

Experimental Signatures of Hilbert-Space Ergodicity: Universal Bitstring Distributions and Applications in Noise Learning

Adam L. Shaw^{1,*}, Daniel K. Mark^{2,*}, Joonhee Choi³, Ran Finkelstein¹, Pascal Scholl¹,
Soonwon Choi^{2,†} and Manuel Endres^{1,‡}

¹*California Institute of Technology, Pasadena, California 91125, USA*

²*Center for Theoretical Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA*

³*Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA*

 (Received 21 September 2024; revised 28 April 2025; accepted 29 May 2025; published 1 July 2025)

Systems reaching thermal equilibrium are ubiquitous. For classical systems, this phenomenon is typically understood statistically through ergodicity in phase space, but translating this to quantum systems is a long-standing problem of interest. Recently, a strong notion of quantum ergodicity has been proposed, namely, that isolated, global quantum states uniformly explore their available state space, dubbed *Hilbert-space ergodicity*. Here, we observe signatures of this process with an experimental Rydberg quantum simulator and various numerical models, before generalizing to the case of a local quantum system interacting with its environment. For a closed system, where the environment is a complementary subsystem, we predict and observe a smooth quantum-to-classical transition in that observables progress from large, quantum fluctuations to small, Gaussian fluctuations as the bath size grows. This transition exhibits universal properties on a quantitative level among a wide range of systems, including those at finite temperature, those with itinerant particles, and random circuits. For an open system, where the environment is uncontrolled, we predict the statistics of observables under largely arbitrary noise channels including those with correlated errors, allowing us to discriminate between candidate error models both for continuous Hamiltonian time evolution and for digital random circuits. This allows for computationally efficient experimental noise learning and, more broadly, is a new avenue for quantitatively classifying the behavior of noisy quantum systems. Ultimately, our results clarify the role of ergodicity in quantum dynamics, with fundamental and practical consequences.

DOI: [10.1103/h6xy-zpx4](https://doi.org/10.1103/h6xy-zpx4)

Subject Areas: Quantum Physics,
Quantum Information,
Statistical Physics

I. INTRODUCTION

The developing study of *quantum thermalization* [1–8] has sought to create a consistent formulation for both statistical and quantum mechanics. In this framework, entanglement entropy between subsystems of a quantum state plays the role of thermal entropy in statistical mechanics. Considering a closed, bipartite quantum system composed of interacting subsystems A and B under a time-independent Hamiltonian, quantum thermalization

suggests that, at late times, local observables in A behave as if they were drawn from a thermal state. Here, B acts as a thermal bath, for which the system Hamiltonian and initial state define an effective temperature [8], behavior observed by numerous experiments [9–14].

In order to explain the ubiquity of this phenomenon, a set of interconnected theories have been studied, most notably the *eigenstate thermalization hypothesis* [8] (ETH). The ETH predicts that the above system-bath behavior occurs even on the level of individual many-body energy eigenstates, and, accordingly, it suggests that quantum thermalization occurs for generic states because, at late times, coherent contributions between energy eigenstates will be effectively dephased. Underlying the ETH is the assumption that energy eigenstates act as random vectors in certain ways in accordance with predictions from random matrix theory [8,15,16], which is generally believed to hold when the Hamiltonian is nonintegrable, meaning it has no exact solution.

*These authors contributed equally to this work.

†Contact author: soonwon@mit.edu

‡Contact author: mendres@caltech.edu

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

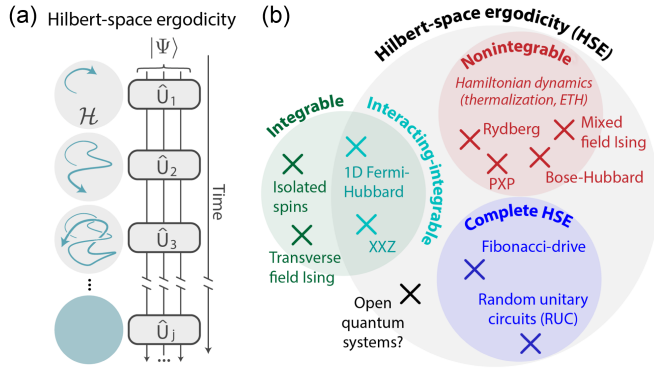


FIG. 1. Hilbert-space ergodicity. (a) Hilbert-space ergodicity (HSE) characterizes unitary quantum evolution under which a global quantum state $|\Psi\rangle$ uniformly explores its Hilbert space $\mathcal{H}$. (b) While originally introduced [17] as *complete* HSE, which characterizes certain symmetry-free dynamics, in recent theoretical work [18], HSE was extended to systems with symmetries and conservation laws. This expanded the framework and consequences of HSE to a host of new models of time-constant Hamiltonian systems, such as those exhibiting nonintegrable dynamics traditionally understood via the eigenstate thermalization hypothesis, as well as certain interacting-integrable systems. In this work, we show experimental evidence consistent with the predictions of HSE with a Rydberg quantum simulator, we generalize the theory to describe and then observe universal behavior of subsystems of closed systems displaying HSE, and, finally, we introduce a framework inspired by HSE for describing the behavior of open quantum systems.

Despite its success, the ETH restricts itself to the expectation values of *local* observables. In contrast, newer theoretical perspectives have begun to shed light on the behavior of *global* quantum systems. In particular, it was proposed that under certain quantum dynamics global quantum states homogeneously explore their available Hilbert space [17,18]. This phenomenon, dubbed *Hilbert-space ergodicity* (HSE), is analogous to ergodicity in classical mechanics, whereby over time a system visits its state space uniformly [Fig. 1(a)]. While HSE was originally [17] restricted only to certain symmetry-free models—such as random unitary circuits (RUCs) and aperiodic sequences of unitaries, for which it was delineated as *complete* HSE—recent theoretical work [18] has extended it to systems with symmetries and conservation laws, including nonintegrable thermalizing systems and certain interacting-integrable systems [Fig. 1(b)].

Importantly, a system obeying HSE entails consequences beyond standard predictions of quantum thermalization and the ETH. For instance, consider a global projective observable which is not described by the ETH, like the probability to measure a specific basis state (e.g., a “bitstring” for a system of qubits). HSE predicts that *relative* probability fluctuations over time precisely follow an exponential (or Porter-Thomas [19,20]) distribution. While this is well known for the case of random quantum circuits [20,21], it

applies to all HSE systems including time-independent, thermalizing Hamiltonian dynamics. Given the power and generality of this framework, it is important to first test its existing conjectures before exploring further generalizations and potential practical applications.

Here, we employ an experimental Rydberg quantum simulator [22–24] undergoing thermalizing dynamics to study both existing and new predictions of HSE. To start, we find temporal fluctuations of bitstring probabilities are consistent with the emergence of the Porter-Thomas distribution, a result also borne out by numerics. Then, we extend the framework of HSE to the case of a system interacting with its environment to make concrete predictions beyond the scope of conventional quantum thermalization, for the case of both closed and open quantum systems.

For closed quantum systems, we consider the typical setting of quantum thermalization: when the environment is a complementary subsystem, which we call the intrinsic bath. For thermalization at infinite effective temperature, we find temporal fluctuations of projective observables follow a universal form as a direct consequence of HSE: the Erlang distribution [25]. Solely parametrized by the dimension of the bath, the Erlang distribution undergoes a smooth quantum-to-classical transition, progressing from the global Porter-Thomas distribution to a narrow Gaussian distribution as the intrinsic bath size increases. Using a precisely defined notion of effective bath dimension [26], we find consistent behavior for systems evolving at finite effective temperature or with conservation laws. Our results are exact and universal (i.e., independent of system details) for ergodic systems both globally and locally. This in contrast to studies of quantum thermalization, which had typically either only bounded [27,28] or predicted the scaling of fluctuations with the total system dimension [29–31]; previous predictions have been made more exact for special cases like explicitly randomized dynamics [32] or the weakly nonintegrable regime [33].

For open quantum systems, we focus our study on global observables. Such a setting is ubiquitous for noisy quantum experiments [34]. The coupling to the environment is generically thought to damp out fluctuations of observables, but we find the manner in which this occurs is highly dependent on the microscopic system-environment interaction for certain observables like bitstring probabilities. Based on our understanding of HSE principles, we devise a formalism to account for this behavior, and, in particular, we find that such fluctuations follow a hypoexponential distribution, a generalization of the Erlang distribution which captures the microscopic effects of the environment. We leverage the accuracy of this approach for practical applications by identifying classes of noise models with distinct signatures in the hypoexponential distribution in order to discriminate candidate models directly with many-body bitstring measurements. Our method can handle bespoke models with correlated noise and with heavy

tails [35], both of which could be detrimental to quantum error correction [35–37].

Overall, our work examines the role of ergodicity in open and closed quantum dynamics and studies the consequent emergence of universal statistical behavior. This leads to uncovering both fundamental and practical ramifications for modern quantum experiments.

II. HILBERT-SPACE ERGODICITY

HSE makes universal predictions about the statistics of bitstring measurements readily available from quantum simulators. In this section, we first briefly recapitulate the theoretical concepts of HSE (see Ref. [18] and Appendix C for further details); in subsequent sections, we discuss its original predictions for the distribution of global bitstring probabilities and finally extend it to the broader settings of subsystems and open systems.

Central to HSE is the *temporal ensemble*, the set of pure states $|\Psi(t)\rangle = \exp(-i\hat{H}t/\hbar)|\Psi(t=0)\rangle$, evaluated at different times t for evolution under a time-independent Hamiltonian $\hat{H}$. Explicitly, if we discretize the time evolution with infinitesimal step size δ and define the time $t_j = \delta j$ for $j \in \mathbb{Z}_+$, the temporal ensemble is the infinite set $\{|\Psi(t_j)\rangle\}_{j \in \mathbb{Z}_+}$. We also define the *diagonal ensemble* [27,28], which is the average over all states in the temporal ensemble:

$$\hat{\rho}_d = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt |\Psi(t)\rangle \langle \Psi(t)| \quad (1)$$

$$= \mathbb{E}_t[|\Psi(t)\rangle \langle \Psi(t)|] = \sum_n |c_n|^2 |E_n\rangle \langle E_n|, \quad (2)$$

where $|E_n\rangle$ are energy eigenstates and $|c_n|^2 \equiv |\langle \Psi(0)|E_n\rangle|^2$ are their corresponding populations, set by the initial state $|\Psi(0)\rangle$.

In the case of time-independent Hamiltonian evolution considered here, the Schrödinger equation conserves the populations $|c_n|^2$, thus restricting the temporal ensemble to a subset of Hilbert space obeying energy conservation. On this manifold, the only remaining degrees of freedom are the complex phases $\arg[\langle E_n|\Psi(t)\rangle]$. It was shown [18] that HSE equates the temporal ensemble with the so-called *random phase ensemble* [18,26,38]. The only assumptions for this relation are the *no-resonance conditions*, a commonly held assumption [18,27–29,39,40] that there are no high-order degeneracies in the energy spectrum (see Appendix C). The no-resonance conditions are believed to apply to any nonintegrable quantum system. Their only known exceptions are systems of noninteracting spins and systems with exact solution by noninteracting fermions, such as the TFIM. These resonances are unstable to nonzero interactions, and it has been proven that almost every local quantum system satisfies the no-resonance condition [41].

Under this condition, we can define the higher-order moments of the temporal ensemble in terms of the well-known moments of the Haar ensemble [18,42]. In particular, the k th moment of the temporal ensemble is, up to corrections which are exponentially small in the system size, equal to

$$\mathbb{E}_t[|\Psi(t)\rangle \langle \Psi(t)|^{\otimes k}] \approx \hat{\rho}_d^{\otimes k} \sum_{\sigma \in S_k} \text{Perm}(\sigma), \quad (3)$$

which, for $\hat{\rho}_d = \hat{I}/D$, are the moments of the Haar ensemble [18,42] (up to exponentially small corrections). Here, $\text{Perm}(\sigma)$ is the permutation operator acting on k copies of the Hilbert space with a permutation σ in the symmetric group S_k which takes any configuration $|i_1, i_2, \dots, i_k\rangle \in \mathcal{H}^{\otimes k}$ and maps it to $|i_{\sigma(1)}, i_{\sigma(2)}, \dots, i_{\sigma(k)}\rangle$. As an example, for $k = 2$,

$$\mathbb{E}_t[|\Psi(t)\rangle \langle \Psi(t)|^{\otimes 2}] \approx \hat{\rho}_d^{\otimes 2} (\hat{I} + \hat{S}), \quad (4)$$

where $\hat{I}$ is the identity and $\hat{S}$ is the swap operator between two copies of Hilbert space.

Crucially, this means that if we pick any random (late) time, then in a statistical sense the properties of the corresponding pure state $|\Psi(t)\rangle$ will be those of a state drawn from the uniform (i.e., Haar) random ensemble, reweighted as

$$|\Psi(t)\rangle \approx \sqrt{D\hat{\rho}_d}|\phi\rangle, \quad |\phi\rangle \sim \text{Haar}. \quad (5)$$

This is the general statement of HSE, that global quantum states uniformly randomly cover their *available* Hilbert space, as constrained by conservation laws, symmetries, or a finite effective temperature, effects which are all encoded by the diagonal ensemble state $\hat{\rho}_d$. This factorization between the universal Haar behavior and the evolution-specific $\hat{\rho}_d$ then allows for straightforward evaluation of high-order, otherwise nontrivial quantities.

Specifically, we consider the probability $p(z, t)$ to measure a given bitstring z over time, where the time-average probability $p_{\text{avg}}(z)$ is given by $p_{\text{avg}}(z) = \langle z|\hat{\rho}_d|z\rangle$. We then are interested not just in the time average, but in temporal *fluctuations* about this mean, i.e., higher-order moments like the variance, etc. If we contract Eq. (3) with $\langle z|^{\otimes k}$ and $|z\rangle^{\otimes k}$, we find $\mathbb{E}_t[p(z, t)^k] = k! p_{\text{avg}}(z)^k$, where each of the $k!$ permutations have equal contribution. If we then divide out the $p_{\text{avg}}(z)$ and instead study a normalized bitstring probability $\tilde{p}(z, t) = p(z, t)/p_{\text{avg}}(z)$, we find that HSE predicts *relative* fluctuations follow a universal form of simply

$$\mathbb{E}_t[\tilde{p}(z, t)^k] = k!. \quad (6)$$

Intuitively, normalization by the mean eliminates the effects of symmetries and conservation laws which constrain HSE and make the $p_{\text{avg}}(z)$ nonequal.

III. TESTING HSE: THE PROBABILITY-OF-PROBABILITIES DISTRIBUTION

Having reviewed the basics of HSE, we now move to test whether these theoretical predictions are consistent with the behavior of an experimental quantum system. In particular, we use an experimental Rydberg quantum simulator [22,23] to implement a time-independent, global quench under the one-dimensional Ising-like Rydberg Hamiltonian, with parameters chosen such that the initial all-zero state $|\Psi_0\rangle = |0\rangle^{\otimes N}$ thermalizes at infinite effective temperature within the Rydberg-blockaded subspace [Fig. 2(a)]—see Appendix A for experimental details. We consider a system size of $N = 10$ qubits, which are

measured after evolving for a time t , producing a global bitstring $z \in \{0, 1\}^N$, and over many measurements we build up the global probability distribution $p(z, t)$. We then normalize these by $p_{\text{avg}}(z)$ (obtained from numerics) to obtain relative fluctuations $\tilde{p}(z, t)$ [Fig. 2(b)]. We note that, for numerical simulations, we consider the full Rydberg Hamiltonian, described in Appendix A, taking into account dynamics outside the blockaded Hilbert space.

A first approach to testing HSE might be to directly calculate the moments $\mathbb{E}_t[\tilde{p}(z, t)^k]$ and compare against the prediction of Eq. (6) order by order. However, we can directly compare at all orders simultaneously if, instead of considering individual moments, we study the distribution which they encode. This so-called PoP is the central object of interest for the remainder of our work [Fig. 2(c)]. Written as a function of the bitstring probability $\tilde{p}_0$, we write the PoP as (where we drop the z for notational simplicity)

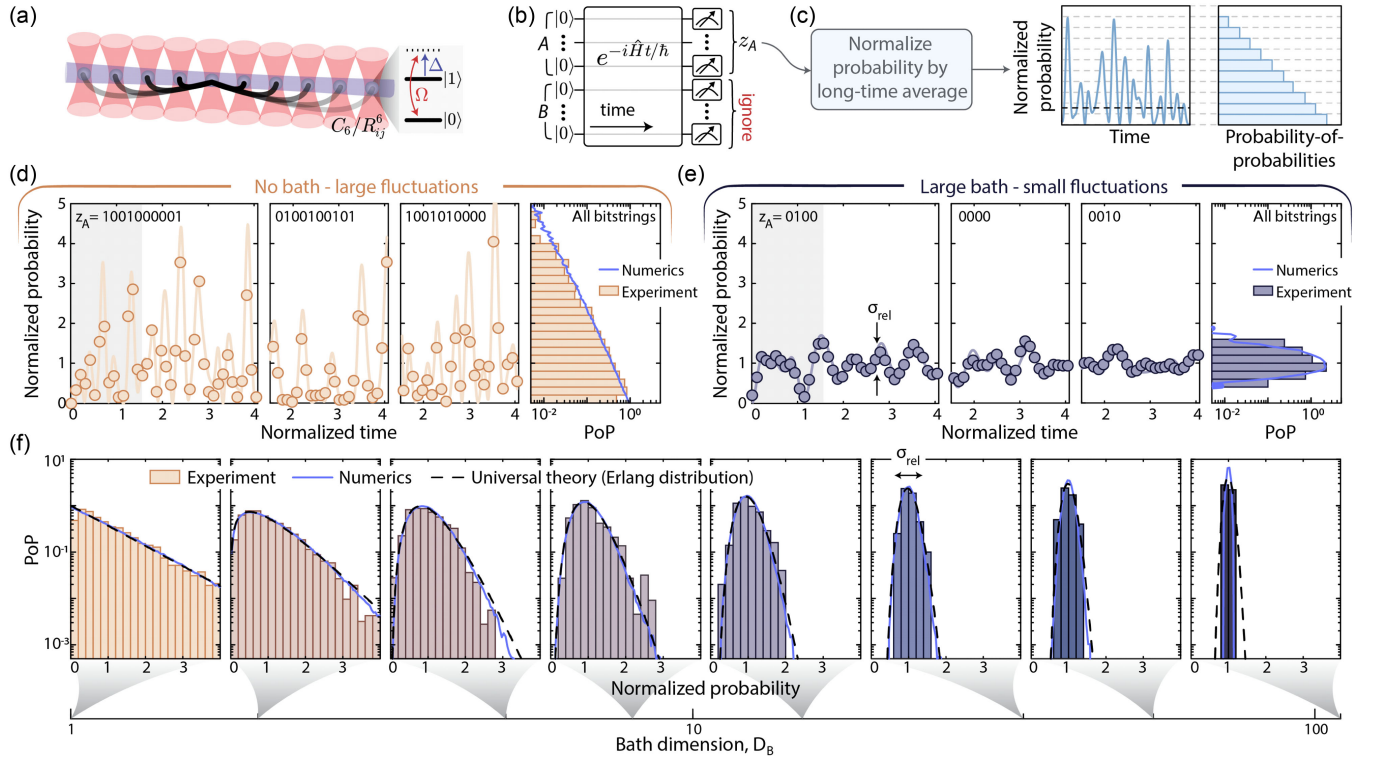


FIG. 2. HSE for system-bath interactions: a smooth quantum-to-classical transition for closed systems. (a) We employ a Rydberg quantum simulator with $N = 10$ atoms to simulate an Ising-like Hamiltonian parameterized by detuning Δ , Rabi frequency Ω , interaction constant C_6 , and interatomic distance R (model details are in Appendix A). (b) The all-zero state is time evolved and measured producing a bitstring; in postprocessing, we subselect the reduced bitstring z_A in subsystem A, ignoring the measurement in subsystem B. (c) To characterize relative fluctuations of the bitstring probabilities, probabilities are normalized by their long-time average and then binned into a histogram (indicated by dashed lines). This is the probability-of-probabilities (PoP) distribution. (d) When A is the entire system (i.e., there is no intrinsic bath, $N_A = 10$ qubits, $N_B = 0$ qubits), normalized bitstring probabilities are highly fluctuating past a short-onset timescale (gray fill before this time, time is rescaled by the Rabi frequency). (e) In contrast, when the bath is large ($N_A = 4$ qubits, $N_B = 6$ qubits), fluctuations are substantially reduced (e.g., typified by the temporal variance σ_{rel}^2), and the PoP distribution is narrow. (f) PoP distributions for many bath Hilbert-space dimensions, D_B , are in excellent agreement with our prediction that the PoP follows a universal form, the *Erlang distribution* (see the text). The Erlang distribution starts as an exponential distribution for $D_B = 1$ (an existing prediction of HSE [18]) before becoming progressively more narrow and Gaussian as D_B increases. D_B values (indicated by gray enlargement from the lower axis) are constrained by the Rydberg blockade mechanism; see Appendix A.

$$P(\tilde{p}') = \mathbb{E}_t\{\delta[\tilde{p}(t) - \tilde{p}']\} \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \delta[\tilde{p}(t) - \tilde{p}'] dt, \quad (7)$$

where δ is the δ function. We construct the experimental PoP distribution by forming a histogram of $\tilde{p}(z, t)$ aggregated over all times, using a finite bin width $\Delta\tilde{p}$.

The key property of the PoP is that the moments of the PoP are equal to the temporal fluctuations of $\tilde{p}(z, t)$ at all orders:

$$\int_0^\infty \tilde{p}'^k P(\tilde{p}') d\tilde{p}' = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \tilde{p}(t)^k dt \equiv \mathbb{E}_t[\tilde{p}(t)^k]. \quad (8)$$

In other words, the PoP encodes the full distribution of the bitstring probability values; this is in contrast to typical quantities like the cross-entropy [20] which encode only second-order fluctuations (and which require extensive classical computation). From Eq. (6), we identify the PoP for normalized bitstring probabilities under HSE as the Porter-Thomas [19,20] (exponential) distribution

$$P(\tilde{p}) = \exp(-\tilde{p}), \quad (9)$$

the same as for the case of random quantum dynamics [20,21].

For several representative choices of z , we observe the experimental $\tilde{p}$ are highly fluctuating [Fig. 2(d), left]. We can then aggregate probabilities from all bitstrings from intermediate evolution times, from which we find the PoP is a nearly exponential distribution [Fig. 2(d), right]. For the experiment here, we focus on intermediate time dynamics such that the global fidelity remains high ($\gtrsim 0.8$ [24]), i.e., such that the influence of the external environment is small. This is critical, as external decoherence can deform the PoP, as is visible in the slightly lowered experimental probability of $P(0) \approx 0.5$ compared to the theoretical prediction of $P(0) = 1$, visible in Fig. 2(d), right. We elucidate the importance of this choice in the latter half of this work, where we study the deformation of the PoP at late times under the influence of the noisy environment. This deformation effect also forces our aggregation over many bitstrings, as we can experimentally sample only a few tens of distinct early times, which is not sufficient to fully resolve the PoP on a single bitstring level. Aggregating over bitstrings does not modify the PoP, because we can treat the probabilities $p(z, t)$ as independent random variables, both over time as well as between different bitstrings [26]. These random variables are predicted to follow the same Porter-Thomas distribution, up to an overall factor of $p_{\text{avg}}(z)$ which is negligible here as the evolution is at infinite effective temperature. Furthermore, in Fig. 8 in Appendix C, we numerically show the agreement between the PoP and the exponential distribution

persists even on the level of a single bitstring sampled over time for clean, noise-free dynamics.

Furthermore, while Eq. (9) is rigorously true only in the infinite time limit, remarkably we find the key signatures of our theory are visible experimentally already at these early times (see also Fig. 9 in Appendix C). Thus, the results in Fig. 2(d) stand as a concrete experimental observation consistent with Eq. (9) and HSE in ergodic Hamiltonian dynamics more generally.

