

Preparing random states and benchmarking with many-body quantum chaos

<https://doi.org/10.1038/s41586-022-05442-1>

Received: 26 March 2021

Accepted: 13 October 2022

Published online: 18 January 2023

 Check for updates

Joonhee Choi^{1,8}, Adam L. Shaw^{1,8}, Ivaylo S. Madjarov¹, Xin Xie¹, Ran Finkelstein¹, Jacob P. Covey^{1,2}, Jordan S. Cotler³, Daniel K. Mark⁴, Hsin-Yuan Huang¹, Anant Kale³, Hannes Pichler^{5,6}, Fernando G. S. L. Brandão¹, Soonwon Choi^{4,7}✉ & Manuel Endres¹✉

Producing quantum states at random has become increasingly important in modern quantum science, with applications being both theoretical and practical. In particular, ensembles of such randomly distributed, but pure, quantum states underlie our understanding of complexity in quantum circuits¹ and black holes², and have been used for benchmarking quantum devices^{3,4} in tests of quantum advantage^{5,6}. However, creating random ensembles has necessitated a high degree of spatio-temporal control^{7–12} placing such studies out of reach for a wide class of quantum systems. Here we solve this problem by predicting and experimentally observing the emergence of random state ensembles naturally under time-independent Hamiltonian dynamics, which we use to implement an efficient, widely applicable benchmarking protocol. The observed random ensembles emerge from projective measurements and are intimately linked to universal correlations built up between subsystems of a larger quantum system, offering new insights into quantum thermalization¹³. Predicated on this discovery, we develop a fidelity estimation scheme, which we demonstrate for a Rydberg quantum simulator with up to 25 atoms using fewer than 10^4 experimental samples. This method has broad applicability, as we demonstrate for Hamiltonian parameter estimation, target-state generation benchmarking, and comparison of analogue and digital quantum devices. Our work has implications for understanding randomness in quantum dynamics¹⁴ and enables applications of this concept in a much wider context^{4,5,9,10,15–20}.

We start by illustrating the concept of pure random state ensembles through a thought experiment: consider a programmable quantum device that evolves an input state $|\psi_0\rangle$ to an arbitrary output state $|\psi_j\rangle$, labelled by the program setting, j (Fig. 1a). If the set of states $|\psi_j\rangle$ —in the limit of many repetitions with different j —is homogeneously distributed over the output Hilbert space, it is termed a Haar-random (or uniform) state ensemble²¹. A distribution of states close to the Haar-random one is shown for a single qubit in Fig. 1b (right).

In practice, approximations to Haar-random state ensembles are generated by certain quantum devices requiring explicit classical randomization, in the sense that output states $|\psi_j\rangle$ are produced by randomly chosen unitary evolution operators $\hat{U}_j$. Examples include random unitary circuits (RUCs)^{8,10}, in which each configuration j is realized by a random choice of single- and two-qubit gates, and stochastic evolution with a dynamically changing Hamiltonian¹², $\hat{H}_j(t)$. In such systems, generation of approximate random state ensembles has been used for benchmarking of large-scale quantum devices^{3,4}, including fidelity estimation as part of quantum advantage^{5,6} and quantum volume tests²². On a more fundamental level, random ensembles provide important insights into studies of complexity growth in quantum systems¹ and understanding the quantum properties of black holes^{2,23}.

However, it is currently unknown how to generate such random ensembles from the simplest form of quantum evolution, that governed by a fixed, time-independent Hamiltonian $\hat{H}$ that is not explicitly randomized, as is the case for dynamics of closed and unperturbed quantum systems. Here, by considering pure state ensembles generated during partial measurement of a larger quantum system (Fig. 1c), we show such random ensembles do in fact emerge under such conditions. These emergent random ensembles enable applications such as device benchmarking, even in systems without explicit local, time-resolved control, which we demonstrate here experimentally using a Rydberg atom simulator^{24–26} with up to 25 atoms.

Observation of emergent randomness

We generically consider Hamiltonian evolution that produces a global quantum state $|\psi\rangle$, which we suppose here describes a set of qubits with basis states $|0\rangle$ and $|1\rangle$. We bipartition the state into two subsystems: a local system of interest A and its complement B (Fig. 1c). Explicitly keeping track of measurement results in B, which are bitstrings of the form, for example, $z_B = 100 \dots 010$, provides a full description of the total system state as

¹California Institute of Technology, Pasadena, CA, USA. ²Department of Physics, The University of Illinois at Urbana-Champaign, Urbana, IL, USA. ³Harvard University, Cambridge, MA, USA.

⁴Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁵Institute for Theoretical Physics, University of Innsbruck, Innsbruck, Austria. ⁶Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Innsbruck, Austria. ⁷Department of Physics, University of California, Berkeley, CA, USA. ⁸These authors contributed equally: Joonhee Choi, Adam L. Shaw. ✉e-mail: soonwon@mit.edu; mendres@caltech.edu

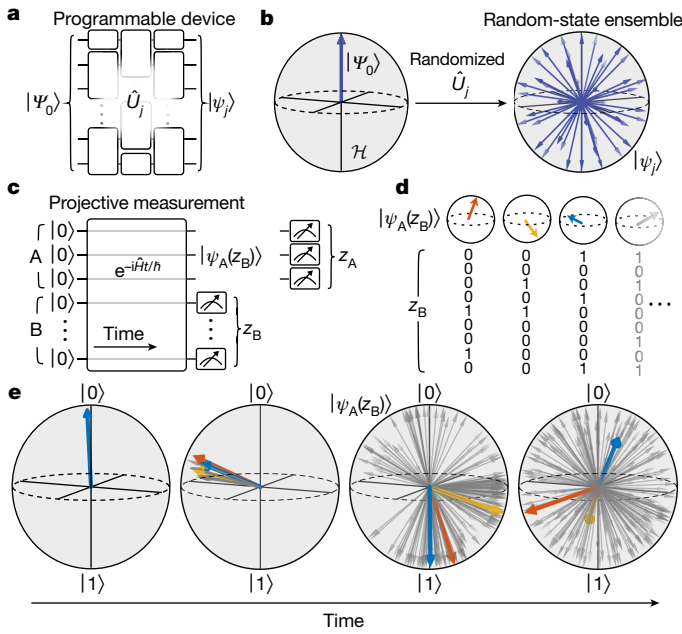


Fig. 1 | Random pure state ensembles from Hamiltonian dynamics.

a, A thought experiment, consisting of a programmable device producing arbitrary quantum states $|\psi_j\rangle$ through unitary operations $\hat{U}_j$, where j enumerates over different program settings. **b**, Repeatedly applying explicitly randomized unitary evolution to an initial state $|\psi_0\rangle$ produces an ensemble of pure quantum states $|\psi_j\rangle$ (blue arrows), which is distributed near-uniformly over the Hilbert space, $\mathcal{H}$ (grey sphere), giving a random state ensemble. **c**, Here we demonstrate a new approach to creating random state ensembles based on only a single instance of time-independent Hamiltonian evolution. An initial product state evolves under a Hamiltonian, $\hat{H}$, before site-resolved projective measurement in the computational basis $\{|0\rangle, |1\rangle\}$. We bipartition the system into two subsystems A and B, and analyse the conditional measurement outcomes in subsystem A, z_A , given a specific result z_B from the complement B. These outcomes are described by the projected ensemble, a pure state ensemble in A, $\{|\psi_A(z_B)\rangle\}$, realized through measurement of B. **d**, As an example for when A consists of a single qubit, conditional single-qubit quantum states $|\psi_A(z_B)\rangle$ are visualized on a Bloch sphere for all possible z_B bitstrings. **e**, Numerical simulations of our experimental system show that the distribution of the conditional pure state ensemble in A changes during evolution into a near-uniform form. Selected states are chosen (as shown in **d**) and highlighted to demonstrate their late-time divergence (right panel) despite nearly overlapping initial conditions (left panel).

$$|\psi\rangle = \sum_{z_B} \sqrt{p(z_B)} |\psi_A(z_B)\rangle \otimes |z_B\rangle, \quad (1)$$

where $p(z_B)$ is the probability of measuring a given bitstring in B and $|\psi_A(z_B)\rangle$ is a pure quantum state in A conditioned on the measurement outcome in B. Thus, for each possible z_B , there is a well-defined pure state in A, the set of all of which is generally not orthogonal. Together these states, $|\psi_A(z_B)\rangle$, and their respective probabilities, $p(z_B)$, form what we term the projected ensemble¹⁴ (Fig. 1d); similar concepts also enter the definition of localizable entanglement²⁷ and in the concept of conditional wavefunctions²⁸. By tracking the time evolution of the projected ensemble through both the states and probabilities that compose it, we can probe for signatures of the ensemble approaching a Haar-random distribution (Fig. 1e).

We stress that this concept is distinct from typical studies of equilibration in quantum many-body systems. There, the central object of interest is the reduced density operator on A, $\hat{\rho}_A = \text{Tr}_B(\hat{\rho})$, found from tracing out B from the full density operator $\hat{\rho}$. The reduced density operator can be constructed by averaging over the projected ensemble

states, $\hat{\rho}_A = \sum_{z_B} p(z_B) |\psi_A(z_B)\rangle \langle \psi_A(z_B)|$, but as such can only provide information on the mean of ensemble observables and never on the actual ensemble distribution itself.

To explain the importance of this distinction and reveal the emergence of random statistical properties of the projected ensemble, we use a Rydberg analogue quantum simulator^{24–26}, implemented with alkaline-earth atoms^{29–32}, which provides high fidelity preparation, evolution and readout²⁶ (Extended Data Fig. 1 and Methods). After a variable evolution time, we perform site-resolved readout in a fixed measurement basis, yielding experimentally measured bitstrings, z , which we bipartition into bitstrings z_A and z_B (Methods).

