

Statistical learning on randomized data to verify quantum state approximate k -designsKaustav Mukherjee^{1,2,*} Sarah Chehade^{1,2} Lorenzo Versini^{3,4} Karim K. Alaa El-Din^{3,4}
Florian Mintert^{3,5} and Rick Mukherjee^{1,2,3}¹Department of Physics and Astronomy, *University of Tennessee*, Chattanooga, Tennessee 37403, USA²UTC Quantum Center, *University of Tennessee*, Chattanooga, Tennessee 37403, USA³Blackett Laboratory, *Imperial College London*, London SW7 2AZ, United Kingdom⁴Atomic and Laser Physics (ALP), Department of Physics, *University of Oxford*, Parks Road, Oxford OX1 3PU, United Kingdom⁵*Helmholtz-Zentrum Dresden-Rossendorf*, Bautzner Landstraße 400, 01328 Dresden, Germany

(Received 3 November 2025; accepted 4 June 2026; published 29 June 2026)

Random ensembles of pure states have proven to be extremely important in various aspects of quantum physics such as benchmarking the performance of quantum circuits, testing for quantum advantage, studying many-body thermalization, and the black hole information paradox. Although generating a truly random quantum ensemble is experimentally challenging, approximate realizations are equally valuable and are known to emerge naturally in a variety of physical models, including Rydberg setups. These are referred to as approximate quantum state designs, and verifying their degree of randomness can be a measurement-intensive task, similar to performing full quantum state tomography on many-body systems. In this theoretical work, we present a measurement scheme and analysis techniques to validate the degree of randomness of a quantum ensemble generated by a simulated experimental setup. This is achieved by translating the information residing in the complex many-body state into a succinct representation of classical data using projective measurements in randomly chosen bases, which is then processed using methods of statistical inference such as maximum-likelihood estimation and neural networks, benchmarked against the predictions of shadow tomography. Our scheme only requires individually addressed single-qubit operations to be performed in order to be employed, making it applicable for a range of physical platforms.

DOI: [10.1103/16yg-xqtd](https://doi.org/10.1103/16yg-xqtd)

I. INTRODUCTION

Random number generation has a wide range of applications, including information security [1], networking [2,3], board games, lotteries, and gambling [4,5]. However, it is very costly and almost impossible to generate a genuinely random string of numbers in the deterministic classical world. For practical purposes, pseudorandom number generators are used where a deterministic computer algorithm creates a sequence of numbers that appear to be statistically random when compared to a fully independent set of random variables [6–9]. Thus, the average value of the pseudorandom distribution will not be distinguishable from the average value of a perfect random distribution. Similarly, in combinatorial mathematics, one can define classical k -designs that are indistinguishable from truly random distributions up to k -moments. [10–13].

Quantum k -designs are extensions of the classical k -design concept in which the probability distributions of unitary operators or quantum states, averaged over a polynomial function up to degree k , are identical to the averages taken over a unique group-invariant measure, in this case the Haar

measure. Quantum k -designs find many applications in quantum information theory, which include estimating fidelity [14], performing state tomography [15], implementing cryptography [16,17], randomized encoding and benchmarking of quantum circuits [18–22] and simulators [23], checking for entanglement [24], generating topological order [25], and studying the black hole information paradox [26,27]. Recent studies also reveal the relevance of quantum k -designs in the thermalization process for many-body systems [28–30]. However, generating quantum designs on any quantum processor is challenging and requires a high degree of control over a large set of noisy parameters [14,15,31–34], as the number of gates grows exponentially with the number of qubits. Alternatively, *approximate quantum state designs* (see Fig. 1) are relatively easy to construct, and by definition, they are distributions of states that approximate their exact counterparts.

The degree of randomness in an ensemble of states is characterized by estimating its closeness to the relevant quantum k -design, a task that naturally depends on the information available about the ensemble of states. However, gathering such information can be an experimentally resource-intensive task, akin to quantum state tomography for large systems [35–38]. Over the years, there has been a push toward efficient quantum state tomography [38–41]. In the same spirit, this work deals with the efficient and accurate characterization of numerically simulated random ensembles of states.