IV. GENERALIZING HSE TO SUBSYSTEMS

While we have tested HSE in the established setting of global observables, the experimental data we have on hand allow us to investigate more than just the PoP of global bitstrings. By bipartitioning the system into two subsystems A and B (of sizes N_A and N_B and Hilbert-space dimensions D_A and D_B , respectively), we can discard measurements in B and examine the PoP formed only from *local* bitstrings measured in A [Fig. 2(b)]. This allows us to directly study the fluctuations of a system coupled to a bath, where the dimension of the bath is readily controllable in postprocessing. Explicitly, we consider the marginal probability, $p(z_A, t)$, to measure a given bitstring z_A in A at a particular time t . As in the global case, we normalize probabilities by their long-time average (obtained from numerics) and then study their *relative* fluctuations, e.g., $\tilde{p}(z_A, t) \equiv p(z_A, t)/p_{\text{avg}}(z_A)$. We then seek to study how the temporal fluctuations of $\tilde{p}(z_A, t)$ differ from the global case.

As an example, we take $N_A = 4$ and $N_B = 6$ and plot $\tilde{p}(z_A, t)$ for a few choices of z_A [Fig. 2(e)]. We see that the bitstring fluctuations are greatly reduced (compared to those from the global system), though still not completely damped out. Accordingly, when we form the PoP by aggregating over multiple experimental measurement times, we see that the resultant PoP is much narrower than in the global case. Indeed, the width of the PoP decreases monotonically with increasing bath dimension D_B [Fig. 2(f)], progressing from a wide exponential distribution for $D_B = 1$ to a narrower, Gaussian-like distribution for $D_B \gg 1$.

To understand this behavior, we apply HSE to derive the behavior of local observables. In particular, under HSE we predict the PoP to be

$$P_{\text{Erlang}}(\tilde{p}; D_B) = \frac{D_B^{D_B}}{(D_B - 1)!} \exp(-D_B \tilde{p}) \tilde{p}^{D_B - 1}, \quad (10)$$

which is the *Erlang distribution*, parametrized solely by the intrinsic bath Hilbert-space dimension D_B (see Appendix C for the full derivation).

Here, we have assumed a large global Hilbert-space dimension $D \rightarrow \infty$ and thermalization at infinite effective temperature. Thus, in the thermodynamic limit, Eq. (10) dictates that temporal fluctuations of observables in subsystem A are entirely controlled by the dimension of the

bath and *not* by the dimension of the subsystem in which they are measured. The Erlang distribution is the convolution of D_B independent exponential distributions (a viewpoint which we return to and generalize in the following section concerning open system dynamics); for $D_B = 1$, the Erlang is simply the exponential distribution, while for $D_B \rightarrow \infty$ it approaches a narrow Gaussian distribution [25]. In other words, when the bath is small, large quantum fluctuations dominate (and recover the expected prediction of the exponential distribution), but when the bath is large, quantum fluctuations are damped out, leaving only small, Gaussian fluctuations. We identify this smooth quantum-to-classical crossover as integral to the transition from quantum mechanical to statistical mechanical descriptions of generic, closed, quantum many-body systems as the number of bath configurations increases.

Across various choices of D_B , we see that the Erlang distribution is in excellent agreement with experimental data across several orders of magnitude [Fig. 2(f)]. We do observe slight deviations from the Erlang distribution in Fig. 2(f) for $D_B \approx D$, as our derivation assumed $D \rightarrow \infty$. Still, using HSE, we can rigorously calculate arbitrary moments of the PoP accounting for finite D and finite temperatures, as we demonstrate explicitly for low-order moments in Appendix C. In particular, we predict the variance (second moment) of the PoP is

$$\sigma_{\text{rel}}^2(z_A) \equiv \frac{1}{D_{B,\text{eff}}(z_A)} - \frac{1}{D_\beta(z_A)}, \quad (11)$$

where $D_{B,\text{eff}}(z_A)$ and $D_\beta(z_A)$ are effective dimensions that appropriately generalize a z_A -dependent notion of Hilbert-space dimension at finite effective temperature for the subsystem B and the global system, respectively (see Appendix C for their detailed definitions). Importantly, at infinite temperature, $D_\beta \rightarrow D$ and $D_{B,\text{eff}} \rightarrow D_B$; then, taking $D \rightarrow \infty$ returns the expected variance of the Erlang distribution, $1/D_B$.

Significantly, the predictions of Eq. (10) for infinite-temperature systems and Eq. (11) for more general systems are not unique to the Rydberg Hamiltonian but are instead universal for *any* quantum many-body system which exhibits HSE [Fig. 1(b)]. To observe this, we plot $\sigma_{\text{rel}}^2 + 1/D_\beta$ for various systems expected to exhibit HSE including our experimental Rydberg system and also numerical simulations of one-dimensional systems including RUCs, a mixed field Ising model, and itinerant Hubbard models (Fig. 3)—see Appendix B for details of the various numerics. In all cases, we find good agreement with the analytic prediction of Eq. (11).

On the contrary, for failure cases which do not exhibit HSE (because they do not fulfill the no-resonance conditions [26]; see Appendix C), such as the transverse field Ising model (TFIM) and systems of noninteracting spins,

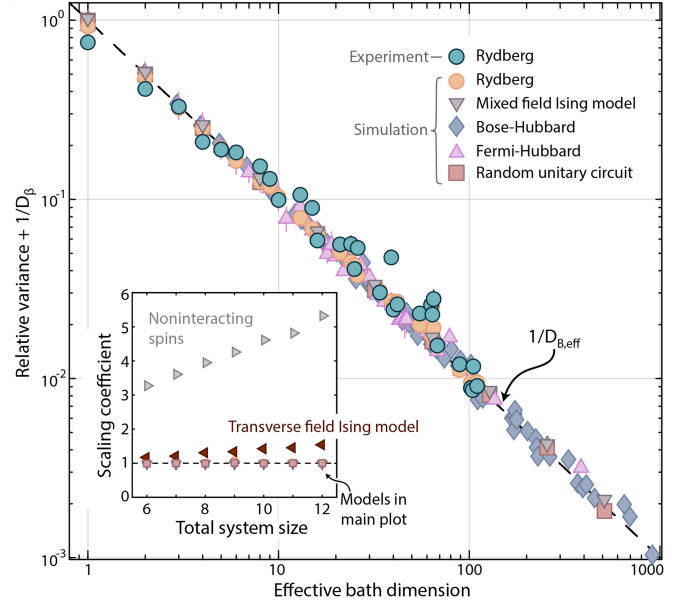


FIG. 3. Universality of the PoP variance. The PoP variance (i.e., the variance of temporal fluctuations) scales inversely with the bath dimension for both the experiment and various one-dimensional numerical models exhibiting Hilbert-space ergodicity, in agreement with Eq. (11). Markers are averaged over all bitstring realizations with an effective bath dimension in an interval from $[i - 0.5, i + 0.5)$, for $i \in \mathbb{Z}$, and error bars are the standard deviations over these realizations. Inset: the scaling coefficient $\langle D_{B,\text{eff}}(\sigma_{\text{rel}}^2 + 1/D_\beta) \rangle$ ($\langle \cdot \rangle$ denotes an average over choices of $D_{B,\text{eff}}$) is uniformly 1 for the models studied in the main plot but deviates as a function of total system size for models for which Hilbert-space ergodicity does not hold.

the prediction of Eq. (11) is violated (Fig. 3 inset). Specifically, the estimated scaling coefficient with bath dimension— $\langle D_{B,\text{eff}}(\sigma_{\text{rel}}^2 + 1/D_\beta) \rangle$, where $\langle \cdot \rangle$ is an average over choices of $D_{B,\text{eff}}$ —is uniformly 1 for systems exhibiting HSE but shows nonuniform and progressively greater deviations from 1 as a function of system size for non-ergodic cases. Both the TFIM and the 1D Fermi-Hubbard model are integrable (the former via Jordan-Wigner transformation [43], the latter by Bethe ansatz [44]), but HSE holds for the latter and not the former because of the no-resonance conditions [26], and accordingly the 1D Fermi-Hubbard model exhibits the expected universal behavior in Fig. 3 while the TFIM does not. These results support the universality of HSE and demonstrate its testability in low-order observables, even incorporating the effects of finite system size.

V. HSE PRINCIPLES FOR OPEN SYSTEMS AND APPLICATIONS IN NOISE LEARNING

So far, we have studied HSE for closed system dynamics through the lens of the PoP distribution; for both local and global systems, we found that the PoP has a universal, quantitative form. Now we ask how this behavior

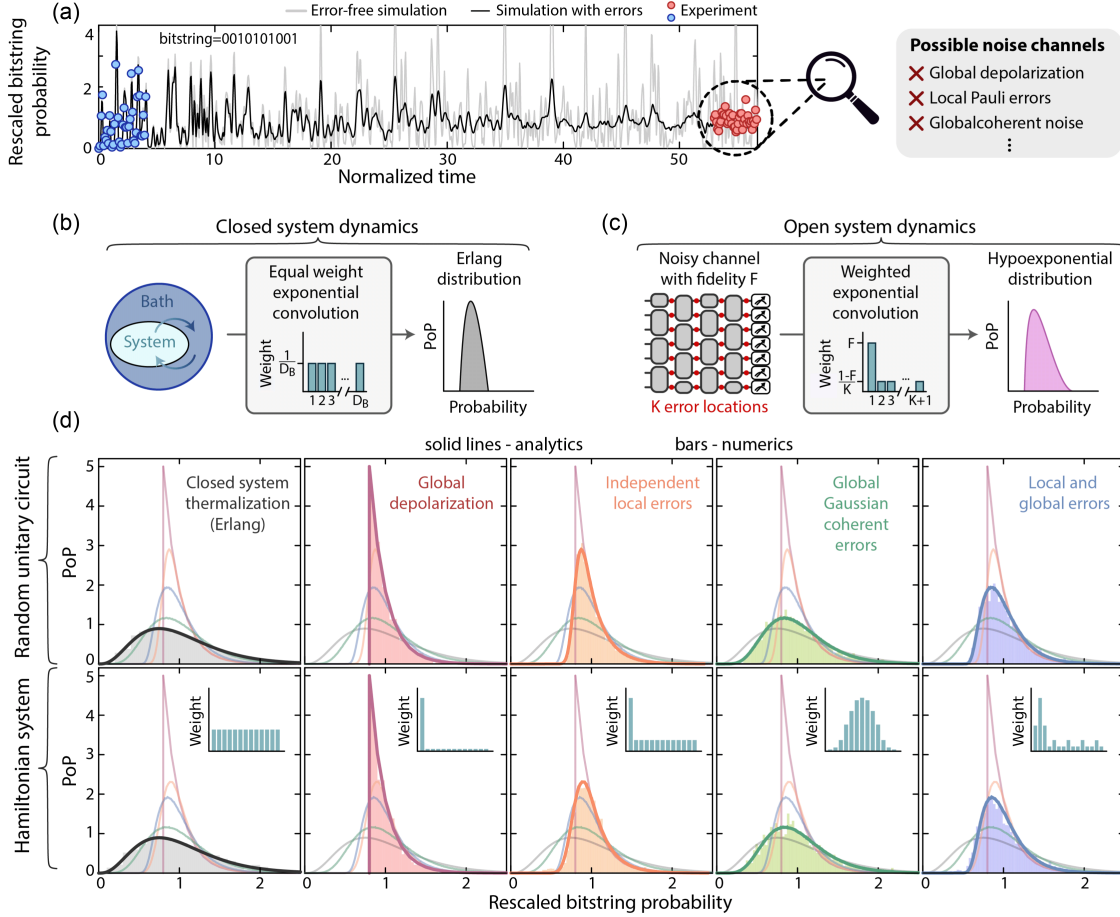


FIG. 4. Distinguishing noise channels using PoP distributions. (a) Interaction with the external environment through open system dynamics also leads to concentration of bitstring probabilities; as an example, we show bitstring probabilities following quench dynamics from experiment (markers), error-free simulation (gray line), and error-model simulation [24] (black line). However, the microscopic error model is highly relevant to the manner of concentration, which allows us to discriminate different possible noise models directly from bitstring measurements. (b) The Erlang distribution observed in the case of closed, infinite-temperature thermalization can be alternatively derived as a convolution of independent exponential distributions with equal weights, one for each bath configuration. (c) We extend this heuristic derivation to the case of open system dynamics, where the weights in the convolution are determined by the specifics of the noise channel, for example, here a local error channel (see the text), resulting in the PoP being a *hypoexponential* distribution. (d) Hypoexponential distributions for four different noise channels, each with distinct hypoexponential weights (insets; see Appendix B for model details), with noise strengths set to result in the same fidelity $F \approx 0.2$. Also shown is the Erlang distribution from closed system thermalization with an intrinsic bath of two qubits. PoP obtained from numerical simulations of RUC (top) and Hamiltonian evolution under a mixed field Ising model (bottom) show excellent agreement with their corresponding analytical predictions (thick, dark lines) but can easily be distinguished from predictions of other bath types (thin, faint lines), both intrinsic and extrinsic.

changes under the influence of an external, uncontrollable environment, as is the setting for all modern quantum experiments.

Under such noisy open system dynamics, the PoP distribution narrows [20]; i.e., fluctuations of observables damp out [Fig. 4(a)], analogous to how fluctuations damped out for local subsystems as the bath dimension increased. We introduce a quantitative description of this open system behavior through an appropriate generalization of the Erlang distribution, leveraging our understanding of baths from closed systems into open ones. This provides a tool to distinguish between the different types of

noise in the system even when the system and environment are far from equilibrium—applying it to experimental and numerical data yields predictions in agreement with *ab initio* noise models of the system.

We first revisit our understanding of the Erlang distribution for closed system thermalization. Under HSE, the global bitstring probability $\tilde{p}(z)$ is randomly sampled from an exponential distribution. The marginal probability $\tilde{p}(z_A)$ is then a sum of D_B uncorrelated exponential random variables, each with equal weight, $\omega = 1/D_B$, as the infinite-temperature bath has no preferred configuration. This process leads precisely to the Erlang distribution; i.e.,

the Erlang distribution is the convolution of D_B exponential distributions [25] [Fig. 4(b)].

We can make a similar assumption for open system dynamics—that for each bath configuration the bitstring distribution is sampled from an independent exponential distribution—but now allow for nonequal weights that are determined by the microscopic details of the extrinsic coupling mechanisms. In other words, for each bath state, we take the system to be in an independently random state. As an example, we consider the case of RUC evolution, where a single error can occur on a random qubit sometime during the evolution [Fig. 4(c)]. For a circuit with N qubits run for a depth D , there are $K = N \times D$ potential error locations, and the circuit executes with a finite global fidelity of F . Thus, with a probability of F , the circuit will execute without error and will sample bitstring probabilities from an exponential distribution. However, if an error does occur at one of the K locations, assuming the circuit is sufficiently scrambling [20], it will sample a separate *independent* exponential distribution. Without the ability to explicitly keep track of all errors, the result is an incoherent sum of independent exponential variables with weights $\omega_1 = F$ and $\{\omega_2, \dots, \omega_{K+1}\} = (1 - F)/K$ [Fig. 4(c)].

For a given set of weights ω_i (which are positive and sum to 1) and associated probability distributions $\{p_i(z)\}$, the overall bitstring probability distribution is given by the weighted sum:

$$p(z) = \sum_i \omega_i p_i(z). \quad (12)$$

When all weights are equal, the associated PoP is the Erlang distribution, but when weights are allowed to be nonequal, we identify the associated PoP as a *hypoexponential distribution* [25,45], or generalized Erlang distribution, with the probability distribution function (PDF) given by

$$P_{\text{Hypo}}(x) = \otimes_i P_{\text{Exp}}(x/\omega_i), \quad (13)$$

where $\otimes$ denotes repeated convolution. While Eq. (13) does not have a closed form for general weights ω_i , it can be analytically simplified for many relevant special cases and can always be numerically solved. Thus, we can quantitatively characterize largely arbitrary noise channels which couple the quantum system to the external environment simply by adjusting the weight vector [Fig. 4(c)]. See Appendix D for explicit analytical constructions of the hypoexponential distribution for various error models and Algorithm 1 therein for an example of how to numerically apply this approach for more bespoke error models. Note that the hypoexponential weights can be related to the Kraus operators of a given noise channel; see Appendix G.

To demonstrate the power of this formalism, we perform numerical simulations of RUC evolution and mixed field Ising model Hamiltonian evolution; explicitly, we study

the global PoP distribution measured at a fixed time, for which the results of HSE also apply [18] (see Fig. 8 in Appendix C). In both cases we apply either (i) global depolarizing errors, (ii) local amplitude damping errors, (iii) global Gaussian coherent errors, or (iv) a combination of errors (ii) and (iii). For each, the error strength is tuned so the final fidelity is $F \approx 0.2$. By *coherent error*, we mean a static, shot-to-shot under- or overrotation in either the qubit gates or the amplitude of the Hamiltonian terms, which here we take to be Gaussian distributed and applied to all qubits globally. Crucially, determining the weight vectors and associated hypoexponential distributions for the different noise models does not require explicit simulation and instead follows directly from the model definitions; see Appendix D. We study these representative noise models, as they are highly relevant to various aspects of quantum science. For instance, global depolarizing error is often assumed in studies of error-correcting codes, local amplitude damping is prevalent in T1-limited systems like many superconducting circuits [46], global Gaussian coherent errors are limiting for certain analog simulations including for Rydberg systems [24], and the combination of local and global errors is likely universal for all quantum platforms.

We note that the weights for the global depolarizing channel are the same as those of the local error channel with $K \rightarrow \infty$, consistent with Ref. [47]. By contrast, however, the PoP for global coherent errors does not converge to the PoP of global depolarizing noise for increasing system sizes at fixed fidelity. Coherent errors could arise from non-Markovian Hamiltonian parameter variations [24,48] or systematic gate miscalibrations [46,49]; discriminating such noise sources from Markovian errors is crucial, as they lead to different functional forms of fidelity decay which may be otherwise difficult to distinguish [24,50], and their mitigation necessitates different models of error correction [35–37,51].

In all cases, we see excellent agreement between the numerical PoP distributions and their corresponding analytical hypoexponential predictions, while being visually distinct from those of other noise channels, and from thermalization with an infinite-temperature bath with $D_B = 4$ [Fig. 4(d)]. Thus, on a qualitative level, we can discriminate between a wide range of noise models through many-body bitstring probability measurements, even when those models lead to an equal final fidelity.

We then turn to apply this method quantitatively to our experimental quantum simulator. We program the same time-independent dynamics as we studied in the case of intrinsic thermalization but now evolve out to late times with $N = 15$ atoms such that the global fidelity is $F \approx 0.18$, as estimated via the F_d cross-entropy variant [24,26]. We then compare the experimental PoP distribution against the predictions for various noise models including our in-house *ab initio* noise model. The use of a finite number of samples modifies the predicted PoP distributions [inset in Fig. 5(a); see Appendix E], which can potentially be circumvented

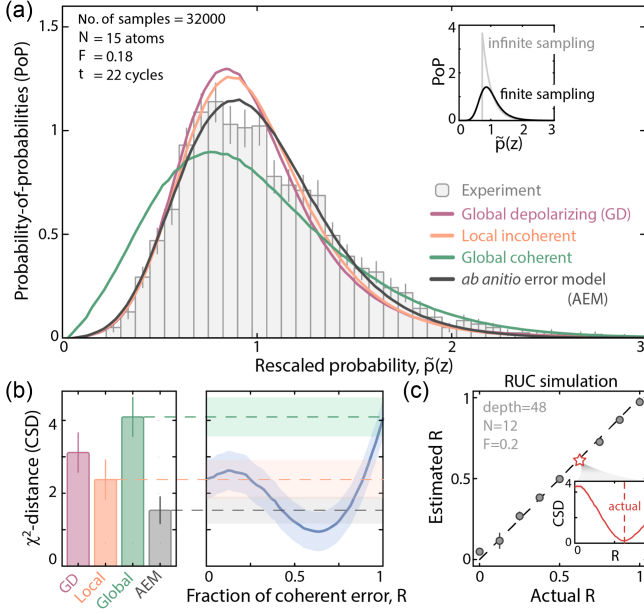


FIG. 5. Characterizing experimental noise. (a) We compare the PoP distribution obtained from the experiment following long-time evolution against various noise models, including our own *ab initio* error model. Model PoPs are nontrivially modified by finite sampling effects (inset, see Appendixes). (b) We compute the symmetric χ^2 distance (CSD) between the experimental distribution and the model PoPs (left), including to a mixture error model composed of global coherent error and local incoherent error with varying global coherent error fraction, R . (c) We perform noisy RUC simulations with varying R and find that estimating R from the CSD minimum is well correlated with the actual R (Appendixes).

by applying postprocessing to rectify the distributions (Fig. 15). Furthermore, even with limited sampling, the different noise models can be robustly discriminated by studying more sample-efficient low-order moments of the PoP (Appendix F).

Importantly, in Appendix F, we also derive the fidelity dependence of the low-order moments of the global PoP under various noise channels. For instance, we find that the second moment of the PoP scales as F^2 for global depolarizing noise but F for global Gaussian coherent noise. These sharp dependencies underscore the necessity of operating the experimental system in a high-fidelity regime when studying the effects of thermalization on the PoP, as we do in the early part of this work.

A qualitative evaluation [Fig. 5(a)] already suggests our *ab initio* noise model is the best match to the experiment. To quantify this statement, we compute the normalized symmetric χ^2 distance (CSD) between the experiment and the model PoPs, defined as

$$\chi^2(\vec{x}, \vec{y}) = \frac{1}{B} \sum_i \frac{(x_i - y_i)^2}{(x_i + y_i)/2}, \quad (14)$$

where B is the number of bins in the PoP, x_i is a bin drawn from experiment, and y_i is a bin drawn from the model distribution. We find the distance is indeed minimized for the *ab initio* noise model [Fig. 5(b)].

Furthermore, we apply this technique to error models with a mixture of global Gaussian coherent error and local incoherent error. We vary the fractional contribution of global coherent error R and analytically construct the model PoP to compare against. We find the lowest CSD is found for $R \approx 2/3$, suggesting room for future improvement of our *ab initio* error model. This technique is applicable to both analog quantum simulators and digital quantum processors. For instance, we simulate noisy RUC dynamics for different choices of R and for each compare the R estimated from the CSD minimum versus the actual R [Fig. 5(c)]. We find the estimated and actual values are well correlated (which improves with deeper circuits; see Fig. 13) across the entire range of R . This result implies that one can quantitatively estimate the relative contributions of coherent versus incoherent errors in noisy circuit dynamics using our methods.

Importantly, our analysis requires no circuit or Hamiltonian simulation and relies only on the emergence of the Porter-Thomas distribution, which applies to both time-dependent and time-independent Hamiltonian evolution, as well as random quantum circuits. Furthermore, it makes no assumption about the underlying noise model besides its weight vector. However, this also means it cannot discriminate between noise models with the same weights, such as distinguishing local bit flips from local phase flips. Thus, it should be considered a tool for efficiently identifying experimentally consistent models with bespoke noise distributions, which can then be further refined via additional measurements to distinguish different models with the same weights. More generally, beyond its potential practical applications, this approach opens a new suite of theoretical analysis for understanding open system dynamics, as, for instance, the weights of the hypoexponential distribution are related to the system's Kraus operators; see Appendix G.

VI. DISCUSSION

The evolution of a quantum system interacting with its environment is of both fundamental and practical interest. Here, we have studied this process through the lens of Hilbert-space ergodicity [18] both for the case of interactions between complementary subsystems and for noisy interactions with an external environment.

For closed system thermalization at infinite temperature, we have analytically, experimentally, and numerically found that the PoP for bitstring measurements is universally the Erlang distribution. As a corollary, we showed experimental evidence consistent with existing predictions of HSE, namely, the emergence of the Porter-Thomas distribution for global systems. Accordingly, there is a smooth quantum-to-classical transition with bath size from large,

Porter-Thomas fluctuations to small, Gaussian fluctuations, highlighting the transition between quantum mechanical and statistical mechanical descriptions of physical systems as the number of bath configurations grows. This transition was observed universally for systems obeying HSE, even for systems which were integrable, which evolved at finite temperature, or which were constrained by conservation laws. While questions remain for calculating the PoP distribution at all orders for finite effective temperatures, we have made significant strides, showing how low-order moments of the PoP may be calculated as a function of an effective bath dimension and showing how the PoP may be obtained in some limits (Appendix D). It remains open how our results extend to systems which do not exhibit Hilbert-space ergodicity, such as some noninteracting integrable systems [43] or potentially nonthermalizing models like many-body localized systems [4,5].

