operator, for example. Here, we focus on a simpler example. PHS is present in the PXP model with no detuning ($\delta = 0$): $H_{\text{PXP}} = \sum_j P X_j$. H_{PXP} anticommutes with the unitary operator $\mathcal{C} \equiv \prod_j \sigma_j^z$: $\mathcal{C} H_{\text{PXP}} \mathcal{C} = -H_{\text{PXP}}$ [56]. This symmetry is present in arbitrary dimensions— we will present numerical evidence for the two dimensional PXP model.

The PXP model is particularly convenient, for a number of reasons:

- (i) The computational basis states $|z\rangle$ are eigenstates of $\mathcal{C}$: $\mathcal{C}|z\rangle = \pm|z\rangle$. This implies that easily-prepared bitstring initial states such as $|\Psi_0\rangle = |\downarrow \dots \downarrow\rangle$ are also eigenstates of $\mathcal{C}$.
- (ii) H_{PXP} is real and thus possesses a time-reversal symmetry: $H_{\text{PXP}} = H_{\text{PXP}}^*$. In particular, its eigenstates are real in the computational basis.

The breakdown of the second no-resonance condition can be seen when we compute the time average of $P^{(2)}(t) = \sum_z p_{\text{avg}}(z) \tilde{p}(z, t)^2$:

$$\mathbb{E}_t [F_{\text{id,d}}(t)] \approx \sum_z \left[2 \left(\sum_E |\langle z|E\rangle \langle E|\Psi_0\rangle|^2 \right)^2 + \left| \sum_E \langle z|E\rangle \langle z|-E\rangle \langle E|\Psi_0\rangle \langle -E|\Psi_0\rangle \right|^2 \right] / p_{\text{avg}}(z). \quad (\text{S2})$$

Using properties (i) and (ii) above, $\sum_E \langle z|E\rangle \langle z|-E\rangle \langle E|\Psi_0\rangle \langle -E|\Psi_0\rangle = p_{\text{avg}}(z)$ and we obtain $\mathbb{E}_t [F_{\text{id,d}}(t)] \approx 3$ instead of the usual value of two. This is illustrated in Fig. S6(a). We note that this result is contingent on conditions (i) and (ii) above — if these conditions were not satisfied, $\mathbb{E}_t [F_{\text{id,d}}(t)]$ will have a non-universal value between two and

three. For example, if we use an initial state that is *not* particle-hole symmetric, such as a low-energy Boltzmann state, $\mathbb{E}_t[F_{\text{id},d}(t)] \approx 2$, and F_d remains valid.

In particular, this implies that our F_d formula will not estimate the fidelity. Applying our simplifying assumptions in Eq. (S6), we obtain: $F_d \approx \frac{4}{3}F - \frac{1}{3}$. Instead, a modified formula should be employed:

$$F_d^{(\text{PH})} = \frac{3 \sum_z q(z) \tilde{p}(z)}{2 \sum_z p(z) \tilde{p}(z)} - \frac{1}{2}. \quad (\text{S3})$$

However, a similar analysis reveals that F_e can be used with no modification. This is illustrated in Fig. S6(b).

a. Real Porter-Thomas distribution

The value of three above is the second moment of the *real Porter-Thomas distribution* [39] (stated, for simplicity, in the $D \rightarrow \infty$ limit):

$$P_{\text{PT},\mathbb{R}}(x) = \frac{\exp(-x/2)}{\sqrt{2\pi x}}. \quad (\text{S4})$$

This distribution can be viewed as the overlap probability of a *real-valued* random vector $|\Psi\rangle$. It has k -th moments: $\int_0^\infty dx x^k P_{\text{PT},\mathbb{R}}(x) = (2k-1)!! \equiv (2k-1)(2k-3)\cdots 3 \cdot 1$.

Intuitively, if both the initial state $|\Psi_0\rangle$ and $|z\rangle$ are even under $\mathcal{C}$, the overlap will be purely real: $\langle z|\Psi(t)\rangle = 2 \sum_{E \geq 0} \langle z|E\rangle \langle E|\Psi\rangle \cos(Et)$. We intuitively expect this to be a random variable and for $p(z, t) = |\langle z|\Psi(t)\rangle|^2$ to follow the real Porter-Thomas distribution. This also holds if $|\Psi_0\rangle$ or $|z\rangle$ are odd under $\mathcal{C}$.

We can prove that $\tilde{p}(z, t)$ (both with fixed time, and fixed configuration z) follows $P_{\text{PT},\mathbb{R}}$, assuming conditions (i) and (ii) above, and a *particle-hole symmetric* k no-resonance condition. The proof is identical to that of Theorem 2, except that instead only considering the permutations of k elements, we consider *perfect matchings* of $2k$ vertices. There are $(2k-1)!!$ such perfect matchings, giving the desired k -th moments. We emphasize that to satisfy both conditions (i) and (ii), both time-reversal and particle-hole symmetry must be present in the Hamiltonian, and the initial state and readout basis states must have definite PHS parity.

