Emergent Randomness and Benchmarking from Many-Body Quantum Chaos

Joonhee Choi, Adam L. Shaw, * Iaylo S. Madjarov, Xin Xie,
Jacob P. Covey, Jordan S. Cotler, Daniel K. Mark, Hsin-Yuan Huang, Anant Kale,
Hannes Pichler, Fernando G.S.L. Brandão, Soonwon Choi, and Manuel Endres

1 California Institute of Technology, Pasadena, CA 91125, USA
2 Department of Physics, The University of Illinois at Urbana-Champaign, Urbana, Illinois 61801-3080, USA
3 Harvard University, Cambridge, MA 02138, USA
4 Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
5 Institute for Theoretical Physics, University of Innsbruck, Innsbruck A-6020, Austria
6 Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Innsbruck A-6020
7 Department of Physics, University of California, Berkeley, CA 94720, USA

Chaotic quantum many-body dynamics typically lead to relaxation of local observables1–7. In this process, known as quantum thermalization, a subregion reaches a thermal state due to quantum correlations with the remainder of the system, which acts as an intrinsic bath5,6. While the bath is generally assumed to be unobserved, modern quantum science experiments have the ability to track both subsystem and bath at a microscopic level5,6. Here, by utilizing this ability, we discover that measurement results associated with small subsystems exhibit universal random statistics following chaotic quantum many-body dynamics, a phenomenon beyond the standard paradigm of quantum thermalization. We explain these observations with an ensemble of pure states, defined via correlations with the bath, that dynamically acquires a close to random distribution. Such random ensembles play an important role in quantum information science, associated with quantum supremacy tests9–13 and device verification14–18, but typically require highly-engineered, time-dependent control for their preparation14,19–26. In contrast, our approach uncovers random ensembles naturally emerging from evolution with a time-independent Hamiltonian. As an application of this emergent randomness, we develop a benchmarking protocol which estimates the many-body fidelity during generic chaotic evolution and demonstrate it using our Rydberg quantum simulator27–29. Our work has wide ranging implications for the understanding of quantum many-body chaos and thermalization1–7 in terms of emergent randomness30 and at the same time paves the way for applications9,12,14–18,21,31 of this concept in a much wider context.

To reveal the emergence of random state ensembles, we employ a Rydberg analog quantum simulator27–29, implemented with alkaline earth atoms32–35, which provides high-fidelity preparation, evolution, and readout29 (Fig. 1, Ext. Data Fig. 1). The system is initialized with ten qubits in their ground state and subsequently evolved with a time-independent Hamiltonian, H, describing an array of strongly interacting qubits; Hamiltonian parameters are tuned to induce quantum thermalization to infinite temperature locally (Methods). After a variable evolution time, we perform site-resolved readout in the z-basis composed of the |0⟩ and |1⟩ qubit states, resulting in global measurement bitstrings z, e.g. of the form z = 1010010101. We post-process these bitstrings by bipartitioning the global system into two regions, A and B, resulting in local bitstrings zA and zB. We then regard A as the local system of interest and its complement B as an intrinsic bath.

The traditional approach to quantum thermalization studies observables in A while ignoring information about the state of the bath. In contrast, here we consider the bath explicitly and ask: how do we describe the properties of A given a specific measurement outcome in B? To this end, the most basic observable is the conditional probability p(zA|zB), which is the probability of measuring a bitstring zA, conditioned on observing a given zB. We first consider the simplest case where A consists of only a single qubit, and plot the time-resolved conditional probability of measuring the qubit in |0⟩ for all different possible results in B (Fig. 1b). At early times, all p(zA=0|zB) traces follow similar trajectories regardless of zB. However, as correlations build up between A and B the conditional probabilities for different zB diverge, moving independently in a seemingly random fashion.

We quantify this behavior by analyzing the spread in conditional probabilities during a fixed time window at a late time (shaded region in Fig. 1b). To this end, we discretize the conditional probabilities into a series of bins and count the number of occurrences in each bin; this forms a histogram P(p) with p ≡ p(zA=0|zB) (Fig. 1c). Interestingly, though individual traces display nontrivial dynamics, we find a nearly uniform histogram across different p values, indicating that the dynamics of conditional probabilities are ergodic. In fact, this uniform histogram is consistent with the most random distribution possible for a single qubit: if we consider a random

* These authors contributed equally to this work
† soonwon@mit.edu
‡ mendres@caltech.edu
Fig. 1 | Emergence of Random Statistics in Chaotic Dynamics. a, We experimentally implement chaotic quantum many-body dynamics with a Rydberg quantum simulator, described by a time-independent Hamiltonian, $H$, starting from an initial product state of ten qubits (Methods). We perform site-resolved, projective measurements yielding an outcome 0 or 1 for finding each qubit in state $|0\rangle$ or $|1\rangle$, respectively. We bipartition the system into two subsystems $A$ and $B$, for which we record bitstrings $z_A$ of length $L_A$ and bitstrings $z_B$ of length $L_B$, for the qubits in the corresponding subsystems. b, We first choose a subsystem $A$ that consists of only a single qubit ($L_A = 1$) and plot probabilities of finding the qubit in state $|0\rangle$, conditioned on specific measurement results $z_B$ in $B$, $p(z_A=0|z_B)$. Trajectories of $p(z_A=0|z_B)$ for all outcomes $z_B$ (color varying from red to blue) are shown as a function of time. We find that the different trajectories develop a significant dependence on $z_B$ at late times, exhibiting apparently chaotic behavior. c, Histogram, $P(p)$, of the measured probabilities $p(z_A=0|z_B)$ within a late-time window centered around $t \approx 0.5 \mu s$ (grey band in b). d, The experimental results are close to the flat distribution of probabilities for observing the $|0\rangle$ state, predicted from an ensemble of uniformly distributed single-qubit states on the Bloch sphere (solid line in c). e, Similarly, the experimentally obtained histograms from larger subsystems of length $L_A = 2, 3$, and 4 agree well with the predictions from uniform ensembles over correspondingly larger Hilbert space dimensions in $A$ (solid lines). Histograms in c, e use results from all contiguous subsystems of a given length $L_A$, with histograms in e additionally averaged over all possible choices of $z_A$. Error bars in c, e denote the standard error of the mean (s.e.m.).

ensemble of pure states uniformly spread over the Bloch sphere, their probabilities of being in state $|0\rangle$ follow a uniform distribution (Fig. 1d, black line in Fig. 1c).

Motivated by the similarity to a uniformly random ensemble for the single-qubit case, we study the distribution of $p(z_A|z_B)$ for larger sizes of $A$ ($L_A = 2, 3$, and 4), with correspondingly larger Hilbert space dimensions, $D_A$ (Fig. 1e). We construct similar histograms of conditional probabilities, and compare against the expectation from a $D_A$-dimensional uniformly random ensemble, given by $\rho_P = (D_A - 1)(1 - p)^D_{A-2}$ (Methods). This distribution, for increasing dimension, progressively transforms from the flat distribution (for $D_A = 2$) to the exponential Porter-Thomas distribution $D_A e^{-D_A p}$. In all cases we see good agreement between the experimentally obtained histograms and the uniformly random ensemble prediction.

To better understand this correspondence, we now consider moments characterizing the shape of the histograms (e.g. the second moment is closely related to the variance), with the $k$th moment defined as $\langle p^k \rangle = \sum_p p^k P(p)$. We plot a rescaled version of these moments, with an additional pre-factor of $D_A \cdots (D_A + k - 1)$, such that the prediction from a uniformly random ensemble becomes $k!$, independent of subsystem dimension (Fig. 2a, Ext. Data Fig. 2). Looking order-by-order, we find that the rescaled experimental moments approximately approach this analytic value of $k!$, although fluctuations become more pronounced with increasing system size in $A$; this is likely due to a decreasing number of possible outcomes from $B$ at a fixed total system size (Methods).

Furthermore, we find that not only is the convergence to $k!$ independent of subsystem sizes, but it is also independent of details of the subsystem selection (Fig. 2b). For instance, the same convergence is seen when $A$ is chosen to be at the border of the global system, at the center, or even when $A$ is discontinuous; in all cases the same behavior appears. We find similar results in fluctuations of conditional two-point correlation functions which approach a universal value of $\sqrt{4/35}$, consistent with the prediction from a uniformly random ensemble of two-qubit states (Ext. Data Fig. 3). Our observations indicate that conditional observables exhibit universal statistics linked to random state ensembles. We stress that this phenomenon is not predictable via a canonical approach to quantum thermalization that ignores system-bath correlations, and hence requires a different theoretical framework, which we present now.