For the case of open system dynamics, the hypoexponential model we introduce allows us to categorize different noise models via the weight vector and produce an analytical prediction for the PoP. This enables experimentally accessible discrimination between candidate noise models purely through their effect on the many-body bitstring PoP, without expensive numerical simulations of quantum dynamics. We have demonstrated how bespoke noise models can be found through a quantitative comparison between model predictions and experimental statistics, as has been explored in the context of random circuit sampling [52–54]. Our technique is computationally inexpensive and applies equally to analog quantum simulators and digital quantum processors. Additional quantitative tests, such as comparing the low-order moments of the PoP distributions, may enhance discrimination power while being more sample-efficient. Fruitful questions remain, such as directly learning the hypoexponential weights via entanglement spectroscopy methods [55,56]. Furthermore, learning the weights of heavy-tailed and correlated noise distributions may guide improvements of quantum error correction schemes [35–37]. As our method takes the fidelity as input, we note it is an explicit application of fidelity-estimation techniques [10,20,24,26,49].

Ultimately, our results probe new types of universal behavior visible to high-precision quantum systems. In turn, the universality we uncover has applications for both improving quantum experiments and studying fundamental quantum dynamics. In total, our work demonstrates the power of applying a unified approach to studying many-body quantum systems, for both open and closed dynamics, through HSE principles.

ACKNOWLEDGMENTS

We acknowledge insightful discussions with Bill Fefferman, Andreas Elben, Gil Refael, and Federica M. Surace and feedback on this manuscript from Elie Bataille, Richard Tsai, Xiangkai Sun, and Gyohei Nomura.

We acknowledge support from the NSF QLCI program (OMA-2016245), the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant No. PHY-1733907), the Center for Ultracold Atoms, an NSF Physics Frontiers Center (NSF Grant No. PHY-1734011), the DOE (DE-SC0021951), the DARPA ONISQ program (W911NF2010021), the NSF CAREER Awards (No. 1753386 and No. DMR-2237244), the AFOSR YIP (FA9550-19-1-0044), and the AFOSR (FA9550-23-1-0625). Support is also acknowledged from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. J. C. acknowledges support from the Terman Faculty Fellowship at Stanford. R. F. acknowledges support from the Troesh postdoctoral fellowship. P. S. acknowledges support from the IQIM postdoctoral fellowship.

APPENDIX A: DETAILS OF THE EXPERIMENT

Our experiment is a Rydberg atom array quantum simulator [23] trapping individual strontium-88 atoms in optical tweezers [57,58]. Up-to-date details of our apparatus may be found in previous works [24,59]. In brief, one-dimensional defect-free [60,61] arrays of atoms are initialized in the $5s^2\ ^1S_0$ state and are cooled on the narrow-line $5s^2\ ^1S_0 \leftrightarrow 5s5p\ ^3P_1$ transition close to their motional ground state. Atoms are prepared into the long-lived $5s5p\ ^3P_0$ clock state with a preparation fidelity of approximately 0.996 per atom, which we then treat as a metastable ground state, $|0\rangle$. Atoms are then driven globally to the $5s61s\ ^3S_1, m_J = 0$ Rydberg state, $|1\rangle$ while the traps are briefly blinked off.

Following Hamiltonian evolution, state projection is performed by autoionizing [62] the Rydberg atoms, leaving them dark to our fluorescence imaging, with a fidelity of approximately 0.9996 per atom. Atoms in the clock state are pumped into the imaging cycle, from which we map atomic fluorescence to qubit state [62,63] with a detection fidelity $\gtrsim 0.9995$ per atom [59,63]. For an initially defect-free array, this results in a series of bitstrings associated with the measured qubit states in the array; note that we discard any experimental bitstrings for which initial rearrangement failed.

1. System Hamiltonian

The Hamiltonian of our system is given by

$$\hat{H}/\hbar = \Omega \sum_i \hat{S}_i^x - \Delta \sum_i \hat{n}_i + \frac{C_6}{a^6} \sum_{i>j} \frac{\hat{n}_i \hat{n}_j}{|i-j|^6}, \quad (\text{A1})$$

which describes a set of interacting two-level systems, labeled by site indices i and j , driven by a laser with Rabi frequency Ω and detuning Δ . The interaction strength is determined by the C_6 coefficient and the lattice spacing a .

Operators are $\hat{S}_i^x = (|1\rangle_i\langle 0|_i + |0\rangle_i\langle 1|_i)/2$ and $\hat{n}_i = |1\rangle_i\langle 1|_i$, where $|0\rangle_i$ and $|1\rangle_i$ denote the electronic ground and Rydberg states at site i , respectively. Both Ω and Δ are tunable.

We calibrate Hamiltonian parameters first via independent measurements and then optimize further through maximum-likelihood estimation based on fidelity estimation. This technique is described in detail in our previous work [10]; in brief, we use the accumulated experimental bitstrings to calculate the many-body fidelity estimator [10,24,26], F_d , through cross-correlation with a numerical simulation. We then vary the simulation parameters to find the optimal fidelity overlap with the experiment.

For the $N = 10$ data at approximately infinite temperature shown in Figs. 2 and 3 in the main text, the Hamiltonian parameters are chosen to be $\Omega/2\pi = 5.3$ MHz, $\Delta/2\pi = 0.6$ MHz, $C_6/2\pi = 254$ GHz $\times \mu\text{m}^6$, and $a = 3.77$ μm (the lattice spacing is calibrated with a laser-based ruler [64]). For the $N = 15$ data at approximately infinite temperature shown in Fig. 5 in the main text, the Hamiltonian parameters are the same, but with $\Omega/2\pi = 5.4$ MHz. For the $N = 10$ data at finite temperature in Fig. 7, the parameters are the same, but with $\Omega/2\pi = 5.3$ MHz and $\Delta/2\pi = 4.6$ MHz.

We typically normalize Hamiltonian evolution times by the Rabi frequency, with a corresponding unit of *cycles*.

2. Data analysis

Through measurement, we accrue a set of time-labeled bitstrings, $\{z\}$. When normalizing bitstring probabilities, we obtain the temporal long time average directly from the *diagonal ensemble* [2,3,26]:

$$\hat{\rho}_d = \sum_n |c_n|^2 |E_n\rangle\langle E_n|, \quad (\text{A2})$$

where $|E_n\rangle$ are energy eigenstates and $c_n = |\langle E_n|\Psi_0\rangle|^2$ are overlaps with the initial state. Any observable, including bitstring probabilities, of $\hat{\rho}_d$ is equal to the same observable obtained from an infinite time average:

$$p_{\text{avg}}(z) = \langle z|\hat{\rho}_d|z\rangle = \int_0^\infty dt |\langle z|\Psi(t)\rangle|^2. \quad (\text{A3})$$

A major feature of our system Hamiltonian is the Rydberg blockade mechanism [23], whereby the energy eigenspectrum is naturally stratified into a number of widely spaced sectors. In particular, the nearest-neighbor interaction is approximately $20\times$ greater than the next-largest energy scale, so cases where neighboring pairs of atoms are both excited to the Rydberg state are greatly suppressed. The resulting effective Hilbert-space dimension D of a system with N qubits is given by $\text{Fib}(N+2) \approx 1.6^N$, where Fib is the Fibonacci function. To be clear, statements in this work that the Rydberg system

thermalizes at infinite effective temperature refer only to thermalization within the blockade-satisfying subspace, not across the entire Hilbert space.

For $N = 10$, we find approximately 99% of all experimental bitstrings are in the blockade-satisfying energy sector at short times; for data shown in the main text, we refine $\{z\}$ by discarding all realizations not in this sector. For the late-time $N = 15$ data, the number of blockade-satisfying bitstrings drops to approximately 75%, which we attribute mainly to spontaneous decay [59] though leave careful investigation to future work. We stress that the number of blockade-satisfying bitstrings decreases only mildly as a function of system size [24] and is not a major impediment to the studies of either closed-system thermalization or open system decoherence that we carry out in this work.

The effective dimension of the intrinsic bath, D_B , is altered by the blockade constraint. Specifically, the blockade must be accounted for at the boundary between subsystems A and B , such that D_B then depends on the bitstring measured in A . As an example, consider the system A as the leftmost atoms in a one-dimensional array with ten total atoms and B as the remaining $N_B = 9$ atoms, such that the global bitstring is written as $z = z_A z_B$. For the subsystem bitstring $z_A = 0$, there is no possible blockade violation across the boundary between A and B , and so the bath dimension is given by $D_B(z_A = 0) = \text{Fib}(N_B + 2) = 89$. However, for the subsystem bitstring $z_A = 1$, the blockade constraint forbids the leftmost qubit in B from also being a 1. Thus, the bath dimension is $D_B(z_A = 1) = \text{Fib}[(N_B - 1) + 2] = 55$. In general, we allow A to be any contiguous block of atoms in the array (meaning B may be discontinuous), and, for instance, in the analysis of Fig. 2(c) in the main text, we aggregate over bitstrings in A with equal effective bath dimension.

The variance for the random unitary circuit in Fig. 3 is taken as the median over 1000 circuit realizations. In Fig. 5, the error model prediction is calculated via a Monte Carlo wave-function approach [10,24,65] from which bitstrings are sampled at an equal number as the experiment. Experimental error bars in Fig. 5 are analytic based on the Poisson sampling errors: For a histogram bin with m counts, its standard error is $\sqrt{m}$ (see Appendix E 3).

APPENDIX B: NUMERICAL SIMULATION PARAMETERS

Here, we summarize the parameters used in numerical simulations throughout the main text. In general, PoP distributions from numerics are aggregated over several hundred cycles of evolution, except where otherwise noted.

1. Error-free simulations

Rydberg—Rydberg simulation parameters are the same as for experiment, as given in a previous section.

MFIM—For the mixed field Ising model simulations in Figs. 3 and 7, we utilize the following Hamiltonian [66]:

$$\hat{H}/\hbar = \sum_{i=1}^N (h_x \hat{S}_i^x + h_y \hat{S}_i^y - \zeta \hat{S}_i^z) + \sum_{i=1}^N J \hat{S}_i^x \hat{S}_{i+1}^x, \quad (\text{B1})$$

where $h_x = 0.41$, $h_y = 0.45$, and $J = 1$. The field $\zeta = 0$ corresponds to infinite-temperature evolution and $\zeta = 0.86$ to finite-temperature evolution. In Fig. 11, we instead set $\zeta = 0$ but tune the initial state as $|\Psi_0\rangle = [\exp(i\theta \hat{S}^x)|0\rangle]^{\otimes N}$ for various angles θ .

TFIM—For the transverse field Ising model simulations in the inset in Fig. 3, we consider a one-dimensional open chain under the following Hamiltonian:

$$\hat{H}/\hbar = -\sum_i \hat{S}_i^x - 3 \sum_{i \neq j} \hat{S}_i^z \hat{S}_j^z. \quad (\text{B2})$$

We use as an initial state of $|\Psi(t=0)\rangle = |+\rangle^{\otimes N}$.

Noninteracting spins—Noninteracting spins studied in the inset in Fig. 3 are initialized in their ground state and then evolved with equal Rabi frequency and no detuning, thus undergoing disconnected, correlated Rabi oscillations.

Bose-Hubbard—We simulate the Bose-Hubbard model with N particles on a one-dimensional open chain with length M :

$$\hat{H}/\hbar = -\Omega \sum_{j=1}^{M-1} (\hat{b}_j^\dagger \hat{b}_{j+1} + \text{H.c.}) + \frac{U}{2} \sum_{j=1}^M \hat{n}_j (\hat{n}_j - 1), \quad (\text{B3})$$

where $\hat{n}_j = \hat{b}_j^\dagger \hat{b}_j$. We choose parameter values $(\Omega, U) = (1, 2.87)$. In our simulations, we have $M = N$, and we take the initial state to be the one with one particle at each site, i.e., $|1, 1, \dots, 1\rangle$.

Fermi-Hubbard—The 1D Fermi-Hubbard model Hamiltonian is

$$H_{\text{FH}} = \sum_{j=1}^N \left[-\Omega \sum_{\sigma=\uparrow, \downarrow} (c_{j,\sigma}^\dagger c_{j+1,\sigma} + \text{H.c.}) + U n_{j,\uparrow} n_{j,\downarrow} \right], \quad (\text{B4})$$

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. We use the parameter value $(\Omega, U) = (1, 1)$. We take as initial state the antiferromagnetic state at half filling, $|\uparrow, \downarrow, \dots\rangle$.

Random unitary circuit—For RUC simulations, we consider a brickwork pattern of $\text{SU}(4)$ gates.

2. Open system simulations

Rydberg Hamiltonian dynamics—Our *ab initio* error model for Rydberg Hamiltonian dynamics has been

described previously [10] and shown to be accurate to experiment in several regimes [10,24,59]. We utilize stochastic wave-function evolution [65] with around a thousand trajectories, incorporating various precalibrated noise sources such as laser noise, atomic motion, and spontaneous decay, among others. We note that many of these noise sources generically are highly non-Markovian with important impact on the many-body fidelity [24]. Furthermore, spontaneous decay channels are potentially nonunitary and predominantly dominated by leakage errors out of the qubit subspace [59].

Spin chain Hamiltonian dynamics: Amplitude damping—To obtain the bottom row in Fig. 4 in the main text, we simulate the open system dynamics of the mixed Field Ising model [Eq. (B1)] in the presence of amplitude damping, or spontaneous decay $|0\rangle\langle 1|$ errors, with the quantum trajectory method [65,67]. In all trajectory methods, we perform a sufficient number of trajectories, typically 1000 for a system size of $N = 12$, to see convergence of the PoP.

Global coherent errors—To simulate global coherent errors, we again perform a trajectory method, where in each trajectory we sample the parameter J in Eq. (B1) from a normal distribution $N(J_0, \sigma_J^2)$, with $J_0 = 1$ and σ_J tuned to achieve the desired fidelity at a given time.

Mixed error sources—To simulate both amplitude damping and global coherent errors, we simply use a quantum trajectory method with both noise sources enabled.

Random unitary circuits: Local errors—We simulate the open system dynamics of a RUC in the presence of amplitude damping errors. We evolve the density matrix by applying a layer of random unitary gates, before applying Kraus operators of the amplitude damping channel (for the top row in Fig. 4) independently at every qubit.

Global coherent errors—In order to model global coherent errors in a RUC, we parametrize each ideal random two-qubit gate as an $\text{SU}(4)$ rotation about some axis and with some angle. In each run, we model the angles of each gate to be systematically over- or underrotated by a factor which is constant over all gates within each run but sampled from a normal distribution over many runs. For the mixed error sources in Fig. 4, we perform this stochastic over- and underrotation with the density matrix evolution method outlined above.

Mixed error sources—For the mixed error sources in Fig. 5, we use a quantum trajectory method with both bit-flip errors and stochastic over- and underrotation errors (outlined above).

APPENDIX C: DERIVING THE ERLANG DISTRIBUTION FROM HILBERT-SPACE ERGODICITY

In this work, we studied closed system dynamics and predicted the distribution functions of bitstring probabilities (or PoP) in a subsystem which is strongly coupled to the

rest of the system that serves as a bath. Our main quantitative predictions were Eq. (10), the Erlang distribution, for such systems at infinite effective temperature and large total system sizes, as well as Eq. (11) for the variance of the PoP at arbitrary temperatures.

Here, we provide a rigorous derivation of the Erlang distribution, repeating some of the statements from Sec. II for completeness. First, we formally define the PoP distribution and the no-resonance conditions upon which Hilbert-space ergodicity depends. Then, we derive the Erlang distribution which describes the PoP of subsystems thermalizing at infinite effective temperature, during closed system dynamics and in the large system size limit. In more technical sections, we then relax some of our assumptions and show how arbitrary order moments of the PoP may be calculated even at finite effective temperatures and finite system sizes. We also note that we provide a different, more heuristic, derivation of the Erlang distribution in Appendix D 3.

1. The probability-of-probabilities distribution

The PoP distribution for a single bitstring, z , is a histogram of the bitstring probabilities, $p(z, t)$, formally over infinite time. For a given probability p_0 , we write it as (where we drop the z for notational simplicity)

$$P(p_0) = \mathbb{E}_t\{\delta[p(t) - p_0]\} \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \delta[p(t) - p_0] dt. \quad (\text{C1})$$

The PoP encodes all temporal fluctuations of the time series $p(t)$ as its moments:

$$\begin{aligned} \int_0^1 p_0^k P(p_0) dp_0 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \int_0^1 p_0^k \delta[p(t) - p_0] dp_0 dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T p(t)^k dt \equiv \mathbb{E}_t[p(t)^k]. \end{aligned} \quad (\text{C3})$$

The PoP is defined here for a single bitstring over all evolution times, but it can alternatively be defined for a single time over all bitstrings, a distinction which we revisit below.

In this study, we are chiefly interested in studying *relative* temporal fluctuations, i.e., normalizing the time series $p(t)$ to have mean 1. For random-circuit dynamics, this is simply the rescaling $\tilde{p} \equiv Dp$, while in Hamiltonian dynamics, this is $\tilde{p}(z) \equiv p(z)/p_{\text{avg}}(z)$, where $p_{\text{avg}}(z) \equiv \mathbb{E}_t[p(z, t)]$ is the time-averaged probability of z . This accordingly rescales the PoPs, which we write as $P(\tilde{p})d\tilde{p}$ (occasionally omitting $d\tilde{p}$). The $p_{\text{avg}}(z)$ may be found from the diagonal ensemble:

$$\hat{\rho}_d = \sum_n |c_n|^2 |E_n\rangle\langle E_n|, \quad (\text{C4})$$

where $|E_n\rangle$ are energy eigenstates and $c_n = |\langle E_n | \Psi_0 \rangle|^2$ are overlaps with the initial state. Any observable, including bitstring probabilities, of $\hat{\rho}_d$ is equal to the same observable obtained from an infinite time average:

$$p_{\text{avg}}(z) = \langle z | \hat{\rho}_d | z \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt |\langle z | \Psi(t) \rangle|^2. \quad (\text{C5})$$

2. The no-resonance conditions

As described in Sec. II, the chief object of our interest is the *temporal ensemble*, the set of pure states, $|\Psi(t)\rangle$, evaluated at different times t for evolution under a time-independent Hamiltonian $\hat{H}$. The temporal ensemble has been previously studied [18,26,38], where at late times it was connected to the *random phase ensemble*, the ensemble of quantum states with well-defined amplitudes in the energy eigenbasis but random phases. This connection depends on so-called *no-resonance conditions* [28], which are colloquially the absence of high-order degeneracies in the energy eigenspectrum. More concretely, a Hamiltonian $\hat{H}$ satisfies the k th no-resonance condition if for any two sets of k eigenvalues $\{E_{\alpha_j}\}_{j=1}^k$ and $\{E_{\beta_j}\}_{j=1}^k$ the equation

$$E_{\alpha_1} + E_{\alpha_2} + \dots + E_{\alpha_k} = E_{\beta_1} + E_{\beta_2} + \dots + E_{\beta_k} \quad (\text{C6})$$

is true if and only if the sets of indices $(\alpha_1, \alpha_2, \dots, \alpha_k)$ and $(\beta_1, \beta_2, \dots, \beta_k)$ are equal up to reordering. For example, the second-no-resonance condition states that the relation $E_a + E_b = E_c + E_d$ is true if and only if $a = c, b = d$ or $a = d, b = c$. That is, there are no “accidental resonances.” Existence of k th no-resonance conditions for any order k guarantees the temporal ensemble is equivalent to the random phase ensemble in the infinite time limit [18]; as shown in Ref. [18], this then is equivalent to the statement that Hilbert-space ergodicity holds. We note that, while the statements of HSE require the no-resonance conditions to hold at all orders, this is not necessary for trivial first-order degeneracies, which can be eliminated by working with representative states from projections onto degenerate subspaces [26].

3. The Erlang distribution from Hilbert-space ergodicity

In Sec. II, we describe how the Porter-Thomas distribution could be derived for global systems using Eq. (3), the moments of the temporal ensemble. Similarly, we can also derive that the rescaled marginal distributions $p(z_A)/p_{\text{avg}}(z_A)$ follow the Erlang distribution, as stated in the main text. To do so, we consider the k th cumulants of $p(z_A)$, which is related to the k th moment but subtracts off trivial contributions from lower-order moments. The k th cumulant of the Erlang distribution can be simply calculated as $\kappa_k = (k-1)!/D_B^{k-1}$, since the Erlang distribution is

the convolution of multiple exponential distributions [25], and cumulants of convolutions are additive [68].

To calculate the k th cumulant, we consider the permutations in Eq. (3). The permutation operator swaps between different copies of the Hilbert space and can be represented by a directed graph indicating the action of the permutation on k elements. Importantly, this graph can be disconnected; e.g., for three copies of the Hilbert space, the permutation operator can leave one unchanged while swapping the other two. However, in this case, the permutation will contain only terms which are related to lower-order moments. The cumulant explicitly subtracts off such lower-order moments, and, thus, it is related only to *cyclic* permutations, i.e., permutations with only a single cycle [68].

Applying this to our observable of interest, we write the general formula for the k th cumulant of the time series of $p(z_A)$, κ_k , as

$$\kappa_k = \sum_{\sigma \in C_k} \sum_{z_B^{(1)}} \cdots \sum_{z_B^{(k)}} \left(\bigotimes_{i=1}^k \langle z_A, z_B^{(i)} | \right) \hat{\rho}_d^{\otimes k} \text{Perm}(\sigma) \times \left(\bigotimes_{i=1}^k | z_A, z_B^{(i)} \rangle \right), \quad (\text{C7})$$

where $\otimes$ represents a repeated Kronecker product and C_k is the subset of S_k containing only cyclic permutations, which importantly has $(k-1)!$ elements. This is a general statement for cumulants of a local bitstring probability, $p(z_A)$, in the large total system size limit; it is much like Eq. (6) but is subselected only to the cyclic permutations and with the inclusion of sums over different choices of z_B in the contraction.

To simplify this equation, it is helpful to first consider its lowest nontrivial form for $k=2$, for which the only contributing permutation is the swap operator $\hat{S}$, with explicit steps

$$\kappa_2 = \sum_{z_B, z'_B} (\langle z_A, z_B | \otimes \langle z_A, z'_B |) \hat{\rho}_d^{\otimes 2} \hat{S}(|z_A, z_B\rangle \otimes |z_A, z'_B\rangle) \quad (\text{C8})$$

$$= \sum_{z_B, z'_B} \langle z_A, z_B | \hat{\rho}_d | z_A, z'_B \rangle \langle z_A, z'_B | \hat{\rho}_d | z_A, z_B \rangle \quad (\text{C9})$$

$$= \sum_{z_B} \langle z_A, z_B | \hat{\rho}_d | z_A, z_B \rangle^2 + \sum_{z_B \neq z'_B} |\langle z_A, z_B | \hat{\rho}_d | z_A, z'_B \rangle|^2 \quad (\text{C10})$$

$$\approx \sum_{z_B} \langle z_A, z_B | \hat{\rho}_d | z_A, z_B \rangle^2. \quad (\text{C11})$$

In Eq. (C10), we have split the sum over different choices of z_B and z'_B into cases where $z_B = z'_B$ and cases where $z_B \neq z'_B$. In Eq. (C11), we have made the infinite-temperature approximation that $\hat{\rho}_d \approx \hat{I}/D$ —up to small

variations along the diagonal due to the varied $p_{\text{avg}}(z)$ —which nullifies the second term, where $z_B \neq z'_B$. In other words, we assume off-diagonal terms are negligible.

Carrying out a similar calculation—and again keeping only terms where all z_B are equal—for the k th cumulant yields

$$\kappa_k \approx (k-1)! \sum_{z_B} \langle z_A, z_B | \hat{\rho}_d | z_A, z_B \rangle^k, \quad (\text{C12})$$

where the coefficient $(k-1)!$ comes precisely from the number of elements in the set of cyclic permutations. To be explicit, by selecting cases where all z_B are equal (because of the infinite-temperature condition), the cyclicity of the permutations does not actually matter for our proof, beyond the fact that there are $(k-1)!$ of them.