Hamiltonian parameters are chosen such that, after a short settling time, the marginal probability, $p(z_A)$, of measuring a given z_A (while ignoring the complementary z_B) agrees with the prediction from $\hat{\rho}_A$ being a maximally mixed state. In the language of quantum thermalization^{13,33–35}, this is equivalent to saying that $\hat{\rho}_A$ has reached an equilibrium at infinite effective temperature with the complement B as an effective, intrinsic bath^{13,36,37}. For a single qubit in A, such a reduced density operator is $\hat{\rho}_A = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$; the qubit has a probability of being in state $|0\rangle$ of $p(z_A = 0) = 1/D_A = 1/2$, where $D_A = 2$ is the local dimension of A. As shown in Fig. 2a, after a short transient period the experimentally measured probabilities $p(z_A = 0)$ (grey squares) equilibrate in agreement with this prediction.

We contrast this equilibration with the dynamics of conditional probabilities, $p(z_A|z_B)$, of measuring a given z_A conditioned on finding an accompanying measurement outcome in the intrinsic bath z_B . We note the marginal probability for finding z_A is the weighted average over conditional probabilities, $p(z_A) = \sum_{z_B} p(z_B)p(z_A|z_B)$. More generally, whereas $p(z_A)$ yields information of the reduced density operator, such conditional probabilities yield signatures of the projected ensemble, as $p(z_A|z_B) = |\langle z_A|\psi_A(z_B)\rangle|^2$. In Fig. 2a, we plot numerically simulated $p(z_A = 0|z_B)$ in grey, with selected traces highlighted in colour to be compared with their corresponding experimental data (circle markers). Importantly, conditional probabilities fluctuate chaotically with sensitive dependence on z_B , even when the marginal probability has reached a steady state. In experiments, fluctuations slowly damp out over time owing to extrinsic decoherence effects from coupling to an external environment at very late time, but these decoherence effects do not appear to affect the late-time marginal probability (Fig. 2a, right).

To analyse fluctuations, we bin conditional probabilities into a histogram $P(p)$ for a time when fluctuations are strong and decoherence effects are small (t_0 , Fig. 2b), as well as at very late time (t_1 , Fig. 2c) when decoherence dominates. At t_0 , the experimental $P(p)$ distribution is essentially flat, as predicted for a Haar-random ensemble, up to finite-sampling fluctuations and weak decoherence effects (Supplementary Information). We show projected states obtained from simulation (Bloch sphere in Fig. 2b), including decoherence, to illustrate how such a flat distribution is generated from a near-uniform ensemble of states. At very late time, t_1 , decoherence reduces the purity of projected states significantly, leading to $P(p)$ becoming concentrated around $1/D_A = 0.5$ (Fig. 2c). This highlights that the agreement between the experimental data and the random ensemble prediction in Fig. 2b,d is a coherent phenomenon of closed quantum system dynamics. We further validate this in Fig. 2d,e by plotting the $P(p)$ for A composed of two and three atoms, with corresponding Hilbert space dimensions of $D_A = 3$ and 5, respectively (Methods). Here the prediction from the Haar-random distribution⁵ is $P(p) = (D_A - 1)(1 - p)^{D_A - 2}$, which we note in the limit $D_A \rightarrow \infty$ becomes the well-known Porter–Thomas distribution³⁸, $P(p) = D_A e^{-D_A p}$, a key signature of the formation of random state ensembles.

The convergence of the projected ensemble to a nearly Haar-random distribution can be temporally resolved by considering moments of the distributions $P(p)$, for which the k th moment is defined as $p^{(k)} = \sum_p p^k P(p)$ (Fig. 3a). Looking order by order, we find that, after rescaling by a factor of $D_A \cdots (D_A + k - 1)$, moments from both experiment and numerics quickly approach $k!$, the analytical result expected from a Haar-random

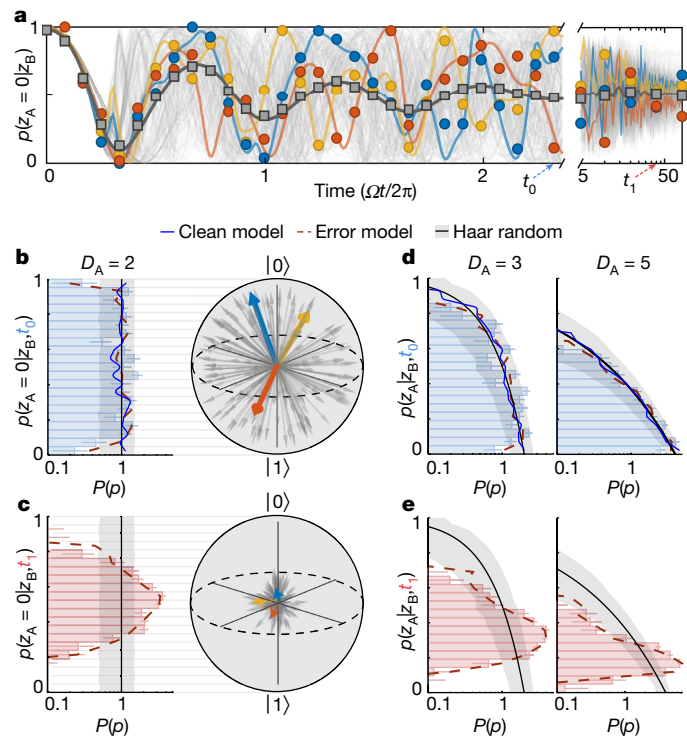


Fig. 2 | Experimental signatures of random pure state ensembles. **a**, We use a ten-atom Rydberg quantum simulator (Extended Data Fig. 1) to perform Hamiltonian evolution leading to quantum thermalization at infinite effective temperature (see main text for details). For a single qubit in A, we plot the probabilities for finding a single-qubit subsystem in state $|0\rangle$ as a function of time. Grey square markers indicate the marginal probabilities $p(z_A = 0)$, which equilibrate to around 0.5 owing to thermalization with B. By contrast, coloured circle markers show conditional probabilities given a specific measured z_B in B, $p(z_A = 0|z_B)$, which show large fluctuations even after the marginal probability reaches a steady state; these then diminish at late times because of extrinsic decoherence effects. Such conditional probabilities yield information about the projected ensemble as $p(z_A|z_B) = |\langle z_A|\psi_A(z_B)\rangle|^2$. Grey lines are simulated trajectories of $p(z_A = 0|z_B)$ for all outcomes z_B , with a few highlighted to be compared with experimental data (colour lines and markers). Decoherence sources (Supplementary Information) are included for simulations after the axis break. **b**, Histograms, $P(p)$, of the probabilities $p(z_A = 0|z_B)$ at intermediate ($\Omega t_0/2\pi = 2.3$) time. The experimental results are close to a flat distribution, consistent with a Haar-random ensemble, as visualized by the simulated distribution of projected states (right). **c**, However, at late ($\Omega t_1/2\pi = 38$) time, decoherence effects have concentrated probabilities around $1/D_A = 0.5$, consistent with the error model simulation showing the reduced lengths of single-qubit states (right). **d, e**, Similar agreement with predictions from random state ensembles is also seen at both early (**d**) and late (**e**) times for larger subsystem sizes of A with higher subsystem dimension, D_A (Methods). In **b–e**, black lines and grey bands are predictions and uncertainties (from finite sampling) of a D_A -dimensional uniform random ensemble; red dashed lines and blue solid lines are from simulations with and without decoherence, respectively (Supplementary Information).

ensemble (Supplementary Information); this is independent of the details of subsystem selection, such as whether A is chosen at the edge, centre or is even discontinuous (Extended Data Fig. 2), and universal values are also found for two-point correlators (Supplementary Information). At very late time, moments decrease, indicating sensitivity to decoherence effects (Fig. 3a, right). Although here we consider solely the projected ensemble equilibrated to infinite effective temperature, signatures of similar universal behaviour are seen numerically for finite effective temperature cases¹⁴ (Supplementary Information).

Importantly, convergence of the k th moment to $k!$ only suggests formation of low-order approximations to the Haar-random ensemble,

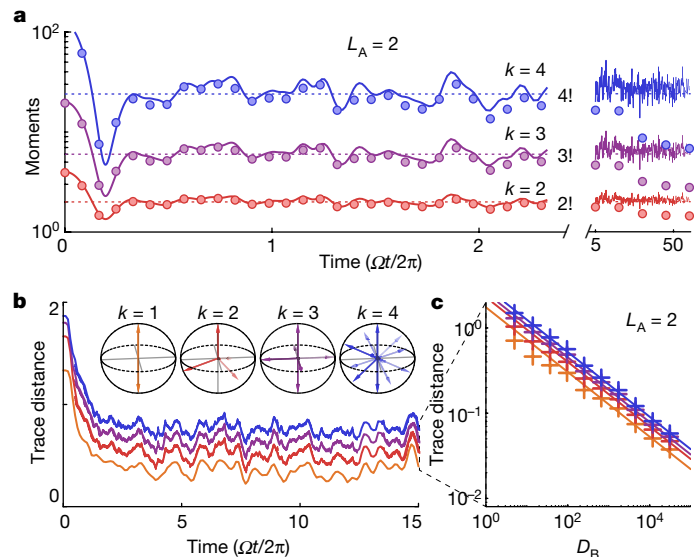


Fig. 3 | Development of emergent randomness. **a**, Rescaled second (red), third (purple) and fourth (blue) moments of the conditional probability distributions in Fig. 2d for subsystem of length $L_A = 2$. Experimental moments saturate to approximately $k!$, the expectation from the uniformly random ensemble (dotted lines) and consistent with numerical simulation (solid lines), before eventually decaying owing to decoherence. **b**, Numerically computed trace distances as a function of time between the $L_A = 2$ projected ensemble and the four lowest order approximations to the uniform random ensemble, the so-called quantum state k -designs, for $k = 1, 2, 3, 4$ (inset). Distances for all k decrease initially before saturating due to finite system size effects (Supplementary Information). If the trace distances up to order k vanish, the ensemble is as random as the k th design, and fluctuations of observables match up to order k , such as the k th moments in **a**. **c**, Late-time distances decrease as approximately $1/\sqrt{D_B}$ (solid lines), the Hilbert space dimension of the effective bath, subsystem B.

so-called quantum state k -designs³⁹. We study the trace distance between the projected ensemble, generated from error-free simulation, and successive k -designs (Fig. 3b), finding the distance decreases for all k th orders as a function of time, before saturating to a value exponentially small in the total system size (Fig. 3c). Similar numerical results are found for the case of RUCs and a Hamiltonian used in ion trap experiments (Extended Data Fig. 3). In an accompanying paper¹⁴, we show that the formation of uniformly random, pure state ensembles in subsystems is a more universal phenomenon.