The Projected State Ensemble

In a canonical approach to quantum thermalization, the state of $A$ is described by a reduced density operator, $\rho_A = \text{tr}_B \{\psi\langle\psi|\}$, where $|\psi\rangle$ is the global state of the system and $\text{tr}_B$ denotes the average over all possible quantum states in $B$. By construction, this framework cannot describe any correlations between the observables in $A$ and the state of its bath $B$, such as the conditional probabilities measured in our experiments.

To make connection with our observations and explicitly retain such correlations, we introduce a different formalism $^{89}$ where every state in $B$ is resolved using a con-
Moments

2!
3!
4!

0 0.2 0.4 0.6

Time (s) A = B =

Late-time moments

2 3 4

Moment index k

2!
3!
4!

LA = 2

LA = 3

k = 4

k = 3

k = 2

a

b

Fig. 2 | Characterization of Emergent Universal Randomness: a, We show rescaled moments of order \( k = 2, 3, \) and 4 for histograms \( P(p) \) of subsystems with length \( L_A = 2 \) and 3 (see Ext. Data Fig. 2 for different \( L_A \) cases). These moments are obtained from histograms at each time via \( \sum_p p^2 P(p) \), and characterize fluctuations in the conditional probabilities \( p(z_B|z_B) \). Values shown are rescaled by a subsystem dependent factor (see main text). After an initial transient period, both experimental (markers) and simulated (solid lines) moments saturate to \( \approx k! \), consistent with the expectation from uniformly random ensembles (dashed lines). b, Late-time moments characterized for various choices of subsystems (see panel on the right); we find a universal convergence to \( \approx k! \), independent of subsystem choice. In b, the time-averaged moment values are obtained from a time window of \( t = 0.5 - 0.7 \mu s \) and the error bars denote their standard deviations. Moments in a are obtained by averaging over subsystems with contiguous number of \( L_A \) qubits, while late-time moments in b are from the exact choices shown in the panel on the right. In a, error bars for the s.e.m. are smaller than the marker size.

### Conditional quantum state

\[
|\psi(z_B)\rangle = \frac{1}{\sqrt{p(z_B)}} P_{z_B} |\psi\rangle,
\]

where \( P_{z_B} = 1_A \otimes |z_B\rangle \) is a projective measurement operator acting only on \( B \) and \( p(z_B) \) is the probability of finding the outcome \( z_B \). The projected state \( |\psi(z_B)\rangle \) is the pure quantum state of subsystem \( A \), conditional on finding a particular outcome \( z_B \) in \( B \). The conditional probability follows as \( p(z_A|z_B) = \langle z_A | \psi(z_B) \rangle^2 \).

We call the full set of states \( |\psi(z_B)\rangle \) for all outcomes \( z_B \), together with their respective probabilities \( p(z_B) \), the **projected ensemble**. We note that ensembles of this type also enter the definition of localizable entanglement. The projected ensemble fully describes \( A \); for instance, the reduced density operator is its average (the first moment), \( \rho_A = \sum_z p(z_B) |\psi(z_B)\rangle \langle \psi(z_B)| \). Crucially, this description allows access to higher-order observables not attainable from the reduced density operator, such as the moments in Fig. 2 which can be formulated as \( \sum_{z_A,z_B} p(z_B) |\langle z_A | \psi(z_B) \rangle|^{2k} \) up to scaling.

Our observations point to such higher-order observables behaving as if the projected ensemble itself becomes approximately uniformly distributed in the Hilbert space after a sufficiently long evolution time. To test this hypothesis, we evaluate the degree of randomness in the projected ensemble by numerically computing the trace distance to quantum state \( k \)-designs. Quantum state \( k \)-designs are pure state ensembles that become progressively more complex with larger \( k \); for example, for the single-qubit case shown in Fig. 3a, state \( k \)-designs appear as increasingly uniform distributions on the Bloch sphere, and realize the uniform ensemble in the infinite \( k \) limit. A state \( k \)-design reproduces observables up to order \( k \) as if they were sampled from the uniform ensemble, e.g., the analytic result of \( k! \) for the moments shown in Fig. 2 can be predicted from a state \( k \)-design (Methods). Thus, a vanishing distance between the state \( k \)-design and the projected ensemble implies that the projected ensemble and the uniform ensemble are statistically indistinguishable for any observables up to order \( k \).

From numerical simulations, we find that the distances decrease, for \( k = 1, 2, 3, \) and 4, as a function of time (Fig. 3b) before saturating to a value exponentially small in total system size (Fig. 3c). If the observed scaling persists, the projected ensemble forms quantum state \( k \)-designs in the thermodynamic limit. Importantly, quantum thermalization to a maximally mixed state only implies convergence to a lowest-order \( 1 \)-design. For example, for a single-qubit subsystem such a reduced density operator can already be captured with only two antipodal states (Fig. 3a, left). Thus, the convergence to higher \( k \)-designs implies that the projected states of the subsystem exhibit universal random behaviors beyond what is described by quantum thermalization of subsystems. In a concurrent theory paper the same behavior is found in other systems, leading to the hypothesis that the formation of approximate quantum state \( k \)-designs in the projected ensemble is a generic feature of chaotic quantum many-body dynamics.

### Application to Fidelity Estimation

While these results are important from a fundamental science perspective, they also open avenues for practical utilization; for instance, \( k \)-designs play a role in various applications from quantum cryptography and supremacy tests to verification of quantum devices. A key question is whether the formation of approximate \( k \)-designs in projected ensembles implies such applications can be realized in a wider set of circumstances than was previously possible, potentially with reduced experimen-
Fig. 3 | Randomness Beyond Thermalization: a, Depiction of single-qubit quantum state $k$-designs on Bloch spheres; the $k$-designs are progressively more complex approximations to the uniform ensemble, which is reached for large $k$. Quantum thermalization to a maximally mixed state in a subsystem is captured by convergence to a lowest-order 1-design. b, Numerically calculated trace distances between the $L_A = 2$ projected ensemble and $k$-designs for $k = 1, 2, 3,$ and 4 (colors matching $k$-designs in a) as a function of time. Distances for all $k$ decrease initially before convergence to a finite value limited by finite system-size effects (Methods). c, These late-time distances (crosses) decrease as $\sim 1/\sqrt{D_B}$ (solid lines), where $D_B$ is the Hilbert space dimension of subsystem $B$, suggesting the emergence of $k$-designs in subsystem $A$ in the thermodynamic limit (see Ext. Data Fig. 4 for different $L_A$). The convergence to higher $k$-designs is a phenomenon beyond conventional thermalization.

ditional complexity. In particular, the generation of approximate quantum state $k$-designs typically requires a high degree of spatial and temporal control in random unitary circuits and a generic mixed-field Ising Hamiltonian (Fig. 4c, Methods). In both cases, we find that $F_c$ approximates the true fidelity even at very early times. We contrast this with another recently introduced fidelity estimator for random unitary circuits, the linear cross-entropy benchmark $F_{\text{XEB}}$, which appears to work only for deep circuits (dotted line in Fig. 4c, left). This effect is consistent with $F_{\text{XEB}}$ being designed to work for a global wavefunction sampled from a 2-design, while $F_c$ only requires the local emergence of a 2-design in the projected ensemble, which we expect to occur at earlier times. We also find that $F_c$ and $F_{\text{XEB}}$ slightly lags behind the true fidelity due to the finite time needed for an error to be detectable in a fixed measurement basis (Methods and Ext. Data Fig. 7). Finally, we note that in the random circuit case $F_c$ is smooth as it is averaged over a number of different circuit realizations, but for the case of Hamiltonian evolution $F_c$ exhibits visible fluctuations due to the utilization of only a single Hamiltonian; the amplitudes of these fluctuations decrease with increasing system size (Methods, Ext. Data Fig. 5).

We now turn to apply this protocol in experiment to benchmark the evolution of our Rydberg quantum simulator. We show results in Fig. 4d where the experimentally obtained $F_c$ (markers) decays approximately exponentially on a time-scale of 3.6(3) $\mu$s. This indicates high fidelity evolution during the time needed for buildup of significant entanglement across the two halves of the target state (blue line). To test the applicability of our protocol carefully, we also show results for a noisy model that incorporates the most dominant sources of experimental errors (Methods). For this ab initio error model, we observe that $F_c$ (red solid line) is a good estimator of the true fidelity $F$ (red dashed line). We also find that $F_c$ from the error model is in good agreement with the experimental $F_c$ and, remarkably, it also predicts the experimental global bitstring probability distribution for all times (Ext. Data Fig. 8), making it a trustworthy reference to compare against. We note that the plotted fidelity estimator is slightly modified to account for both spatial inversion symmetry and Rydberg blockade (Methods).