Then, to further simplify Eq. (C12), we assume that any z_B dependence can be ignored, and we can approximate

$$\sum_{z_B} \langle z_A, z_B | \hat{\rho}_d | z_A, z_B \rangle^k \approx \langle z_A | \text{tr}_B(\hat{\rho}_d) | z_A \rangle^k / D_B^{k-1} \quad (\text{C13})$$

$$= p_{\text{avg}}(z_A)^k / D_B^{k-1}. \quad (\text{C14})$$

This is an assumption that the basis $\{z_B\}$ is *energy nonrevealing*; see Appendix D 4.

Thus, when considering cumulants of relative quantities, i.e., dividing the k th cumulant by $p_{\text{avg}}(z_A)^k$, we find that the k th cumulant is $(k-1)!/D_B^{k-1}$. This is precisely equal to the k th cumulant of the Erlang distribution. Since all cumulants match, all moments must match, too, and we conclude that the marginal probabilities $p(z_A)$ follow an Erlang distribution for systems thermalizing at infinite temperature, up to corrections exponentially small in the system size.

In this derivation, we have made two approximations. First, we have assumed that the basis z_B is energy non-revealing, and, second, we have assumed that the off-diagonal terms of the form $\langle z_A z_B | \hat{\rho}_d | z_A z'_B \rangle$ are negligible. Remarkably, we may perform an exact calculation of the k th moments $\mathbb{E}_t[p(z_A, t)^k]$ to estimate the effect of these approximations. We outline these more technical results below.

4. Universal variance of subsystem bitstring probabilities, including $1/D$ corrections

We study the variance of bitstring probabilities by using explicit expressions for the second moment of the temporal ensemble when Hilbert-space ergodicity applies [18,26]. Specifically, we study the relative temporal fluctuation $\sigma_{\text{rel}}^2(z_A)$, and we find an expression that has a natural interpretation as an effective dimension of the finite-temperature bath.

We define the relative temporal fluctuation $\sigma_{\text{rel}}^2(z_A)$ of the probability of a particular bitstring z_A in subsystem A as follows:

$$\sigma_{\text{rel}}^2(z_A) \equiv \frac{\mathbb{E}_t[p(z_A, t)^2] - \mathbb{E}_t[p(z_A, t)]^2}{\mathbb{E}_t[p(z_A, t)]^2} = \frac{\mathbb{E}_t[p(z_A, t)^2]}{\mathbb{E}_t[p(z_A, t)]^2} - 1. \quad (\text{C15})$$

Here, $p(z_A, t) = \text{tr}[(|z_A\rangle\langle z_A| \otimes \hat{I}_B)|\Psi(t)\rangle\langle\Psi(t)|]$ is the marginal probability of measuring z_A at time t , where $\hat{I}_B$ is the identity operator acting on subsystem B . The denominator of Eq. (C15) can be simplified to

$$\mathbb{E}_t[p(z_A, t)] = \text{tr}\left[|z_A\rangle\langle z_A| \otimes \hat{I}_B \left(\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt |\Psi(t)\rangle\langle\Psi(t)|\right)\right] = \langle z_A | \text{tr}_B(\hat{\rho}_d) | z_A \rangle \equiv \text{tr}_B[\hat{\rho}_d(z_A)], \quad (\text{C16})$$

where in the second-last equality $\text{tr}_B(\hat{\rho}_d)$ is a mixed state on subsystem A and in the last equality we instead consider the quantity $\hat{\rho}_d(z_A) \equiv (|z_A\rangle\langle z_A| \otimes \hat{I}_B) \hat{\rho}_d (|z_A\rangle\langle z_A| \otimes \hat{I}_B)$, which is a mixed state on the subsystem B .

Meanwhile, the numerator of Eq. (C15) can be evaluated from the second moment of the temporal ensemble [18,26]. This yields three terms: the permutation terms from Eq. (3), as well as a correction term illustrated in Fig. 6:

$$\mathbb{E}_t[p(z_A, t)^2] = \mathbb{E}_t[(\langle z_A | \text{tr}_B(|\Psi(t)\rangle\langle\Psi(t)|) | z_A \rangle)^2] \quad (\text{C17})$$

$$= \text{tr}[(|z_A\rangle\langle z_A| \otimes \hat{I}_B)^{\otimes 2} \hat{\rho}_d^{\otimes 2}] + \text{tr}[(|z_A\rangle\langle z_A| \otimes \hat{I}_B)^{\otimes 2} \hat{\rho}_d^{\otimes 2} \hat{S}] - \text{tr}[(|z_A\rangle\langle z_A| \otimes \hat{I}_B)^{\otimes 2} \hat{\rho}_d^{(2)}] \quad (\text{C18})$$

$$= \text{tr}_B[\hat{\rho}_d(z_A)]^2 + \text{tr}_B[\hat{\rho}_d(z_A)^2] - \sum_E q_{z_A}^2(E), \quad (\text{C19})$$

where $\hat{\rho}_d^{(2)} = \sum_E |E, E\rangle\langle E, E| \hat{\rho}_0^{\otimes 2} |E, E\rangle\langle E, E|$ is a two-copy generalization of the diagonal ensemble and in Eq. (C19) we have defined $q_{z_A}(E) \equiv \sum_{z_B} |\langle z_A z_B | E \rangle|^2 |\langle E | \Psi_0 \rangle|^2$ (here, we have dropped the subscript on energy eigenstates for notational simplicity). The inclusion of $\hat{\rho}_d^{(2)}$ prevents double counting of certain terms in the sum of permutations (see Fig. 6) and is explicitly included here, while previously [e.g., Eq. (3)] we disregarded it as negligible in the large system size limit.

Combining all of these simplified results, we obtain the relative fluctuation as follows:

$$\sigma_{\text{rel}}^2(z_A) = \frac{\text{tr}_B[\hat{\rho}_d(z_A)]^2 + \text{tr}_B[\hat{\rho}_d(z_A)^2] - \sum_E q_{z_A}^2(E)}{\text{tr}_B[\hat{\rho}_d(z_A)]^2} - 1 \quad (\text{C20})$$

$$\equiv \frac{1}{D_{B,\text{eff}}} - \frac{1}{D_\beta}, \quad (\text{C21})$$

where

$$D_{B,\text{eff}}^{-1} \equiv \frac{\text{tr}_B[\hat{\rho}_d(z_A)^2]}{\text{tr}_B[\hat{\rho}_d(z_A)]^2}, \quad (\text{C22})$$

$$D_\beta^{-1} \equiv \frac{\sum_E q_{z_A}^2(E)}{\text{tr}_B[\hat{\rho}_d(z_A)]^2}. \quad (\text{C23})$$

Here, $D_{B,\text{eff}}^{-1}$ is the purity of the conditional reduced density matrix $\hat{\rho}_d(z_A)/\text{tr}_B[\hat{\rho}_d(z_A)]$. It measures the effective size of the bath; hence, we dub it the *effective bath dimension*. Meanwhile, D_β^{-1} is the purity of the classical distribution

[26] (of size D) over the energy eigenstates, which has typical value $O(1/D)$. At infinite temperature, we can approximate $\hat{\rho}_d(z_A) \approx \hat{I}_B/D$ and $q_{z_A}(E) \approx D_B/D^2$, where D and D_B are the Hilbert-space dimension of the total system and subsystem B , respectively, leading to $D_\beta \rightarrow D$ and $D_{B,\text{eff}} \rightarrow D_B$.

Applying these infinite-temperature approximations, we immediately arrive at

$$\sigma_{\text{rel}}^2(z_A) = \frac{1}{D_B} - \frac{1}{D}. \quad (\text{C24})$$

Explicitly, in Fig. 3 in the main text, in the main plot for the Rydberg, mixed field Ising model, and random unitary circuits, the x axis is D_B , as those systems are evolved at infinite effective temperature, while for the Fermi-Hubbard and Bose-Hubbard systems the x axis is $D_{B,\text{eff}}$ to account for particle number conservation. Furthermore, when calculating the scaling coefficients for all models in the inset in Fig. 3, we use $D_{B,\text{eff}}$ instead of D_B —this is to be maximally fair to all models in case there are slight non-infinite-temperature conditions.

The prediction of Eq. (C21) is also matched by both numerical and experimental data at finite effective temperature. In Fig. 7, we plot the relative variance $\sigma_{\text{rel}}^2(z_A)$ versus either the bath dimension D_B or the effective bath dimension $D_{B,\text{eff}}$ [Eq. (C22)] for finite-temperature evolution with both Rydberg (experiment and simulation) and mixed field Ising model (simulation) systems. We observe the relative variance scales inversely with the effective bath dimension, as expected.

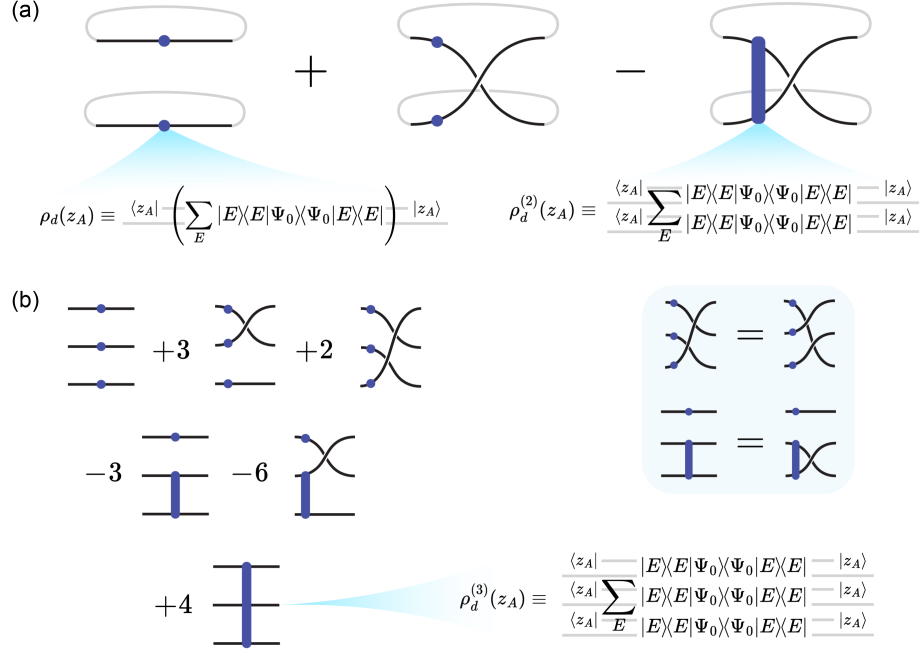


FIG. 6. Diagrams for subsystem bitstring probability fluctuations. We diagrammatically organize our computations for the second and third moments $\mathbb{E}_t[p(z_A, t)^k]$, which we show are equal to the low moments of the Erlang distribution at infinite temperature. (a) We organize the terms in the second moment of the temporal ensemble diagrammatically, which translates to the equation $\mathbb{E}_t[p(z_A, t)^2] = \text{tr}[\hat{\rho}_d(z_A)]^2 + \text{tr}[\hat{\rho}_d(z_A)^2] - \text{tr}[\hat{\rho}_d^{(2)}(z_A)]$ [Eq. (C19)]. The blue dot represents $\hat{\rho}_d(z_A)$, which is an operator on $\mathcal{H}_B$, and can be interpreted as the initial state, dephased in the energy eigenbasis, and projected onto a subsystem bitstring $|z_A\rangle$. The blue bar represents $\hat{\rho}_d^{(2)}(z_A)$, which acts on the double Hilbert space $\mathcal{H}_B^{\otimes 2}$, and is a subleading correction term to the first two terms. Each diagram is ultimately traced out (gray lines), which we subsequently omit for simplicity. (b) The diagrammatic framework greatly simplifies the computation of the third moment [Eq. (C25)]. The six permutations of S_3 give rise to three terms, since permutations with the same cycle type produce the same terms (blue box). Correction terms may be expressed in terms of diagrams with blue bars. Furthermore, diagrams where the lines with a blue bar are swapped are equivalent. Therefore, there are only two inequivalent diagrams involving $\hat{\rho}_d^{(2)}(z_A)$. The number of equivalent diagrams, along with double-counting coefficients from the twirling identity, gives rise to the coefficients -3 and -6 . Finally, there is a higher-order diagram that accounts for triple counting, involving $\hat{\rho}_d^{(3)}(z_A)$.

To be explicit, for each bitstring z_A we calculate the corresponding $D_{B,\text{eff}}$ and D_β numerically, which will be different generically for every z_A . For the experimental data, we select only blockade-satisfying bitstrings, as we do not have sufficiently many experimental measurements to resolve probabilities with high confidence outside this space. However, we note, in general, that rescaling the bath dimension naturally takes into account the Rydberg blockade, as it is essentially a finite-temperature effect. In other words, while we can operate in an infinite effective temperature regime within the blockade-satisfying subspace, within the global Hilbert space the system is at effectively low temperature; this temperature is naturally

accommodated via the bath dimension rescaling for each z_A .

5. Higher moments of subsystem bitstring probabilities

Equation (C21) shows that the relative temporal variance, i.e., the second moment of the PoP, approximately satisfies $\sigma_{\text{rel}}^2 = 1/D_B - 1/D$. This agrees with the second moment of the Erlang distribution, up to the $1/D$ term. Here, we explicitly compute the third moment, demonstrate that it is similarly consistent with the Erlang distribution, and discuss leading-order corrections.

Higher moments of the temporal ensemble may be evaluated [18,26]. We find

$$\begin{aligned} \mathbb{E}_t[p(z_A, t)^3] = & \text{tr}[\hat{\rho}_d(z_A)]^3 + 3\text{tr}[\hat{\rho}_d(z_A)]\text{tr}[\hat{\rho}_d(z_A)^2] + 2\text{tr}[\hat{\rho}_d(z_A)^3] - 3\text{tr}[\hat{\rho}_d(z_A)]\text{tr}[\hat{\rho}_d^{(2)}(z_A)] \\ & - 6\text{tr}[\hat{\rho}_d^{(2)}(z_A)]\{\hat{I} \otimes \hat{\rho}_d(z_A)\} + 4\text{tr}[\hat{\rho}_d^{(3)}(z_A)], \end{aligned} \quad (\text{C25})$$

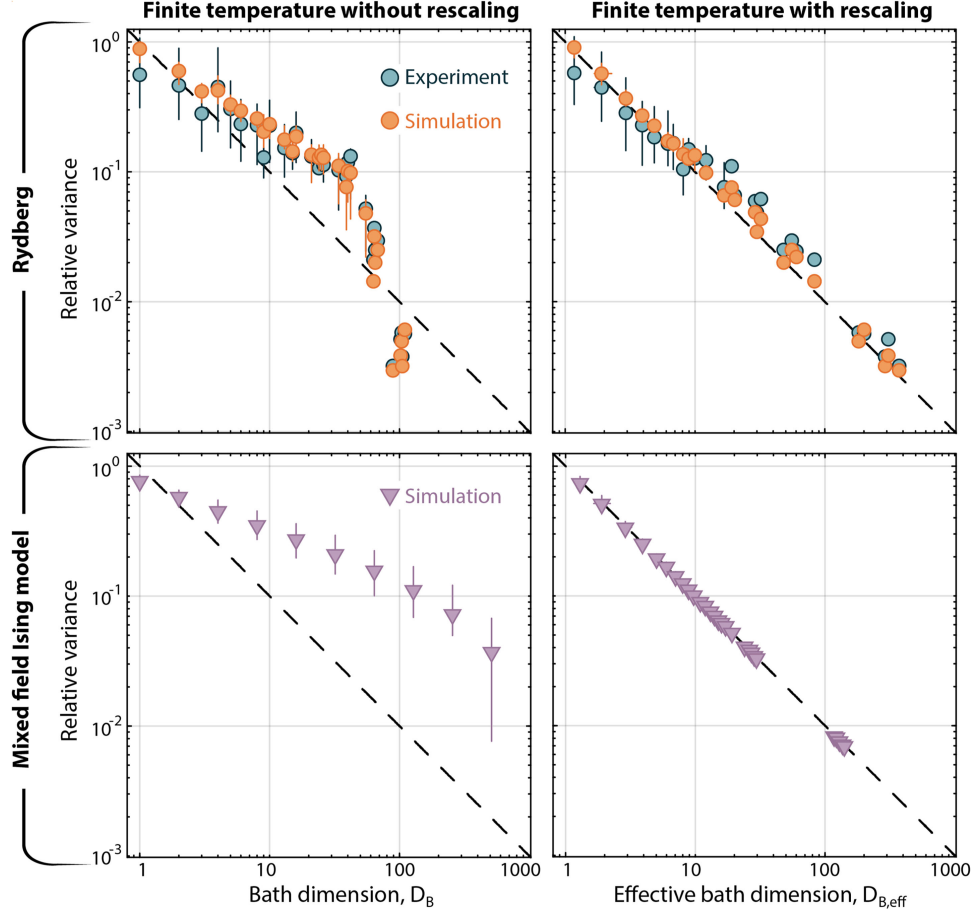


FIG. 7. Rescaling by effective bath dimension for finite-temperature systems. We show relative variance plots for Rydberg (both experiment and simulation) as well as for the mixed field Ising model (simulation). In each case, the quench evolution is at a finite effective temperature. In the left plots, we show the relative variance (without the addition of a $1/D$ term) versus the bath dimensional, showing nonuniversal behavior. In the right plots, we show the relative variance [plus the $1/D_\beta$ term, Eq. (C23)] versus the effective bath dimension $D_{B,\text{eff}}$ [Eq. (C22)], which recovers the scaling expected from Eq. (C21). In all plots, the markers are averaged over all bitstring realizations with an effective bath dimension in an interval from $[i - 0.5, i + 0.5]$, for $i \in \mathbb{Z}$, and error bars are the standard deviations over these realizations.

where $\hat{\rho}_d^{(k)}(z_A) \equiv (\langle z_A | \otimes \hat{I}_B)^{\otimes k} \sum_E |c_E|^{2k} |E\rangle \langle E|^{\otimes k} (|z_A\rangle \otimes \hat{I}_B)^{\otimes k}$ are higher-order correction terms. We diagrammatically illustrate this expression in Fig. 6(b). We emphasize that this expression is exact, and further approximations are often required to make headway.

The relevant quantity we wish to compare is the third cumulant κ_3 :

$$\kappa_3 \equiv \frac{\mathbb{E}_t[p(z_A, t)^3]}{\mathbb{E}_t[p(z_A, t)]^3} - 3 \frac{\mathbb{E}_t[p(z_A, t)^2]}{\mathbb{E}_t[p(z_A, t)]^2} + 2 = 2 \frac{\text{tr}[\hat{\rho}_d(z_A)^3]}{\text{tr}[\hat{\rho}_d(z_A)]^3} - 6 \frac{\text{tr}\{\hat{\rho}_d^{(2)}(z_A)[\hat{I} \otimes \hat{\rho}_d(z_A)]\}}{\text{tr}[\hat{\rho}_d(z_A)]^3} + 4 \frac{\text{tr}[\hat{\rho}_d^{(3)}(z_A)]}{\text{tr}[\hat{\rho}_d(z_A)]^3} \quad (\text{C26})$$

$$\approx \frac{2}{D_B^2} - \frac{6}{D_B D} + \frac{4}{D^2}, \quad (\text{C27})$$

where in the last step we have used the same approximations above at infinite temperature, that $\hat{\rho}_d(z_A) \approx \hat{I}_B/D$ and $q_{z_A}(E) \approx D_B/D^2$ (and note that $\text{tr}[\hat{\rho}_d^{(3)}(z_A)] = \sum_E q_{z_A}(E)^3$). As with the variance, this agrees with the third cumulant of

the Erlang distribution (which is simply $2/D_B^2$) up to corrections $O(1/D)$. It is also close, but not exactly equal to $2(1/D_B - 1/D)^2$, i.e., replacing $1/D_B$ with $1/D_B - 1/D$ as the parameter in the Erlang distribution.

We note that both Eqs. (C24) and (C27) agree with the (average) second and third cumulants of Haar random states [32,42], i.e., for marginal distributions $p_\psi(z_A) \equiv \langle z_A | \text{tr}_B(|\psi\rangle\langle\psi|) | z_A \rangle$ of Haar random states $|\psi\rangle$,

$$\mathbb{E}_{|\psi\rangle \sim \text{Haar}}[D_A^2 p_\psi^2(z_A)] - 1 = \frac{D + D_A}{D + 1} - 1 \approx \frac{1}{D_B} - \frac{1}{D}, \quad (\text{C28})$$

$$\mathbb{E}_{|\psi\rangle \sim \text{Haar}}[D_A^3 p_\psi^3(z_A) - 3D_A^2 p_\psi^2(z_A) + 2] = \frac{D^3 + 3D^2 D_A + 2DD_A^2}{D(D+1)(D+2)} - 3\frac{D + D_A}{D + 1} + 2 \approx \frac{2}{D_B^2} - \frac{6}{D_B D} + \frac{4}{D^2}. \quad (\text{C29})$$

We next note that the substitutions $\hat{\rho}_d(z_A) \approx \hat{I}_B/D$ and $q_{z_A}(E) \approx D_B/D^2$ leading up to Eq. (C27) are heuristic in nature and they may not appropriately capture the relevant higher-order fluctuations to the desired level of precision (here, up to $1/D^2$). Nevertheless, they appear to give the correct answer; we have verified that Eqs. (C24) and (C27) agree with a more careful calculation which treats the eigenbasis $\{|E\rangle\}$ as a random basis and averages over such bases with an integral over Haar-random unitaries.

6. Equivalence of PoP distributions over bitstrings or over time and time dependence of the PoP

So far, we have shown the global PoP forms the Porter-Thomas distribution when aggregated over many measurement times. However, under Hilbert-space ergodicity, Refs. [18,26] state that the Porter-Thomas distribution should also be apparent when forming the PoP over many choices of bitstrings at a fixed measurement time.

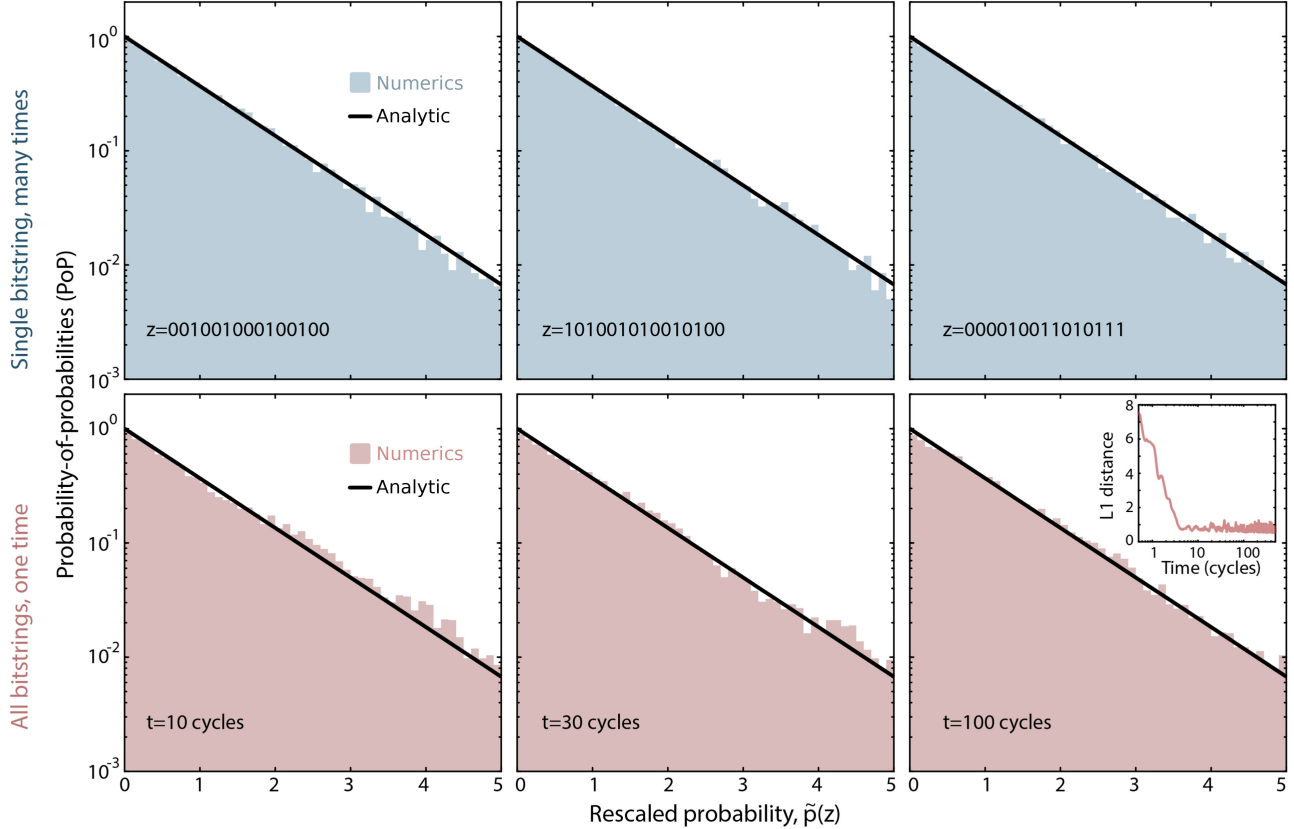


FIG. 8. Observation of Porter-Thomas statistics on the level of a single bitstring or time. PoP distributions obtained from numerical simulations for an $N = 15$ Rydberg atom array with the same parameters as used in the main text. Top: PoP distributions for the probability to observe a single bitstring, aggregated over evolution times ranging from 0 to 500 cycles and normalized by the average temporal probability. Bottom: PoP distributions aggregated over all 2^{15} bitstrings at a single fixed time, with probabilities normalized by their average temporal probability. In all cases, the observed PoP distributions agree with the exponential Porter-Thomas distribution, as is expected from theoretical arguments related to Hilbert-space ergodicity [18]. In the case of averaging over all bitstrings at a fixed time, this agreement is observed to set in past approximately 4 cycles, as quantified by the L1-norm between the numerical and analytic PoP as a function of time (inset). Note that in all plots we consider not just the lowest-energy nonblockaded sector, but the entire Hilbert space.