The characterization of approximate quantum k -designs investigated here involves two key steps: (1) sampling the

*Contact author: Kaustav-Mukherjee@utc.edu

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

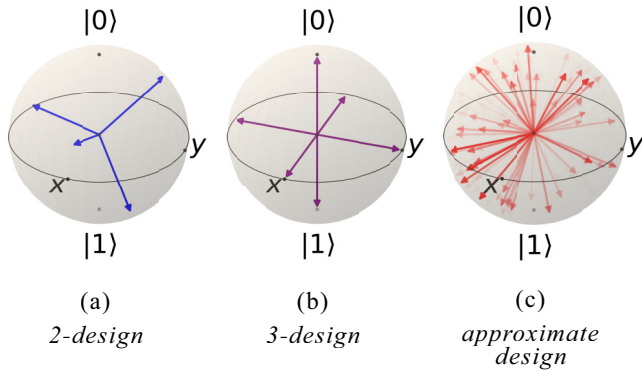


FIG. 1. The figure depicts the quantum k -designs for a single qubit in a Bloch sphere for $k = 2, 3$ in panels (a) and (b), respectively, and an ensemble of states that approximates an arbitrary k -design in panel (c). Less transparent vectors in panel (c) occur with higher probability within the ensemble of states.

ensemble of states using projective measurements in the random bases and (2) applying statistical learning methods on the measurement data for postprocessing. Using a prototypical platform of interacting spins in a Rydberg setup, we numerically construct the ensemble of states by implementing a protocol that involves splitting the system into two subsystems: A and B , realizing an approximate quantum state design in A and treating B as an auxiliary system used for random ensemble generation [28]. Measuring subsystem B in the computational basis projects the joint system into a corresponding ensemble of postmeasurement states on subsystem A . We then propose randomly selecting the basis of projective measurement for subsystem A while measuring subsystem B in the computational basis. It is well established that utilizing the statistical correlations arising from randomized measurements forms a flexible toolbox useful for estimating various quantum properties of the many-body quantum state, such as purity, entanglement, and out-of-time-ordered correlations [42]. Next, we leverage tools of machine learning, which are based on statistical inferences, to estimate the parameters that describe the ensemble of states. Using both maximum-likelihood estimation (MLE) and restricted Boltzmann machines (RBM), we demonstrate that this routine significantly reduces the number of measurements needed to accurately represent the ensemble of states compared to either full state tomography or a frequentist method as implemented in Ref. [28]. In addition, we show that the maximum-likelihood approach outperforms even efficient shadow tomography [40,43,44] for all cases tested, while the restricted Boltzmann machines achieve this performance in the high-data regime. Finally, we note that our results are applicable to any quantum platform that can generate random ensembles of states.

II. THEORY

A. Constructing approximate quantum state k -design using projected ensembles

A uniform distribution of all pure states in a Hilbert space is defined by the Haar ensemble [16]. Averaging arbitrary functions across the ensemble is impossible in practice, due to the finite number of samples that can be collected. Instead,

it is possible to use quantum state designs that reproduce the universal properties of randomness of the Haar ensemble up to a certain order. However, these designs are typically constructed approximately using random unitary quantum circuits [31,33,45–47], which require sufficient depth of the quantum circuit, resulting in polynomial scaling with the number of qubits, thereby contributing to the overall infidelity of the operation.

Another approach to generating approximate quantum state designs involves performing local measurements on many-body states that evolve under a time-dependent Hamiltonian in quantum many-body systems [28,29]. This approach of constructing approximate k -designs, which is also referred to as the *projected ensemble method*, is advantageous due to its generality, as it can be readily implemented in experimental platforms that realize Ising models [28]. We consider a system of N spins, which are partitioned into two subsystems: A and B , with N_A and N_B spins, respectively, such that $N_A \ll N_B$. In this work, we numerically simulate the works [28,29] to generate the data for approximate quantum state k -designs, which are then characterized using statistical learning methods.