2. Violating the no-resonance condition: Non-interacting models

The second no-resonance condition is also violated in non-interacting models. In non-interacting models, energy eigenvalues are given in terms of single-particle eigenvalues. For concreteness, consider a non-interacting Bose-Hubbard model, Eq. (S1) with $U = 0$. Let the eigenvalues of the single-particle Hamiltonian be ϵ_k . An N -particle eigenstate is labelled by the occupation $\{n_k\}$ of the single-particle modes k , with energy $\sum_k n_k \epsilon_k$. This violates the 2 no-resonance condition: For any i, j, k, l , there are eigenvalues $E_1 = \epsilon_i + \epsilon_j$, $E_2 = \epsilon_k + \epsilon_l$ and $E_3 = \epsilon_i + \epsilon_k$, $E_4 = \epsilon_j + \epsilon_l$ such that $E_1 + E_2 = E_3 + E_4$. Therefore, our Theorem does not apply to non-interacting models, reflected in Fig. S6(c). Consequently, neither F_d or F_e estimate the many-body fidelity (Fig. S6(d)). However, integrable and non-integrable dynamics may be quantitatively similar at short times, at which times the integrability breaking terms have not had sufficient time to appreciably change the dynamics. Since F_d is designed to work for these short quenches, F_d appears to estimate F in this case, as seen in Fig. S6(d).

3. Small effective dimension: Low temperature quench

Our protocol also breaks down when the effective dimension D_β is small. This happens in low temperature quenches, for example. In the extreme case that $|\Psi_0\rangle$ is an eigenstate, $D_\beta = 1$. When $|\Psi_0\rangle$ is a low temperature state, it's dynamics are effectively constrained to a low-energy subspace. Following Theorem 2, in Fig. S6(e), the variable $\tilde{p}_z(t)$ approximately follows a Porter-Thomas distribution (S12) with the effective dimension of the outcome z , $D_\beta(z)$. Meanwhile, the distributions $\tilde{p}_t(z)$ show large fluctuations over time, again because of the small effective dimension.

Our protocol's systematic error and fluctuations scale inversely with D_β . Therefore, it is less applicable to low temperature quenches. In Fig. S6(f), we see a large discrepancy between F_d and F and large fluctuations over time. However, since the systematic error of F_e is less sensitive to D_β [its $O(D_\beta^{-1})$ term vanishes], F_e estimates F , when averaged over a time window.