Demonstration of device benchmarking

A key question is whether the formation of approximate k -designs in the projected ensemble enables associated applications such as device benchmarking with only global, time-independent control. Decoherence changes observables of the projected ensemble (Figs. 2 and 3); can this quantitatively signal the onset of decoherence in a quantum device?

We affirmatively answer this question by benchmarking the evolution of our experimental system under a time-independent Hamiltonian. We stress that our approach would be impossible with access only to the reduced density operator as it is relatively insensitive to decoherence (Fig. 2a). As a toy example, we consider the case of a single error occurring at time t_{err} during unitary evolution. The effect of this error then propagates outward⁴⁰, generically transforming the evolution output state and affecting measurement outcomes in subsystem A (Extended Data Fig. 4). Using the fact that the projected ensemble forms an approximate 2-design^{4,5,9,15,19,20} (Supplementary Information), we devise a fidelity estimator F_c to quantify the effect of this

error (Methods) as a rescaled cross-correlation between measurement probabilities in the experimental and ideal conditions:

$$F_c = 2 \frac{\sum_z p_0(z) p(z)}{\sum_z p_0^2(z)} - 1, \quad (2)$$

where $p(z)$ and $p_0(z)$ are the experimental and theoretical probabilities of observing a global bitstring z , respectively. We numerically confirm that, shortly after we apply an instantaneous phase rotation error on one qubit, our estimator approximates the many-body state overlap, $F_c \approx F = \langle \psi | \hat{\rho} | \psi \rangle$, between the ideal state $|\psi\rangle$ and the erroneous state $\hat{\rho}$ (Extended Data Fig. 4b, Methods and Supplementary Information).

To evaluate F_c experimentally, we formulate an empirical, unbiased estimator:

$$F_c \approx 2 \frac{\frac{1}{M} \sum_{i=1}^M p_0(z_{\text{exp}}^{(i)})}{\sum_z p_0^2(z)} - 1, \quad (3)$$

where M is the number of measurements and $z_{\text{exp}}^{(i)}$ is the experimentally measured bitstring at the i th repetition. Although this reformulation still requires calculation of a reference theory comparison, we note that the required number of experimental samples to accurately approximate F_c scales favourably with system size N . Concretely, the standard deviation of F_c is estimated to be $\sigma(F_c) \approx 1.04^N / \sqrt{M}$ (Extended Data Fig. 5), yielding an improved sample complexity in comparison with other existing methods^{39,41}. This implies that we do not need to fully reconstruct the experimental probability distribution for fidelity estimation of large quantum systems.

We test our benchmarking protocol for errors occurring continuously with a Rydberg quantum simulator of up to $N = 25$ atoms. We estimate the fidelity of our experimental device, $F_{c,\text{exp}}$, by correlating measured bitstrings to results from error-free simulation as a function of evolution time. In addition, we use an ab initio error model with no free parameters that mimics the experimental output (Supplementary Information), from which we extract both the fidelity estimator, $F_{c,\text{model}}$, and the model fidelity, $F_{\text{model}} = \langle \psi(t) | \hat{\rho}_{\text{model}}(t) | \psi(t) \rangle$ (Fig. 4a).

In Fig. 4b, we compare F_{model} , $F_{c,\text{exp}}$ and $F_{c,\text{model}}$ for a system of ten atoms. We observe $F_{c,\text{model}} \approx F_{\text{model}}$, validating the efficacy of the estimator under realistic error sources. In addition, we find $F_{c,\text{exp}} \approx F_{c,\text{model}}$ and that full bitstring probability distributions show good agreement between the error model and the experiment (Supplementary Information), indicating that our ab initio error model is a good description of the experiment.

We further apply this method to estimate the fidelity for generating states with a maximum half-chain entanglement entropy in larger systems. We first use error-free simulation to calculate the half-chain entanglement entropy growth as a function of system size, finding that the entanglement saturates at a time t_{ent} , which is linear in system size (Fig. 4c and Methods). We evaluate the fidelity estimator F_c for N ranging from 10 to 25, each at their respective t_{ent} , again finding good agreement between experiment and our ab initio error model (Extended Data Fig. 6) in the range for which our error model is readily calculable (Fig. 4d). We note an estimated fidelity of 0.49(2), where the number in brackets is the standard error in the last digit, for generating a state with maximum half-chain entanglement entropy for $N = 25$.

We numerically show that F_c also applies for erroneous evolution using other quantum devices (Extended Data Fig. 3). In the case of RUCs, F_c accurately estimates the fidelity at much shorter evolution times than do existing methods such as linear cross-entropy benchmarking^{3,5}, which can be explained through the early time-formation of the projected ensemble (Methods).

Applications of benchmarking

Our protocol enables various applications, including evaluating the relative performance of analogue and digital quantum devices, in situ

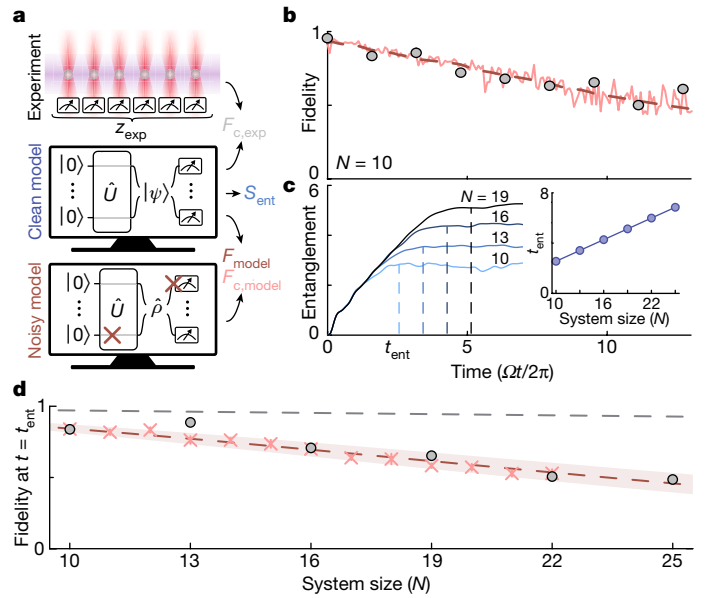


Fig. 4 | Fidelity estimation of an analogue Rydberg quantum simulator.

a, To estimate experimental fidelity, we repeatedly perform Hamiltonian evolution, each time performing a projective measurement to accrue an ensemble of measured bitstrings z_{exp} . We then correlate the measured bitstrings with an error-free simulation of the dynamics to calculate the fidelity estimator, $F_{c,\text{exp}}$. To validate our fidelity estimation method, we compare the error-free simulation against results from an ab initio error model (Supplementary Information), to calculate the model fidelity F_{model} and accompanying estimator $F_{c,\text{model}}$. **b**, Experimental benchmarking of a Rydberg quantum simulator for $N = 10$ atoms with blocked Hilbert space dimension $D = 144$. Shown are $F_{c,\text{exp}}$ (grey markers), the fidelity F_{model} (dashed red line) and $F_{c,\text{model}}$ (solid pink line). **c**, The half-chain entanglement entropy (calculated from the error-free simulation) increases before saturating at a time, t_{ent} , which grows linearly with system size (inset). **d**, Fidelity estimated at t_{ent} , showing estimator F_c from experiment (grey markers) up to $N = 25$ and from error model (pink crosses) up to $N = 22$. In addition, we show a fit to the model fidelity, given as $F_0^N \exp[-\gamma(N)t_{\text{ent}}(N)]$ (red dashed line), where F_0 is the single-atom preparation fidelity and $\gamma(N)$ is the many-body fidelity decay rate of our Rydberg simulator (Methods and Extended Data Fig. 6). The fidelity estimation uses only less than 10^4 experimentally sampled bitstrings per data point. See Methods for description of error bars.

Hamiltonian parameter estimation and benchmarking the fidelity of preparing various target states. First, to compare analogue and digital quantum evolution, we evaluate the fidelity achieved at t_{ent} for both analogue quantum simulators and digital quantum computers (for which t_{ent} is defined in terms of gate depth, see Methods). We find our system has an equivalent effective, state preparation and measurement (SPAM)-corrected, two-qubit cycle fidelity of 0.987(2) for the gate-set used in ref.⁵ and 0.9965(5) for a gate-set based on two-qubit SU(4) gates⁴ (Extended Data Fig. 7 and Methods).

Next, to perform Hamiltonian parameter estimation, we measure F_c while varying Hamiltonian parameters in the simulation; when numerical parameters do not match with those in experiment, F_c will decay quickly. To capture this effect in a single quantity we plot the normalized, time-integrated F_c (Fig. 5a). For each Hamiltonian parameter, a sharp maximum emerges (Supplementary Information), showing good agreement with precalibrated values (dashed lines and shaded areas). Parameter estimation also works when applied to learn local, site-dependent terms of a disordered Hamiltonian (Fig. 5b), notably without any local control during readout.