Using the projective measurement approach, an ensemble of pure states is generated by performing local measurements on subsystem B for given projected states in complementary subsystem A . The naive brute-force approach to characterize the ensemble of states would be to perform measurements in all directions (X, Y, Z), on both subsystems A and B . This allows for reconstruction of the full many-body state, which captures the complete information about the entire system. However, this procedure is prohibitively expensive, especially for larger systems. Alternatively, one may measure both the subsystems, solely in the computational or Z basis as performed in the original experiment [28]. Measuring the entire system in the computational basis results in the loss of useful information (such as phase), which can be useful for characterizing the ensemble of pure states. Instead, we propose measuring each spin in subsystem A in a randomly chosen basis from X, Y , or Z , while measuring subsystem B uniformly in the Z basis. This measurement scheme results in bitstrings $r_A \in \{r_A^{(1)}, r_A^{(2)}, \dots\}$ for $r_A^{(i)} \in \{x_A^{(i)}, y_A^{(i)}, z_A^{(i)}\}$ for subsystem A and z_B for subsystem B . Each bitstring contains a string of bits 0 or 1 representing the state of the individual spin when measured in the chosen basis. The scheme of measurements used in A is referred to as *measurement in a randomly chosen basis*, as discussed in more detail in Appendix B. The many-body state of the full system is denoted by $|\Psi\rangle$, and the state of subsystem A conditioned on a particular bitstring z_B being measured is given by

$$|\Psi_A(z_B)\rangle = (\mathbb{1}_A \otimes \langle z_B |) |\Psi\rangle / \sqrt{p(z_B)}, \quad (1)$$

where $p(z_B) = \langle \Psi | (\mathbb{1}_A \otimes |z_B\rangle\langle z_B|) | \Psi \rangle$ is the probability to measure a specific bitstring z_B . The marginal probability of measuring the subsystem A in a particular state r_A for a given choice of measurement basis is given by

$$\begin{aligned} p(r_A) &= \sum_{z_B} p(z_B) p(r_A | z_B), \\ &= \sum_{z_B} p(z_B) |\langle r_A | \Psi_A(z_B) \rangle|^2, \end{aligned} \quad (2)$$

TABLE I. Summary of different approaches for characterizing quantum state ensembles. The key methodological features of the frequentist, MLE, CRBM, and shadow-tomography approaches are shown in the top three rows. Each method differs in the measurement basis, statistical representation of the state, and the scaling of the sampling requirements. The quoted error scalings refer to different notions of statistical error. Namely, for shadow tomography, the $1/\sqrt{|\mathcal{D}|}$ scaling corresponds to the root-mean-square error of unbiased estimators constructed from independent measurement outcomes. For MLE, the $1/|\mathcal{D}|$ scaling refers to asymptotic mean-squared error (MSE) under standard regularity assumptions, rather than a nonasymptotic bound on state distance. For CRBM, the reported scaling reflects empirical convergence observed in practice rather than a general theoretical guarantee. See text and Appendix D for more details. The optimal dataset size for characterizing quantum state ensembles with each approach is obtained from Fig. 4.

Feature	Frequentist	Max. likelihood	CRBM	Shadow tomography
Measurement basis (A)	Z basis	Random basis	Random basis	Random basis
State representation $ \Psi_A(z_B)\rangle$	$\sum \sqrt{p_m} m\rangle$ $p_m \rightarrow \text{prob. of } m\rangle$	$\sum_{z'_A} \sqrt{\frac{p(z'_A z_B)}{p(z_B)}} z'_A\rangle$ $p(z'_A, z_B) \rightarrow \text{Joint prob.}$	$\sqrt{p_\theta(z)} e^{i\phi_\mu(z)}$ $(p_\theta, \phi_\mu) \rightarrow (\text{Amp.}, \text{phase})$	$\hat{\rho}_A^{(n)} = \bigotimes_i (3 a_i^{(n)}\rangle\langle a_i^{(n)} - I)$ $ a_i^{(n)}\rangle \rightarrow \text{Measurement}_{i \text{ qubit}}^{n \text{ shot}}$
Estimation error	–	$O(1/ \mathcal{D})$ [67] (asymptotic MSE)	$O(1/\sqrt{ \mathcal{D} })$ [68] (empirical)	$O(1/\sqrt{ \mathcal{D} })$ [28] (estimator rms)
Optimal dataset size				
($k = 2$)	–	45	576	3000
($k = 3$)	–	72	1152	3000

where $p(r_A|z_B)$ is the conditional probability distribution. Thus, the ensemble of states is a collection of all such conditional states of subsystem A along with its corresponding probabilities to measure a given bitstring z_B and is written as

$$\mathcal{E} = \{(|\Psi_A(z_B)\rangle, p(z_B))\} \quad \forall z_B \in [0, 2^{N_B}). \quad (3)$$