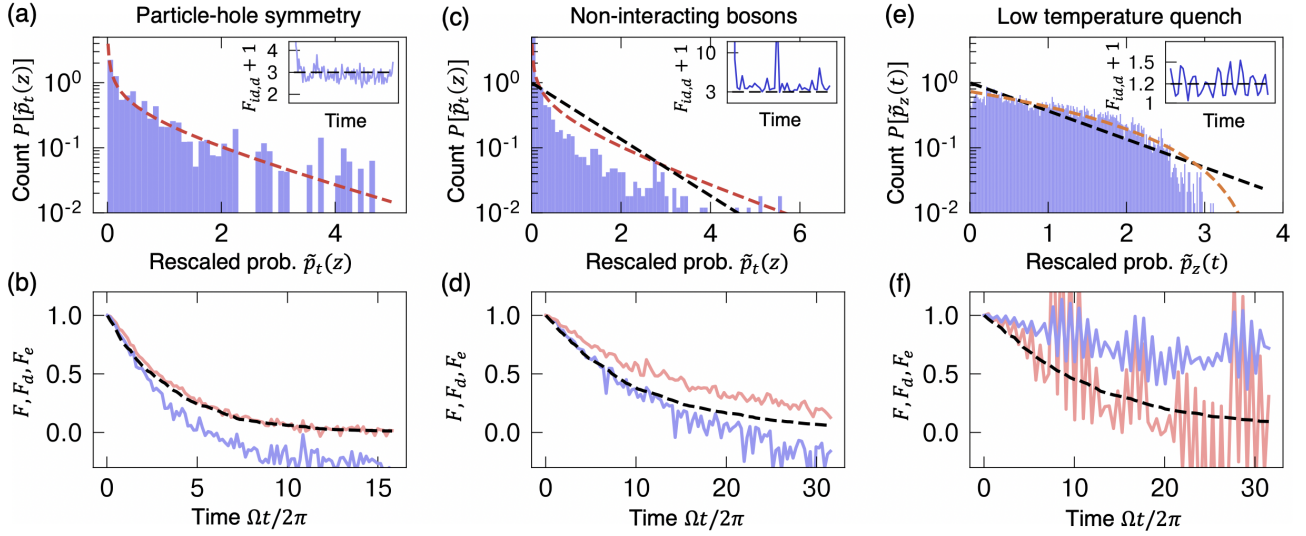


Figure S6. Failure cases of benchmarking protocol: Violation of no-resonance conditions and low effective dimension. (a) The particle-hole symmetry in the 2d PXP model violates the k no-resonance conditions (S3). Under special conditions (satisfied here in the 2d PXP model quenched from $|0 \cdots 0\rangle$), $\tilde{p}(z, t)$ follows the *real* Porter-Thomas distribution, and $F_{id,a} + 1$ has an equilibrium value of 3 (inset). (b) The F_d formula (blue) breaks down, and must be modified as in Eq. (S3). In contrast, F_e (red) can be used without modification. (c) Non-interacting models also violate the no-resonance conditions. Here, we simulate a quench of the Bose-Hubbard model with no interaction $U = 0$. The $\tilde{p}(z, t)$ follows neither PT distributions (red and black), and $F_{id,a} + 1$ has large temporal fluctuations (inset). (d) As a consequence, neither F_d nor F_e approximate F , although F_d is a good estimate at short times. (e) A low-temperature quench has a small effective dimension. We realize this with a Gibbs initial state $|\Psi_0\rangle = \sum_E \exp(-\beta/2E) |E\rangle$, with $\beta = 4$ in a Bose-Hubbard model with 8 particles on 8 sites. $\tilde{p}_z(t)$, for an arbitrarily chosen z , follows a Porter-Thomas distribution with a small dimension (orange): here $D_\beta(z) \approx 3.75$. (f) As a consequence, F_d (blue) and F_e (red) have large temporal fluctuations, and F_d has a large systematic error. F_e is less susceptible to a small D_β , and is on average close to F .

4. Slow equilibration: Quantum many-body scar states

Some many-body systems exhibit long-lived oscillations which delays equilibration of the quenched state. In the presence of these oscillations, normalizing the measurement outcomes $p(z)$ with the *equilibrium* distribution $p_{\text{avg}}(z)$ gives incorrect estimators.

One prominent example of long-lived oscillations is the recent discovery of *quantum many-body scar states* [56, 57]. Originally experimentally discovered in a Rydberg quantum simulator, this phenomenon can be captured with the 1D PXP model:

$$H_{1D\text{-PXP}} = \Omega \sum_{j=2}^{L-1} P_{j-1}^\downarrow X_j P_{j+1}^\downarrow + X_0 P_1^\downarrow + P_{L-1}^\downarrow X_L + \delta \sum_j n_j, \quad (\text{S5})$$

where $P_j^\downarrow \equiv |\downarrow\rangle\langle\downarrow|_j$ is the projector onto the atomic ground state and $n_j \equiv |\uparrow\rangle\langle\uparrow|_j$ is the projector on the Rydberg excited state. A quench of the Neel state $|\downarrow\uparrow\downarrow\uparrow \cdots \downarrow\uparrow\rangle$ was observed to exhibit long-lived oscillations — in particular, persistent revivals of observables such as the magnetization and many-body fidelity. These are attributed to a tower of eigenstates, energy eigenstates with which $|\Psi_0\rangle$ has considerable overlap [Fig. S7(a), inset]. Due to special properties of these eigenstates — most notably low half-chain entanglement entropy — these states are dubbed *many-body scar states*. While the phenomenon is most apparent with zero detuning, it persists with non-zero detuning. In our results in Fig. S7(a,b), we use a non-zero detuning $\delta = 0.2\Omega$ to break the particle-hole symmetry.

These revivals eventually decay in time. Therefore, $\tilde{p}_z(t)$ follows a PT distribution at late times [Fig. S7(a)]. However, these long-lived oscillations are reflected in F_d and F_e [Fig. S7(b)] — F_d is a closer estimate to F than F_e .

5. Slow equilibration: Inhomogeneous initial state

If our initial state is strongly inhomogeneous, it will exhibit hydrodynamic oscillations before equilibrating [58]. At short times, the outcome probabilities $p(z, t)$ will be far from the relatively homogeneous $p_{\text{avg}}(z)$. We illustrate this