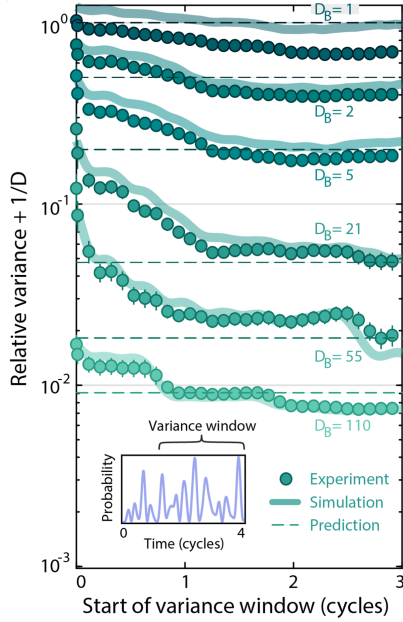


FIG. 9. Time-resolved saturation of the temporal variance. We plot $\sigma_{\text{rel}}^2 + 1/D$ for both experiment and numerical simulations evolved at infinite temperature, for various choices of bath dimension D_B . The variance is calculated in a temporal window of varying length, where the endpoint is fixed at four Rabi cycles of evolution, the latest time measured experimentally, while the start of the window is varied (inset). We see for both the numerics and the experiment that the variance settles to the prediction of $1/D_B$ for start times of approximately 1.5 cycles. We note that when simulating out to many hundreds of cycles—to emulate the infinite time limit—the convergence is observed without excluding any early time transient behavior (not shown), but we do not have the experimental resolution to evolve to such late times without incurring fidelity loss.

In Fig. 8, we present numerical evidence for this by simulating the Rydberg Hamiltonian with no decoherence out to several hundred cycles of evolution for $N = 15$ atoms. We then form the global PoP either by (a) aggregating normalized probabilities across many times for a single bitstring or (b) aggregating normalized probabilities over many bitstrings at a fixed time. Explicitly, in both cases, the normalization is done by dividing probabilities by their respective long-time averages. Note that in the main text for the Rydberg systems we consider only the nonblocked Hilbert space, as it is difficult to gather sufficient experimental statistics to reconstruct bitstring probabilities in the higher-energy sectors, but in Fig. 8 we consider all bitstrings, not just those drawn from the nonblocked subspace.

An important aspect visible in the inset in Fig. 8 is that the fixed-time PoP converges close to the Porter-Thomas distribution over a few cycles of evolution. In Fig. 2 in the main text—where we observe that the experimental global PoP is roughly consistent with the Porter-Thomas distribution—we aggregated over both bitstring index and measurement times

from 1.5 to 4 cycles of evolution to gather sufficient statistics. Evolving out to longer times experimentally modifies the PoP distribution due to interaction with the external environment, as discussed at length in the main text. To verify we study sufficiently late dynamics in the experiment, we consider σ_{rel}^2 calculated in a temporal window of varying length but fixed end point at four Rabi cycles of evolution (Fig. 3), the latest time measured experimentally. We find σ_{rel}^2 appears roughly converged for a start time of approximately 1.5 cycles, for which we observe both the experimental and numerical σ_{rel}^2 closely matches the $1/D_B - 1/D$ prediction (see Fig. 9). We note some disagreement for the experimental data at $D_B = 1$, which again we attribute to early signatures of decoherence.

7. Failure cases

As shown in the inset in Fig. 3 of the main text, not all models follow our analytic predictions, in particular, those that do not fulfill the no-resonance conditions such as the transverse field Ising model and systems of noninteracting spins. As a further example of this failure, we plot the PoP aggregated over global bitstring probabilities from many times for the transverse field Ising model (Fig. 10). Probabilities are normalized by their long time average and are measured out to several hundred cycles. We observe clear nonexponential signatures in the PoP.

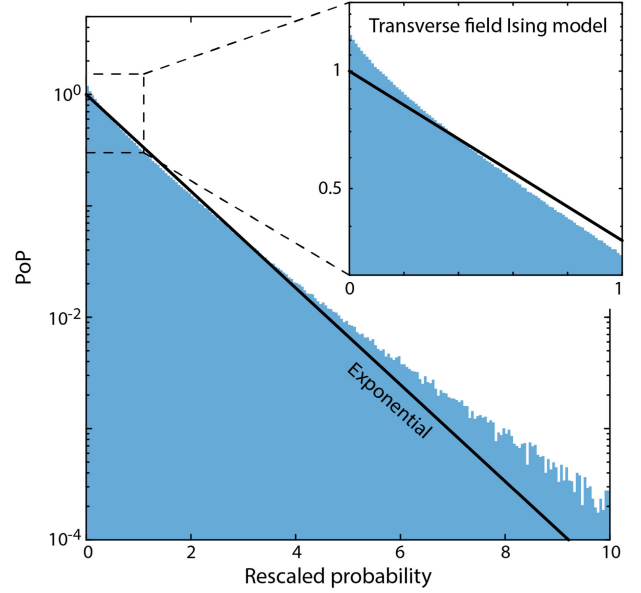


FIG. 10. Non-Porter-Thomas behavior of the transverse field Ising model. We evolve an $N = 12$ atom system under the transverse field Ising model out to several hundred cycles, as in the inset in Fig. 3, and aggregate all global bitstring probabilities over all bitstrings and at different times (to obtain better statistics), normalized by their respective long time averages. The resulting PoP shows nonexponential and nonuniversal signatures. Accounting for such effects may allow us to better describe models not exhibiting Hilbert-space ergodicity.

APPENDIX D: HYPOEXPONENTIAL DISTRIBUTIONS FOR PROBABILITY OF PROBABILITIES

In Appendix C, we formally derived the emergence of the Erlang distribution from results in Hilbert-space ergodicity for closed system dynamics. Using the intuition gained from that derivation, in this appendix, we build models of how the PoP should behave under a variety of conditions, including open system dynamics, explicitly comparing them to numerical simulations. We find that the PoP qualitatively and quantitatively is changed in the presence of different types of noise, which we use in the main text both to discriminate between noise channels as well as learn their relative strengths.

We first give an overview of how the exponential or Porter-Thomas distribution emerges from ergodic unitary dynamics. We then introduce hypoexponential distributions, which parametrize all the PoP distributions studied in this work, including the Erlang distribution, which describes closed system dynamics at infinite temperature. For closed system dynamics at finite temperature, we describe the hypoexponential weights and discuss the conditions on the bath size under which these weights are the eigenvalues of the system (or bath) reduced density matrix.

We then move on to the case where our system is coupled to an external environment, describing our models for local errors, global depolarizing noise, Gaussian coherent errors, and a mixture of local and Gaussian coherent errors.

In Appendix E, we describe the effects of finite sampling on empirically estimating the PoP. This naturally leads us to more sample-efficient methods of characterizing the PoP via its cumulants (Appendix F), which, by a chain of approximate equalities summarized in Eq. (G3), quantify the decoherence strength (Kraus operators) of the quantum channel.

In this section, all PDFs will be supported on the non-negative real line $x \geq 0$.

1. Exponential (Porter-Thomas) distribution from random probability distributions

Our basic building block will be the exponential distribution $\text{Exp}(\lambda)$, with PDF

$$P_{\text{Exp}}(x, \lambda) dx = \lambda \exp(-\lambda x) dx. \quad (\text{D1})$$

We refer to variables which follow the exponential distribution with $\lambda = 1$ (or Porter-Thomas distribution) as “standard exponential variables.”

The Porter-Thomas distribution is a universal feature of random quantum states [19–21]. To see this, note that the probability distributions $p(z) = |\langle z | \Psi \rangle|^2$ obtained from Haar-random states are random probability distributions $p(z)$. To be precise, such distributions are drawn from the *flat Dirichlet distribution*, which is a distribution over the space of probability distributions (the probability simplex).

Such random distributions can be generated by i.i.d. sampling each probability value $p(z) \sim \text{Exp}(1/D)$ and then normalizing all entries such that $\sum_z p(z) = 1$. This final step introduces weak correlations $\text{cov}[p(z), p(z')] = \delta_{z,z'}/2D - 1/2D^2$, which vanish in the limit of large Hilbert-space dimension D .

We are primarily interested in the marginal distributions $P[p]$ of individual probability values, namely, the distribution of the values of $p(z)$. This is known to be the *Beta distribution* $\text{Beta}(1, D-1)$ [25,69], which converges to $\text{Exp}(1/D)$ in the large system limit $D \rightarrow \infty$. Crucially, the marginal distribution $P[p(z)]$ can be observed from a single *typical* distribution $p(z)$, and one does not have to sample many Dirichlet random distributions. Since the entries $p(z)$ of a typical distribution $\{p(z)\}$ are very weakly correlated, the histogram of their values, or PoP, is also approximately exponentially distributed. In this context, the exponential distribution is referred to as the Porter-Thomas distribution.

In the above discussion, we have motivated the Porter-Thomas distribution through Haar-random states. However, the Porter-Thomas distribution is present in states which can be practically produced. Specifically, it is present in the states output from RUCs [20]. It is convenient to rescale the probability values as $Dp(z)$, which follow the standard exponential distribution (i.e., with mean 1).

Furthermore, as shown in Sec. II, the Porter-Thomas distribution appears in states obtained from unitary Hamiltonian evolution [26], i.e., $|\Psi(t)\rangle = \exp(-i\hat{H}t)|\Psi_0\rangle$ for an arbitrary initial state $|\Psi_0\rangle$. We prove [18] this in the limit of exponentially long times and observe that in practice the required time is small, growing only weakly with system size [26]. Specifically, we consider the probabilities $p(z, t) \equiv |\langle z | \Psi(t) \rangle|^2$ and their time-average values $p_{\text{avg}}(z) \equiv \mathbb{E}_t[p(z, t)]$. Hamiltonian dynamics differs from RUC dynamics by energy conservation, which introduces nontrivial $p_{\text{avg}}(z)$. The probabilities $p(z, t)$ fluctuate about their average value, but the rescaled probabilities $\tilde{p}(z) \equiv p(z)/p_{\text{avg}}(z)$ follow the Porter-Thomas distribution.

2. Hypoexponential distributions

As described in the main text, the hypoexponential distribution is the probability distribution of a weighted sum of independent standard exponential variables. For a given set of weights ω_i (which are positive and sum to 1) and associated i.i.d. Dirichlet probability distributions $\{p_i(z)\}$, the overall distribution is given by the weighted sum:

$$p(z) = \sum_i \omega_i p_i(z). \quad (\text{D2})$$

When the number of weights is small (i.e., for a small bath dimension), the weights can be related to the eigenvalues of the density matrix $\hat{\rho}$. Their properties may be

inferred from each other (Appendix F 1). For example, the state purity is related to the second moment of the weights $\text{tr}(\hat{\rho}^2) = \sum_i \omega_i^2$, which we utilize heavily in Appendix F.

The corresponding PoP distribution is given by $\text{Hypo}(\{\omega_i\})$ (note that the hypoexponential distribution is conventionally parametrized in terms of the inverse weights ω_i^{-1} , which are known as *rates*). When the ω_i are distinct, the hypoexponential distribution has PDF [45]

$$P_{\text{Hypo}}(x) = \otimes_i P_{\text{Exp}}(x/\omega_i) \\ = \sum_i \omega_i^{-1} \exp(-x/\omega_i) \prod_{j \neq i} \frac{\omega_i}{\omega_i - \omega_j}, \quad (\text{D3})$$

where $\otimes$ denotes convolution, an operation of functions defined by $(f \otimes g)(x) = \int dy f(y)g(x-y)$. Equation (D3) does not hold when there are degenerate ω_i , which is often the case for our purposes. Hence, we rarely directly use Eq. (D3) (see Ref. [45] for an expression valid for degenerate ω_i).

Different types of noise give different weights ω_i . Below, we state types of noise we consider, their weights ω_i , and their resultant PoP distributions.

3. Intrinsic bath at infinite temperature

We first consider the PoP distribution in the context of closed system quantum thermalization: A subsystem effectively thermalizes due to its entanglement with the rest of the system, which serves as a bath. In this section, we let the bath be at infinite effective temperature. The quantities of interest are the subsystem bitstring probabilities $p(z_A)$, which are marginal probabilities $p(z_A) = \sum_{z_B} p(z_A \cdot z_B)$. As discussed in the main text, we find that the PoP of the marginal probabilities are described by Erlang distributions. In Appendix C, we gave a rigorous derivation of this, including small deviations from the perfect Erlang distribution. Here, we give an intuitive, heuristic derivation.

Consider an infinite-temperature state in the absence of symmetries for simplicity. We take the average probabilities $p_{\text{avg}}(z)$ to be trivially $1/D$. Therefore, the joint probabilities $p(z_A \cdot z_B)$ are essentially i.i.d. exponential variables, and the (rescaled) sum $D_A p(z_A)$ is the average of D_B i.i.d. standard exponential variables, i.e., $\omega_i = 1/D_B$ for $i = \{1, 2, \dots, D_B\}$. This special case of the hypoexponential distribution is precisely the Erlang distribution, parametrized by a positive integer k (here, $k = D_B$). Because the weights are degenerate, its PDF cannot be directly obtained from Eq. (D3) but can be independently derived as

$$P_{\text{Erlang}}(x; k) = \otimes_{i=1}^k P_{\text{Exp}}(kx) = \frac{k^k x^{k-1}}{(k-1)!} \exp(-kx), \quad (\text{D4})$$

where in our case of interest the variable $x = D_A p(z_A)$ and $k = D_B$.

4. Intrinsic bath at finite temperature

Having established the PoP distribution for an intrinsic bath at infinite effective temperature, a natural next step is to investigate the PoP distribution at finite effective temperatures. Specifically, we consider the following setting: A system is partitioned into a larger subsystem A and smaller one B , which we treat as the “bath.”

We first restrict ourselves to a special type of measurement basis z_A , whose conditions are explained later. Here, we claim that, when the bath B is small, the PoP of the bitstrings on A is the hypoexponential distribution with coefficients λ_i^{-1} , where $\{\lambda_i\}_{i=1}^{D_B}$ are the eigenvalues of the reduced density matrix $\hat{\rho}_A$ (or $\hat{\rho}_B$). That is,

$$\tilde{p}(z_A) \equiv p(z_A)/p_{\text{avg}}(z_A) \sim \text{Hypo}(\{\lambda_i\}), \quad (\text{D5})$$

where $p(z_A) = \sum_{z_B} p(z_A \cdot z_B)$ and $p_{\text{avg}}(z_A) = \sum_{z_B} p_{\text{avg}}(z_A \cdot z_B)$ are the marginal distributions of z_A .

This result may be understood through the lens of *projected ensembles* [10,70]. These are the ensembles of states defined on a subsystem Hilbert space (here, $\mathcal{H}_B$) by measuring a global state $|\Psi\rangle$ in its complement (A). For concreteness, we let $|\Psi\rangle = |\Psi(t)\rangle$ be generated by many-body dynamics for time t . Each of the exponentially many measurement outcomes z_A projects $|\Psi\rangle$ into a distinct pure state $|\Psi(z_A)\rangle \in \mathcal{H}_B$. The statistical properties of projected ensembles translate to properties of the PoP.

In companion work [18], we study projected ensembles at finite temperature, relevant to our setting here. When two conditions are satisfied, $\hat{\rho}_d(z_A)$ is approximately the reduced density matrix $\hat{\rho}_B(t) \equiv \text{tr}_A[|\Psi(t)\rangle\langle\Psi(t)|]$, which gives the stated claim Eq. (D5). These two conditions are (i) that the measurement basis should be “energy non-revealing” and (ii) the bath size must be sufficiently small.

Effect of the measurement basis—The result Eq. (D5) applies to specific measurement bases $|z_A\rangle$ that are energy nonrevealing, i.e., do not have significant correlations with the Hamiltonian that generated the state $|\Psi(t)\rangle$ (and which defines the notion of temperature). One example of an energy nonrevealing basis is the Z_A basis for the mixed field Ising model with fields and couplings in the X and Y directions, where the measurement outcomes z_A are uncorrelated with the energy density. For other bases, such as the X_A basis for the mixed field Ising model discussed above, such correlations may exist. Specifically, due to energy conservation of the overall state, measurement outcomes with higher values of energy (on subsystem A) on average lead to projected states with lower energy (on subsystem B), and vice versa. Therefore, the projected ensemble $\{|\Psi(z_A)\rangle\}$ may not be taken to be samples from *the same distribution*, independent of z_A . Among other consequences [18], the simple prediction Eq. (D5) does not hold in this setting.

Effect of bath size—When the measurement basis is energy nonrevealing, we may conclude that the *time-averaged* reduced density matrix $\mathbb{E}_t[\hat{\rho}_B(t)]$ is proportional

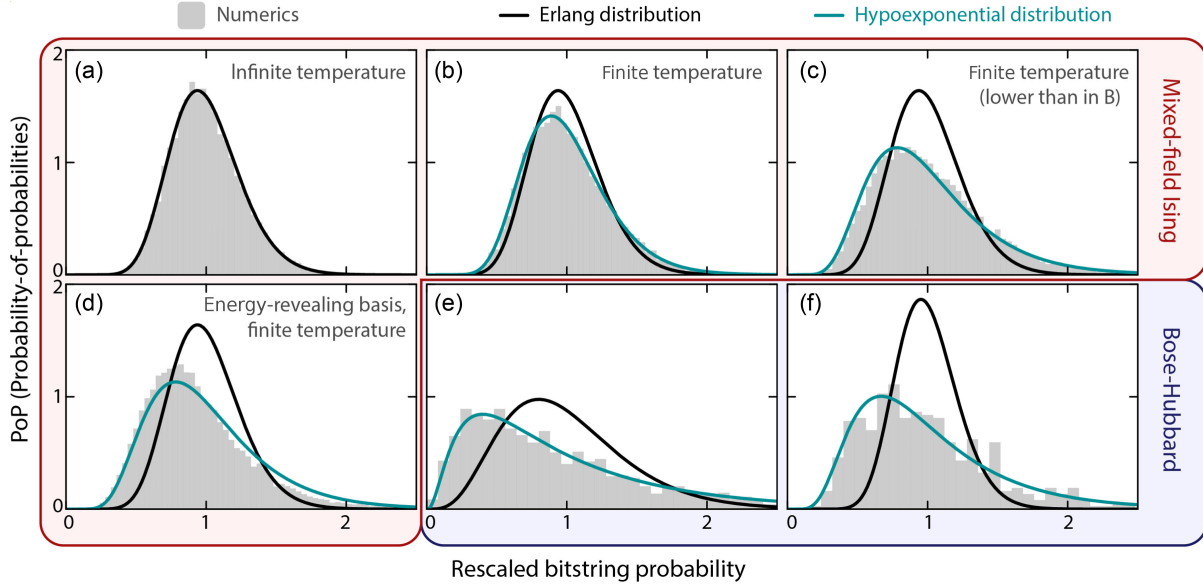


FIG. 11. Hypoexponential distributions for finite-temperature closed system dynamics. In (a)–(d), we plot the PoP of the conditional probabilities of $\tilde{p}(z_A) = p(z_A)/p_{\text{avg}}(z_A)$ formed by states time evolved by the MFIM. Here, the system is a one-dimensional $N = 18$ open chain, and the subsystem B is four sites at the center of the chain. (a)–(c) The temperature is tuned by rotating the initial state $|\Psi_0\rangle = [\exp(i\theta\hat{S}_x)|0\rangle]^{\otimes N}$ at various angles $\theta = 0, -0.6, -1.2$. In (a)–(c), B is measured in the $\{z_B\}$ basis. Only at infinite temperature does the PoP agree with the Erlang distribution (green), while at finite temperatures, the PoP agrees with the hypoexponential distribution [Eq. (D5)] instead. In (d), B is measured in the energy-revealing $\{x_B\}$ basis, showing small deviations from the hypoexponential distribution. In (e) and (f), we form the PoP from the state $|1, 1, \dots, 1\rangle$ evolved by the Bose-Hubbard model [Eq. (B3)] with $M = 10$ sites and $N = 10$ particles, measured in the particle number basis, and with a subsystem of three sites. Because of the number conservation of the evolution, we restrict to the measurement outcomes where z_B has (e) four and (f) five particles and form the symmetry-resolved density matrix accordingly. The PoPs agree better with the hypoexponential prediction instead of the Erlang distribution with the appropriate symmetry-resolved subsystem dimension.

to any of the $\hat{\rho}_d(z_A)$. However, $\hat{\rho}_B(t)$ has fluctuations over time, and, therefore, the eigenvalues of $\mathbb{E}_t[\hat{\rho}_B(t)]$ (the weights of the hypoexponential PoP) may not agree with the eigenvalues of $\hat{\rho}_B(t)$. To see this, it is instructive to consider the case of infinite temperature, where we model the global state as a Haar-random state. Indeed, for Haar-random states, the eigenvalues of the reduced state follow the so-called Marchenko-Pastur distribution [71,72], which is supported on the interval $[D_B^{-1}(1 - \sqrt{D_B/D_A})^2, D_B^{-1}(1 + \sqrt{D_B/D_A})^2]$ and is controlled by the parameter $D_B/D_A = 2^{2|B|-N}$. Meanwhile, the time-averaged reduced state is equal to $\hat{I}_B/D_B$, with all eigenvalues equal to $1/D_B$. When $|B| = N/2$, the second moment of the Marchenko-Pastur distribution is $\sum_j \lambda_j^2 \approx 2/D_B$, while the flat spectral distribution has second moment $1/D_B$, a clear disagreement. This feature is not particular to our results and instead reflects the conditions under which the time-dependent reduced density matrix may be approximated by a time-independent thermal state. Thus, the size of the bath should be sufficiently small; specifically, $|B|$ should be a constant amount less from half the system size $N/2$.