The k th moment of the ensemble $\mathcal{E}$ is defined as follows:

$$\rho_{\mathcal{E}}^{(k)} = \sum_{z_B} p(z_B) (|\Psi_A(z_B)\rangle\langle\Psi_A(z_B)|)^{\otimes k}, \quad (4)$$

while the k th moment of the Haar ensemble for a Hilbert space of dimensionality d is defined as $\rho_{\text{Haar}}^{(k)} = \int_{\psi \sim \text{Haar}(d)} d\psi (|\psi\rangle\langle\psi|)^{\otimes k}$, which has a closed analytical form as detailed in Appendix A. Here, $|\psi\rangle$ refers to random pure states that are obtained from the unitarily invariant (Haar) measure on the unit sphere in d -dimensional Hilbert space. The degree of closeness between $\mathcal{E}$ and the Haar ensemble is quantified using the 1-norm, i.e., the trace distance, between the two ensembles, which is defined as

$$\delta_{(k)} = \frac{1}{2} \text{Tr} \left(\sqrt{(\rho_{\mathcal{E}}^{(k)} - \rho_{\text{Haar}}^{(k)})^\dagger (\rho_{\mathcal{E}}^{(k)} - \rho_{\text{Haar}}^{(k)})} \right). \quad (5)$$

Thus, for $\delta_{(k)} \ll 1$, the many-body state $|\Psi\rangle$ yields an ensemble of pure states on the subsystem whose probability distribution is indistinguishable from the Haar ensemble up to the k th moment, thereby forming an approximate quantum state k -design. The efficient and accurate characterization of (approximate) quantum state k -designs, which is related to estimating the value of $\delta_{(k)}$, depends on the type of information encoded in the ensemble of states $\mathcal{E}$. This, in turn, relies on the actual details of the measurement protocol.

In this work, we use four different protocols to characterize quantum designs: (1) frequentist method, (2) shadow tomography [40], (3) complex-restricted Boltzmann machine (CRBM) [48,49], and (4) MLE, of which the latter two demonstrate greater efficiency with respect to data acquisition. Each of the four protocols uses a distinct ansatz to reconstruct the state $|\Psi_A(z_B)\rangle$ for each z_B from repeated measurement

outcomes: (1) The frequentist method reconstructs the state by determining the coefficients of $|\Psi(z_B)\rangle$ based on probabilities of outcomes in the computational basis; (2) shadow tomography estimates the density operator from measurement outcomes in random bases leveraging unbiased statistics (U -statistics; see Appendix D 4); (3) the CRBM employs a bipartite neural network to parametrize the state from measurements in a randomly chosen basis; and (4) the MLE directly optimizes the coefficients of $|\Psi(z_B)\rangle$ from measurement outcomes in random basis. Table I summarizes the measurement basis, state representation $|\Psi_A(z_B)\rangle$, and estimation errors, highlighting how the error decreases with the size of the measurement dataset for each protocol. The underlying theory of these four protocols and their implementation for this work are described in detail in Appendix D. In the next section, we describe the theory behind the Rydberg simulator used to generate the approximate k -designs similar to Refs. [28,29].