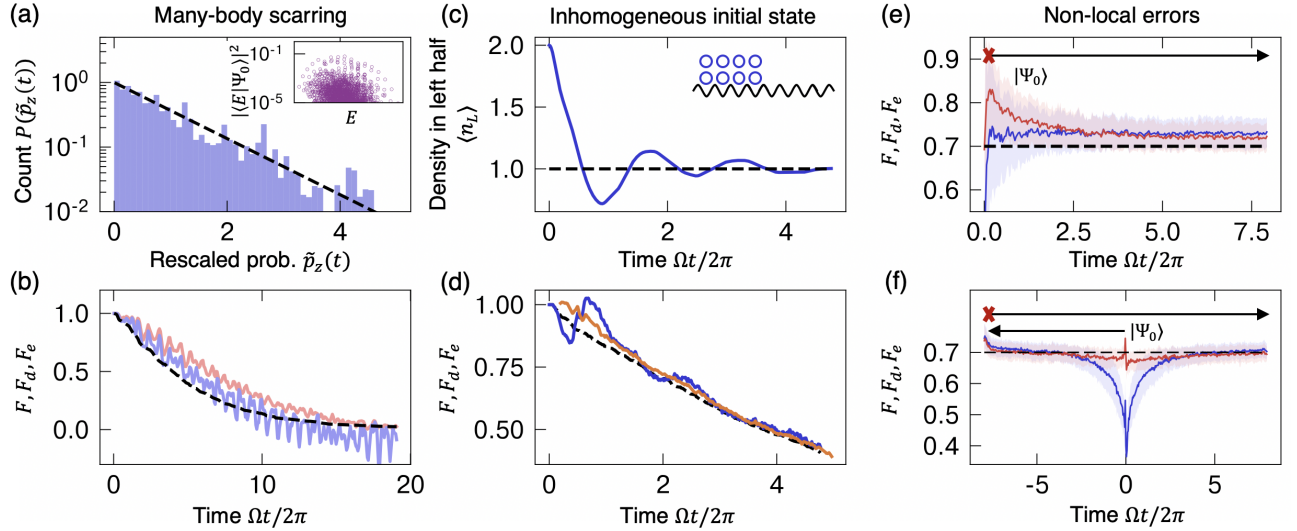


Figure S7. Failure cases of benchmarking protocol: Slow equilibration and non-local errors. (a) In the 1D PXP model, a quench of the Neel state $|\downarrow\uparrow\downarrow\uparrow\dots\rangle$ undergoes slow thermalization due to large overlap with a “tower” of eigenstates (inset). $p_z(t)$ follows a PT distribution. (b) However, both F_d (blue) and F_e (red) show large oscillations and do not estimate F well (c) If an inhomogeneous state is quench evolved, it undergoes slow hydrodynamic relaxation, evidenced by the decaying oscillations of the average density in the left half of a Bose-Hubbard chain (blue). (d) Under this quench, both F_d (blue) and F_e (not plotted) exhibit oscillations. A modified F_e (orange) using a moving-averaged p_{avg} yields a closer estimate of the fidelity. (e) F_d and F_e response to an initial error. At short times F_d (blue) is a better estimate of F than F_e (red). Solid lines are average values over an ensemble of disordered Hamiltonians, with shaded areas indicating variance. (f) We apply a local error on a *backward time-evolved* state $|\Psi(-\tau)\rangle$ to engineer a non-local error on the initial state. In this case F_e (red) estimates F , while F_d (blue) does not, consistent with Appendix. D.

with the initial state $|\Psi_0\rangle = |2, 2, 2, 2, 0, 0, 0, 0\rangle$ on a Bose-Hubbard model with 8 bosons on 8 sites. We measure its relaxation with the average density in the left half $\langle n_L \rangle \equiv \langle n_1 + n_2 + n_3 + n_4 \rangle / 4$, which exhibits decaying oscillations in Fig. S7(c). These oscillations are reflected in F_d and F_e , which show large deviations from F (Fig. S7(d)). We obtain better performance by normalizing $p(z, t)$ and $q(z, t)$ by a moving average $p'_{\text{avg}}(z) \equiv \int_{t-\Delta t/2}^{t+\Delta t/2} dt p(z, t)$, for a short interval Δt . We find that the modified F_e , $F'_e(t) \equiv (\sum_z q(z, t) \tilde{p}(z, t) - 1) / (\sum_z p(z, t) \tilde{p}(z, t) - 1)$ is a closer approximation to F .

6. Non-local error on low entanglement states

In Appendix D, the random MPS model predicts that when the error size N_A is much smaller than the bond dimension χ , F_d is a better approximation to F at short times than F_e . This condition is satisfied in typical open system dynamics, since χ grows exponentially in time while N_A only grows linearly in time.

To verify our claim, we apply a single error at the start of the quench ($t = 0$) and compute F_d and F_e . Averaging over a large number of Bose-Hubbard Hamiltonians with disordered hopping strengths, we verify that F_d on average estimates F at very short times (albeit with large variance) (blue, Fig. S7(e)), while F_e (red) has a longer relaxation time.

To engineer a situation in the opposite limit, where $\chi \ll N_A$, we backwards time-evolve the initial state, then apply the local error on $|\Psi(-\tau)\rangle$. We then forward evolve the ideal and perturbed states. When $|t|$ is small, this is equivalent to a non-local error $V(t + \tau)$ on the low-entanglement states $|\Psi(t)\rangle$. In this regime, as predicted, F_d (blue, Fig. S7(f)) deviates from F — reaching a minimum value of $2F - 1$, consistent with Eq. (S2) — while F_e is close to F (red).