In Fig. 11, we confirm some of our predictions. In Figs. 11(a)–11(c), we plot the PoP of a state time evolved under the MFIM [Eq. (B1)], with a total system size

of $N = 18$ qubits and a bath size of $N_B = 4$ qubits. We tune the effective temperature by rotating the initial state $|\Psi_0\rangle = [\exp(i\theta\hat{S}_x)|0\rangle]^{\otimes N}$. At infinite temperature [Fig. 11(a), $\theta = 0$], the PoP is well described by the Erlang distribution. Tuning through finite temperature [Figs. 11(b) and 11(c), $\theta = -0.6, -1.2$], the PoP is instead well described by the hypoexponential distribution [Eq. (D5)]. Measuring in the energy-revealing $\{x_B\}$ basis [Fig. 11(d)], the PoP shows small deviations from Eq. (D5). To illustrate the ubiquity of our prediction, we illustrate this with a finite-temperature state evolved under the Bose-Hubbard model. In the presence of number conservation $\langle \hat{N} \rangle = N_{\text{tot}}$, we must form the projected ensemble by restricting to measurement outcomes with fixed number $\hat{N}_A$ and forming the according symmetry resolved density matrix $\hat{\rho}_B(\hat{N}_A) \propto \sum_{\{z_A | n(z_A) = \hat{N}_A\}} (\langle z_A | \otimes \hat{I}_B) |\Psi(t)\rangle \times \langle \Psi(t) | (|z_A\rangle \otimes \hat{I}_B)$, as detailed in Refs. [18,70].

5. Extrinsic bath: Local errors

We now consider the appearance of the hypoexponential under coupling to an extrinsic bath, i.e., with weights controlled by a given noise model.

First, we consider the case of local errors. As illustrated in Fig. 4 in the main text, our model for the PoP obtained

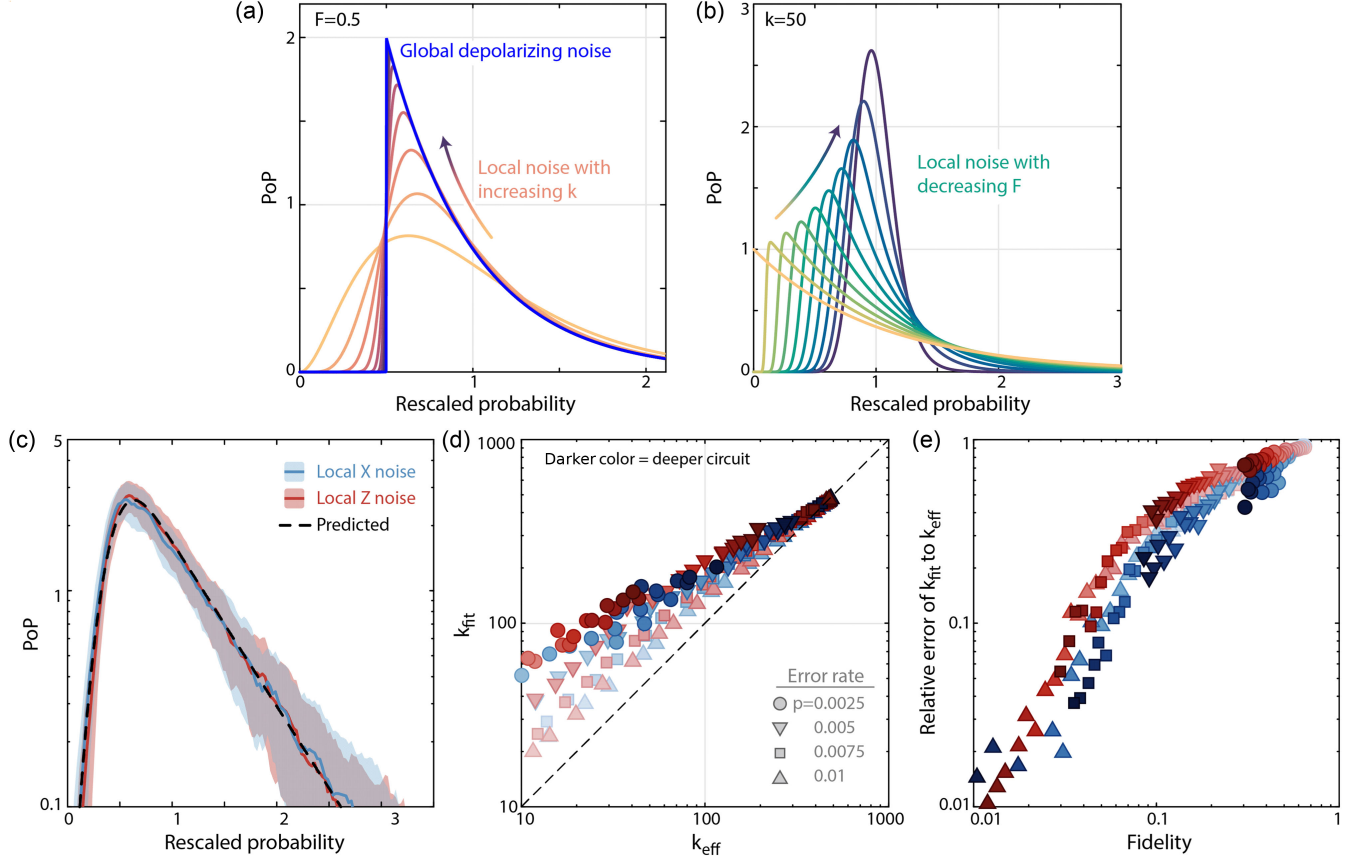


FIG. 12. PoP for local noise. (a) PoP distributions for local noise [see Eq. (D6)] with $F = 0.5$ and k varying from 2 to 10 000 logarithmically. As k increases, the PoP distribution converges to the prediction for global depolarizing noise [see Eq. (D8)] at the same fidelity, consistent with Ref. [47]. (b) PoP distribution for local noise with $k = 50$ and F varying from 1 to 0.1. (c) PoP distribution for either local Pauli-X errors or Pauli-Z errors, both of which are roughly equally well described by the same effective k value. (d) Fitted value of k versus value predicted from Eq. (D7) for noisy simulation with either Pauli-X or Pauli-Z errors evaluated for varying circuit depths. (e) The fitted value agrees best with the predicted value in the low-fidelity regime, as quantified by the relative error between k_{eff} and k_{fit} . In (c)–(e), results are aggregated over ten random circuit instances, blue points represent a local Pauli-X error model, and red points represent a local Pauli-Z error model.

from local errors with k locations is that it is the convolution of one exponential variable (speckle pattern) with large weight F (the “ideal trajectory”) and k exponential variables with small weights $(1 - F)/k$. In this case, the PDF is given by

$$P_{\text{Local Err}}(x; F, k) = P_{\text{Exp}}(x/F) \otimes P_{\text{Erlang}}[x/(1 - F); k] \\ = \left(\frac{k}{1 - F^{-1} + k} \right)^k \gamma \left(k, \frac{F - 1 + kF}{F(1 - F)} x \right) \\ \times \frac{e^{-x/F}}{F(k - 1)!}, \quad (\text{D6})$$

where $\gamma(k, x) \equiv \int_0^x z^{k-1} e^{-z} dz$ is the lower incomplete Gamma function. In Figs. 12(a) and 12(b), we plot these PDFs for various values of k and F .

This model holds for any type of local error. In Fig. 4(b) in the main text, we illustrate the PoP with amplitude damping errors, i.e., spontaneous decay of the $|1\rangle$ state to

$|0\rangle$, while in Fig. 12 we show it for local Pauli noise. In practice, however, the various assumptions made in the main text for the local noise channel may not hold, for instance, because errors are not necessarily completely orthogonal and because multiple errors can occur in a single shot of the experiment. These largely result in a value of k which is not simply the spacetime volume of the circuit. To address this, in practice we use an effective predicted k value defined as

$$k_{\text{eff}} = \frac{(1 - F)^2}{\kappa_2 - F^2}, \quad (\text{D7})$$

where κ_2 is the second cumulant of the experimentally measured PoP distribution (as discussed in greater detail in Appendix F—see Table I). Using Eq. (D7), we can predict the k value for a given measured PoP distribution at a given fidelity. In Fig. 12(c), we show the good agreement between Eq. (D6) and numerical simulations using an

TABLE I. Low-order cumulants of PoPs for various selected noise channels.

Noise channel	κ_2	κ_3
Global depolarization	F^2	$2F^3$
Local incoherent	$F^2 + (1 - F)^2/k$	$2(F^3 + (1 - F)^3/k^2)$
Global Gaussian coherent	$F/\sqrt{2}$	$2F^2/\sqrt{3}$

effective k found from Eq. (D7). Furthermore, in Fig. 12(d), we compare k_{eff} from Eq. (D7) against k value of best fit for various error rates and evolution depths and generally find that the two are in good agreement, particularly in the low-fidelity regime [Fig. 12(e)].

We stress that this model is limited. As an example, it stipulates that the weights of the nonideal trajectories be equal. This is not true if more than one error can occur during the dynamics: The probability of the trajectory decreases exponentially with the number of errors, while the number of such distinct trajectories increases exponentially. Furthermore, even the weights of single-error events may not be uniform. For example, the probability of amplitude damping errors depends on the magnetization of the state at each layer, which may change during the dynamics. These effects may lead to small deviations between our model and PoPs generated by realistic many-body dynamics, which may affect high-precision tasks such as determining the fraction of coherent-to-local error in Figs. 5(b) and 5(c) in the main text. However, such limitations may be solved with straightforward model-level refinements (such as explicitly adjusting weights to account for multiple error occurrences).

6. Extrinsic bath: Global depolarizing noise

Global depolarizing noise is a commonly used toy model for the effects of noise on quantum states. This channel maps states $\hat{\rho} \mapsto F\hat{\rho} + (1 - F)\hat{I}/D$, where $\hat{I}$ is the identity and F is a parameter which is approximately the fidelity of the channel. Under this channel, the bitstring distribution $p(z) \mapsto Fp(z) + (1 - F)/D$, and, hence, the PoP distribution of $x = Dp(z)$ is shifted to the right and rescaled by $1/F$, with PDF

$$P_{\text{GD}}(x; F) = \begin{cases} 0 & \text{for } 0 \leq x < 1 - F, \\ F^{-1} \exp\left(-\frac{x - (1 - F)}{F}\right) & \text{for } x \geq 1 - F. \end{cases} \quad (\text{D8})$$

As seen in Fig. 12, $P_{\text{Local Err}}(x; F, k) \rightarrow P_{\text{GD}}(x; F)$ as $k \rightarrow \infty$, consistent with Ref. [47], which states that the output distribution of random circuit sampling under local noise converges to that of “global white noise” when scaling the circuit spacetime volume.

7. Extrinsic bath: Gaussian coherent errors

Next, we discuss the case of Gaussian coherent errors, which describes shot-to-shot variation of a global parameter.

For a circuit model, we model this error by systematically under- or overrotating every two-qubit gate by a multiplicative, Gaussian-distributed factor. For Hamiltonian evolution, a global Hamiltonian parameter, such as the Rabi frequency Ω , is taken to be a Gaussian random variable. Such a noise channel is linked to shot-to-shot parameter fluctuations observed in Hamiltonian evolution [24] and certain models of control errors. As found in Ref. [24], evolution for a time t under these parameter values produces a continuous family of pure states $|\Psi(t, \Omega)\rangle$ which mutually overlap as $|\langle\Psi(t, \Omega)|\Psi(t, \Omega')\rangle|^2 \propto \exp[-N(\Omega - \Omega')^2 t^2 / 2\sigma_\Omega^2]$. When Ω is normally distributed, i.e., with $P(\Omega) = \exp[-(\Omega - \Omega_0)^2 / (2\sigma'^2)] / \sqrt{2\pi\sigma'^2}$ for some σ' , we obtain a normally distributed mixture $\hat{\rho}(t) = \int d\Omega P(\Omega) |\Psi(t, \Omega)\rangle \langle\Psi(t, \Omega)|$.

Our model for the PoP converts this continuous mixture of states into a discrete sum of probability distributions, with discrete weights ω_i . It is natural to let the weights follow a Gaussian profile when σ_F is sufficiently large (see below for deviations)

$$\omega_i^{\text{coh}} = \frac{1}{\mathcal{Z}(\sigma_F)} \exp\left[-\frac{i^2}{2\sigma_F^2}\right], \quad (\text{D9})$$

where i ranges from $-\infty$ to ∞ . These weights are parametrized by the width σ_F , and $\mathcal{Z}(\sigma_F) \equiv \sum_{i=-\infty}^{\infty} \exp[-i^2 / (2\sigma_F^2)]$ is a normalization constant. This sum is equal to the Jacobi theta function [73] $\theta_3\{\exp[-1/(2\sigma_F^2)]\}$, which is well approximated by $\sqrt{2\pi}\sigma_F$ when $\sigma_F \gg 1$ but approaches 1 as $\sigma_F \rightarrow 0$. Incidentally, since $\omega_0 = 1/\mathcal{Z}(\sigma_F)$ has the largest height, it corresponds to the fidelity F , i.e., the maximum overlap $\langle\Psi|\hat{\rho}|\Psi\rangle$ for any pure state $|\Psi\rangle$. Therefore, we may parametrize the PoP in terms of F instead, by solving for σ_F such that $\theta_3\{\exp[-1/(2\sigma_F^2)]\} = F^{-1}$.

A large fidelity corresponds to a small σ , and our model of discrete Gaussian weights shows quantitative differences in this regime. As an illustration, the second moment of the PoP can be computed analytically and differs from the Gaussian hypoexponential model at small σ [Appendix F and Fig. 16(c)]. This may be addressed by a non-Gaussian profile of weights in the narrow limit.

As a technical point, the weights $\omega_{-i}^{\text{coh}} = \omega_i^{\text{coh}}$ are degenerate, and, therefore, we must modify Eq. (D3) to

$$P_{\text{Gauss Coh}}(x; \{\omega_i\}) = P_{\text{Exp}}(x/\omega_0) \otimes \left(\bigotimes_{i=1}^{\infty} P_{\text{Erlang}}(x/\omega_i; 2) \right) \quad (\text{D10})$$

$$= \omega_0^{-1} e^{-x/\omega_0} \prod_{j>0} \frac{\omega_0^2}{(\omega_0 - \omega_j)^2} + \sum_{i=1}^{\infty} \left[\omega_i^{-2} \left(x - \sum_{\substack{j \in \mathbb{Z} \\ j \neq i}} \frac{\omega_i \omega_j}{\omega_i - \omega_j} \right) e^{-x/\omega_i} \frac{\omega_i}{\omega_i - \omega_0} \prod_{\substack{j \in \mathbb{Z} \\ j \neq i}} \frac{\omega_0}{\omega_0 - \omega_j} \right]. \quad (\text{D11})$$

8. Extrinsic bath: Mixture of local and global coherent errors

Finally, we turn to modeling the effects of both local and global coherent errors simultaneously. We take the errors to be independent of each other, and the weights are simply products of the local and global weights $\omega_{ij} = \omega_i^{\text{loc}} \times \omega_j^{\text{coh}}$. We further notice that, in numerical simulations of such a mixture of errors, the fidelity factorizes into the fidelities arising from incoherent and coherent errors [Fig. 16(a)], i.e., $F = F_{\text{coh}} \times F_{\text{loc}}$, i.e., $\log(F_{\text{coh}}) + \log(F_{\text{loc}}) = \log(F)$. This allows us to define the “fraction” of coherent errors as $f = \log(F_{\text{coh}})/\log(F)$. We note that such a factorization is also approximately apparent for our *ab initio* Rydberg error model [24].

For completeness, we outline our algorithm used in Figs. 5(b) and 5(c) to generate the PoP arising from a mixture of local and global errors based on our model, with a finite number of samples. The effects of a finite number of samples are detailed in a later section. We use this algorithm to generate the local and global coherent finite-sample PoPs in

Fig. 5(a) as well as to provide a reference distribution to calculate the χ^2 distances in Figs. 5(b) and 5(c). Note that, in the case of RUCs, the average probability is $p_{\text{avg}}(z) = 1/D$, and the rescaled empirical probabilities will take finite values jD/M only for positive integer j .

9. Other noise sources

We briefly remark that our analysis here is by no means complete. While we have analyzed commonly considered noise sources, other sources such as qubit leakage and readout error will yield different PoP distributions. For example, readout error may be treated in a similar fashion to local error. However, here there are only N possible error locations, and as such the parameter k may not be taken to be large. This is particularly notable for our moment analysis in Eq. (F9). We leave a detailed study of this, and other noise models, to future work. However, we expect that, in general, accounting for novel noise models will simply amount to judicious corresponding choice of the hypoexponential weight vector.

Algorithm 1. Numerical algorithm to generate model PoP under both local and global coherent noise.

Data: Fidelity F , fraction f , number of samples M , Hilbert-space dimension D , number of trajectories k , averaged probabilities $p_{\text{avg}}(z)$, and bin width $\Delta\tilde{p}$

Result: Empirical PoP $\vec{n}$

$F_{\text{loc}} \leftarrow F^f$;
 $F_{\text{coh}} \leftarrow F^{1-f}$;
 $\sigma_F \leftarrow \text{Solution of } \theta_3\{\exp[-1/(2\sigma_F^2)]\} = F_{\text{coh}}^{-1}$;
 $n(\sigma_F) \leftarrow F_{\text{coh}}$;
 $p(z) \leftarrow 0$;
for j in $\{-\lceil 5\sigma_F \rceil, -\lceil 5\sigma_F \rceil + 1, \dots, \lceil 5\sigma_F \rceil\}$; /* Infinite sum truncated at $\pm\lceil 5\sigma_F \rceil$ */
 do
 $p_{\text{Rand}}(z) \leftarrow \text{Dirichlet random distribution}$;
 $p(z) \leftarrow p(z) + F_{\text{inc}} n(\sigma_F) \exp[-j^2/(2\sigma_F^2)] p_{\text{Rand}}(z)$
 for n in $\{1, 2, \dots, k\}$ **do**
 $p_{\text{Rand}}(z) \leftarrow \text{Dirichlet random distribution}$;
 $p(z) \leftarrow p(z) + \frac{(1-F_{\text{inc}})}{k} n(\sigma_F) \exp[-j^2/(2\sigma_F^2)] p_{\text{Rand}}(z)$;
 end
 end
 $p(z) \leftarrow p(z) \times p_{\text{avg}}(z) / \|p(z) \times p_{\text{avg}}(z)\|_1$; /* Multiply mixture of Dirichlet random distributions by p_{avg} and normalize */
 $n_{\text{trial}} \leftarrow 100$;
 $\vec{n} \leftarrow \vec{0}$; /* Initialize mean for finite sample PoP */
for j in $\{1, 2, \dots, n_{\text{trial}}\}$ **do**
 $\{z_1, \dots, z_M\} \leftarrow M$ samples from $p(z)$;
 $N_z \leftarrow \#(z_i = z)$;
 $\vec{n} \leftarrow \vec{n} + \text{Hist}(\{\frac{N_z}{M p_{\text{avg}}(z)}\}; \Delta\tilde{p}) / n_{\text{trial}}$ /* Form histogram of rescaled empirical frequencies $N_z/[M p_{\text{avg}}(z)]$, with bin width $\Delta\tilde{p}$, add to mean vector $\vec{n}$. Each entry is the expected number of bitstrings with empirical frequency in bin. */
end

APPENDIX E: EFFECTS OF FINITE SAMPLING ON THE PROBABILITY OF PROBABILITIES

It is important to understand the effect of a finite number of experimental samples on the probability of probabilities discussed in this work. Finite sampling must be carefully treated, since its effects on the PoP can be drastic: The simplest estimation of the PoP introduces not only statistical noise, but also systematic bias. Here, we first discuss how experimental and model PoPs are compared in the main text and then describe how the model PoP is affected by finite sampling by effectively smearing out the distribution. Finally, we suggest a method to potentially invert this finite sampling effect. In the next section, we show that low-order moments of the PoPs can be estimated in an unbiased way which is robust to a finite number of samples.

1. Comparing experiment and finite-sampled model PoP

In the main text, we compare the experimental empirical PoP against PoPs obtained from analytical models, accounting for the finite sampling of the experiment. Specifically, we accrue roughly 32000 blockade-satisfying [23] bitstrings from the experimental evolution, which is $m \approx 20.6 \times$ the Hilbert-space dimension of $D = 1597$, and compare against analytical PoPs which are modified to account for this degree of finite sampling (described below). For our *ab initio* error model, we perform the full quantum simulation, then sample an equivalent number of bitstrings from the resultant state, and then repeat this sampling many times to find the mean finite-sampling prediction.

For the analytical noise models, no direct quantum simulation is necessary. We first take the analytical model PoP and sample a given normalized probability vector from it. We then sample this probability vector as many times as the experiment was sampled and reconstruct an empirical model PoP. We repeat this procedure many times and take the average over iterations to define the numerically sampled model PoP. We emphasize that, in doing so, no actual quantum simulation was needed, simply repeated resampling. The results of this resampling are in excellent agreement with analytical predictions where available (discussed below), as in Fig. 14.

For our comparison between the model PoPs and RUC numerics in Fig. 5(c) in the main text, we again employ the standardized χ^2 distance (CSD) as our distance and compare our noisy RUC simulations against model PoPs. For each point in Fig. 5, we average over 100 RUC circuit realizations. We show additional results in Fig. 13, where we see the estimation correlation improves with greater circuit depth at an equal fidelity, which we attribute to (a) limited scrambling in shallower circuits and (b) not accounting for higher-order terms in the analytic noise channels; we note that the second of these effects is

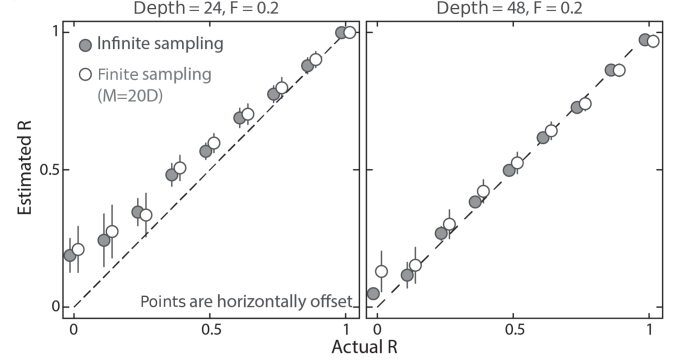


FIG. 13. Noise learning improves with greater depth. We perform RUC numerics with varying fraction of global coherent error to local incoherent error, as in Fig. 5(c) in the main text, with parameters chosen such that the final fidelity is 0.2 after either depth 24 (left) or depth 48. For shorter depth circuits, the estimator correlation becomes worse, which we believe is due to (a) limited scrambling and (b) not accounting for higher-order terms in the analytic noise channels (i.e., nondilute effects of the local error channel). We also show the results for either infinite or finite sampling, which are horizontally offset from each other to better distinguish, which does not seem to have a dramatic effect on the estimation correlation.

solvable with model-level refinements, discussed in Appendixes D 5 and D 7. We further find that finite sampling does not introduce significant bias in the estimator correlation, only increasing its statistical fluctuations.

In the following subsections, we describe an analytic description of finite sampling for the PoP distribution and potential ways to undo these effects.

2. Multinomial distribution and Poisson sampling

Taking M i.i.d. samples from a probability distribution $p(z)$ yields a number n_z of occurrences of each bitstring z . The vector $\vec{n} \equiv (n_{z_1}, \dots, n_{z_D})$ of such numbers is a random variable following the *multinomial distribution* $\vec{n} \sim \text{Multinomial}[M, p(z)]$ with PDF

$$P[\vec{n}; M, p(z)] = \frac{M!}{n_{z_1}! \dots n_{z_D}!} p(z_1)^{n_{z_1}} \dots p(z_D)^{n_{z_D}}, \quad (\text{E1})$$

where we have indexed the bitstrings from 1 to D . This is known *i.i.d. sampling* [74].

While Eq. (E1) is an exact statistical description of the outcome of sampling, it is analytically easier to use the Poisson sampling model, which assumes that each number n_z is an independent Poisson random variable with mean $Mp(z)$, i.e., $n_z \sim \text{Poi}[\mu = Mp(z)]$ [74]. Since each number variable is independent, the constraint $\sum_z n_z = M$ is not enforced; in the multinomial distribution, there are weak anticorrelations between number variables n_z and $n_{z'}$. The Poisson sampling approximation becomes increasingly accurate with increasing sample number M and serves as

a useful model for sampling [74]. Nevertheless, the full multinomial distribution is necessary for quantitatively accurate formulas in some cases which we indicate.