Finally, F_c can be used to benchmark the fidelity of preparing various quantum states of interest by preparing a target state and then quenching the Hamiltonian to evolve the prepared state at infinite effective

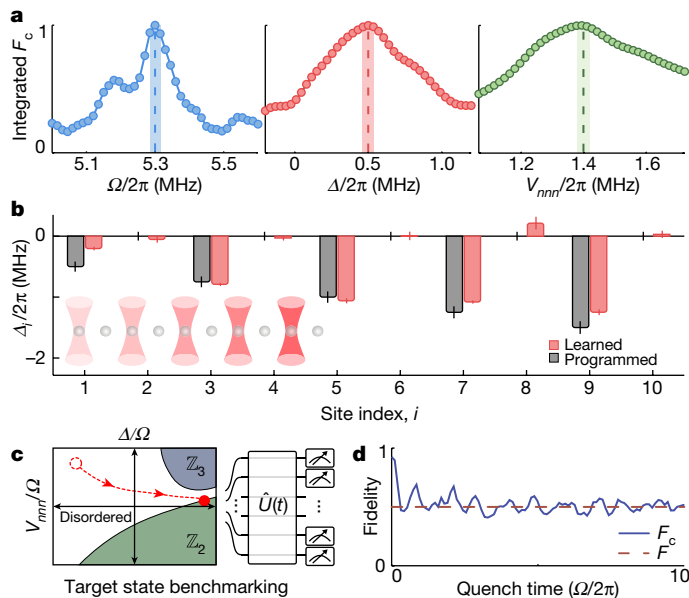


Fig. 5 | Hamiltonian learning and target-state benchmarking. **a**, Normalized, time-integrated F_c as a function of the global Rabi frequency, detuning and the next-nearest-neighbour interaction strength in the Rydberg model (Methods); this normalized F_c is maximized only when the correct parameters are used. Vertical dashed lines and shaded areas denote independently calibrated values and their uncertainties. **b**, Programmed (grey bars) and learned (red bars) local Hamiltonian parameters for an arbitrary, site-dependent detuning field imposed by an intensity-dependent lightshift from locally addressable optical tweezers (inset, red funnels). **c**, The one-dimensional Rydberg ground state phase diagram⁴² and an adiabatic path (red dotted line) to produce a specific target state near the Ising quantum phase transition. **d**, Our F_c benchmark (solid line) can estimate the fidelity F (dashed line) of the prepared noisy state consisting of an equal probability mixture of the ground and first excited states by evolving at infinite effective temperature after preparation. Here we numerically demonstrated the target state benchmarking for a ground state of system size $N = 15$ (Methods).

temperature (Fig. 5c and Methods). As a numerical proof-of-principle, we show results for such target-state benchmarking to prepare a ground state near the Ising quantum phase transition in the Rydberg model⁴² (Fig. 5c,d), where the noisy state is an equal probability mixture of the ground and first excited states. After a short disordered quench, the estimator F_c reveals the fidelity of the prepared state, offering a way to perform in situ optimization of many-body state preparation; further examples are shown in Extended Data Fig. 8.

In conclusion, we have uncovered emergent randomness arising from partial measurement of an interacting many-body system. Subsequently, we have shown a widely applicable fidelity estimation scheme that works at shorter evolution times and with reduced experimental complexity compared with existing approaches, and have demonstrated applications in quantum device comparison, Hamiltonian parameter estimation and benchmarking the fidelity of preparing interesting quantum states. The concept of emergent randomness could provide a new framework for quantum thermalization, chaos and complexity growth⁴³. Open questions remain, such as a deeper understanding of the finite effective temperature case¹⁴ (Supplementary Information) and uncovering the signatures of non-ergodic dynamics in integrable or localized systems^{33–35,44,45}. Such developments could enable a more flexible and standardized way of performing quantum fidelity estimation in a wide variety of quantum hardware, including trapped ions⁴⁶, superconducting qubits³, photonic systems⁴⁷, and cold atoms and molecules in optical lattices⁴⁸. Ultimately, emergent random ensembles could find a broader range of applications, including

quantum advantage tests^{5,6,15–17,47}, in situ Hamiltonian learning^{5,49} and optimization of target quantum state preparation.

Note added in proof: During the course of the revision, a new fidelity estimator has been introduced⁵⁰; we present a comparison in Supplementary Information.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-05442-1>.

- Brandão, F. G. S. L., Chemsyany, W., Hunter-Jones, N., Kueng, R. & Preskill, J. Models of quantum complexity growth. *PRX Quantum* **2**, 30316 (2021).
- Hayden, P. & Preskill, J. Black holes as mirrors: quantum information in random subsystems. *J. High. Energy Phys.* **2007**, 120 (2007).
- Neill, C. et al. A blueprint for demonstrating quantum supremacy with superconducting qubits. *Science* **360**, 195–199 (2018).
- Cross, A. W., Bishop, L. S., Sheldon, S., Nation, P. D. & Gambetta, J. M. Validating quantum computers using randomized model circuits. *Phys. Rev. A* **100**, 32328 (2019).
- Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
- Wu, Y. et al. Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.* **127**, 180501 (2021).
- Emerson, J., Weinstein, Y. S., Saraceno, M., Lloyd, S. & Cory, D. G. Pseudo-random unitary operators for quantum information processing. *Science* **302**, 2098–2100 (2003).
- Harrow, A. W. & Low, R. A. Random quantum circuits are approximate 2-designs. *Commun. Math. Phys.* **291**, 257–302 (2009).
- Dankert, C., Cleve, R., Emerson, J. & Livine, E. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A* **80**, 12304 (2009).
- Brandão, F. G. S. L., Harrow, A. W. & Horodecki, M. Local random quantum circuits are approximate polynomial-designs. *Commun. Math. Phys.* **346**, 397–434 (2016).
- Ohliger, M., Nesme, V. & Eisert, J. Efficient and feasible state tomography of quantum many-body systems. *New. J. Phys.* **15**, 15024 (2013).
- Nakata, Y., Hirsch, C., Koashi, M. & Winter, A. Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics. *Phys. Rev. X* **7**, 21006 (2017).
- Kauffman, A. M. et al. Quantum thermalization through entanglement in an isolated many-body system. *Science* **353**, 794–800 (2016).
- Cotler, J. et al. Emergent quantum state designs from individual many-body wavefunctions. Preprint at <https://arxiv.org/abs/2103.03536> (2021).
- Boixo, S. et al. Characterizing quantum supremacy in near-term devices. *Nat. Phys.* **14**, 595–600 (2018).
- Boulund, A., Fefferman, B., Nirkhe, C. & Vazirani, U. On the complexity and verification of quantum random circuit sampling. *Nat. Phys.* **15**, 159–163 (2019).
- Haferkamp, J. et al. Closing gaps of a quantum advantage with short-time Hamiltonian dynamics. *Phys. Rev. Lett.* **125**, 250501 (2020).
- Brydges, T. et al. Probing Rényi entanglement entropy via randomized measurements. *Science* **364**, 260–263 (2019).
- Elben, A. et al. Cross-platform verification of intermediate scale quantum devices. *Phys. Rev. Lett.* **124**, 10504 (2020).
- Huang, H. Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020).
- Harrow, A. W. The church of the symmetric subspace. Preprint at <https://arxiv.org/abs/1308.6595> (2013).
- Jurcevic, P. et al. Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Sci. Technol.* **6**, 025020 (2021).
- Piroli, L., Sünderhauf, C. & Qi, X. L. A random unitary circuit model for black hole evaporation. *J. High. Energy Phys.* **2020**, 63 (2020).
- Bernien, H. et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017).
- Browaeys, A. & Lahaye, T. Many-body physics with individually controlled Rydberg atoms. *Nat. Phys.* **16**, 132–142 (2020).
- Madjarov, I. S. et al. High-fidelity entanglement and detection of alkaline-earth Rydberg atoms. *Nat. Phys.* **16**, 857–861 (2020).
- Popp, M., Verstraete, F., Martín-Delgado, M. A. & Cirac, J. I. Localizable entanglement. *Phys. Rev. A* **71**, 42306 (2005).
- Goldstein, S., Lebowitz, J. L., Mastrodonato, C., Tumulka, R. & Zanghi, N. Universal probability distribution for the wave function of a quantum system entangled with its environment. *Commun. Math. Phys.* **342**, 965–988 (2016).
- Norcia, M. A., Young, A. W. & Kaufman, A. M. Microscopic control and detection of ultracold strontium in optical-tweezer arrays. *Phys. Rev. X* **8**, 41054 (2018).
- Cooper, A. et al. Alkaline-Earth atoms in optical tweezers. *Phys. Rev. X* **8**, 41055 (2018).
- Saskin, S., Wilson, J. T., Grinkemeyer, B. & Thompson, J. D. Narrow-line cooling and imaging of ytterbium atoms in an optical tweezer array. *Phys. Rev. Lett.* **122**, 143002 (2019).
- Covey, J. P., Madjarov, I. S., Cooper, A. & Endres, M. 2000-times repeated imaging of strontium atoms in clock-magic tweezer arrays. *Phys. Rev. Lett.* **122**, 173201 (2019).
- Rigol, M., Dunjko, V. & Olshanii, M. Thermalization and its mechanism for generic isolated quantum systems. *Nature* **452**, 854–858 (2008).
- Abanin, D. A., Altman, E., Bloch, I. & Serbyn, M. Colloquium: many-body localization, thermalization, and entanglement. *Rev. Mod. Phys.* **91**, 21001 (2019).

35. Ueda, M. Quantum equilibration, thermalization and prethermalization in ultracold atoms. *Nat. Rev. Phys.* **2**, 669–681 (2020).
36. Popescu, S., Short, A. J. & Winter, A. Entanglement and the foundations of statistical mechanics. *Nat. Phys.* **2**, 754–758 (2006).
37. del Rio, L., Hutter, A., Renner, R. & Wehner, S. Relative thermalization. *Phys. Rev. E* **94**, 22104 (2016).
38. Porter, C. E. & Thomas, R. G. Fluctuations of nuclear reaction widths. *Phys. Rev.* **104**, 483–491 (1956).
39. Ambainis, A. and Emerson, J. Quantum t-designs: t-wise independence in the quantum world. In *Proceedings Twenty-Second Annual IEEE Conference on Computational Complexity* 129–140 (IEEE, 2007).
40. Khemani, V., Vishwanath, A. & Huse, D. A. Operator spreading and the emergence of dissipative hydrodynamics under unitary evolution with conservation laws. *Phys. Rev. X* **8**, 31057 (2018).
41. Flammia, S. T. & Liu, Y. K. Direct fidelity estimation from few Pauli measurements. *Phys. Rev. Lett.* **106**, 230501 (2011).
42. Slagle, K. et al. Microscopic characterization of Ising conformal field theory in Rydberg chains. *Phys. Rev. B* **104**, 235109 (2021).
43. Cotler, J., Hunter-Jones, N. & Ranard, D. Fluctuations of subsystem entropies at late times. *Phys. Rev. A* **105**, 022416 (2022).
44. Nandkishore, R. & Huse, D. A. Many-body localization and thermalization in quantum statistical mechanics. *Annu. Rev. Condens. Matter Phys.* **6**, 15–38 (2015).
45. Turner, C. J., Michailidis, A. A., Abanin, D. A., Serbyn, M. & Papić, Z. Weak ergodicity breaking from quantum many-body scars. *Nat. Phys.* **14**, 745–749 (2018).
46. Monroe, C. et al. Programmable quantum simulations of spin systems with trapped ions. *Rev. Mod. Phys.* **93**, 025001 (2021).
47. Zhong, H. S. et al. Quantum computational advantage using photons. *Science* **370**, 1460–1463 (2020).
48. Gross, C. & Bloch, I. Quantum simulations with ultracold atoms in optical lattices. *Science* **357**, 995–1001 (2017).
49. Giovannetti, V. Quantum-enhanced measurements: beating the standard quantum limit. *Science* **306**, 1330–1336 (2004).
50. Mark, D. K., Choi, J., Shaw, A. L., Endres, M. and Choi, S. Benchmarking quantum simulators using quantum chaos. Preprint at <https://arxiv.org/abs/2205.12211> (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Description of the experiment