We emphasize that such a situation is unlikely to arise in practical situations, and therefore F_d is more appropriate for short time quench dynamics than F_e . We illustrate this in an example of target state benchmarking in Fig. S8(b).

7. Non-ergodicity due to symmetry and readout in symmetry-resolved basis

We lastly discuss another case where F_d does not estimate F in the presence of symmetry and a readout basis that has well-defined symmetry quantum numbers.

In the presence of a symmetry, quantum states ergodically explore the Hilbert space of their symmetry sector. In generic readout bases, the presence of such symmetries does not impede F_d . However, if every state in the readout basis belongs to one symmetry sector, F_d will not be able to detect relative phase errors, or phase decoherence, between the different symmetry sectors. Physically, we can understand this as non-ergodicity of certain information about the state: the phase information never propagates into the readout basis. Mathematically, this does not affect our Theorems 1 and 2. However, our ansatz that $q(z) = Fp(z) + (1-F)p_\perp(z)$ is no longer valid. Instead, consider two states, expressed as superpositions of states in different symmetry sectors: $|\Psi\rangle = \sum_j c_j |\Psi_j\rangle$ and $|\Psi'\rangle = \sum_j d_j |\Psi'_j\rangle$. Then for a bitstring z in sector j , the ansatz should be modified to (assuming the quench processes for $|\Psi\rangle$ and $|\Psi'\rangle$ are identical):

$$q(z) = F_j p(z) + (1 - F_j) p_\perp(z), \quad F_j \equiv |\langle \Psi_j | \Psi'_j \rangle|^2 \quad (\text{S6})$$

Under this ansatz, F_d does not reproduce the true fidelity $F = \left| \sum_j c_j^* d_j \langle \Psi_j | \Psi'_j \rangle \right|^2$. Applying our ansatz, and recalling that $p_{\text{avg}}(z) \equiv \mathbb{E}[p(z, t)]$, $F_{\text{XEB,d}} \equiv \sum_z q(z) \tilde{p}(z) - 1$ takes the value $2 \sum_j |d_j|^2 F_j - 1$

Consider the simplest case, $|\Psi_j\rangle = |\Psi'_j\rangle$. Then $F_j = 1$, and $F_{\text{XEB,d}} = 2 \sum_j |d_j|^2 - 1 = 1$, regardless of the value of c_j, d_j . Therefore, F_d is unable to estimate the fidelity. The infidelity arising from unequal populations $|c_j|^2$ of the symmetry sectors may be captured by a suitably modified formula. However, infidelity arising from different phases of c_j is intrinsically inaccessible from such measurements: if $d_j = \exp(i\theta_j) c_j$, then $q(z) = p(z)$ identically.

Many analog simulators have readout bases that belong to a given symmetry sector. For example, itinerant particles in optical lattices readout configuration outcomes that have well-defined particle number, and their natural Hamiltonians also conserve particle number. However, states without well-defined particle number are not natural in such settings.

Such states that are in multiple symmetry sectors, however, may arise in other settings. Consider our trapped ion model H_{TI} (S4). This Hamiltonian conserves parity $\prod_j \sigma_j^z$, and the computational basis states have well-defined parity. Furthermore, it is easy to prepare states without well-defined parity: for example, one starts with the all-downs state $|\downarrow, \dots, \downarrow\rangle$, and applies a local rotation on a single site. Therefore, such a family of states cannot be benchmarked with the F_d formula. As another example, such a scenario can arise in a spin chain, with a quench Hamiltonian that conserves magnetization $\sum_j \sigma_j^z$. Bitstrings in the computational basis have well-defined magnetization. In both cases, however, this can be remedied by measuring in the X -basis, in which outcomes do not have well-defined parity nor magnetization. This is explicitly demonstrated in a trapped-ion spin model in Fig. S8(e,f).

Appendix G: Greedy algorithm for multiparameter estimation

1. Algorithm

In this section we detail our greedy algorithm for simultaneous estimation of multiple Hamiltonian parameters, used in Fig. 4(c). Our algorithm is as follows. We are given a Hamiltonian $H(\boldsymbol{\theta})$ with n unknown parameters $\boldsymbol{\theta} = \{\theta_j\}_{j=1}^n$. We want to determine the actual experimental parameters $\{\theta_j^*\}$, assuming that we can prepare a fiducial state $|\Psi_0\rangle$ with high fidelity.