Finally, our benchmarking protocol allows for the in situ estimation of multiple Hamiltonian parameters simultaneously using measured bitstring samples only. Concretely, we can systematically vary the Hamiltonian implemented in simulation and monitor the resulting $F_c$ to find the maximum likely Hamiltonian parameters actually realized by experiment.
ample, the target state wavefunctions from simulation can be parameterized by the Rabi frequency, Ω, as $|\psi(t, \Omega)\rangle = e^{-i t H_R(t)/\hbar}|0\rangle \otimes |0\rangle^{10}$. When the value of Ω does not match the Rabi frequency used in the experiment, the target state $|\psi(t, \Omega)\rangle$ will have smaller overlap with the experimental state, and the fidelity estimator $F_c(t, \Omega) \approx |\langle\psi(t, \Omega)\rangle_{\text{experimental}}|^{2}$ will decay more quickly. To capture this effect in a single quantity we plot normalized, time-integrated $F_c$ as a function of the various Rydberg Hamiltonian parameters in Fig. 4e. For each parameter only a single sharp maximum emerges, showing good agreement with our recalibrated values, up to error bars coming from typical variations in the control parameters (Methods).

**Outlook**

We have experimentally observed emergent randomness in the statistics of conditional measurement results following chaotic quantum many-body evolution. Remarkably, this is a universal feature that cannot be predicted by a canonical approach to quantum thermalization; instead, our results suggest a more general description of subsystems in terms of an ensemble of pure states, which we call the projected ensemble. We numerically find that this ensemble takes on a random form, consistent with our experimental observations. Such emergent randomness implies quantum thermalization to an infinite temperature ensemble as its lowest-order prediction, but is a more general phenomenon. We have subsequently shown an application in fidelity estimation for chaotic dynamics, applicable to much shorter evolution times and with reduced experimental complexity compared to existing approaches. The concept of emergent randomness could provide a new framework for quantum thermalization, chaos, and complexity growth $^{41}$–$^{44}$, $^{41}$–$^{44}$, and it would be interesting to study the signatures of non-ergodic dynamics of integrable or localized systems $^{42}$ in the projected ensemble. Such generalizations could enable a more flexible and standardized way of performing quantum fidelity estimation in a wide variety of quantum hardware, including trapped ions $^{44}$, superconducting qubits $^{29}$, photonic systems $^{13}$, solid-state spins $^{45}$, $^{46}$, and cold atoms and molecules in optical lattices $^{42}$. Ultimately, emergent random ensembles could find a broader range of applications, including quantum supremacy tests $^{9}$–$^{13}$, certified random number generation $^{47}$ or quantum-enhanced Hamiltonian parameter estimation. In conclusion, our results could lead to a new research frontier based on the notion of emergent randomness, enabling deeper insights into complex quantum systems and their applications.
[40] Roberts, D.A. and Yoshida, B. *Chaos and complexity by design*. Journal of High Energy Physics **2017**, 121 (2017).

[41] Cotler, J., Hunter-Jones, N. and Ranard, D. *Fluctuations of subsystem entropies at late times*. arXiv:2010.11922v2 (2020).

[42] Gross, C. and Bloch, I. *Quantum simulations with ultracold atoms in optical lattices*. Science **357**, 995–1001 (2017).

[43] Turner, C.J., Michailidis, A.A., Abanin, D.A., Serbyn, M. and Papić, Z. *Weak ergodicity breaking from quantum many-body scars*. Nature Physics **14**, 745–749 (2018).

[44] Monroe, C. et al. *Programmable Quantum Simulations of Spin Systems with Trapped Ions*. arXiv:1912.07845 (2019).

[45] Zwanenburg, F.A. et al. *Silicon quantum electronics*. Reviews of Modern Physics **85**, 961–1019 (2013).

[46] Awschalom, D.D., Hanson, R., Wrachtrup, J. and Zhou, B.B. *Quantum technologies with optically interfaced solid-state spins*. Nature Photonics **12**, 516–527 (2018).

[47] Acín, A. and Masanes, L. *Certified randomness in quantum physics*. Nature **540**, 213–219 (2016).

[48] Giovannetti, V. *Quantum-Enhanced Measurements: Beating the Standard Quantum Limit*. Science **306**, 1330–1336 (2004).

[49] Madjarov, I.S. et al. *An Atomic-Array Optical Clock with Single-Atom Readout*. Physical Review X **9**, 041052 (2019).

[50] Barredo, D., de Leseleuc, S., Lienhard, V., Lahaye, T. and Browaeys, A. *An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays*. Science **354**, 1021–1023 (2016).

[51] Endres, M. et al. *Atom-by-atom assembly of defect-free one-dimensional cold atom arrays*. Science **354**, 1024–1027 (2016).

[52] Vaillant, C.L., Jones, M.P.A. and Potvliege, R.M. *Long-range Rydberg–Rydberg interactions in calcium, strontium and ytterbium*. Journal of Physics B: Atomic, Molecular and Optical Physics **45**, 135004 (2012).

[53] Harrow, A.W. *The Church of the Symmetric Subspace*. arXiv:1308.6595 (2013).

[54] Kim, H., Ikeda, T.N. and Huse, D.A. *Testing whether all eigenstates obey the eigenstate thermalization hypothesis*. Physical Review E **90**, 052105 (2014).

[55] Mølmer, K., Castin, Y. and Dalibard, J. *Monte Carlo wave-function method in quantum optics*. Journal of the Optical Society of America B **10**, 524 (1993).
**Ext. Data Fig. 1 | Experimental system and parameter feedback:**

**a,** Illustration of a Rydberg quantum simulator consisting of strontium-88 atoms trapped in optical tweezers (red funnels). All atoms are driven by a global transverse control field (purple horizontal beam) at a Rabi frequency $\Omega$ and a detuning $\Delta$ (right panel). The interaction strength is given as $C_6/R^6_{ij}$ with an interaction constant $C_6$ and atomic separations $R_{ij}$ between two atoms at site $i$ and $j$.

**b,** Schematic of the experimental feedback scheme. We automatically interleave data taking with feedback to global control parameters and systematic variables through a home-built control architecture (Methods); in particular, we feedback to the clock laser frequency (to maintain optimal state preparation fidelity), the Rydberg laser alignment, the Rydberg detuning $\Delta$, and the Rabi frequency $\Omega$.

**c,** Example of the interleaved automatic Rabi frequency stabilization over the course of $\approx 20$ hours with no human intervention. Feedback is comprised of performing single-atom Rabi oscillations, fitting the observed Rabi frequency, and updating the laser amplitude, rather than simply stabilizing the laser amplitude against a photodiode reference. While the Rabi frequency setpoint (orange squares) changes over the course of the sequence (due to long-time instabilities like temperature drifts), the measured Rabi frequency (blue circles) stays constant to within $\leq 0.3\%$, with a standard deviation of $0.15\%$. This same stability is seen over the course of multiple days with nearly continuous experimental uptime.
Ext. Data Fig. 2 | Conditional probability histograms, moments, and experiment-theory correlations: As shown in Figs. 1, 2, and 4 the same kind of characterization is performed for projected ensembles of different contiguous subsystem lengths \( L_A \). The grey zoom-in triangles in each subplot denote the approximate time points chosen for the characterization; the early-time histogram is a time-average over the same window as in Fig. 1. We note that at late times \( t > 1 \mu s \) the experimental moments decay as decoherence processes gradually suppress fluctuations in the conditional probabilities (time scale is logarithmic after the break). The red dashed lines in the late-time histograms (top right corner of each subpanel) indicate the probability values expected for a maximally mixed state of a given subsystem \( A \), i.e., \( p = 1/D_A \) where \( D_A \) is the Hilbert space dimension of \( A \).
Ext. Data Fig. 3 | Fluctuations of conditional two-point correlations in projected state ensembles:  

Two-point correlations based on conditional projected states of two qubits separated at distance $d$.  

b, Fluctuations of two-point correlation functions of conditional probabilities, $\sigma_{\text{corr}}(z_1, z_2)$, characterized with a ten-qubit Rydberg-atom array simulator (Methods, see Eq. (11)). Experimental (markers) and numerical (solid lines) results increase and saturate to near the theoretical prediction of $\sigma_{\text{corr}} = \sqrt{\frac{4}{35}}$ from the uniformly random ensemble. At late times $t > 1 \mu s$, the universal feature gradually vanishes due to accumulated errors and decoherence processes in the experiment (time scale is logarithmic after the break). Colors and markers used are: $d = 2$, blue circle; $d = 4$, orange square; $d = 6$, purple triangle; $d = 8$, red triangle.  

c, Individual plots for the four different pair separations shown in b.
Ext. Data Fig. 4  |  Trace distance between the projected ensemble and quantum state $k$-designs: **a**, **b**, System size scaling of late-time trace distances from three different subsystem sizes of $L_A = 1, 2,$ and $3$. For fixed subsystem $A$ size, we increase the Hilbert space dimension of outer subsystem $B$, $D_B$, and compute trace distances to the quantum state $k$-designs for four different orders, $k = 1$ (orange), 2 (red), 3 (purple), and 4 (blue). Sampling weighting factors of $p(z_B)$ and $p^k(z_B)/\sum_{z_B} p^k(z_B)$ are used in **a** and **b**, respectively, in the construction of a $k$th-order projected ensemble. See Methods for simulation details.