The details of our experiment have been summarized previously^{26,30,32,51}; in brief, we use an array of optical tweezers to trap 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 $\langle n \rangle \approx 0.3$ (corresponding to approximately 3 μ K). For all data shown, we rearrange the initially stochastically filled array to a defect-free array^{52,53} of atoms spaced by 3.75 μ m, discarding extras. Atoms are initialized to the $5s5p^3P_0$ (698 nm) clock state through a combination of coherent drive and incoherent pumping, for a total preparation fidelity of 0.997(1) per atom. We treat the clock state as a metastable qubit ground state, $|0\rangle$, and subsequently drive to the $5s61s^3S_1$, $m_j = 0$ (317 nm) Rydberg state, $|1\rangle$. Following Hamiltonian evolution, state readout is performed using the auto-ionizing transition $5s61s^3S_1$, $m_j = 0 \leftrightarrow 5p_{3/2}61s_{1/2}$ (408 nm, $J = 1$, $m_j = \pm 1$), which rapidly ionizes atoms in the Rydberg state with high fidelity (approximately 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

$$\hat{H}/\hbar = \Omega \sum_i \hat{S}_i^x - \Delta \sum_i \hat{n}_i + \frac{C_6}{a^6} \sum_{i>j} \frac{\hat{n}_i \hat{n}_j}{|i-j|^6} \quad (4)$$

which describes a set of interacting two-level systems, labelled 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.

For measurements observing the emergence of random ensembles, we use $\Omega/2\pi = 4.70(1)$ MHz, $\Delta/2\pi = 0.90(3)$ Hz and $a = 3.75(5)$ μ m, with an experimentally measured next-nearest-interaction strength of $V_{nnn}/2\pi = C_6/(2a)^6 = 1.40(2)$ MHz, yielding an estimated C_6 coefficient of $2\pi \times 249(20)$ GHz μ m⁶. Under this condition, we confirm numerically that the initial all-zero state rapidly thermalizes to an infinite effective temperature thermal ensemble locally within the constrained subspace where no two adjacent atoms are simultaneously in the Rydberg state^{24–26}. Benchmarking measurements are performed with $\Omega/2\pi = 5.30(1)$ MHz, $\Delta/2\pi = 0.50(3)$ MHz, which again thermalizes to an infinite effective temperature thermal ensemble.

As the experimental data shown throughout the main text require 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-made 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 π -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 $5s^2^1S_0 \leftrightarrow 5s5p^3P_1$ (461 nm) transition²⁶.

For the Rydberg alignment, detuning and Rabi frequency, we rearrange the array to non-interacting atoms spaced by 15.1 μ 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 with a numerical simulation across all positions to identify the point of furthest phase and thus maximal intensity. For the Rydberg detuning, we measure the resonance condition at $\Omega t = 13\pi$ 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 (Extended Data Fig. 1).

Data analysis

Our state readout is described in detail in ref. ²⁶; it features single-site detection that discriminates atoms in the clock state, $|0\rangle$, versus the Rydberg state, $|1\rangle$, through a combination of fluorescence imaging and Rydberg auto-ionization. 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 and (3) after the sequence has finished. We postselect for image triplets for which 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 in which 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 bitstring of the qubit states. After taking many shots we accrue an ensemble of such bitstrings $\{z\}$. For randomness measurements, a total of approximately 120,000 shots are used (approximately 3,000 shots per time point). For benchmarking measurements a total of approximately 40,000 shots are used for generating the time trace at $N = 10$ in Fig. 4b (approximately 3,700 shots per time point). Approximately 4,000 total shots are used for the N -scaling plot in Fig. 4d, in which the number of shots for a given system size is approximately given by $M \approx 3,000 + 250N$.

Error bars in Figs. 2 and 3 are calculated via bootstrapping methods, and are often smaller than the marker sizes. In Fig. 4, error bars on experimental quantities are calculated via extrapolation from subsampling of the total number of experimentally measured bitstrings to estimate the sample complexity at a given N (Extended Data Fig. 5). The error bars on $F_{c,model}$ from the ab initio error model stem from typicality errors associated with the temporal fluctuation of our estimator (Supplementary Information). Error bars on the programmed parameters in Fig. 5b come from uncertainty in local detuning intensity, whereas error bars on the learned parameters are standard deviations arising from performing the simultaneous parameter optimization 30 times with randomized starting initial conditions.

Our system Hamiltonian is naturally stratified into a number of energetically widely spaced sectors due to the Rydberg blockade^{24–26}. In particular, the nearest-neighbour interaction is approximately 20 times greater than the next largest energy scale, so cases for which neighbouring pairs of atoms are both excited to the Rydberg state are greatly suppressed. For $N = 10$, we find around 99% of all experimental bitstrings are in the blockade-satisfying energy sector at short times ($t < 1$ μ s) but this probability starts to decrease at late times ($t > 1$ μ s) owing to experimental imperfections—we refine $\{z\}$ by discarding all realizations not in this sector. We note, however, that all simulations are performed in the full Hilbert space.

For calculations involving conditional probabilities, we bipartition each bitstring 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 D_A 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.

Derivation of the fidelity estimator F_c

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

$$p_0(z) = p_0(z_A|z_B)p_0(z_B) \quad (5)$$

$$p(z) = p(z_A|z_B)p(z_B), \quad (6)$$

for complementary subsystems A and B. We consider the simplest case of a single local error $\hat{V}$ occurring at time t_{err} during time evolution and assume that the time-evolved error operator $\hat{V}(\tau) = \hat{U}(\tau)\hat{V}\hat{U}(\tau)^\dagger$ is supported within subsystem A. Here $\tau = t - t_{\text{err}}$ is the time past the occurrence of the error and $\hat{U}(\tau)$ is the time-evolution operator from t_{err} 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

$$F_c = 2 \frac{\sum_z p_0(z)p(z)}{\sum_z p_0(z)^2} - 1 \quad (7)$$

$$= 2 \frac{\sum_{z_B} p_0^2(z_B) \sum_{z_A} p(z_A|z_B)p_0(z_A|z_B)}{\sum_{z_B} p_0^2(z_B) \sum_{z_A} p_0^2(z_A|z_B)} - 1 \quad (8)$$

$$\approx \frac{\sum_{z_B} p_0^2(z_B)(F_{\text{XEB}}(z_B) + 1)}{\sum_{z_B} p_0^2(z_B)} - 1 \quad (9)$$

$$= \sum_{z_B} q(z_B) F_{\text{XEB}}(z_B) \quad (10)$$

where $q(z_B) = \frac{p_0^2(z_B)}{\sum_{z_B} p_0^2(z_B)}$ and

$$F_{\text{XEB}}(z_B) = (D_A + 1) \sum_{z_A} p(z_A|z_B)p_0(z_A|z_B) - 1 \quad (11)$$

is the z_B -dependent, linear cross-entropy benchmark⁵ in subsystem A, and D_A is the Hilbert space dimension of A. From equation (8) to equation (9), we used the second-order moment of the projected ensemble in an error-free case

$$\frac{1}{D_A} \sum_{z_A} p_0^2(z_A|z_B) \approx \frac{2!}{D_A(D_A + 1)} \quad (12)$$

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

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

$$\begin{aligned} & \sum_{z_B} q(z_B) p_0(z_A|z_B) p(z_A|z_B) \\ &= \sum_{z_B} q(z_B) \langle \psi_A(z_B) | \hat{P}_{z_A} | \psi_A(z_B) \rangle \langle \psi_A(z_B) | \hat{V}^\dagger(\tau) \hat{P}_{z_A} \hat{V}(\tau) | \psi_A(z_B) \rangle \\ &= \text{tr} \left\{ (\hat{P}_{z_A} \otimes \hat{V}^\dagger(\tau) \hat{P}_{z_A} \hat{V}(\tau)) \cdot \sum_{z_B} q(z_B) (|\psi_A(z_B)\rangle \langle \psi_A(z_B)|)^{\otimes 2} \right\} \\ &\approx \frac{\text{tr} \{ (\hat{P}_{z_A} \otimes \hat{V}^\dagger(\tau) \hat{P}_{z_A} \hat{V}(\tau)) \cdot (\hat{1} + \hat{S}_A) \}}{D_A(D_A + 1)} \\ &= \frac{1 + |\langle z_A | \hat{V}(\tau) | z_A \rangle|^2}{D_A(D_A + 1)} \end{aligned} \quad (13)$$

where $\hat{P}_{z_A} = |z_A\rangle \langle z_A|$ is the z -basis projector onto a specific bitstring z_A in A, $q(z_B)$ is the probability weighting factor, $\hat{1}$ is the identity operator

and $\hat{S}_A$ is the swap operator acting on subsystem A for the duplicated Hilbert space $\mathcal{H}_A^{\otimes 2}$. To obtain the fourth line, we used

$$\sum_{z_B} q(z_B) (|\psi_A(z_B)\rangle \langle \psi_A(z_B)|)^{\otimes 2} \approx \frac{\hat{1} + \hat{S}_A}{D_A(D_A + 1)}, \quad (14)$$

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

Inserting equation (13) into equation (10), we obtain

$$\begin{aligned} F_c &\approx \frac{1}{D_A} \sum_{z_A} |\langle z_A | \hat{V}(\tau) | z_A \rangle|^2 \\ &= \frac{1}{D} \sum_z |\langle z | \hat{V}(\tau) | z \rangle|^2, \end{aligned} \quad (15)$$

where the equality on the second line holds because one can always multiply the identity $\frac{1}{D_B} \sum_{z_B} \langle z_B | z_B \rangle^2 = 1$ 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 equation (15) explains how F_c estimates the many-body fidelity with a good accuracy. The right-hand side of equation (15) describes the return probability of $\hat{V}(\tau)$ (also known as the Loschmidt echo) averaged over all possible initial states in the fixed measurement basis. Under chaotic time evolution, the propagated error operator $\hat{V}(\tau)$ becomes scrambled, and it is exponentially unlikely with 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, for example, $\hat{V}(\tau) = c_0 \hat{1} + \sum_s c_s(\tau) \hat{\sigma}_s$ 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 $\hat{V}$ 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 $\hat{V}$ is diagonal in the measurement basis, for example dephasing error along the z axis, and if the error occurs shortly before the bitstring measurements, the return probability in equation (15) will be close to unity despite the fact 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 $\hat{V}(\tau)$ becomes scrambled in the operator basis and F can be approximately estimated (Extended Data Fig. 4). In other words, even in the case of the diagonal errors, our formula becomes valid after a finite delay time.