3. Finite-sample estimate of PoP distribution

The Poisson sampling model will be essential to helping us understand the effects of finite sampling on reconstructing the PoP distribution. Here, we wish to estimate the distribution of probabilities $P(p)$ from empirical data. The challenge lies in the fact that the probabilities $p(z)$ are difficult to empirically estimate—the *plug-in estimate* $\hat{p}(z) \equiv n_z/M$ is a rescaled Poisson variable and is unbiased but may exhibit large fluctuations. For simplicity of discussion, we first consider the case where $p_{\text{avg}}(z) = 1/D$ (as in RUCs) before tackling the general case. In this case, the possible values of $\hat{p}(z)$ are discrete; they are integer multiples of $1/M$, and the rescaled probability $\tilde{p}(z)$ is simply $Dp(z)$. Accordingly, we express our PoP as $P(Dp)$, which is system-size independent.

A standard practice to empirically estimate the PoP is to simply take the histogram of $\hat{p}(z)$. That is, one takes M samples from $p(z)$, measuring each bitstring n_z times. The empirical frequencies $\hat{p}(z) = n_z/M$ take discrete values in multiples of $1/M$ [when $p_{\text{avg}}(z)$ is nontrivial, the empirical rescaled frequencies are continuous] and estimates the PoP by the fraction of bitstrings with a given empirical frequency $\hat{P}(D\hat{p}) \equiv N_{\hat{p}}/D$.

For each bitstring z with probability $p(z)$, the empirical frequency n_z/M is a rescaled Poisson random variable. For a given empirical frequency $\hat{p}$, the number of bitstrings $N_{\hat{p}}$ with that empirical frequency is also a Poisson random variable with expected value

$$\hat{P}(D\hat{p} = nD/M) = \int_0^\infty d(Dp) P(Dp) p_{\text{Poi}}(n; pM), \quad (\text{E2})$$

where $p_{\text{Poi}}(n; pM) \equiv e^{-pM} (pM)^n / n!$ is the probability of a Poisson variable taking value n when the expected number of counts is $\mu = pM$. This prediction determines the empirical PoP with a finite number of samples. We emphasize that Eq. (E2) is the expected value of the finite-sampling PoP, and a given experimental PoP will show statistical fluctuations about this mean. In practice, occurrences for large Dp occur with sufficiently small probability that the number of potential values may be truncated at some level. Then, as long as M/D is large enough, there will be enough outcomes that each PoP count is approximately independent (i.e., the normalization constraint is negligible), acting as a Poisson random variable with $\sqrt{N_{\hat{p}}}$ uncertainty.

When the factor $p_{\text{avg}}(z)$ is not constant, the empirical rescaled probability $\hat{p}/p_{\text{avg}}(z)$ takes on continuous values,

and we must perform histogram binning instead. Here, the expected value of the histogram is given by

$$\hat{P}(\hat{p}) \approx \sum_{n=0}^{\infty} \frac{1}{D} \sum_z \int d\tilde{p} P(\tilde{p}) p_{\text{Poi}}[n; Mp_{\text{avg}}(z)\tilde{p}] \delta \left(\hat{p} - \frac{n}{Mp_{\text{avg}}(z)} \right). \quad (\text{E3})$$

The empirical PoP $\hat{P}(\hat{p})$ is akin to (but not exactly) a convolution of the underlying PoP with the Poisson PDF. It solely depends on the underlying PoP, $P(Dp)$, as well as the ratio of samples to Hilbert-space dimension, $m \equiv M/D$. In the limit $m \rightarrow \infty$, the estimate converges to the true value $\hat{P}(D\hat{p}) \rightarrow P(Dp)$. Finite values of m show potentially substantial deviations between $\hat{P}(D\hat{p})$ and $P(Dp)$. In Fig. 14, we illustrate the effects of finite sampling when the underlying PoP distribution is Porter-Thomas or arising from global depolarization errors [Eq. (D8)]. The Porter-Thomas (PT) distribution is accurately reconstructed with modest ratios $m \gtrsim 1$. However, the global depolarization distribution shows a sharp discontinuity at $Dp = 1 - F$, which is smoothed out under finite sampling. Even large sampling ratios such as $m = 30$ show considerable deviations.

These two cases can be analytically solved. The PT distribution yields the following finite-sampling PoP:

$$\hat{P}(D\hat{p}) d(D\hat{p}) = \frac{1}{P_{>0}} \left(\frac{m}{1+m} \right)^{1+mD\hat{p}} d(D\hat{p}), \quad (\text{E4})$$

where $P_{>0} = m/(1+m)$ is the probability of observing a nonzero value of $\hat{p}$ in Eq. (E2), elaborated on below. Meanwhile, the global-depolarization PoP gives

$$\begin{aligned} \hat{P}(D\hat{p}) d(D\hat{p}) &= \frac{1}{P_{>0}} F^{-1} e^{F^{-1}-1} \left(\frac{Fm}{1+Fm} \right)^{1+mD\hat{p}} \\ &\times \frac{\Gamma[1+mD\hat{p}, \frac{1-F}{F}(1+Fm)]}{\Gamma(1+mD\hat{p})} d(D\hat{p}), \end{aligned} \quad (\text{E5})$$

where $P_{>0} = 1 - e^{m(F-1)}$, $\Gamma(k, x) \equiv \int_x^\infty z^{k-1} \exp(-z) dz$ is the upper incomplete Gamma function, and $\Gamma(k)$ is the Gamma function.

In the literature, the probability $\hat{P}(\hat{p} = 0)$ is typically not included in the empirical PoP, and we must renormalize the resulting histogram by the factor $P_{>0}^{-1}$. Indeed, if one does include $\hat{P}(\hat{p} = 0)$, the finite-sampling deviations from the PT distribution become more evident, the factor $P_{>0}^{-1}$ should not be included in Eqs. (E4) and (E5), and Eq. (E2) is true without modification.

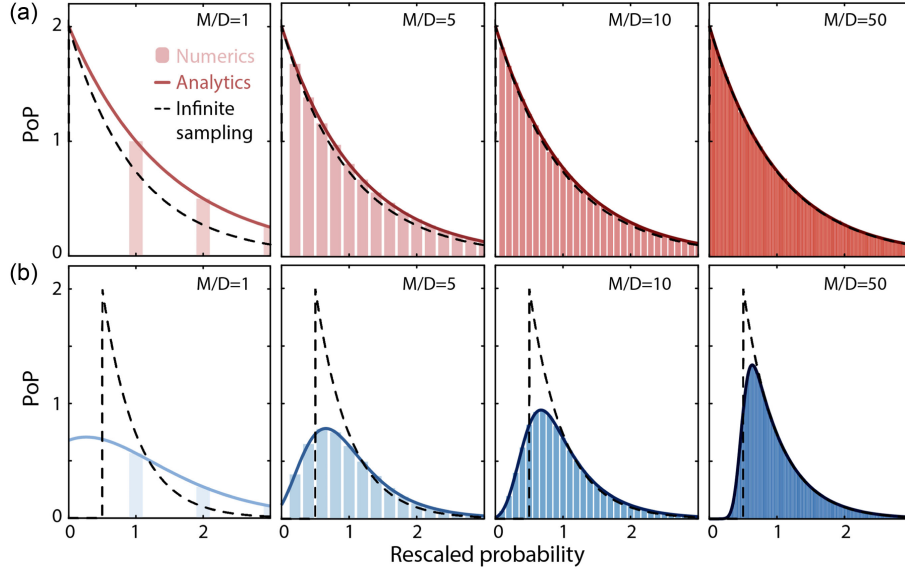


FIG. 14. Finite sampling effects on the PoP. The PoP is nontrivially modified when reconstructed from a finite number of samples, as generally considered in Eq. (E2). However, not all PoPs are as susceptible as others; for instance, (a) the Porter-Thomas distribution [Eq. (E4)] shows relatively little sensitivity to M/D (the ratio of number of measured bitstrings divided to the total Hilbert-space dimension), while (b) the PoP for global depolarization noise with $F = 0.5$ [Eq. (E5)] shows greater sensitivity due to the discontinuous PDF. In both cases, we see good agreement between numerical finite sampling and the analytical predictions Eqs. (E4) and (E5).

4. Inverting finite sample effects

Proof of principle: Inverting with Laguerre polynomials—Remarkably, for any value of m , it is theoretically possible to invert Eq. (E2) with Laguerre polynomials [making a technical assumption that the PoP $P(Dp)$ is square integrable, which is true in any finite system]. We expand $P(Dp)$ in terms of Laguerre polynomials:

$$P(Dp) = \sum_{k=0}^{\infty} c_k L_k(2mDp) \exp(-mDp), \quad (\text{E6})$$

where $L_k(x)$ is the k th Laguerre polynomial, satisfying the orthogonality relation $\int_0^{\infty} dx e^{-x} L_n(x) L_m(x) = \delta_{nm}$, and c_k are the unknown coefficients to be determined. Using the above orthogonality relation, c_k can be extracted from the empirical PoP, $\hat{P}(D\hat{p}_n)$, with the formula

$$c_k = 2 \sum_{n=0}^k \binom{k}{n} (-2)^n \hat{P}(D\hat{p}_n), \quad (\text{E7})$$

where $\hat{p}_n = n/M$ are the discrete values of the empirical frequencies. However, this formula is not practical. It is not robust to any amount of statistical or measurement error, since it weights the measured frequencies $\hat{P}(D\hat{p} = Dn/M)$ with exponentially increasing and oscillatory weights in n . Nevertheless, this serves as a theoretical basis that the ideal PoP can be reconstructed from any finite-sampling PoP. For a more practical procedure, we employ a regularized least-squares approach.

Inversion with regularized least squares—Our practical approach to undo the effects of finite sampling is based on regularized least squares. We note that this method is not used for any data shown in the main text, and we instead present it for completeness and as a potential way to further improve our results.

We discretize the domain of the PoP into a discrete number of histogram bins centered at $\tilde{p}_j$ and with width $\Delta\tilde{p}$. For concreteness, it suffices to stop at $\tilde{p} = 5$. We also discretize the finite-sampling PoP into discrete and finite histogram bins. When $p_{\text{avg}}(z) = 1/D$, its domain is naturally discrete, but this is not so for nontrivial $p_{\text{avg}}(z)$. Working in the former case for simplicity, we can rewrite Eq. (E2) as a linear equation:

$$\hat{\mathbf{P}} = \mathcal{M}\mathbf{P}, \quad (\text{E8})$$

$$\mathcal{M}_{n,j} = \exp(-\tilde{p}_j M) \frac{(\tilde{p}_j M)^n}{n!} (\Delta\tilde{p}). \quad (\text{E9})$$

When $p_{\text{avg}}(z)$ is nontrivial, one must perform histogram binning on both the range and the domain, and the matrix $\mathcal{M}$ is instead

$$\mathcal{M}_{i,j} \approx \frac{1}{D} \sum_z \sum_{n=\lceil M(\hat{p}_i - \Delta\hat{p}/2)p_{\text{avg}}(z) \rceil}^{\lceil M(\hat{p}_i + \Delta\hat{p}/2)p_{\text{avg}}(z) \rceil} P(\tilde{p}_j) p_{\text{Poi}}[n; M p_{\text{avg}}(z) \tilde{p}_j]. \quad (\text{E10})$$

Our result Eq. (E7) essentially gives an explicit result for $\mathcal{M}^{-1}$ in the Laguerre polynomial basis. Unfortunately,

$\mathcal{M}^{-1}$ is extremely ill conditioned and cannot accommodate the shot noise in $\hat{\mathbf{P}}$. If one naively attempts to solve this linear equation, the solution is wildly oscillatory and large. In order to remedy this, we simply regularize our solution and solve the least-squares equation with a penalty term:

$$\mathbf{P}_{\text{est}} = \text{argmin}_{\mathbf{y}} [(\mathcal{M}\mathbf{y} - \hat{\mathbf{P}})^T A (\mathcal{M}\mathbf{y} - \hat{\mathbf{P}}) + \lambda \mathbf{y}^T B \mathbf{y}], \quad (\text{E11})$$

where we have presented this problem in its full generality and A and B are arbitrary symmetric positive-semidefinite matrices that we can suitably choose to regularize the regression. When A and $B = \hat{I}$, the first term is the well-known squared error, and the second is the L_2 norm. However, we may modify A and B to enforce other constraints, such as smoothness of the solution. This is known as ridge, or Tikhonov, regression [75] and has explicit solution

$$\mathbf{P}_{\text{est}} = (\mathcal{M}^T A \mathcal{M} + \lambda B)^{-1} \mathcal{M}^T A \hat{\mathbf{P}}. \quad (\text{E12})$$

As an illustration, we demonstrate this in Fig. 15. The parameter λ is tunable, and one must choose an appropriate value. While it can be deduced analytically, here, we present a user-friendly and intuitive method. In Fig. 15(c), we plot the squared error $\|\mathcal{M}\mathbf{P}_{\text{est}}(\lambda) - \hat{\mathbf{P}}\|_2^2$ as a function of the parameter λ . This shows two parameter regions. When λ is small, the squared error is nearly independent of λ . In this regime, the solution is overfitted, and enforcing the constraint effectively regularizes the solution without significant effect on the squared error. When λ is large, the squared error grows quickly with λ . In this regime, the solution is overdamped and the constraint is too strong. We

find that operating at the midpoint between both regimes is often optimal. Despite this, its reconstructed solutions are often imperfect, though potentially improvable through addition of additional constraints on the regression (like forcing the PoP to be always positive). Therefore, in the main text, we rely on direct comparisons of the finite-sampling PoP, outlined below.

APPENDIX F: CUMULANTS OF PROBABILITY-OF-PROBABILITIES DISTRIBUTIONS

As established above, due to finite-sampling effects, it can be challenging to empirically estimate the PoP. In the main text, we have compared directly to a finite sampled version of model PoP predictions, and in addition we have discussed potential theoretical means of reversing the finite-sampling effect. Here, we discuss how we can also estimate properties of the PoP to greater precision, without reconstructing the PoP distribution itself. Specifically, the moments of the PoP can be efficiently estimated, which we may use to learn about the type of noise present in the system. Furthermore, the moments for different error models can have different functional dependencies on the fidelity F , meaning that analyzing how moments change over multiple values of fidelity may improve predictive power.

From Appendix D, the probability distribution can be written as $\tilde{p}(z) = \sum_i \omega_i \tilde{p}_i(z)$, where $\tilde{p}(z) \equiv p(z)/p_{\text{avg}}(z)$ is the rescaled probability and $\tilde{p}_i(z)$ are independent Porter-Thomas (i.e., Dirichlet random) distributions. Remarkably, some properties of the coefficients $\{\omega_i\}$ can be estimated from empirical samples from $p(z)$, even *without knowledge* of the constituents $\tilde{p}_i(z)$. Specifically, cumulants of the PoP are related to moments of $\{\omega_i\}$.

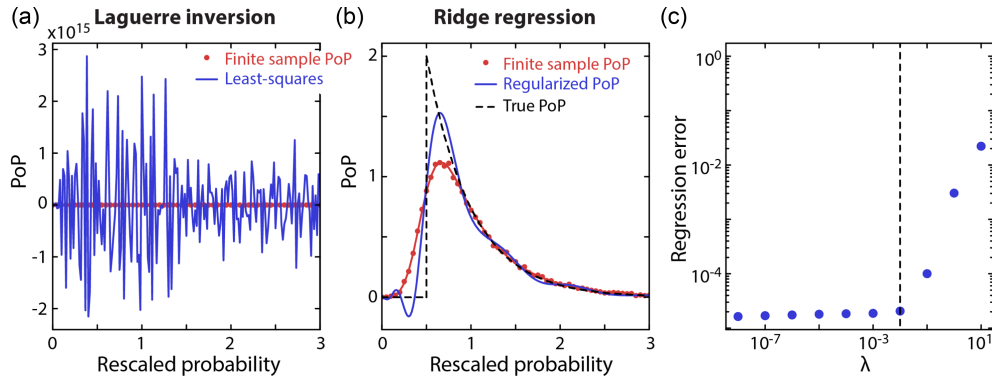


FIG. 15. Strategies to reverse finite-sampling effects on the PoP. We simulate the finite-sampling effects on a global depolarization PoP, with a Hilbert-space dimension of $D = 2^{15}$ and $M = 20D$ samples. (a) Regular least-squares regression [Eq. (E8)] is numerically unstable and leads to large, unphysical solutions (note scale of values). (b) Regularized regression penalizes this divergence and gives an approximate reconstruction (blue line) of the infinite-sampling PoP (black dashed line) from the finite-sampling PoP (red points). (c) The free parameter λ is chosen by plotting the regression error as a function of λ , displaying two regimes where the solution is respectively overfitted and overdamped. We choose λ at the transition between these regimes. Nevertheless, the reconstructed PoP is imperfect; adding further constraints, that, for instance, force the PoP to be always positive, may improve agreement, but we leave such avenues to future work.

As above, we first discuss the case when the factors p_{avg} are constant, before generalizing. The k th (raw, i.e., non-central) moment of the PoP $P(Dp)$ is given by

$$m_k \equiv D^{-1} \sum_z [Dp(z)]^k \approx \int (Dp)^k P(Dp) d(Dp), \quad (\text{F1})$$

where the approximation follows by treating each value $Dp(z)$ as an independent sample from the PoP $P(Dp)$. We then consider the cumulants κ_k . For example, the first three cumulants are

$$\kappa_1 = m_1, \quad (\text{F2})$$

$$\kappa_2 = m_2 - m_1^2, \quad (\text{F3})$$

$$\kappa_3 = m_3 - 3m_2m_1 + 2m_1^3. \quad (\text{F4})$$

A key property of the cumulants is that they add for sums of independent random variables [68]. That is, for $\tilde{p} = \sum_i \omega_i \tilde{p}_i$, $\kappa_k(\tilde{p}) = \sum_i \omega_i^k \kappa_k(\tilde{p}_i)$. Since each $\tilde{p}_i$ is an independent Porter-Thomas distribution, they have identical cumulants $\kappa_k = (k-1)!$. Therefore, these k th cumulants of the PoP are equal to the raw k th moment of $\{\omega_i\}$:

$$\frac{\kappa_k(p)}{(k-1)!} \approx \sum_i \omega_i^k. \quad (\text{F5})$$

As an illustration, we concretely demonstrate this with $k = 2$. We have

$$\kappa_2(p) = D \sum_z p(z)^2 - 1 \approx \sum_i \omega_i^2. \quad (\text{F6})$$

To see this, note that $\sum_z p(z)^2 = \sum_z \omega_i \omega_j \sum_{ij} p_i(z) p_j(z)$. Since $p_i(z)$ and $p_j(z)$ are independent Porter-Thomas distributions, they are approximately orthogonal, in the sense $D \sum_z p_i(z) p_j(z) = 1 + \delta_{ij} + O(D^{-1/2})$. Using this relation, we obtain

$$D \sum_z p(z)^2 = 1 + \sum_i \omega_i^2 + O(D^{-1/2}), \quad (\text{F7})$$

where we have used the fact that $\sum_i \omega_i = 1$. Remarkably, due to this approximate orthogonality, the second moment of the PoP depends on only the coefficients ω_i , up to exponentially small fluctuations. When $p_{\text{avg}}(z)$ is non-trivial, we should replace $Dp(z)$ with $\tilde{p}(z) \equiv p(z)/p_{\text{avg}}(z)$ and the average over z as a weighted mean $\mathbb{E}_z[\cdot] = \sum_z p_{\text{avg}}(z)[\cdot]$. This modifies the moments to $m_k = \sum_z p(z)^{k-1}/p_{\text{avg}}(z)^{k-1}$.

1. Relation of hypoexponential moments to mixed state moments

The moments give key information about the mixed state. In the context of closed system dynamics, we found that the moments of the weight vector satisfy

$$\sum_i \omega_i^k \approx \text{tr}(\hat{\rho}_A^k), \quad (\text{F8})$$

as long as the bath B is sufficiently small.

We conjecture that the above equation also holds in the context of open system dynamics, as long as the fidelity is sufficiently high. Indeed, Eq. (F8) is true for Haar-random measurement bases $\{|z\rangle\}$, up to subleading $O(D^{-1/2})$ corrections. If the eigenstates of $\hat{\rho}$ look “random” with respect to the computational basis $|z\rangle$, we get the desired equivalence.

2. Moments under different noise sources

The moments $\sum_i \omega_i^k$ behave differently under different noise channels. In Table I, we summarize the low-order cumulants (κ_2 and κ_3) of the PoP predicted for various noise channels and their dependence on the fidelity F . By independently measuring the fidelity F (using, for example, cross-entropy benchmarking [24,26,49]) and the cumulants $\kappa_k(p)$, one would be able to discriminate between such noise models.

Consider first a local error channel, we have $\{\omega_i^{\text{loc}}\} \approx \{F, (1-F)/k, (1-F)/k, \dots, (1-F)/k\}$. Therefore, its second moments

$$\sum_i (\omega_i^{\text{loc}})^2 = F^2 + \frac{(1-F)^2}{k} \xrightarrow{k \rightarrow \infty} F^2. \quad (\text{F9})$$

This relationship generalizes to higher moments: The largest weight F dominates, and we have $\sum_i (\omega_i^{\text{loc}})^k \approx F^k$.

We next consider global coherent errors. Here, the weights follow a Gaussian distribution

$$\omega_i^{\text{coh}} = \exp[-i^2/(2\sigma_F^2)]/\mathcal{Z}(\sigma_F) \quad (\text{F10})$$

with a width σ_F that depends on the fidelity F . When $\sigma_F \gg 1$, the normalization factor is well approximated by $\mathcal{Z}(\sigma_F) \approx \sqrt{2\pi\sigma_F^2}$. Performing the sum, we have

$$\sum_i (\omega_i^{\text{coh}})^2 \approx \frac{1}{4\pi\sigma_F^2} = \frac{F}{\sqrt{2}}. \quad (\text{F11})$$

Note the different scaling with F . In Fig. 16(b), we verify these scaling behaviors for our two types of errors in an RUC simulation.

Note that the above prediction breaks down when $F \approx 1$. Using our Gaussian hypoexponential model above, the discrete Gaussian sums can be explicitly evaluated,

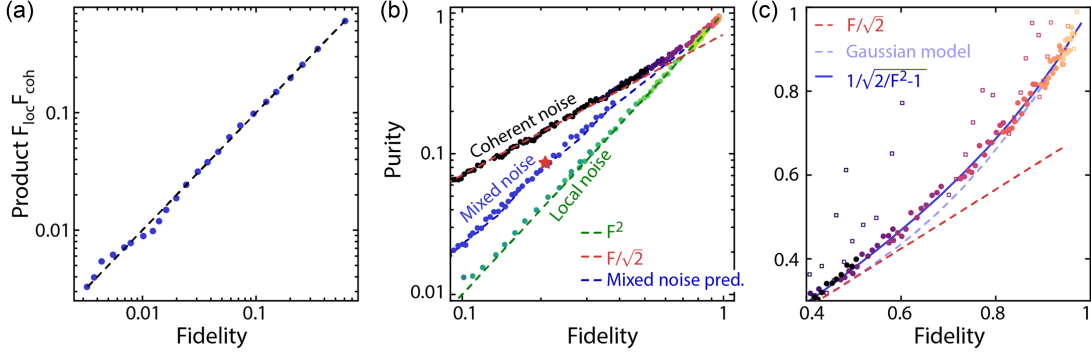


FIG. 16. RUC simulations of mixed local and coherent errors. (a) We first verify that errors factorize over a large range of fidelities. Here, a single instance of a random unitary circuit is presented, at a fixed proportion of local error and coherent error. As a function of circuit depth, its fidelity F decreases. We simulate the same circuit in the presence of only local or coherent errors at the same rate, obtaining the “local” and “coherent” fidelities F_{loc} and F_{coh} , respectively. The overall fidelity is well approximated by $F \approx F_{\text{loc}} F_{\text{coh}}$. This makes the concept of an error source “fraction” meaningful. (b) Plots of the purity, or second moment, $D \sum_z p(z)^2$ for noisy RUC dynamics in the presence of global coherent or local noise, or a mixture of the two noise sources (Table I). The purity due to global coherent noise asymptotically behaves as $F_{\text{coh}}/\sqrt{2}$, while the purity due to local noise behaves as F_{loc}^2 . In the presence of both types of noise, the purity behaves as $F_{\text{loc}}^2 F_{\text{coh}}/\sqrt{2}$ (blue dashed line), enabling efficient estimation of the noise content from the second moment. Our mixed error simulation is a depth-96 RUC with 12 qubits, using the same error rates as in Fig. 5(c) in the main text (precise depth marked with a red star), while our local and global coherent errors are simulations of RUCs at several error rates. Data points displayed are at depths of 24 and above. (c) Purity from global coherent noise at high fidelities. Here, we see deviations from the asymptotic behavior $F/\sqrt{2}$ (red dashed line). In particular, it takes a value of 1 at $F = 1$. At intermediate values $0.5 < F < 0.9$, the purity deviates from the prediction from our Gaussian-hypoexponential model (blue dashed line), instead agreeing with $(2/F^2 - 1)^{-1/2}$ (blue solid line), an expression obtained from a calculation of the state purity. This indicates that our model for the hypoexponential distribution [Eq. (D9)] should be refined in this regime. In addition, to illustrate the importance of circuit depth, for the open markers we indicate the purity at depths 24 and below. These deviate from our prediction due to a lack of anticoncentration in shallow circuits.

giving $\sum_i (\omega_i^{\text{coh}})^2 = \theta_3(\exp[-1/\sigma_F^2]) / \theta_3\{\exp[-1/(2\sigma_F^2)]\}^2$, where θ_3 is a Jacobi theta function. Meanwhile, using methods developed in Ref. [24], one can perform an exact calculation of the state purity in this setting, which gives $\text{tr}(\hat{\rho}^2) = (2/F^2 - 1)^{-1/2}$. As seen in Fig. 16(c), these expressions show deviations when $F \approx 1$, with numerical data from RUC simulations more consistent with the latter expression.