Ext. Data Fig. 5  |  Typicality analysis of $F_c$: **a**, Normalized fidelity estimation errors as a function of interrogation time $\tau$ after a local, instantaneous error occurs at time $t_0$ (see the inset of **a**). The error strength is chosen such that the many-body overlap is reduced to $\sim 0.5$. To investigate typical statistical errors in fidelity estimation, we simulate an ensemble of chaotic Hamiltonians at a fixed system size of $N = 16$ for two error occurrence times of $t_0 = 0$ (left) and $t_0 = 5$ (right) (see Methods). The mean and standard deviation of the normalized fidelity estimation errors are plotted as the solid line and grey band, respectively. **b**, Standard deviation of normalized fidelity estimation errors, characterized at two different post-error interrogation times ($\tau = 8$; open markers, $\tau = 30$; closed markers) as a function of the total Hilbert space dimension $D$ for the two error occurrence times of $t_0 = 0$ (left) and $t_0 = 5$ (right). At the late interrogation time of $\tau = 30$, we find that the typical error from the distinct choice of a Hamiltonian follows a scaling of $\sigma[(F_c - F)/F] \sim 1/\sqrt{D}$ (solid lines), as also found for deep random unitary circuits (Methods). In **b**, error bars denote the standard deviation of $\sigma[(F_c - F)/F]$ in a $[\tau - 1, \tau + 1]$ window.
Ext. Data Fig. 6 | Finite sampling analysis for $F_c$. a, b, Statistical fluctuations of the fidelity estimator $F_c(M)$ evaluated with a finite number of $M$ bitstring samples, $\sigma_N(M, \tau)$, at early (a, $\tau = 1$) and late (b, $\tau = 20$) times after an error has occurred at time $t_0$, plotted for three different system sizes, $N = 10, 15, 20$ (darker colors with increasing $N$). A generic mixed-field Ising model is employed with an initial state error ($t_0 = 0$) that results in a many-body overlap of 0.5 (Methods). In a, b, the dotted lines are guides-to-eye showing a $1/\sqrt{M}$ scaling. c, Time- and system size-dependent prefactor, $A(N, \tau)$, of the statistical fluctuation, $\sigma_N(M, \tau) = A(N, \tau)/\sqrt{M}$. The prefactor, $A(N, \tau)$, is plotted as a function of $N$ for $\tau = 1$ (blue), 2 (purple), 5 (green), and 20 (red). The solid lines are phenomenological fits of $A(N, \tau) = A_0 r(\tau)^{N-N_0}$ with $A_0 = 1.4$ and $N_0 = 7$. Inset: $r(\tau)$ is plotted for the four different post-error times $\tau$ with two error occurrence times $t_0$: $t_0 = 0$; circle and $t_0 = 5$; star. We observe that for $t_0 = 0$, $r(\tau)$ is close to 1 at early times and approaches $r(\tau) \approx 1$ over time, while for $t_0 = 5$, $r(\tau) \approx 1$ for all times.

Ext. Data Fig. 7 | Systematic delay in fidelity estimation performed with a fixed measurement basis: Systematic error in the fidelity estimator $F_c$ (red) and the linear cross-entropy benchmark $F_{\text{XEB}}$ (blue) relative to the true fidelity $F$ for a ten-qubit random quantum circuit. Noisy quantum dynamics are simulated with random, local bit-flip and phase-flip errors at four different single-qubit error rate per circuit depth, $\gamma_{\text{err}} = 0.0025$ (circle), 0.005 (square), 0.0075 (diamond), and 0.01 (triangle). The dashed grey lines are the phenomenological error scaling of $\exp(N\gamma_{\text{err}} \tau_d)$ with an $\gamma_{\text{err}}$-independent delay time of $\tau_d = 0.8$ and $N = 10$. The saturation of both $F_c/F$ and $F_{\text{XEB}}/F$ at large circuit depth implies that at late times there is a delayed response in fidelity estimation performed with a fixed measurement basis, i.e., $F = \exp(-N\gamma_{\text{err}} t)$, but $F_c = \exp(-N\gamma_{\text{err}} (t - \tau_d))$ (Methods).

Ext. Data Fig. 8 | Quantitative evaluation of ab initio error model: Kullback-Leibler (KL) divergence of measurement outcome probability distribution between the ten-qubit Rydberg experiment and an error-free model (blue square) and an ab initio error model (red circle) (Methods). When the experimental data is benchmarked against an error-free model, the KL divergence increases over time since the presence of decoherence effects and control imperfections are not taken into account in the error-free model. In the case of the error model, however, we find a significant reduction in the KL divergence between the experiment and error model at all times. Error bars denote the s.e.m.
METHODS

I. DESCRIPTION OF THE EXPERIMENT

The details of our experiment have been summarized previously\textsuperscript{29,33,35,49}, in brief, we use an array of 35 optical tweezers to trap \( \approx 18 \) individual strontium-88 atoms. Initially in the \( 5s^2 \, 1S_0 \) state, atoms are cooled on the narrow-line \( 5s^2 \, 1S_0 \leftrightarrow 5s5p \, 3P_1 \) (689 nm) transition close to their motional ground state, with an average transverse occupation number of \( \langle n_i \rangle \approx 0.3 \) (corresponding to \( \approx 3 \) \( \mu \)K). For all data shown, we rearrange the initially stochastically filled array to a defect-free array\textsuperscript{50,51} of ten atoms spaced by 3.3 \( \mu \)m, discarding extras. Atoms are initialized to the \( 5s5p \, 3P_0 \) (698 nm) \textit{clock state} through a combination of coherent drive and incoherent pumping, for a total preparation fidelity of 0.997(1). We treat the clock state as a metastable qubit ground state, \( |0\rangle \), and subsequently drive to the \( 5s61s \, 3S_1, m_I=0 \) (317 nm) Rydberg state, \( |1\rangle \). Following Hamiltonian evolution, state readout is performed using the auto-ionizing transition \( 5s61s \, 3S_1, m_I=0 \leftrightarrow 5p_{3/2}61s_{1/2} \) (408 nm, \( J=1, m_J=\pm 1 \)) which rapidly ionizes atoms in the Rydberg state with high fidelity (\( \approx 0.999 \)), leaving them dark to our fluorescent imaging. Atoms in the clock state are pumped into the imaging cycle, allowing us to directly map atomic fluorescence to qubit state.