Statistical error scaling from a finite number of bitstring samples

We quantify the typical statistical error from approximating our fidelity estimator via equation (3) in two steps. First, we use our ab initio error model to simulate the quantum evolution of the Rydberg Hamiltonian for system sizes from $N = 10$ to 22, from which we can calculate the exact value of F_c compared to error-free numerics. We then sample a finite number of M samples from the probability distributions produced from the error model simulation, apply equation (3), and plot the standard deviation of F_c as a function of M (Extended Data Fig. 5a). We see a characteristic scaling of $\sigma(F_c) = A/\sqrt{M}$, where A is the sample complexity, expected to scale exponentially with N , and σ denotes the

Article

standard deviation. We perform a similar process directly on our experimental data by repeatedly subsampling the experimentally measured bitstrings to estimate the scaling of the standard deviation. We plot both error model and experimental results in Extended Data Fig. 5b as a function of N , where the time is set as the N -dependent entanglement saturation time. By fitting the experimental (error model) results, we find $A \approx a^N$, with $a = 1.037(2)$ ($a = 1.039(2)$).

Predicting fidelity at the entanglement time

Calculating the entanglement time. As can be seen in Fig. 4c, entanglement growth in our Rydberg quantum simulator is generally characterized by two distinct regions: a size-independent linear increase, followed by saturation at an N -dependent value. To systematically capture this behaviour and predict the entanglement saturation time for arbitrary N , we apply the following protocol. We first calculate the entanglement growth for system sizes ranging from $N = 10$ to 22. We then fit all profiles with a functional form of

$$S_{\text{ent}}(N, t) = \begin{cases} m_1 t & t \leq t_c(N) \\ m_1 t_c(N) + m_2(N)(t - t_c(N)) & t > t_c(N) \end{cases}$$

with free parameters m_1 , $m_2(N)$ and $t_c(N)$, but with the explicit constraint that m_1 must be the same for all system sizes. From this we find $t_c(N)$, and we further define $t_{\text{ent}}(N) = C t_c(N)$, where in the Rydberg case we set $C = 1.35$ to make sure the time we study is firmly in the saturated entanglement regime (as can be verified visually in Fig. 4c). The secondary slope m_2 is used because even past t_{ent} there is still some slight growth to the entanglement entropy, which becomes more noticeable for larger N . This behaviour is attributed to slow coupling to the non-blockaded Hilbert space as the blockade constraint is only approximate. The entanglement saturation time is then fit as a linear function of system size, yielding $t_{\text{ent}}(N) = \alpha_0 + \alpha_1 N$; for our particular Hamiltonian parameters we find $\alpha_0 = -0.0580(2)$ and $\alpha_1 = 0.05404(1)$, both in units of microseconds.

For the case of finding the entanglement saturation depth, d_{ent} , for the case of RUCs considered in Extended Data Fig. 7, we apply essentially the same procedure. We study two different digital circuit implementations. In the first, with a gate-set based on ref. ⁵, the RUC circuit is composed of alternating one- and two-qubit gates; the one-qubit gates are randomly chosen $\pi/2$ rotations along the $\hat{x}$, $\hat{y}$ and $\hat{x} + \hat{y}$ directions, whereas the two-qubit gates are 'fSym'. In the second, the RUC is composed entirely of two-qubit SU(4) gates (without global swap operations⁴). For the first implementation, we set $C = 1$, whereas in the second we set $C = 1.7$, to better guarantee that the chosen depth is in the saturated entanglement regime. Open boundary conditions are used in accordance with the experimental Rydberg system, and thus there are two possible gate topologies (that is, in the first depth applying the 'fSym' gate to qubits 1–2, 3–4 and so on, or 2–3, 4–5 and so on)—we explicitly average over an equal number of randomized realizations of each topology when calculating the entanglement entropy growth. As in the Rydberg case, we fit the entanglement saturation depth as a linear function of the number of qubits in the RUC, yielding $d_{\text{ent}} = \beta_0 + \beta_1 N_{\text{RUC}}$. For the gate-set based on ref. ⁵ we find $\beta_0 = -0.395(17)$ and $\beta_1 = 0.557(1)$, and for the gate-set based on SU(4) gates⁴ we find $\beta_0 = -3.18(77)$ and $\beta_1 = 2.261(51)$.

Estimating fidelity decay. In Fig. 4b, we see that the decay profile of the model fidelity, F_{model} , for our Rydberg simulator is approximately exponential, which we confirm by error model simulations with system sizes ranging from $N = 10$ to 22 in Extended Data Fig. 6a. For each system size, we fit the fidelity decay profile as

$$F(N, t) \propto \exp(-\gamma(N)t), \quad (16)$$

where $\gamma(N)$ is the fidelity decay rate. We find that, for the system size range considered here, $\gamma(N)$ scales approximately linearly with N , from which we fit $\gamma(N) = \gamma_0 + \gamma_1 N$; for our particular Hamiltonian parameters

and noise sources, we find $\gamma_0 = 0.12(4)$ and $\gamma_1 = 0.017(3)$, both in megahertz (Extended Data Fig. 6b).

In Fig. 4d, we use the fitted $\gamma(N)$ explicitly to predict the fidelity scaling of our Rydberg simulator at the N -dependent entanglement saturation time, t_{ent} , as a function of system size. Concretely, we plot (red dashed line):

$$F_{\text{model}}(N, t_{\text{ent}}(N)) = F_0^N \exp(-\gamma(N)t_{\text{ent}}(N)), \quad (17)$$

where $F_0 = 0.997(1)$ is the single-atom preparation fidelity determined experimentally. The shaded red region in Fig. 4d depicts the error from the fit uncertainty of $\gamma(N)$.

For the RUC case, the fidelity decay for a given system size, N_{RUC} , and depth, d , is modelled as a simple product over constituent two-qubit cycle fidelity, F_{cycle} , yielding

$$F_{\text{RUC}}(N, d) = F_{\text{cycle}}^{(N_{\text{RUC}}-1)d/2}, \quad (18)$$

where the exponent on the right-hand side reflects the fact that we apply, on average, $(N_{\text{RUC}} - 1)/2$ two-qubit gates in parallel per depth.

Comparing digital and analogue devices. We wish to directly compare the evolution fidelity of our analogue Rydberg quantum simulator against that of a digital device implementing a RUC with equivalent entanglement entropy at the entanglement saturation time. However, due to the Rydberg blockade mechanism, as well as symmetries of our Hamiltonian (Supplementary Information), an equal number of atoms in the Rydberg simulator, N , and qubits in the RUC, N_{RUC} , will not saturate to the same half-chain entanglement entropy.

To overcome this, in Extended Data Fig. 7b we plot the entanglement entropy, $S(S_{\text{RUC}})$, achieved at $t_{\text{ent}}(d_{\text{ent}})$ for the Rydberg simulator (RUC) as a function of $N(N_{\text{RUC}})$. For the Rydberg simulator, we fit $S(N) = \sigma_0 + \sigma_1 N$ with $\sigma_0 = 0.16(4)$ and $\sigma_1 = 0.26(3)$. For the RUC, we use the prediction of $S_{\text{RUC}}(N_{\text{RUC}}) = \eta_0 + \eta_1 N_{\text{RUC}}$ with $\eta_0 = -\log_2(e)/2 \approx -0.72$ and $\eta_1 = 1/2$ being exact values with no error bars, as defined in ref. ⁵⁴ (where e is Euler's number and where we have used the $\log_2$ entanglement entropy convention). To find the equivalent N_{RUC} for a given N , we then simply equate $S_{\text{RUC}}(N_{\text{RUC}}) = S(N)$, yielding $N_{\text{RUC}} = (\sigma_1 N + (\sigma_0 - \eta_0))/\eta_1 = 0.52N + 1.76$.

With this system size equivalence established we can directly compare the SPAM-corrected Rydberg and RUC systems at their respective entanglement time and depth, to find the equivalent RUC two-qubit cycle fidelity that would match the Rydberg quantum simulator's evolution fidelity. By evolution fidelity, we refer to the fidelity at the entanglement time, up to preparation errors, which based on our validated error model is approximately given by $\exp(-\gamma(N)t_{\text{ent}}(N))$ from equation (17). We equate $F_{\text{RUC}}(N_{\text{RUC}}, d_{\text{ent}}) = \exp(-\gamma(N)t_{\text{ent}}(N))$ and then solve for F_{cycle} . As shown in Extended Data Fig. 7c, for the gate-set used in ref. ⁵, we find $F_{\text{cycle}} = 0.987(2)$, whereas for the SU(4) circuit we find $F_{\text{cycle}} = 0.9965(5)$, which is nearly independent of system size. Error bars originate from the uncertainty on the parameters of γ , t_{ent} , d_{ent} , F_0 and S .