In the presence of mixed noise, we further observe that since the weights factorize as $\omega_{ij}^{\text{mixed}} = \omega_i^{\text{loc}} \omega_j^{\text{coh}}$, so do the cumulants, in particular, $\kappa_2 \approx F_{\text{loc}}^2 F_{\text{coh}}/\sqrt{2}$. Coupled with knowledge of the fidelity $F = F_{\text{loc}} F_{\text{coh}}$, one could extract the fidelities of each noise type. In Fig. 16(b), we show proof-of-principle numerical evidence for this behavior in a RUC.

3. Estimator of moments

Remarkably, the moments can be efficiently estimated [76] with the following estimator:

$$\hat{m}_k = D^{k-1} \sum_z \frac{(n_z)^k}{M^k} \approx D^{k-1} \sum_z p(z)^k, \quad (\text{F12})$$

where $n^k \equiv n(n-1)\dots(n-k)$ is the k th falling power of n and n_z is the number of occurrences of the bitstring z from M measurements. This was shown to be an unbiased

estimator of m_k , with optimal sample complexity scaling. Specifically, Ref. [76] showed that the k th Rényi entropy (obtained from m_k) can be estimated to additive error δ with $M = \Theta(D^{1-1/k}/\delta^2)$ samples, where D is the Hilbert-space dimension (more generally, the number of possible measurement outcomes). In the same work, they showed that this scaling is optimal. Furthermore, note that in the original reference [76] the denominator of Eq. (F12) is M^k instead of M^k . The former is the unbiased estimator in the Poisson sampling setting, while the latter is the unbiased estimator under realistic sampling. When $p_{\text{avg}}(z)$ is nontrivial, the moment estimator is $\hat{m}_k = \sum_z (n_z)^k / [M^k p_{\text{avg}}(z)^{k-1}] \approx \sum_z p(z)^{k-1} / p_{\text{avg}}(z)^{k-1}$.

We are interested in Eq. (F12) because the low moments of the PoP can be estimated with far fewer samples [e.g., $M = O(\sqrt{D})$ for the second moment] than the entire PoP, which requires $M = O(D)$. It is illustrative to consider the simplest nontrivial case, the second moment m_2 . Our estimator $\hat{m}_2 = D \sum_z n_z(n_z - 1) / [M(M - 1)]$ essentially counts the number of times a bitstring z is seen more than once. Accordingly, this estimates the collision probability $D \sum_z p(z)^2$, here rescaled to be typically an $O(1)$ quantity. The sample complexity is almost optimal: Estimating the second moment to multiplicative accuracy δ requires only $\sqrt{D}/\delta^2$ measurements, a manifestation of the birthday paradox. For our purposes, it is also useful to estimate

its sample requirements for *additive* accuracy. Using properties of the multinomial distribution, we find that the expected uncertainty in the second moment m_2 is

$$(\Delta \hat{m}_2)^2 = \left[\frac{2M^2 - M}{M^4} D m_2 + 4 \frac{M^3}{M^4} m_3 - 2 \frac{M^2(2M - 3)}{M^4} m_2 \right] \quad (\text{F13})$$

$$\approx \frac{2D m_2}{M^2} + \frac{4m_3}{M} - \frac{4m_2^2}{M}, \quad (\text{F14})$$

where the last term arises from anticorrelations between n_z and $n_{z'}$ which are not present in the Poisson sampling model. Since m_2 and m_3 are $O(1)$ quantities, this result indicates that, for an additive error $|\hat{m}_2| = \delta$, the sample complexity behaves as $O[\max(\sqrt{D}/\delta, 1/\delta^2)]$, where the crossover between the two scaling behaviors occurs at $M \approx D$.

To recover the result of Ref. [76] for the required number of samples to achieve a multiplicative error δ , namely, $\Theta(\sqrt{D}/\delta^2)$, we focus on the second term in Eq. (F14) and then use the inequalities $m_3 = D^2 \sum_z p(z)^3 \leq D^2 (\sum_z p(z)^2)^{3/2} = \sqrt{D} m_2^{3/2}$ and $m_2 \geq 1$ to give the desired result that $(\Delta \hat{m}_2)^2 \sim m_3/M \leq \sqrt{D} m_2^2/M$. The former inequality comes from the relation between L_p norms, $L_p \leq L_q$ for $p \geq q$, relating m_2 and m_3 to the L_2 and L_3 norms of the probability vector $p(z)$, while the latter inequality is a straightforward application of the

Cauchy-Schwarz inequality. Finally, note that this result is valid only in the asymptotic regime, where the second term in Eq. (F14) dominates the first term. This is valid at a crossover value of $M \approx D$, before which the sample complexity scales as $M \sim \sqrt{D}/\delta$ instead.

APPENDIX G: RELATIONSHIP BETWEEN THE HYPOEXPONENTIAL WEIGHTS AND THE CHANNEL KRAUS OPERATORS

In addition to their relative ease of measurement, the moments of the hypoexponential weight vector $\vec{w}$ can also be theoretically analyzed and related to quantities that characterize the strength of the noise channel. Given our analysis, it is natural to ask how the weight vector $\vec{w}$ reflects the underlying quantum channel. We find that the properties of $\vec{w}$ are related to the strength of the channel, but there are additional terms that we expect to be subleading. We denote the channel Φ , such that the output state is $\hat{\rho} \equiv \Phi(|\Psi_0\rangle\langle\Psi_0|) \equiv \sum_i K_i |\Psi_0\rangle\langle\Psi_0| K_i^\dagger$, where K_i are Kraus operators that describe the channel, satisfying $\sum_i K_i^\dagger K_i = \hat{I}$. $\{K_i\}$ can be uniquely chosen to be orthogonal [77], in the sense that $\text{tr}(K_i K_j^\dagger) = D q_i \delta_{ij}$, such that $\sum_i q_i = 1$.

We first consider the second moment $\sum_i w_i^2$, which, using Eq. (F8), approximates the purity $\text{tr}[\Phi(|\Psi_0\rangle\langle\Psi_0|)^2]$. We approximate this purity with the Haar-averaged value:

$$\mathbb{E}_{\psi \sim \text{Haar}} \{ \text{tr}[\Phi(|\psi\rangle\langle\psi|)^2] \} = \frac{\sum_{i,j} (\text{tr}(K_i K_j^\dagger) \text{tr}(K_j K_i^\dagger) + \text{tr}(K_i K_i^\dagger K_j K_j^\dagger))}{D(D+1)} \quad (\text{G1})$$

$$= \frac{D}{D+1} \left(\sum_i q_i^2 + \text{tr}[\Phi(\hat{I}/D)^2] \right), \quad (\text{G2})$$

where we have used the orthogonality conditions above. The first term is the “purity” of the noise channel strengths p_i , while the second term is related to whether the channel is unital or not [78]. If the channel is unital, then $\Phi(\hat{I}/D) = \hat{I}/D$ so the second term has size $1/D$, which we expect to be exponentially smaller than the purity of the channel. Meanwhile, if the channel is nonunital—in an extreme case, consider the amplitude damping channel such that $\Phi(\hat{I}/D) = |0\rangle\langle 0|$ —then the second term has size 1, which is larger than $\sum_i q_i^2$. In other words, the purity of the state has two contributions, given by the decoherence strength of the channel and the purity of the steady state $\Phi(\hat{I}/D)$. Assuming that $\Phi(\hat{I}/D)$ is unital or close to maximally mixed, however, we conclude that the purity of a typical input state reflects how noisy the channel is.

Higher moments give additional terms which we may not simplify, such as $\sum_{hij} \text{tr}(K_h K_i^\dagger K_j) + \text{H.c.}$ for $k = 3$. However, we may assume that such quantities are exponentially suppressed, since the orthogonal operators K_i are obtained by the (reshaped) eigenvectors of the Choi matrix which describes the channel [77]. If these eigenvectors have sufficiently complicated bipartite structure, we expect the contributions from such terms to be exponentially suppressed. We may repeatedly apply this argument to all orders k and find that the (expected) dominant contribution is $\sum_i q_i^k$. This also bounds the typicality error that is the difference between the moment $\text{tr}(\Phi(|\Psi\rangle\langle\Psi|)^k)$ of a single, typical random state $|\Psi\rangle$ and the ensemble average. Indeed, we also expect our conclusion to hold as long as the initial state $|\Psi_0\rangle$ has

sufficiently small overlap with the eigenstates $\{|E\rangle\}$. Therefore, we have

$$\frac{\kappa_k(p)}{(k-1)!} \approx \sum_i \omega_i^k \approx \text{tr}[\Phi(|\Psi_0\rangle\langle\Psi_0|)^k] \approx \sum_i q_i^k, \quad (\text{G3})$$

which, therefore, allows us to conclude that $\omega_i \approx q_i$, as long as the number of nonzero weights is less than the Hilbert-space dimension, i.e., $\#\{\omega_i\} \ll D$.

-
- [1] M. Srednicki, *Chaos and quantum thermalization*, *Phys. Rev. E* **50**, 888 (1994).
- [2] J. M. Deutsch, *Quantum statistical mechanics in a closed system*, *Phys. Rev. A* **43**, 2046 (1991).
- [3] M. Rigol, V. Dunjko, and M. Olshanii, *Thermalization and its mechanism for generic isolated quantum systems*, *Nature (London)* **452**, 854 (2008).
- [4] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, *Colloquium: Many-body localization, thermalization, and entanglement*, *Rev. Mod. Phys.* **91**, 021001 (2019).
- [5] R. Nandkishore and D. A. Huse, *Many-body localization and thermalization in quantum statistical mechanics*, *Annu. Rev. Condens. Matter Phys.* **6**, 15 (2015).
- [6] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, *From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics*, *Adv. Phys.* **65**, 239 (2016).
- [7] M. Ueda, *Quantum equilibration, thermalization and prethermalization in ultracold atoms*, *Nat. Rev. Phys.* **2**, 669 (2020).
- [8] J. M. Deutsch, *Eigenstate thermalization hypothesis*, *Rep. Prog. Phys.* **81**, 082001 (2018).
- [9] 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).
- [10] 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).
- [11] G. Clos, D. Porras, U. Warring, and T. Schaetz, *Time-resolved observation of thermalization in an isolated quantum system*, *Phys. Rev. Lett.* **117**, 170401 (2016).
- [12] F. Kranzl, A. Lasek, M. K. Joshi, A. Kalev, R. Blatt, C. F. Roos, and N. Yunger Halpern, *Experimental observation of thermalization with noncommuting charges*, *PRX Quantum* **4**, 020318 (2023).
- [13] H. Kim, Y. J. Park, K. Kim, H.-S. Sim, and J. Ahn, *Detailed balance of thermalization dynamics in Rydberg-atom quantum simulators*, *Phys. Rev. Lett.* **120**, 180502 (2018).
- [14] C. Neill, P. Roushan, M. Fang, Y. Chen, M. Kolodrubetz, Z. Chen, A. Megrant, R. Barends, B. Campbell, B. Chiaro et al., *Ergodic dynamics and thermalization in an isolated quantum system*, *Nat. Phys.* **12**, 1037 (2016).
- [15] J. Wang, M. H. Lamann, J. Richter, R. Steinigeweg, A. Dymarsky, and J. Gemmer, *Eigenstate thermalization hypothesis and its deviations from random-matrix theory beyond the thermalization time*, *Phys. Rev. Lett.* **128**, 180601 (2022).
- [16] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, *Random-matrix theories in quantum physics: Common concepts*, *Phys. Rep.* **299**, 189 (1998).
- [17] S. Pilatowsky-Cameo, C. B. Dag, W. W. Ho, and S. Choi, *Complete Hilbert-space ergodicity in quantum dynamics of generalized Fibonacci drives*, *Phys. Rev. Lett.* **131**, 250401 (2023).
- [18] D. K. Mark, F. Surace, A. Elben, A. L. Shaw, J. Choi, G. Refael, M. Endres, and S. Choi, *A maximum entropy principle in deep thermalization and in Hilbert-space ergodicity*, *arXiv:2403.11970*.
- [19] C. E. Porter and R. G. Thomas, *Fluctuations of nuclear reaction widths*, *Phys. Rev.* **104**, 483 (1956).
- [20] 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*, *Nat. Phys.* **14**, 595 (2018).
- [21] S. Mullane, *Sampling random quantum circuits: A pedestrian's guide*, *arXiv:2007.07872*.
- [22] 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).
- [23] A. Browaeys and T. Lahaye, *Many-body physics with individually controlled Rydberg atoms*, *Nat. Phys.* **16**, 132 (2020).
- [24] A. L. Shaw, Z. Chen, J. Choi, D. K. Mark, P. Scholl, R. Finkelstein, A. Elben, S. Choi, and M. Endres, *Benchmarking highly entangled states on a 60-atom analogue quantum simulator*, *Nature (London)* **628**, 71 (2024).
- [25] L. M. Leemis and J. T. McQueston, *Univariate distribution relationships*, *Am. Stat.* **62**, 45 (2008).
- [26] D. K. Mark, J. Choi, A. L. Shaw, M. Endres, and S. Choi, *Benchmarking quantum simulators using ergodic quantum dynamics*, *Phys. Rev. Lett.* **131**, 110601 (2023).
- [27] A. J. Short, *Equilibration of quantum systems and subsystems*, *New J. Phys.* **13**, 053009 (2011).
- [28] P. Reimann, *Foundation of statistical mechanics under experimentally realistic conditions*, *Phys. Rev. Lett.* **101**, 190403 (2008).
- [29] M. Srednicki, *The approach to thermal equilibrium in quantized chaotic systems*, *J. Phys. A* **32**, 1163 (1999).
- [30] C. Nation and D. Porras, *Off-diagonal observable elements from random matrix theory: Distributions, fluctuations, and eigenstate thermalization*, *New J. Phys.* **20**, 103003 (2018).
- [31] C. Nation and D. Porras, *Ergodicity probes: Using time-fluctuations to measure the Hilbert space dimension*, *Quantum* **3**, 207 (2019).
- [32] M. Bauer, D. Bernard, and T. Jin, *Universal fluctuations around typicality for quantum ergodic systems*, *Phys. Rev. E* **101**, 012115 (2020).
- [33] T. Kiendl and F. Marquardt, *Many-particle dephasing after a quench*, *Phys. Rev. Lett.* **118**, 130601 (2017).
- [34] J. Preskill, *Quantum computing in the NISQ era and beyond*, *Quantum* **2**, 79 (2018).
- [35] B. D. Clader, C. J. Trout, J. P. Barnes, K. Schultz, G. Quiroz, and P. Titum, *Impact of correlations and heavy tails on*

- quantum error correction, *Phys. Rev. A* **103**, 052428 (2021).
- [36] C. Cafaro and S. Mancini, *Quantum stabilizer codes for correlated and asymmetric depolarizing errors*, *Phys. Rev. A* **82**, 012306 (2010).
- [37] M. S. Carroll, J. R. Wootton, and A. W. Cross, *Subsystem surface and compass code sensitivities to non-identical infidelity distributions on heavy-hex lattice*, [arXiv:2402.08203](#).
- [38] Y. Nakata, P. S. Turner, and M. Murao, *Phase-random states: Ensembles of states with fixed amplitudes and uniformly distributed phases in a fixed basis*, *Phys. Rev. A* **86**, 012301 (2012).
- [39] N. Linden, S. Popescu, A. J. Short, and A. Winter, *Quantum mechanical evolution towards thermal equilibrium*, *Phys. Rev. E* **79**, 061103 (2009).
- [40] J. Riddell and N. Pagliaroli, *No-resonance conditions, random matrices, and quantum chaotic models*, [arXiv:2307.05417](#).
- [41] Y. Huang, *Extensive entropy from unitary evolution*, [arXiv:2104.02053](#).
- [42] A. W. Harrow, *The church of the symmetric subspace*, [arXiv:1308.6595](#).
- [43] P. Pfeuty, *The one-dimensional Ising model with a transverse field*, *Ann. Phys. (N.Y.)* **57**, 79 (1970).
- [44] F. H. L. Essler, H. Frahm, F. Göhmann, A. Klümper, and V. E. Korepin, *The One-Dimensional Hubbard Model* (Cambridge University Press, Cambridge, England, 2005), ISBN 978-0-521-80262-8.
- [45] S. Nadarajah, *A review of results on sums of random variables*, *Acta Appl. Math.* **103**, 131 (2008).
- [46] J. J. Burnett, A. Bengtsson, M. Scigliuzzo, D. Niepce, M. Kudra, P. Delsing, and J. Bylander, *Decoherence benchmarking of superconducting qubits*, *npj Quantum Inf.* **5**, 1 (2019).
- [47] 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](#).
- [48] E. Paladino, Y. Galperin, G. Falci, and B. Altshuler, *1/f noise: Implications for solid-state quantum information*, *Rev. Mod. Phys.* **86**, 361 (2014).
- [49] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, *Quantum supremacy using a programmable superconducting processor*, *Nature (London)* **574**, 505 (2019).
- [50] Y. Cai, Y. Tong, and J. Preskill, *Stochastic error cancellation in analog quantum simulation*, [arXiv:2311.14818](#).
- [51] B. M. Terhal, *Quantum error correction for quantum memories*, *Rev. Mod. Phys.* **87**, 307 (2015).
- [52] Y. Liu, M. Otten, R. Bassirianjahromi, L. Jiang, and B. Fefferman, *Benchmarking near-term quantum computers via random circuit sampling*, [arXiv:2105.05232](#).
- [53] J.-S. Kim, L. S. Bishop, A. D. Córcoles, S. Merkel, J. A. Smolin, and S. Sheldon, *Hardware-efficient random circuits to classify noise in a multiqubit system*, *Phys. Rev. A* **104**, 022609 (2021).
- [54] R. Harper, S. T. Flammia, and J. J. Wallman, *Efficient learning of quantum noise*, *Nat. Phys.* **16**, 1184 (2020).
- [55] C.-M. Chung, L. Bonnes, P. Chen, and A. M. Läuchli, *Entanglement spectroscopy using quantum Monte Carlo*, *Phys. Rev. B* **89**, 195147 (2014).
- [56] H. F. Song, S. Rachel, C. Flindt, I. Klich, N. Laflorencie, and K. Le Hur, *Bipartite fluctuations as a probe of many-body entanglement*, *Phys. Rev. B* **85**, 035409 (2012).
- [57] A. Cooper, J. P. Covey, I. S. Madjarov, S. G. Porsev, M. S. Safronova, and M. Endres, *Alkaline-earth atoms in optical tweezers*, *Phys. Rev. X* **8**, 041055 (2018).
- [58] M. A. Norcia, A. W. Young, and A. M. Kaufman, *Microscopic control and detection of ultracold strontium in optical-tweezer arrays*, *Phys. Rev. X* **8**, 041054 (2018).
- [59] P. Scholl, A. L. Shaw, R. B.-S. Tsai, R. Finkelstein, J. Choi, and M. Endres, *Erasure conversion in a high-fidelity Rydberg quantum simulator*, *Nature (London)* **622**, 273 (2023).
- [60] M. Endres, H. Bernien, A. Keesling, H. Levine, E. R. Anschuetz, A. Krajenbrink, C. Senko, V. Vuletic, M. Greiner, and M. D. Lukin, *Atom-by-atom assembly of defect-free one-dimensional cold atom arrays*, *Science* **354**, 1024 (2016).
- [61] D. Barredo, S. de Leseleuc, V. Lienhard, T. Lahaye, and A. Browaeys, *An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays*, *Science* **354**, 1021 (2016).
- [62] I. S. Madjarov, J. P. Covey, A. L. Shaw, J. Choi, A. Kale, A. Cooper, H. Pichler, V. Schkolnik, J. R. Williams, and M. Endres, *High-fidelity entanglement and detection of alkaline-earth Rydberg atoms*, *Nat. Phys.* **16**, 857 (2020).
- [63] J. P. Covey, I. S. Madjarov, A. Cooper, and M. Endres, *2000-times repeated imaging of strontium atoms in clock-magic tweezer arrays*, *Phys. Rev. Lett.* **122**, 173201 (2019).
- [64] A. L. Shaw, R. Finkelstein, R. B.-S. Tsai, P. Scholl, T. H. Yoon, J. Choi, and M. Endres, *Multi-ensemble metrology by programming local rotations with atom movements*, *Nat. Phys.* **20**, 195 (2024).
- [65] K. Mølmer, Y. Castin, and J. Dalibard, *Monte Carlo wave-function method in quantum optics*, *J. Opt. Soc. Am. B* **10**, 524 (1993).
- [66] H. Kim, T. N. Ikeda, and D. A. Huse, *Testing whether all eigenstates obey the eigenstate thermalization hypothesis*, *Phys. Rev. E* **90**, 52105 (2014).
- [67] A. J. Daley, *Quantum trajectories and open many-body quantum systems*, *Adv. Phys.* **63**, 77 (2014).
- [68] J. Novak and M. LaCroix, *Three lectures on free probability*, [arXiv:1205.2097](#).
- [69] N. Balakrishnan and V. B. Nevzorov, *A Primer on Statistical Distributions* (John Wiley & Sons, Ltd, New York, 2003), ISBN 978-0-471-72222-9.
- [70] 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).
- [71] H. Shapourian, S. Liu, J. Kudler-Flam, and A. Vishwanath, *Entanglement negativity spectrum of random mixed states: A diagrammatic approach*, *PRX Quantum* **2**, 030347 (2021).

- [72] M. Žnidarič, *Entanglement of random vectors*, *J. Phys. A* **40**, F105 (2006).
- [73] NIST digital library of mathematical functions, <https://dlmf.nist.gov/>.
- [74] C. L. Canonne, *A Survey on Distribution Testing: Your Data is Big. But is it Blue?*, Graduate Surveys Vol. 9 (Theory of Computing Library, 2020).
- [75] R. W. Hoerl, *Ridge regression: A historical context*, *Technometrics* **62**, 420 (2020).
- [76] J. Acharya, A. Orlitsky, A. T. Suresh, and H. Tyagi, *Estimating Renyi entropy of discrete distributions*, [arXiv:1408.1000](https://arxiv.org/abs/1408.1000).
- [77] F. Verstraete and H. Verschelde, *On quantum channels*, [arXiv:quant-ph/0202124](https://arxiv.org/abs/quant-ph/0202124).
- [78] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, *Estimating the coherence of noise*, *New J. Phys.* **17**, 113020 (2015).