The Hamiltonian of this system is well approximated by

\[
H = \Omega \sum_i S_i^x - \Delta \sum_i n_i + \frac{C_6}{a^6} \sum_{i>j} n_i n_j / |i-j|^6
\]

which describes a set of interacting two-level systems, labeled by site indices \( i \) and \( j \), driven by a laser with Rabi frequency \( \Omega \) and detuning \( \Delta \). The interaction strength is determined by the \( C_6 \) coefficient and the lattice spacing \( a \). Operators are \( S_i^x = (|1\rangle_i \langle 0|_i + |0\rangle_i \langle 1|_i)/2 \) and \( n_i = |1\rangle_i \langle 1|_i \), where \( |0\rangle_i \) and \( |1\rangle_i \) denote the electronic ground and Rydberg states at site \( i \), respectively. For measurements observing the emergence of random ensembles (Figs. 1 and 2, Ext. Data Figs. 2 and 3), we use \( \Omega/2\pi = 4.7 \) MHz, \( \Delta/2\pi = 0.9 \) MHz, \( a = 3.3 \) \( \mu \)m, and an experimentally measured interaction coefficient of \( C_6 = 126(2) \) GHz \( \mu \)m\(^6\). Under this condition, we confirm numerically that the initial all-zero state rapidly thermalizes to an infinite-temperature thermal ensemble locally within the constrained subspace where no two adjacent atoms are simultaneously in the Rydberg state\textsuperscript{27-29}. Benchmarking measurements (Fig. 4, Ext. Data Fig. 8) are performed with \( \Omega/2\pi = 6.4 \) MHz, \( \Delta/2\pi = 0 \) MHz. We note that the \( C_6 \) value is independently estimated by measuring the interaction-induced energy shift \( V = C_6 / R^6 \) at various distances \( R \) between two atoms prepared in the Rydberg state. The measured \( C_6 \) is roughly a factor of two smaller than expected from theoretical quantum defects\textsuperscript{52}, though this discrepancy is attributable to a modest (\( \approx 1 \) V/cm) DC electric field, which appears stable over time, but which we do not actively cancel.

As the experimental data shown throughout the main text requires both high statistics (taken over the course of multiple days) and very fine parameter control, we periodically perform automatic feedback to several experimental parameters using a home-built control architecture. Specifically these are: 1) the clock state resonance frequency to ensure maximal preparation fidelity, 2) the Rydberg laser beam alignment, 3) the Rydberg resonance frequency, and 4) the Rydberg Rabi frequency. For the clock frequency, we apply a \( \pi \)-pulse on the clock transition to identify the resonance and perform state-resolved readout by ejecting all ground state atoms from the trap with an intense pulse of light on the \( 5s^2 \, 1S_0 \leftrightarrow 5s5p \, 1P_1 \) (461 nm) transition\textsuperscript{29}.

For the Rydberg alignment, detuning, and Rabi frequency, we rearrange the array to \( \approx 8 \) \textit{non-interacting} atoms spaced by 13.3 \( \mu \)m. During alignment we raster the Rydberg beam across the array sampling different position-dependent Rabi frequencies, and thus evolving to different position-dependent phases. We compare the resultant signal across all positions to a simulation to identify the point of furthest phase, and thus maximal intensity. For the Rydberg detuning, we measure the resonance condition at \( \Omega t = 13\pi \) in order to narrow the resonance feature. For the Rabi frequency, we take a series of time points between \( 13\pi < \Omega t < 17\pi \), and fit the resulting Rabi oscillations. After each feedback experiment, the relevant parameter is automatically updated for subsequent measurements (Ext. Data Fig. 1).

II. DATA ANALYSIS

Our state readout is described in detail in Ref.\textsuperscript{29}; it features single-site detection which discriminates atoms in the clock state, \( |0\rangle \), versus the Rydberg state, \( |1\rangle \), through a combination of fluorescent imaging and Rydberg autoionization. We take a total of three images: 1) after the array is initially loaded to perform rearrangement, 2) after the rearrangement is completed to verify the initial state is correct, 3) after the sequence has finished. We post-select for image triplets where the proper rearrangement pattern is visible in image (2), and calculate the survival of each atom by comparing site occupations in image (2) to image (3). This array of survival signals is then converted into the qubit basis. For instance, in typical experiments where atoms are rearranged into defect-free arrays of ten atoms, we calculate the binary survivals for each atom, and then make the mapping ‘atom survived’\( \rightarrow |0\rangle \) and ‘atom did not survive’\( \rightarrow |1\rangle \), yielding a ten letter bitstring of the qubit states. After taking many shots we accrue an ensemble of such bitstrings, \{ \}. For randomness measurements, a total of \( \approx 120000 \) shots are used (\( \approx 30000 \) shots per time point), while for benchmarking measurements a total of \( \approx 64000 \) shots are used (\( \approx 4500 \) shots per time point).
Our system Hamiltonian is naturally stratified into a number of energetically widely spaced sectors due to Rydberg blockade\textsuperscript{27–29}. In particular, the nearest-neighbor interaction is approximately greater than the next largest energy scale, so cases where neighboring pairs of atoms are both excited to the Rydberg state are greatly suppressed. In practice we find \( \approx 99\% \) of all experimental bitstrings are in the no-blockade energy sector at short times (\( t < 1 \mu s \)) but the no-blockade probability starts to decrease at late times (\( t > 1 \mu s \)) due to experimental imperfections - we refine \( \{ z \} \) by discarding all realizations not in this sector.

For calculations involving conditional probabilities, we bipartition each element of \( \{ z \} \) into subsystems \( A \) and \( B \) with bitstrings \( z_A \) and \( z_B \) respectively. When considering the statistics of conditional probabilities, we note that the blockade interaction can reduce the dimensionality of the Hilbert space of subsystem \( A \) if the boundary qubits in \( B \) are in the Rydberg state. To isolate a set of conditional states having the same Hilbert space dimension, for a given choice of subsystem \( A \) and \( B \), we only consider bitstrings \( z_A \) and \( z_B \) if the qubits in \( B \) bordering \( A \) are in the \( |0 \rangle \) state.

III. PROJECTED ENSEMBLE AND COMPARISON TO QUANTUM STATE 
\( k \)-DESIGNS

To characterize the randomness of the projected ensemble, we compare it to the so-called Haar-random ensemble which represents an ensemble of uniformly random pure states defined in the Hilbert space of subsystem \( A \). Specifically, the similarity between the two ensembles can be characterized via the \( \ell_1 \)-norm that measures a ‘distance’ in operator space (referred to as the ‘trace distance’ in the main text),

\[
\ell_1^{(k)} = \| \rho_A^{(k)} - \rho_{\text{Haar}}^{(k)} \|_1.
\]  

(4)

Here \( \| \cdot \|_1 \) represents the absolute sum of singular values, \( \rho_A^{(k)} \) is the \( k \)th moment of the projected ensemble consisting of local states \( |\psi(z_B)\rangle \) in subsystem \( A \), and \( \rho_{\text{Haar}}^{(k)} \) is the \( k \)th moment of the Haar-random ensemble

\[
\rho_{\text{Haar}}^{(k)} = \int_{\text{Haar}(D_A)} d\psi \ |\psi\rangle \langle \psi| \otimes ^k,
\]  

(5)

and \( \rho_{\text{Haar}} \) is the \( k \)th moment of the Haar-random ensemble

\[
\rho_{\text{Haar}}^{(k)} = \int_{\text{Haar}(D_A)} d\psi \ |\psi\rangle \langle \psi| \otimes ^k.
\]  

(6)

From the Schur-Weyl duality, Eq. (6) can be simplified to

\[
\rho_{\text{Haar}}^{(k)} = \sum_{\pi \in S_k} \text{Perm}_{H_A^{\otimes k}}^{(k)}(\pi) \nu^{(k)}_{H_A^{\otimes k}}(\pi) D_A(D_A + 1) \cdots (D_A + k - 1),
\]  

(7)

where \( \text{Perm}_{H_A^{\otimes k}}(\pi) \) permutes the \( k \) copies of the Hilbert space \( H_A \) of subsystem \( A \) according to a member, \( \pi \), in the permutation group of \( k \) elements, and \( D_A \) is the dimension of \( H_A \).

As shown in Fig. 3 and Ext. Data Fig. 4, we find that for various subsystem lengths, \( L_A \), the \( \ell_1 \)-norm distance decreases for all orders \( k = 1, 2, 3 \), and \( 4 \) with increasing Hilbert space dimension, \( D_B \), of subsystem \( B \). Non-zero \( \ell_1 \)-norm distances, as well as statistical fluctuations observed in the moments at finite \( D_B \), are associated with a finite sample size effect in the constructed projected ensemble. For a finite-size Haar-random ensemble, the trace distance to the uniform ensemble scales as \( \approx 1/\sqrt{D_B} \), consistent with the scaling result of the projected ensemble\textsuperscript{30}. These results are obtained with \( \Omega/2\pi = 4.7 \text{ MHz} \), \( \Delta/2\pi = 0.5 \text{ MHz} \), \( C_0 = 126 \text{ GHz } \mu\text{m}^6 \), and \( a = 3.3 \text{ \mu m} \); the discrepancy between this \( \Delta \) and that imposed in experiment does not result in a significantly different distance to the \( k \)-designs for the experimental system size of ten qubits.