Hamiltonian parameter estimation

The Hamiltonian parameter estimation scheme presented in Fig. 5 works by comparing measurements from experiment against an ensemble of target states from numerical simulations, each evolved with a different set of Hamiltonian parameters. For example, we can define a family of target states, which are parameterized by the Rabi frequency, Ω , as $|\psi(t, \Omega)\rangle = e^{-it\hat{H}(\Omega)/\hbar} |0\rangle^{\otimes N}$. 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 $F_c(t, \Omega) \approx \langle \psi(t, \Omega) | \hat{\rho}(t) | \psi(t, \Omega) \rangle$ will decay more quickly. To identify the best Rabi frequency yielding the highest fidelity, we define the time-integrated estimator defined as $\bar{F}_c(\Omega) = \frac{1}{T} \int_0^T dt F_c(t, \Omega)$, where T is the maximum interrogation time, and plot $\bar{F}_c(\Omega)$ as a function of Ω

(Fig. 5a). We analyse the sensitivity of this method in the Supplementary Information.

Target state benchmarking

Our fidelity estimation protocol can be used both to estimate the fidelity of performing some quantum evolution (Fig. 4) and to estimate the fidelity of preparing a target quantum state of interest (Fig. 5). In this modality, we assume the target state is prepared with some non-unity fidelity due to experimental imperfections, after which we apply an infinite effective temperature quench Hamiltonian and observe the resulting dynamics.

In Fig. 5d, the ideal state is the ground state at $\Delta/\Omega = 3$ and $V_{\text{nnn}}/\Omega = 0.26$, close to the phase transition between the disordered and $\mathbb{Z}_2$ -ordered states of the Rydberg Hamiltonian⁴². The imperfect state is taken to be an incoherent mixture composed of 50% each of the ground and first excited states. This state is then quenched with a Hamiltonian with parameters $\Omega/2\pi = 5.3$ MHz, $\Delta/2\pi = 2.8$ MHz, $C_6/2\pi = 254$ GHz μm^6 and $a = 3.75$ μm , with $2\pi \times \pm 1$ MHz random on-site disorder drawn from a uniform distribution.

In Extended Data Fig. 8, we numerically demonstrate fidelity estimation of various target states such as a cluster state, a Haar-random state of a two-dimensional Rydberg quantum simulator and a symmetry-protected topological ground state.

Specifically, in Extended Data Fig. 8a, we estimate the state preparation fidelity of a one-dimensional cluster state defined as

$$|\psi\rangle_{\text{cluster}} = \prod_{i=1}^{N-1} (\text{CZ})_{i,i+1} |+\rangle^{\otimes N} \quad (19)$$

where $(\text{CZ})_{i,i+1}$ is the two-qubit, controlled-Z gate acting on two adjacent qubits at site i and $i+1$, and $|+\rangle$ is the equal superposition of the $|0\rangle$ and $|1\rangle$ states. The imperfect quantum state is prepared by applying a global phase rotation to the ideal state such that the state overlap becomes $F = 0.5$. We then use an infinite effective temperature quench Hamiltonian given as $\hat{H}/\hbar = \hbar_x \sum_i (\hat{S}_i^x - 1.79\hat{S}_i^y + 4.64\hat{S}_i^x \hat{S}_{i+1}^x)$ to learn the state overlap via our F_c formula.

In Extended Data Fig. 8b, we estimate the state preparation fidelity of a pure Haar-random state generated from Rydberg atoms in a 4×4 two-dimensional square array. The imperfect state is prepared by applying a local phase rotation to a central qubit, yielding $F = 0.5$. For subsequent quench dynamics, identical Hamiltonian parameters are used as in the one-dimensional Rydberg benchmarking case (Fig. 4).

Finally, in Extended Data Fig. 8c, we estimate the state preparation fidelity of a symmetry-protected topological ground state prepared in a Rydberg ladder array realizing the Su–Schrieffer–Heeger topological model, following the approach of ref.⁵⁵. The imperfect state is prepared with a local phase error yielding a state overlap of $F = 0.5$. It is subsequently benchmarked by infinite effective temperature evolution

with the quench Hamiltonian chosen to be the combination of the identical interaction Hamiltonian, a random on-site disorder of strength 1 MHz and a detuned global drive with a Rabi frequency of 2 MHz and a detuning of 0.5 MHz.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Code availability

The code that supports the findings of this study is available from the corresponding authors upon reasonable request.

- Madjarov, I. S. et al. An atomic-array optical clock with single-atom readout. *Phys. Rev. X* **9**, 41052 (2019).
- Barredo, D., de Leseleuc, S., Lienhard, V., Lahaye, T. & Browaeys, A. An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays. *Science* **354**, 1021–1023 (2016).
- Endres, M. et al. Atom-by-atom assembly of defect-free one-dimensional cold atom arrays. *Science* **354**, 1024–1027 (2016).
- Page, D. N. Average entropy of a subsystem. *Phys. Rev. Lett.* **71**, 1291–1294 (1993).
- de Léséleuc, S. et al. Observation of a symmetry-protected topological phase of interacting bosons with Rydberg atoms. *Science* **365**, 775–780 (2019).

Acknowledgements We acknowledge experimental help from P. Scholl during the revision of this manuscript, as well as discussions with A. Deshpande and A. Gorshkov. We acknowledge funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF grant no. PHY-1733907), the NSF CAREER award (no. 1753386), the AFOSR YIP (no. FA9550-19-1-0044), the DARPA ONISQ programme (no. W911NF2010021), the Army Research Office MURI program (no. W911NF2010136), the NSF QLCI program (no. 2016245), the DOE (grant no. DE-SC0021951), the US Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator (grant no. DE-AC02-05CH11231) and F. Blum. J.C. acknowledges support from the IQIM postdoctoral fellowship. A.L.S. acknowledges support from the Eddleman Quantum graduate fellowship. R.F. acknowledges support from the Troesh postdoctoral fellowship. J.P.C. acknowledges support from the PMA Prize postdoctoral fellowship. H.P. acknowledges support by the Gordon and Betty Moore Foundation. H.-Y.H. is supported by the J. Yang & Family Foundation. A.K. acknowledges funding from the Harvard Quantum Initiative (HQI) graduate fellowship. J.S.C. is supported by a Junior Fellowship from the Harvard Society of Fellows and the US Department of Energy under grant contract no. DE-SC0012567. S.C. acknowledges support from the Miller Institute for Basic Research in Science.

Author contributions J.C., A.L.S., S.C. and M.E. conceived the idea and experiment. J.C. and A.L.S. performed the experiments and data analysis. J.C., A.L.S., J.S.C., D.K.M., H.-Y.H., H.P., F.G.S.L.B., S.C. and M.E. contributed to the underlying theory. J.C., A.L.S., I.S.M., X.X., R.F., J.P.C. and A.K. contributed to building the experimental set-up and data taking. J.C., A.L.S., S.C. and M.E. wrote the manuscript with input from all authors. S.C. and M.E. supervised this project.

Competing interests The authors declare no competing interests.

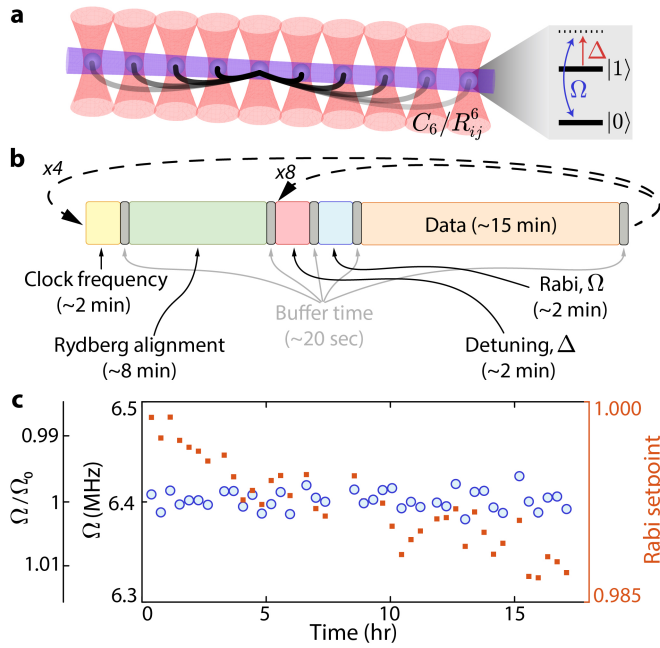
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-05442-1>.

Correspondence and requests for materials should be addressed to Soonwon Choi or Manuel Endres.

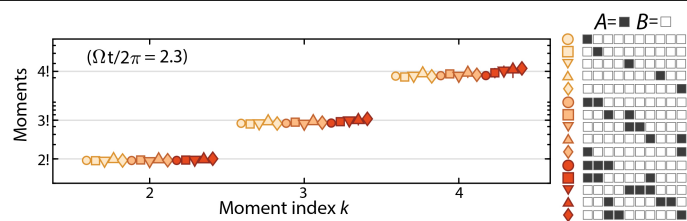
Peer review information Nature thanks Benoît Vermersch and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

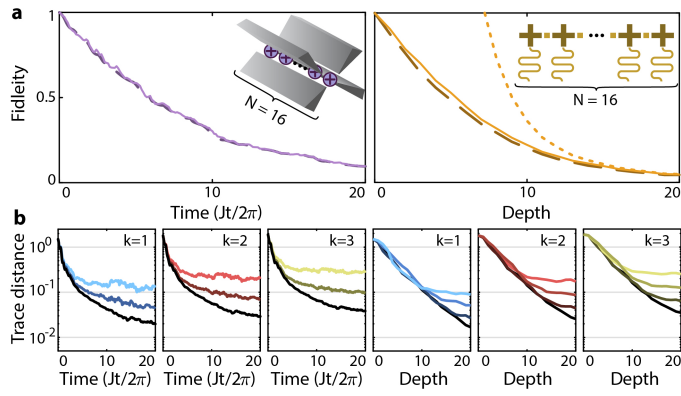


Extended Data Fig. 1 | Experimental system and parameter feedback.