B. Generating random ensemble states with Rydberg simulator

Rydberg atoms are neutral atoms with their valence electron placed in a highly excited state, possessing large dipole moments [50,51]. The characteristic strong dipole-dipole interaction between neighboring Rydberg atoms has been exploited for a wide variety of applications that include many-body physics [52–54], quantum simulation [55–58], and quantum computing [59,60]. In this work, a linear chain of trapped atoms is considered, which are initialized to a state with all atoms in their ground state $|g\rangle$. The laser with detuning Δ and Rabi frequency Ω excites the individual atoms to their Rydberg state, denoted by $|e\rangle$. The system evolves according to the time-independent Hamiltonian

$$\hat{H} = \frac{\hbar\Omega}{2} \sum_{i=1}^N \hat{\sigma}_i^x - \hbar\Delta \sum_{i=1}^N \hat{n}_i + \hbar \frac{C_6}{a^6} \sum_{j>i} \frac{\hat{n}_i \hat{n}_j}{|i-j|^6}, \quad (6)$$

where $\hat{\sigma}_i^x = |g\rangle_i\langle e|_i + |e\rangle_i\langle g|_i$ and $\hat{n}_i = |e\rangle_i\langle e|_i$. Two atoms in their Rydberg states interact with each other via the van der Waals interaction, whose strength is given by the van

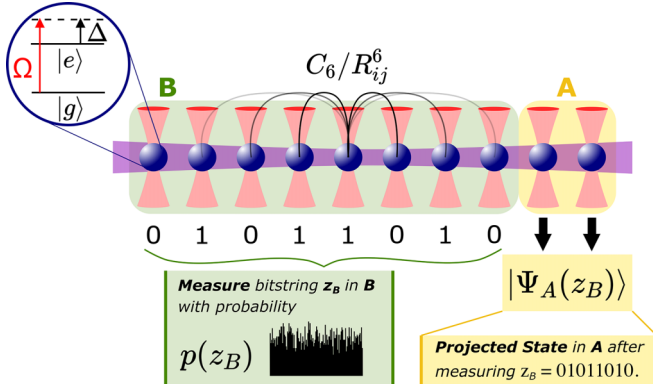


FIG. 2. Simulated Rydberg setup: $N = 10$ Rydberg atoms (blue spheres) trapped by optical tweezers (red cones), which are globally driven by a laser beam (purple) characterized by Rabi frequency Ω and detuning Δ (shown as inset). The first eight atoms are treated as subsystem B whose projective measurements in the computational basis provide the bitstring z_B . The remaining two atoms form the subsystem A described by the wave function $|\Psi_A(z_B)\rangle$, which depends on the outcome z_B .

der Waals coefficient C_6 . Simultaneous excitations of nearest neighboring atoms to their Rydberg states are suppressed due to the Rydberg blockade mechanism [61]. This effect allows us to work with the reduced Hilbert space, which makes the numerical simulations more tractable.

C. Numerical simulation of the Rydberg experiment [28]

For the numerical simulation, we consider a system of $N = 10$ atoms, partitioned into two subsystems A and B with $N_A = 2$ and $N_B = 8$ in Fig. 2. The many-body system is initialized in the ground state $|ggg\cdots\rangle$, which has zero energy expectation value with respect to $\hat{H}$. The initial state evolves under $\hat{H}$ to a final state $|\Psi(t)\rangle = e^{-\frac{i}{\hbar}\hat{H}t}|g\cdots g\rangle$ at time t , well before any decoherence process, such as spontaneous decay of the Rydberg state, becomes significant. A natural consequence of the interacting many-body Rydberg system is the transition from the initially ordered state to chaotic dynamics, characterized by the extensive occupation of a broad manifold of eigenstates across the Hilbert space. This is depicted in the conditional probabilities shown in Fig. 3. The evolved state of the many-body system $|\Psi(t)\rangle$ at time t is expressed as $\sum_{z_B} \sqrt{p(z_B)} |\Psi_A(z_B)\rangle \otimes |z_B\rangle$, which provides the dataset for the ensemble and is used for characterization in this work. For the numerical implementation of the simulator, we closely follow the experimental setup described in Ref. [28], which serves as our benchmarking platform, using the following parameters: lattice spacing $a = 3.3 \mu\text{m}$ apart, $C_6 = 126 \text{ GHz } \mu\text{m}^6$, $\Delta/(2\pi) = 0.9 \text{ MHz}$, and Rabi frequency $\Omega/(2\pi) = 4.7 \text{ MHz}$ [28]. Practical implementation of measurement in randomly chosen bases with single-qubit rotations in Rydberg systems can be achieved in current experiments [62–64].