Our procedure is as follows:

1. Evolve $|\Psi_0\rangle$ for a fixed time.
2. Measure the configurations to approximate the distribution $q(z) = \langle z | \rho(t) | z \rangle$.
3. For some guess $\{\theta_j\}$, compute $p(z, \{\theta_j\}) = |\langle z | \exp(-iH(\{\theta_j\})t) | \Psi_0 \rangle|^2$. Compute $F_d(\{J_j\})$.
4. Choose a random permutation $\pi \in S_n$. For each j , maximize $F_d(\{\theta_j\})$ over $\theta_{\pi(j)}$, keeping all other parameters fixed.
5. Repeat 4. for all $j = 1, \dots, n$.
6. Repeat 3-5. until F_d converges.
7. Optionally repeat all steps with different initial guesses of $\{\theta_j\}$.

This procedure is justified as long as F_d does not vary too quickly in the range of $\{\theta_j\}$ considered.

In Fig. 4(c), we use this procedure to determine an unknown disorder potential atop a known systematic Hamiltonian. Our greedy algorithm works well as long as the disorder potential is small, and is also robust to errors in the experimental evolution.

Appendix H: Target state benchmarking

In Fig. 4(a) of our main text, we demonstrate target state benchmarking to determine the unknown phase of a GHZ-like state in a 2D array of Rydberg atoms. We take the prepared state to have a phase $\phi^* = \pi/4$. We wish to measure ϕ^* by measuring the fidelity of the prepared state to target states with various values of ϕ . Using our protocol, we estimate the fidelity by measuring F_d after quench evolution.

In this section, we provide further demonstrations of target state benchmarking. We first consider the simplest setting of coherent local errors, which we demonstrate in the three other platforms studied. We then study benchmarking between states that are related by non-local deformations. We consider this setting for two applications: determining the Hamiltonian parameters corresponding to a ground state, as well as estimating the temperature of a low-energy state by measuring its overlap with the ground state.

1. Coherent local errors

Here, we demonstrate target state benchmarking in the Bose-Hubbard, Fermi-Hubbard, and trapped-ion spin chain model, using the same Hamiltonian and readout basis discussed in Section E.

We consider locally rotated initial states: for the Bose-Hubbard model we consider a unity-filling Mott insulator state on 8 sites deformed by a local rotation ϕ on the middle two sites:

$$|\Psi(\phi)\rangle_{\text{BH}} = \cos \phi |1, 1, 1, 1, 1, 1, 1, 1\rangle + i \sin \phi |1, 1, 1, 2, 0, 1, 1, 1\rangle . \quad (\text{S1})$$

For the Fermi-Hubbard model we consider the anti-ferromagnetic state with 8 particles on 8 sites, again rotated along the middle bond:

$$|\Psi(\phi)\rangle_{\text{FH}} = \cos \phi |\uparrow, \downarrow, \uparrow, \downarrow, \uparrow, \downarrow, \uparrow, \downarrow\rangle + i \sin \phi |\uparrow, \downarrow, \uparrow, d, 0, \downarrow, \uparrow, \downarrow\rangle , \quad (\text{S2})$$

where $d = \uparrow\downarrow$ denotes a doubly-occupied site, and 0 denotes an empty site. Lastly, for the trapped-ion model we consider the fully polarized state of 12 qubits, rotated by a pair of spin-flips across the middle bond

$$|\Psi(\phi)\rangle_{\text{TI}} = \cos \phi |\downarrow, \dots, \downarrow, \downarrow, \downarrow, \dots, \downarrow\rangle + i \sin \phi |\downarrow, \dots, \downarrow, \uparrow, \downarrow, \dots, \downarrow\rangle . \quad (\text{S3})$$

A single bit-flip produces a coherent superposition across states with different parity. These states do not mix under the quench dynamics and the resulting fidelity cannot be estimated by F_d , if the readout basis is the computational basis. In contrast, if we measured in the X basis, F_d would be accurate, as discussed in (Section F 7). If instead we had flipped two spins, our state would remain in the even-parity sector, and F_d would estimate F with any readout basis.

In all three cases, we compute the fidelity $F = |\langle \Psi(\phi^*) | \Psi(\phi) \rangle|^2 = \cos(\phi - \phi^*)^2$. In Fig. S8(a,c,e), we show that F_d , evaluated at a fixed point in time, estimates the fidelity F in all three cases. In Fig. S8(b,d,f), we show that F_d takes a short delay time to learn the value of F to a target state $\phi = \frac{\pi}{2}$ — only a relatively short quench evolution of $t = O(1/J)$ is needed for this protocol. In the trapped ion model, F_d shows slightly larger deviations from F . This is because F_d fluctuates in time about F : this is indicated by the blue band in Fig. S8(e) and larger fluctuations Fig. S8(f).