IV. MOMENTS AND TWO-POINT CORRELATIONS OF UNIFORMLY RANDOM ENSEMBLES

Moments. Suppose we have a Haar-random state \(|\psi\rangle \) in a \( D_A \)-dimensional Hilbert space, which we measure in the standard \( z \)-basis. Then the probability of measuring a bitstring \( z \) is \( r_\psi(z) = |\langle z |\psi\rangle|^2 \). We can compute the expectation value of the \( k \)th moment of \( r_\psi(z) \) over the Haar-random ensemble as

\[
E_{\psi \sim \text{Haar}(D_A)}[r_\psi(z)^k] = \frac{k!}{D_A(D_A + 1) \cdots (D_A + k - 1)}.
\]  

(8)

Using these moments, we can construct a probability density function

\[
P_H(p) \equiv E_{\psi \sim \text{Haar}(D_A)}[\delta(r_\psi(z) - p)]
\]  

(9)

where \( \delta \) is the Dirac delta function, yielding \( r_\psi(z) = p \) over the Haar-random ensemble. Since \( J_0 dp P_H(p) p^k = E_{\psi \sim \text{Haar}(D_A)}[r_\psi(z)^k] = 1/(D_A^k + k - 1) \) for \( k = 0, 1, 2, \ldots \), we find

\[
P_H(p) = (D_A - 1)(1 - p)^{D_A - 2}.
\]  

(10)

Correlation functions. In Ext. Data Fig. 3, we measure the square root of a weighted average of squared two-point correlators in the \( z \)-basis using the projected ensemble of a two-qubit subsystem, given as

\[
\sigma_{\text{corr}}(z_1, z_2) = \sqrt{\sum_{z_B} p(z_B) C(z_1, z_2 | z_B)^2}
\]  

(11)

where

\[
C(z_1, z_2 | z_B) = \langle \psi(z_B) | Z_1 Z_2 | \psi(z_B) \rangle - \langle \psi(z_B) | Z_1 | \psi(z_B) \rangle \langle \psi(z_B) | Z_2 | \psi(z_B) \rangle.
\]  

(12)
Here, $Z_i$ is the Pauli-$Z$ operator at site $i$, $p(z_B)$ is the probability of observing a bitstring $z_B$ from outer subsystem $B$, and the expectation values on the right-hand side are computed with the projected two-qubit state $|\psi(z_B)\rangle$. To compare the projected ensemble result with a Haar-random ensemble, we use Eq. (7) for $D_A = 4$ to compute the analytical value of such a correlator for the Haar-random ensemble:

$$\sqrt{\mathbb{E}_{\psi \sim \text{Haar}(4)} \left[(|\langle \psi | Z_1 Z_2 | \psi \rangle - \langle \psi | Z_1 | \psi \rangle \langle \psi | Z_2 | \psi \rangle|^2\right]} = \sqrt{\frac{4}{35}}.$$  

(13)

V. RELATION BETWEEN $F_c$ AND MANY-BODY FIDELITY $F$

Our fidelity estimator $F_c$ (Eq. (2)) can be understood by expressing the best string probabilities for ideal and noisy evolutions, $p_0(z)$ and $p(z)$ respectively, in terms of conditional and marginal probabilities as

$$p_0(z) = p_0(z_A z_B) p_0(z_B)$$  \hspace{1cm} (14)
$$p(z) = p(z_A|z_B) p(z_B),$$  \hspace{1cm} (15)

for complementary subsystems $A$ and $B$. We consider the simplest case of a single local error $V$ occurring at time $t_0$ during time-evolution, and assume that the time-evolved error operator, $V(\tau) = U(\tau) V U(\tau)^\dagger$, is supported within subsystem $A$. Here $\tau = t - t_0$ is the time past the occurrence of the error and $U(\tau)$ is the time-evolution operator from $t_0$ to $t$. This implies that the measurement outcome in $B$ is not affected by the error, giving $p(z) = p(z_A|z_B) p_0(z_B)$ because $p(z_B) = p_0(z_B)$. Under these conditions, we can rewrite $F_c$ as

$$F_c = \frac{2 \sum_z p_0(z) p(z)}{\sum_z p_0(z) - 1}$$  \hspace{1cm} (16)
$$= \frac{2 \sum_{z_B} p_0^2(z_B) \sum_z p(z_A|z_B) p_0(z_A|z_B) - 1}{\sum_{z_B} p_0^2(z_B) \sum_z p_0^2(z_A|z_B)}$$  \hspace{1cm} (17)
$$\approx \frac{\sum_{z_B} p_0^2(z_B) F_{\text{XEB}}(z_B) + 1)}{\sum_{z_B} p_0^2(z_B)} - 1$$  \hspace{1cm} (18)
$$= \sum_{z_B} q(z_B) F_{\text{XEB}}(z_B)$$  \hspace{1cm} (19)

where $q(z_B) = \frac{p_0^2(z_B)}{\sum_{z_B} p_0^2(z_B)}$ and $F_{\text{XEB}}(z_B) = (D_A + 1) \sum_z p(z_A|z_B) p_0(z_A|z_B) - 1$  \hspace{1cm} (20)

is the $z_B$-dependent, linear cross-entropy benchmark\textsuperscript{69} in subsystem $A$, and $D_A$ is the Hilbert space dimension of $A$. From Eq. (17) to Eq. (18), we used the second-order moment of the projected ensemble in an error-free case

$$\frac{1}{D_A} \sum_{z_A} p_0^2(z_A|z_B) \approx \frac{2!}{D_A(D_A + 1)}$$  \hspace{1cm} (21)

based on our experimental and numerical observations of emergent local randomness during chaotic quantum dynamics (see Fig. 2 and Eq. (8)).

The validity of the relation $F_c \approx F$ can be analytically understood based on the assumption that the projected ensemble of $|\psi(z_B)\rangle$ approximately forms a quantum state 2-design. To see this explicitly, we consider

$$\sum_{z_B} q(z_B) p_0(z_A|z_B) p_0(z_A|z_B)$$
$$= \sum_{z_B} q(z_B) \langle \psi(z_B)| P_{z_A} \psi(z_B)| \langle \psi(z_B)| V(\tau)^{\dagger} P_{z_A} V(\tau) |\psi(z_B)\rangle$$
$$= \text{tr} \left\{ (P_{z_A} \otimes V(\tau)^{\dagger} P_{z_A} V(\tau)) \cdot \sum_{z_B} q(z_B)(|\psi(z_B)\rangle \langle \psi(z_B)|)^{\otimes 2} \right\}$$
$$\approx \frac{\text{tr} \left\{ (P_{z_A} \otimes V(\tau)^{\dagger} P_{z_A} V(\tau)) \cdot (1 + S_A) \right\}}{D_A(D_A + 1)}$$  \hspace{1cm} (22)

where $P_{z_A} = |z_A\rangle \langle z_A|$ is the $z$-basis projector onto a specific bitstring $z_A$ in $A$, $q(z_B)$ is the probability weighting factor, 1 is the identity operator, and $S_A$ is the swap operator acting on subsystem $A$ for the duplicated Hilbert space $\mathcal{H}_A^{\otimes 2}$. In order to obtain the fourth line, we used

$$\sum_{z_B} q(z_B)(|\psi(z_B)\rangle \langle \psi(z_B)|)^{\otimes 2} \approx \frac{1 + S_A}{D_A(D_A + 1)},$$  \hspace{1cm} (23)

where the right-hand side is due to the projected ensemble forming an approximate quantum state 2-design (see Eq. (7)). We note that the weighting factors $q(z_B)$ are different than those used for the majority of the manuscript; however, we numerically find that approximate 2-designs form regardless of which weighting factor is used (Ext. Data Fig. 4).

Inserting Eq. (22) into Eq. (19), we obtain

$$F_c \approx \frac{1}{D_A} \sum_{z_A} \langle z_A | V(\tau) | z_A \rangle^2$$
$$= \frac{1}{D} \sum_z \langle z | V(\tau) | z \rangle^2,$$  \hspace{1cm} (24)

where the equality on the second line holds because one can always multiply the identity $\frac{1}{D^2} \sum_{z_B} \langle z_B | z_B \rangle^2 = 1$ with the Hilbert space dimension of the complement $D_B = D/D_A$ with $D$ being the Hilbert space dimension of the entire system.