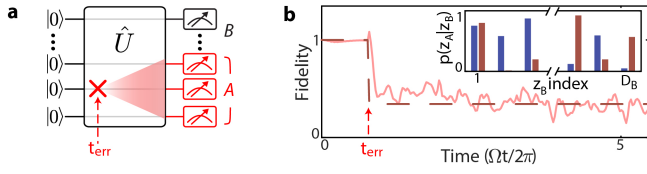
a, Illustration of a Rydberg quantum simulator consisting of strontium-88 atoms trapped in optical tweezers (red funnels). All atoms are driven by a global transverse control field (purple horizontal beam) at a Rabi frequency Ω and a detuning Δ (right panel). The interaction strength is given as C_6/R_{ij}^6 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 Δ , and the Rabi frequency Ω . **c**, Example of the interleaved automatic Rabi frequency stabilization over the course of ≈ 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 $< 0.3\%$, with a standard deviation of 0.15%. This same stability is seen over the course of multiple days with nearly continuous experimental uptime.



Extended Data Fig. 2 | Universality of moments of the projected ensemble. k th moments of the conditional probability distributions in Fig. 2b,d, evaluated at intermediate time ($\Omega t/2\pi = 2.3$) and for a variety of choices of subsystems (see panel on the right); we find a universal convergence to $\approx k!$, independent of subsystem choice, suggesting that a subsystem's projected ensemble converges to the uniform random ensemble irrespective of the details of placement, or connectivity. Error bars are the standard deviation over temporal fluctuations in moments near the evaluated time, as shown in Fig. 3a.

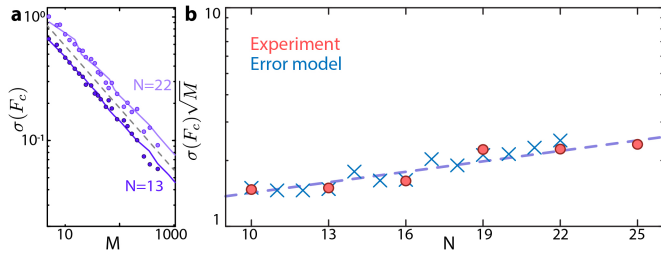


Extended Data Fig. 3 | Emergent randomness and benchmarking in other quantum systems. **a**, Fidelity estimation for the case of a trapped ion quantum simulator governed by chaotic Hamiltonian evolution (left) and a quantum computer implementing a random unitary circuit (RUC) (right); see Supplementary Information for simulation details. In both cases, we plot the many-body fidelity (dashed line), as well as our fidelity estimator, F_c (solid line); for the RUC case we also plot the more conventional linear cross-entropy-benchmark, F_{XEB} (dotted line). We find that F_c approximates the fidelity at much earlier times than F_{XEB} . **b**, Numerically computed trace distances between the projected ensemble of a two-qubit subsystem and the corresponding k -design. Results are shown for multiple different total system sizes: 10, 13, 16 for the trapped ion case, and 10, 12, 14, 16 for the RUC case, with darker colors corresponding to larger total system sizes.

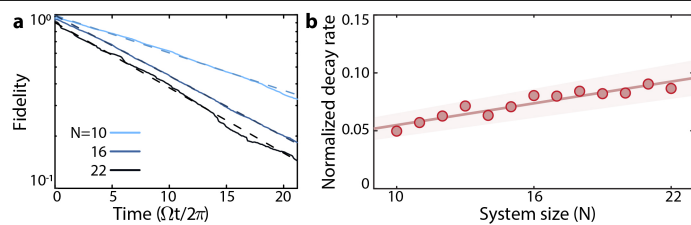


Extended Data Fig. 4 | Detecting errors during quantum evolution.

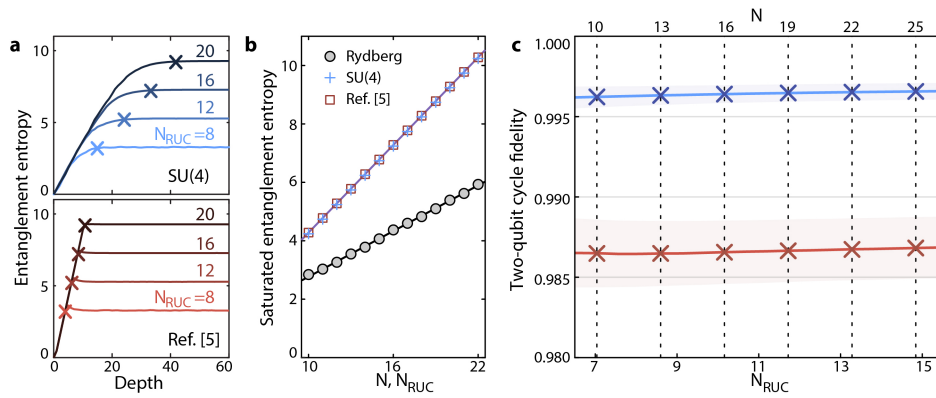
a, Schematic of noisy time evolution with an error occurring at time t_{err} . The influence of the local error propagates outward, affecting the measurement outcomes non-locally at a later time. **b**, Errors during evolution can be detected by correlating the measurement outcomes with an error-free, ideal evolution case. We numerically tested this by applying a local, instantaneous phase error to the middle qubit of an $N = 16$ atom Rydberg simulator at time $\Omega t_{\text{err}}/2\pi \approx 1$. The proposed fidelity estimator, F_c (solid line), accurately approximates the many-body overlap (dashed line) between states produced with and without errors, after a slightly delayed time. Inset: Conditional probability distributions in A before (blue) and after (red) the error, showing decorrelation.



Extended Data Fig. 5 | Finite sampling analysis for F_c . **a**, Statistical fluctuations of the fidelity estimator, F_c , at $N=13$ (dark purple) and $N=22$ (light purple), computed both using our ab initio error model (solid lines) and experiment (markers) evaluated with a finite number of M bitstring samples. Data are consistent with a $1/\sqrt{M}$ scaling, shown here as a guide to the eye (grey dashed line) **b**, Sample complexity of the fidelity estimator, evaluated at the N -dependent entanglement saturation time for the error model (blue crosses), and for the experimental data in Fig. 4d (red circles). A fit to the experimental data (dashed line) with functional form $\sigma(F_c)\sqrt{M} = a^N$ yields an estimate of $a = 1.037(2)$ (a similar fit to the error model yields an estimate of $a = 1.039(2)$).

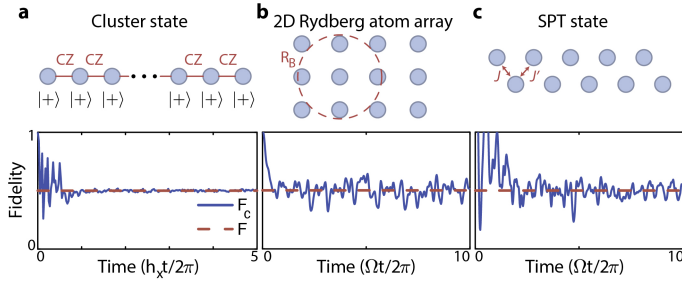


Extended Data Fig. 6 | Predicting fidelity scaling. **a.** We use our ab initio error model (which includes state preparation errors) to predict the fidelity decay rate as a function of system size. For various system sizes we plot the model fidelity (solid lines), as well as fits to exponential decay with an unconstrained value at $t=0$ (dashed lines), which we see are consistent with the time-dependent fidelity. **b.** For the range of system sizes for which our error model is readily calculable, we see the fidelity decay rate normalized by the Rabi frequency, $\gamma(N)\Omega/2\pi$ (markers), is consistent with a linear function of system size (red line). The shaded region comes from uncertainty in the fit parameters.



Extended Data Fig. 7 | Comparison to digital quantum devices executing random circuits. **a**, Numerical simulations of a one-dimensional digital quantum device implementing a random unitary circuit (RUC). Two different digital gate implementations are tested: a configuration based on the gate-set used in ref. ⁵ (bottom), and a configuration where each cycle is composed of parallel two-qubit SU(4) gates (top)⁴. Cross markers indicate when the half-chain entanglement entropy saturates. **b**, Due to the Rydberg blockade mechanism, as well as symmetries of the Rydberg Hamiltonian (Supplementary Information), an equal number of atoms in the Rydberg simulator, N , and qubits in the RUC, N_{RUC} , will not saturate to the same half-chain entanglement entropy. However, we can still find an equivalence by plotting the saturated entanglement entropy for the RUC (blue crosses for the SU(4) gate-set, open

red squares for gate-set from ref. ⁵) and for the Rydberg simulator (grey markers) as a function of their respective system sizes. We fit the results for the Rydberg simulator (black line), and plot the analytic prediction for the RUC⁵⁴ (purple line), from which we can write an equivalent N_{RUC} as a function of N , in the sense of maximum achievable entanglement entropy (Methods). **c**, For a given N (and equivalent N_{RUC}), we plot the SPAM-corrected, two-qubit cycle fidelity for an equivalently-sized RUC to match the evolution fidelity of our Rydberg simulator at the time/depth when entanglement saturates. Red lines, markers and crosses are for the gate-set of ref. ⁵, while blue are for the SU(4) gate-set. Shaded regions come from the error on fitting the various N -dependent parameters which enter this calculation (Methods).



Extended Data Fig. 8 | Applications to target state benchmarking.

a, Benchmarking of a one-dimensional cluster state, **b**, a pure Haar-random state benchmarked in a two-dimensional square Rydberg atom array, and **c**, a symmetry-protected topological (SPT) ground state prepared in a Rydberg ladder array realizing the Su-Schrieffer-Heeger topological model⁵⁵. In **a**, CZ denotes a controlled-Z gate and $|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$. In **b**, R_B denotes the Rydberg blockade radius within which more than a single excitation is not allowed^{24–26}. In **c**, J and J' are the alternating coupling strengths of a two-leg ladder array, respectively. In all cases, $N = 16$ qubits are used, and imperfect quantum states are prepared via phase rotations such that the many-body fidelity overlap becomes 0.5 (red dashed line). Additionally, chaotic evolution is performed such that the initial state is at infinite effective temperature to apply our F_c formalism (blue solid lines) (Methods).