2. Comparing non-locally deformed states

We next demonstrate verifying the preparation of a non-trivial ground state in two ways by computing the fidelities of states related by non-local deformations. We first demonstrate verifying the position in the phase diagram by computing the fidelities between a prepared critical ground state and the ground states of the same Hamiltonian at other parameter values. We then demonstrate estimating the temperature of the prepared state by considering a prepared low-temperature Boltzmann state of the Hamiltonian at the critical point and estimating its fidelity (overlap) to the ground state.

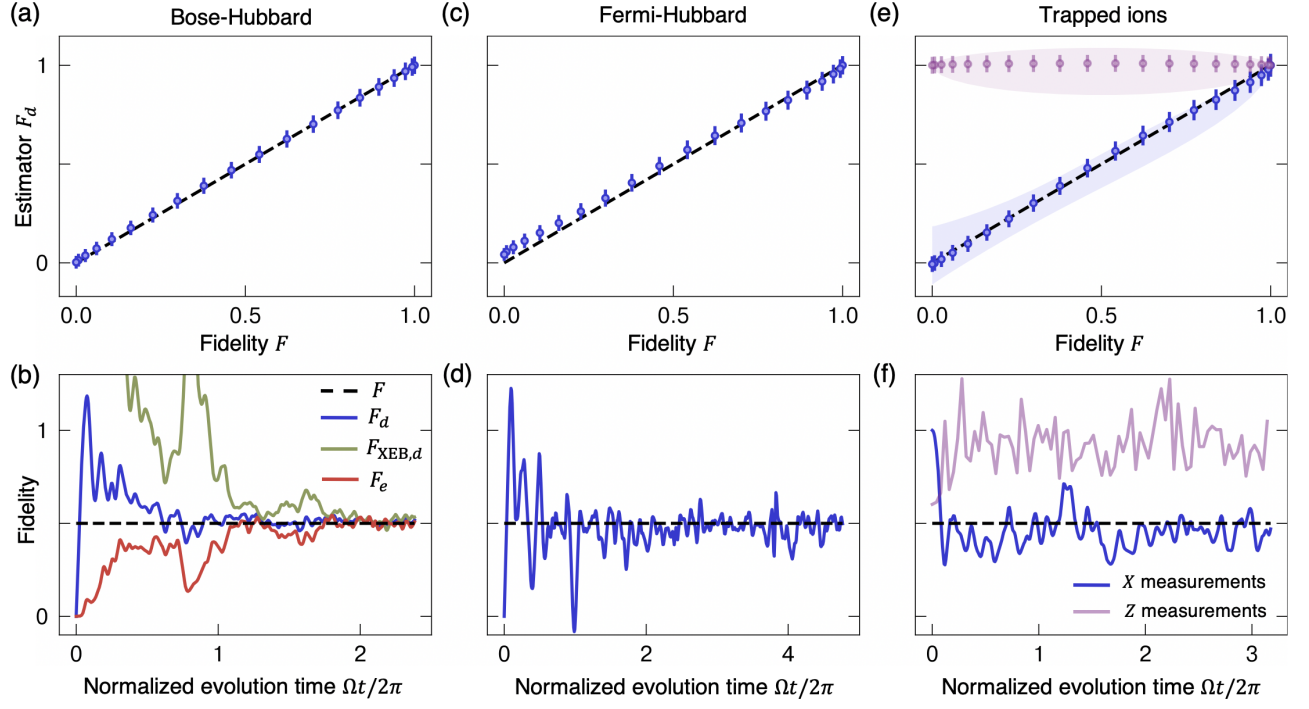


Figure S8. Target state benchmarking of locally-perturbed states. We apply local perturbations to initial states in a (a,b) Bose-Hubbard, (c,d) Fermi-Hubbard, and (e,f) trapped-ion spin chain model. We consider a class of states parametrized by an angle ϕ . We take states with $\phi^* = \pi/4$ as the simulated prepared state, and compare it to target states with variable ϕ . As a function of ϕ , the fidelity between the prepared and target states vary from 0 to 1. By quench evolving these states, we estimate their fidelity with F_d . In the top row, we demonstrate that the estimator F_d at a fixed, sufficiently late time, faithfully reproduces the fidelity F . Error bars denote the statistical uncertainty from a finite number $M = 1000$ of samples. In the bottom row, we demonstrate the required evolution time for F_d to learn the true fidelity F , with the target state at $\phi = \pi/2$. In (b), we compare the different estimators discussed in the text: F_d (blue), F_e (green), and $F_{\text{XEB},d}$ (red) in a Bose-Hubbard model. The difference between $F_{\text{XEB},d}$ and F_d highlights the role of the denominator $\sum_z p_{\text{avg}}(z) \tilde{p}(z)^2$ at short times. In (e), we show that evaluating F_d at a fixed time shows slight deviations from F . This is due to the small effective dimension $D_\beta \approx 11$ in our trapped ion model: F_d fluctuates about F over time, as seen in (f) and indicated with the blue shaded region in (e). The purple markers and shaded regions in (e,f) indicate readout in the computational basis. Due to constraints from symmetry, F_d does not estimate F (in fact, it is constant in our setting). In (f), the purple curve is evaluated with a target state at $\phi = \pi/2 + \varepsilon$ for a small ε . This is because the target state at $\phi = \pi/2$ occupies only one out of two symmetry sectors, hence is a fine-tuned, pathological case.