The relation in Eq. (24) explains how $F_c$ estimates the many-body fidelity with a good precision. The right-hand side of Eq. (24) describes the return probability of $V(\tau)$ (also known as Loschmidt echo) averaged over all possible initial states in the fixed measurement basis. Under chaotic time evolution, the propagated error operator $V(\tau)$ becomes scrambled, and it is exponentially unlikely in the size of $A$ that a computational state remains unchanged.
Therefore, non-vanishing contributions to $F_c$ arise only when the error operator is partly proportional to the identity, e.g., $V(\tau) = c_0 1 + \sum_s c_s(\tau)\sigma_s$ with $c_0 \neq 0$, where $s$ enumerates over all possible Pauli strings supported in $A$. In such a case, $F_c \approx |c_0|^2$ approximates the probability that $V$ did not affect the many-body wavefunction, hence $F_c \approx F$. This statement becomes exact if the local qubit on which the error occurs is maximally entangled with the rest of the system at the time of the error. Our analysis can be straightforwardly generalized to more than one error, either located nearby or distant, as long as their joint support $A$ leads to a random ensemble approximately close to the state 2-design.

Finally, we comment on the conditions in which $F_c$ may significantly deviate from $F$. If $V$ is diagonal in the measurement basis, e.g., dephasing error along the $z$ axis, and if the error occurs shortly before the bitstring measurements, the return probability in Eq. (24) will be close to unity despite that the many-body fidelity may be decreased significantly. Our method can fail in this special case. However, if $F_c$ is evaluated after some delay time from the error, then $V(\tau)$ becomes scrambled in the operator basis, and $F$ can be approximately estimated. In other words, even in the case of the diagonal errors, our formula becomes valid after a finite delay time $\tau_d$ (Ext. Data Fig. 7).

VI. BENCHMARKING OF RANDOM UNITARY CIRCUIT AND CHAOTIC HAMILTONIAN

In Fig. 4c, we present simulated benchmarking results for both digital and analogue quantum systems based on a random unitary circuit (RUC) and a generic mixed-field Ising model exhibiting chaotic behaviors, respectively; below we describe details for simulations of their quantum dynamics in the presence of errors.

**Random unitary circuit:** The dynamics of a one-dimensional RUC are simulated using random two-qubit unitary gates, $U$, sampled from the Haar measure. Specifically, the time evolution of an $N$-qubit system starting from the all-zero initial state, $|0\rangle^\otimes N$, can be described as

$$|\psi(t)\rangle = U_0U_{-1}\cdots U_2U_1|0\rangle^\otimes N, \quad (25)$$

where $U_{odd} = \{U_1, U_3, \cdots\}$ and $U_{even} = \{U_2, U_4, \cdots\}$ are the odd- and even-time unitaries modeled as

$$U_{odd} = \prod_{i=1}^{N/2} U_{2i-1,2i}, \quad U_{even} = \prod_{i=1}^{N/2} U_{2i,2i+1} \quad (26)$$

with periodic boundary condition, and $t$ is the circuit depth. Note that at each circuit depth $t$, we randomly sample two-qubit random unitaries $U_{ij}$ for qubits at site $i$ and $j$ to realize a many-body unitary $U$ giving rise to chaotic dynamics.

**Chaotic Hamiltonian:** We simulate a mixed-field Ising Hamiltonian that leads to thermalization of the all-zero initial state to infinite temperature:

$$H_{QIMF}/\hbar = 2\pi \left[ \sum_{i=1}^{N} 0.22S_i^x + 0.25S_i^y + \sum_{i=1}^{N-1} 0.55S_i^x S_{i+1}^x \right], \quad (27)$$

where $S_i^{x,y} \approx c$ are the spin-1/2 operators at site $i$. The Hamiltonian coefficients are adapted from Ref. 54, and are consistent with those used in an accompanying theory paper 30; however, coefficients are globally rescaled such that the early-time growth rate of half-cut entanglement entropy is matched to that of RUC dynamics. The many-body quantum state at time $t$ is obtained by solving for $|\psi(t)\rangle = e^{-iH_{QIMF}/\hbar}|0\rangle^\otimes N$.

**Noisy time evolution:** To emulate noisy quantum devices, we employ a stochastic evolution method with a simple noise model that considers local bit-flip and phase-flip errors represented by the Pauli operators $\sigma^{x,y,z}$. These local errors are stochastically applied to individual qubits at a single-qubit error rate of $\gamma_{err}$ per unit time (equivalent to a single circuit depth); the particular error applied is chosen randomly from the set of $\sigma^{x,y,z}$. To compare the RUC and Hamiltonian on an equal footing, we run noisy simulations at the same $\gamma_{err}$ between the two cases. Repeating the noisy dynamics simulations more than $\sim 10000$ times at a fixed $\gamma_{err}$, we obtain good approximations of the density matrices, $\rho_{RUC}$ and $\rho_{Ham}$, from which we extract the true fidelity $F = \langle \psi | \rho | \psi \rangle$ as well as our fidelity estimator $F_c$ and the linear cross-entropy benchmark

$$F_{XEB} = (D + 1) \sum_z p_0(z)p(z) - 1, \quad (28)$$

where $D = 2^N$ is the Hilbert space dimension of a whole $N$-qubit system, $z$ is the global bitstring of length $N$, and $p_0$ and $p$ are the bitstring probabilities in the z-basis without and with errors, respectively.

VII. TYPICALITY OF BENCHMARKING AND ASSOCIATED ERROR SCALING

The shown benchmarking results for RUCs are averaged over an ensemble of RUCs realized via sampling of two-qubit unitaries $U$ (Fig. 4c, Ext. Data. Fig. 7). For $F_{XEB}$, a single choice of an RUC out of such an ensemble already yields accurate results in the sense that, in the limit of large circuit depth, the error from using only a single RUC scales as $1/\sqrt{D}$ with $D$ being the Hilbert space dimension of a global system 10. We now investigate if a similar typicality scaling holds for benchmarking of Hamiltonian dynamics (Fig. 4c, d). To this end, we consider an ensemble of Hamiltonians generated from $H_{QIMF}$ by adding a series of site-dependent transverse fields:

$$H_{QIMF}(\mathbf{J})/\hbar = H_{QIMF}/\hbar + 2\pi \sum_{i=1}^{N} J_i S_i^x, \quad (29)$$
where $H_{\text{QIMF}}(\mathbf{J})$ is a Hamiltonian in the ensemble parameterized by site-dependent transverse fields, $\mathbf{J} = (J_1, \ldots, J_i, \ldots, J_N)$. We sample the $J_i$ uniformly from $[-0.5, 0.5]$ with a restriction, $\sum_{i=1}^{N} J_i = 0$, imposed so all target states thermalize to effective infinite temperature. With the $H_{\text{QIMF}}(\mathbf{J})$, we generate an ensemble of parametrized target states, $|\psi(t, \mathbf{J})\rangle = e^{-iH_{\text{QIMF}}(\mathbf{J})t/\hbar}|\psi\rangle \otimes |\psi\rangle \otimes |\psi\rangle \otimes \cdots$, where we consider a single error occurring at a time $t_0$, induced by a single-site rotation around the $x$-axis, on top of otherwise error-free evolution. We denote the erroneous states as $|\psi(t, \mathbf{J})\rangle_e$. The rotation angle is fixed across $\mathbf{J}$-choices with an amplitude that yields a many-body overlap $F(\mathbf{J}) = |\langle \psi(t, \mathbf{J})|\psi(t, \mathbf{J})\rangle_e|^2 \approx 0.5$. We now estimate this many-body overlap using the bitstring probabilities for these two states, yielding a time- and $\mathbf{J}$-dependent estimator $F_c(t, \mathbf{J})$. To quantify relative fluctuations, we introduce the relative difference $d(\tau, \mathbf{J}) = (F_c(t_0 + \tau, \mathbf{J}) - F(\mathbf{J}))/F(\mathbf{J})$, where $\tau = t - t_0$ is the evolution time after the error. We show the ensemble average of $d(\tau, \mathbf{J})$ as a function of $\tau$ (Ext. Data. Fig. 5a) for two different choices of $t_0$: $t_0 = 0$ and $t_0 = 5$. We observe that the ensemble average settles quickly to zero within a $10^{-2}$ level; however, we find fluctuations around this mean value arising from different choices of $\mathbf{J}$. To quantify these fluctuations, we evaluate the standard deviation of $d(\tau, \mathbf{J})$ over $\mathbf{J}$ and find that it decreases with system size (Ext. Data. Fig. 5b).

At late times $\tau$, we find a $1/\sqrt{D}$ scaling similar to the case of RUCs. These results imply that, at sufficiently late time, our method becomes increasingly more precise in the limit of large system size for a single typical Hamiltonian evolution.