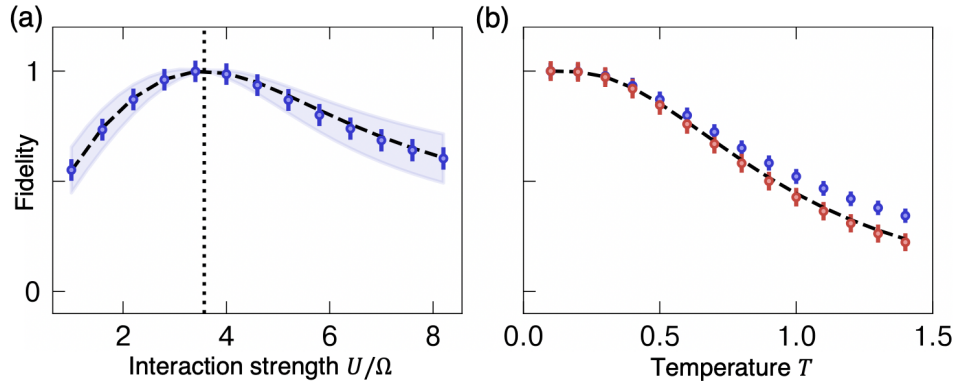


Figure S9. Target state benchmarking by verifying preparation of a ground state in a system. We consider two tasks: (a) estimating the corresponding parameter value and (b) estimating the temperature of the prepared state. In (a), we estimate the fidelity of the prepared state — assumed to be the ground state at the critical value of $(U/\Omega)_{\text{crit}} \approx 3.6$ — against the ground states at various values of U/Ω . We find that F_d (blue markers) is close to the true fidelity F (dashed line). The blue shaded region indicates the size of fluctuations of F_d over time. In (b), we estimate the fidelity of the prepared state — assumed to be the Boltzmann state at various temperatures T — against the ground state at $(U/\Omega)_{\text{crit}}$. We find that F_d (blue) slightly over-estimates F , while F_e (red) closely estimates F , consistent with our analysis in Section. C. In both plots, error bars indicate the statistical uncertainty associated with 1000 measurement samples.

We simulate the ground state of the Bose-Hubbard model (S1) tuned to the Mott insulator-superfluid transition point at $(U/\Omega)_{\text{crit}} \approx 3.6$ on a system of 10 bosons on 10 lattice sites [67]. We wish to benchmark the fidelity of our prepared state against its desired ground state. We consider two purposes: estimating the best parameter U/Ω of the prepared ground state, and estimating the temperature of the prepared state.

To do so, we consider quench evolving our prepared state via a modified Bose-Hubbard Hamiltonian, and comparing it to simulated data of quenched ground states at various values of U/Ω . Here, care must be taken for two reasons. Firstly, the quench Hamiltonian should be chosen such that the ground states under investigation do not have too low energy with respect to the quench Hamiltonian. Secondly, the family of states we consider do not simply differ by local errors. Rather, the family of ground states are related by possibly non-local transformations. This is outside of the scope of our heuristic derivation. However, we numerically observe that the F_d formula still approximates F under judicious choices of the quench Hamiltonian.

In order for the ground states in both phases to have sufficiently ergodic quench dynamics, we consider a quench with a disordered Hamiltonian: $H = H_{\text{BH}}[(U/\Omega)_{\text{crit}}] + \sum_{j=1}^8 \Delta_j n_j$, where $\Delta_j \sim \text{Unif}[-3, 3]$. In Fig. S9(a), F_d approximates the true ground state overlap in a system of 8 bosons on 8 lattice sites.

We next demonstrate estimating the temperature of an experimentally prepared state. Assuming we have correctly determined the parameter U/Ω , we simulate the F_d values between the quench-evolved ground state and quench-evolved Boltzmann states $\rho = Z(T)^{-1} \exp(-H_{\text{BH}}/T)$, where $Z(T)$ is the partition function at temperature T . The fidelity is simply $Z(T)^{-1} \exp(-E_0/T)$. In this case, F_d [Fig. S9(b), blue], is close to the true fidelity. However, because of a relatively small effective dimension $D_\beta \approx 5$, F_d is larger than F , consistent with the prediction in Eq. (S12). This systematic shift can be manually subtracted. Alternatively, the estimator F_e [Eq. (S3)] is more robust to small values of D_β and closely estimates the fidelity [Fig. S9(b), red].