VIII. STATISTICAL ERROR SCALING FROM A FINITE NUMBER OF BITSTRING SAMPLES

In actual experiments we are never able to find the exact value of $F_c$, but are instead left to approximate it from a finite number of bitstring samples. Specifically, let us suppose that a total of $M$ bitstrings $\{z_1, z_2, \ldots, z_M\}$ are sampled at a fixed time for a given system size $N$. Then using the $M$ bitstring samples, $F_c$ is estimated as

$$F_c^{(M)} = 2 \frac{\sum_{i=1}^{M} p_0(z_i)}{\sum_{i=1}^{M} p_0(z_i)^2} - 1. \quad (30)$$

Note that the sum in the numerator is performed over the sampled bitstrings, with their corresponding probabilities evaluated using the theoretical bitstring probability distribution $p_0$, similar to sampling of $F_{\text{XEB}}$. We use the superscript $M$ on $F_c$ to denote that the fidelity estimation is based on a finite number of $M$ samples, with $\lim_{M \to \infty} F_c^{(M)} = F_c$.

We quantify the typical statistical error, $\sigma_N(M, t)$, in estimating $F_c$ with only $M$ samples via the standard deviation of $(F_c^{(M)} - F_c)$ over multiple repetitions of drawing $M$ samples at a fixed time $t$ from a $N$-qubit system. Similar to the procedure adapted for typicality scaling, we consider an error occurring at a time $t_0$ such that the many-body overlap after the error is $F \approx 0.5$. We find a dependence of $\sigma_N(M, t) = A(N, \tau)/\sqrt{M}$, where $A(N, \tau)$ is the prefactor depending on system size $N$ and $\tau$ is the time after the error (Ext. Data. Fig. 6). We find a weakly exponential dependence $A(N, \tau) \propto r(\tau)^N$, with a base $r(\tau)$ that is close to unity. For the case of $t_0 = 0$, we find that $r(\tau)$ starts slightly above one but settles to one for late $\tau$, while for $t_0 = 5$, we find that $r(\tau)$ is very close to one at all $\tau$ (inset of Ext. Data. Fig. 6c).

IX. NOISY SYSTEM DYNAMICS MODELING AND BENCHMARKING OF RYDBERG-ATOM QUANTUM SIMULATOR

Noisy system dynamics in the Rydberg model. The chaotic dynamics of a one-dimensional Rydberg-atom array presented in this study can be described by time evolution under the time-independent Hamiltonian $H$ (see Eq. (3)). In the ideal case without any environmental noise and control imperfections, we assume homogeneous global control and a defect-free ordered array; both Rabi frequencies $\Omega_i = \Omega$ and detunings $\Delta_i = \Delta$ are assumed to be site-independent and atoms are equally spaced with lattice spacing $a$. In experiments, however, quantum systems are not perfectly isolated and can undesirably interact with the randomly fluctuating environment; this causes control parameters in the Hamiltonian to drift over time and leads to a decay of quantum correlations between qubits. To this end, we design an ab initio error model by carefully considering realistic error sources affecting our experiment. Specifically, we introduce random fluctuations in both control parameters and atomic positions such that at each site $i$, atoms experience inhomogeneous, time-dependent control field strengths as well as positional disorder, modeled as

$$\Omega_i = \Omega + \delta\Omega_i(t) + \delta\Omega_i \quad (31)$$

$$\Delta_i = \Delta + \delta\Delta_i(t) + \delta\Delta_i \quad (32)$$

$$R_{ij} = a|i - j| + \delta R_{ij}. \quad (33)$$

Here $\delta\Omega_i(t)$ and $\delta\Delta_i(t)$ are the time-dependent, global fluctuations given by experimentally measured laser intensity and phase noise with no free parameters, respectively. Similarly, $\delta\Omega_i$, $\delta\Delta_i$, and $\delta R_{ij}$ are the time-independent, local fluctuations of Rabi frequency, Doppler shifts, and atomic positions, with $i$ and $j$ denoting site indices. In addition to these perturbations in the Hamiltonian parameters, we also take into account spontaneous decay of the Rydberg state as well as state-preparation-and-measurement errors that account for both initial loss of an atom and small infidelities in atomic state imaging and Rydberg state detection.

We simulate these errors in noisy quantum evolution via the Monte Carlo wavefunction method. We find that the resultant global bitstring measurement proba-
bilities show a high degree of correlation with the experimental measurement outcome probability distribution, indicating that the error model reliably reproduces our experimental system dynamics (Ext. Data Fig. 8). The accurate modeling of noisy dynamics in our system is further corroborated by good agreement with the fidelity estimator $F_c$ (Fig. 4d).

**Evaluation of $F_c$ in the Rydberg model.** As described in Methods Sec. II, Data Analysis, the bitstring probability distributions in our system are characterized in a constrained Hilbert space $H_c$, where the simultaneous excitation of two neighboring atoms to the Rydberg state is not allowed. In order to benchmark our quantum device in the subspace $H_c$, we modify the fidelity estimator $F_c$ as

$$F_c(t) = B(t)B_0(t)\left(2\frac{\sum_{z\in H_c}\hat{p}(z)\hat{p}_0(z)}{\sum_{z\in H_c}\hat{p}(z)^2} - 1\right). \tag{34}$$

Here $B(t)$ are $B_0(t)$ are the total probabilities of being in the subspace $H_c$ at time $t$ in noisy and clean evolutions, respectively, and $\hat{p}$ and $\hat{p}_0$ are the corresponding bitstring probabilities normalized in $H_c$. We numerically confirm that Eq. (34) is a good approximation in the strong Rydberg blockade regime, provided that noisy evolution results in negligible many-body overlap in the manifold outside $H_c$. In addition, we take into account the spatial inversion symmetry of the Rydberg Hamiltonian. Specifically, our Hamiltonian commutes with the left-right inversion operator $Q$, which swaps two atoms at sites $i$ and $N - i + 1$ for every $i$, due to the global uniformity implicit in the one-dimensional Rydberg model. Under this symmetry, we find that $|z_e\rangle = |z+|z\rangle$ and $|z_o\rangle = \frac{|z-|z\rangle}{\sqrt{2}}$ are the even- and odd-parity eigenstates of $Q$ with eigenvalues of 1 and $-1$, respectively. Here, $|z\rangle$ and $|\bar{z}\rangle$ are the $z$-basis bitstring and its mirrored version, i.e., $|z\rangle = |z_1z_2...z_N\rangle$ and $|\bar{z}\rangle = |z_Nz_{N-1}...z_1\rangle$ where $z_i$ is either 0 or 1 at site $i$. Since our initial state $|\psi_0\rangle = |0\rangle^\otimes N$ is the even-parity eigenstate of $Q$, i.e., $Q|\psi_0\rangle = |\psi_0\rangle$, the resulting many-body states after a quench always reside within the even-parity sector, reducing the effective Hilbert space dimension approximately by a factor of two. Furthermore, since projective measurement yields either $|z\rangle$ or $|\bar{z}\rangle$ with probabilities $1/2$ due to the parity eigenbasis $\{|z_e\rangle, |z_o\rangle\}$, their contributions to the fidelity estimation formula need to be adjusted; in consideration of this inversion symmetry, we rewrite both $\hat{p}(z)$ and $\hat{p}_0(z)$ using the parity eigenbasis and assume that both $|z\rangle$ and $|\bar{z}\rangle$ measurement outcomes originate from the same $|z_e\rangle$ in the even-parity sector. This effectively corresponds to checking the bitstring correlation in the even-parity sector of the parity eigenbasis. Having incorporated these modifications, we numerically confirm that the adjusted fidelity estimator shows good agreement with the true fidelity defined from the full Hilbert space (Fig. 4d).

**ACKNOWLEDGEMENTS**

We acknowledge discussions with Abhinav Deshpande and Alexey Gorshkov as well as funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1733907), the NSF CAREER award (1753386), the AFOSR YIP (FA9550-19-1-0044), the DARPA ONISQ program (W911NF2010021), the Army Research Office MURI program (W911NF2010136), the NSF QLCI program (2016245), and Fred Blum. JC acknowledges support from the IQIM postdoctoral fellowship. HH acknowledges support from the J. Yang & Family Foundation. HH is supported by the Gordon and Betty Moore Foundation. PMA acknowledges support from the PMA Prize postdoctoral fellowship. HP acknowledges funding from the Harvard Quantum Initiative (HQI) graduate fellowship. ALS acknowledges support from the Miller Institute for Basic Research in Science.