Now, whether a chaotic final state can provide a projective ensemble of states that statistically represent a quantum state k -design can be addressed in two different ways. First, a random ensemble of states is known to possess certain

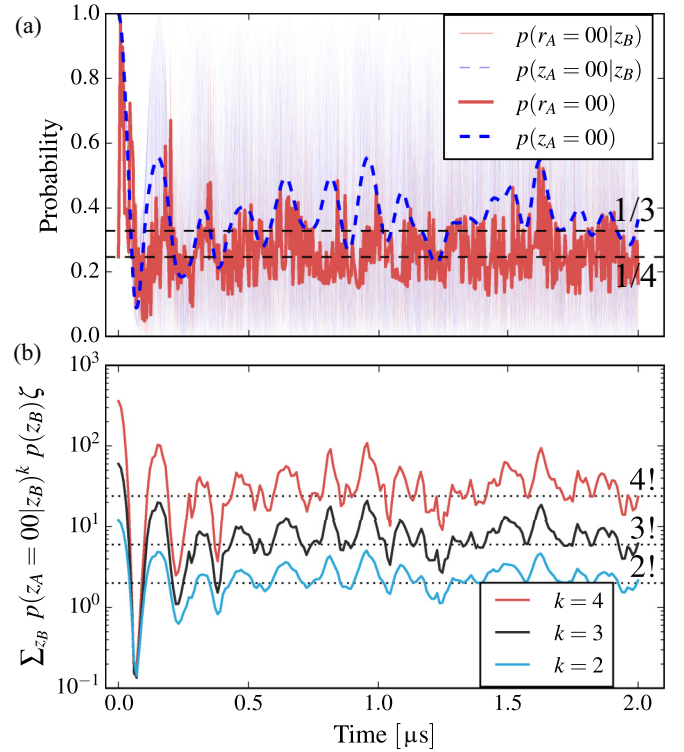


FIG. 3. (a) Numerical simulation of the Rydberg dynamics with partition size $N_A = 2$ and system size $N = 10$: Conditional probabilities $p(z_A = 00|z_B)$ (blue) and $p(r_A = 00|z_B)$ (red) are plotted as a function of time for finding A in state $|00\rangle$ given a measurement z_B . Lighter lines correspond to the plot of $p(z_A = 00|z_B)$ and $p(r_A = 00|z_B)$ over time for each of the possible z_B outcomes, while the oscillating bold lines represent marginal probability $p(z_A = 00)$ (blue) and $p(r_A = 00)$ (red), reaching a steady-state value of 0.33 and 0.25, respectively. (b) Plot of the convergence of higher moments of the conditional probability distribution $p(z_A = 00|z_B)$ for numerically generated measurements with system size $N_A = 2$ and subsystem B size $N_B = 8$, rescaled by $\zeta = D_A \times (D_A + 1) \cdots (D_A + k - 1)$ as indicated in the main text.

universal statistical properties such as the steady-state behavior of the marginal probability distributions $p(r_A)$ [see Eq. (2)] at long times [29]. Figure 3 (a) ascertains the expected convergence of the first moment (mean) to $1/D_A$, i.e., $1/3$ in the Z basis and $1/4$ in a randomly chosen basis. Here, D_A denotes the dimensionality of the Hilbert space for the subsystem A . For a subsystem of two qubits, the blockade interaction prohibits Rydberg excitation next to each other, thereby resulting in $D_A = 3$ possible states in the Z basis, i.e., $z_A \in \{|00\rangle, |01\rangle, |10\rangle\}$. In contrast, in a randomly chosen basis, all four states contribute, leading to $D_A = 4$ with $r_A \in \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, despite the presence of the Rydberg blockade. Figure 3 (b) shows the subsequent higher moments of the conditional probability distributions, which are defined as $\sum_{z_B} p(z_B) p(z_A = 00|z_B)^k$ to values $k!/(D_A \times (D_A + 1) \cdots (D_A + k - 1))$ at long times. The curves in panel (b) are rescaled by a factor $\zeta = (D_A \times (D_A + 1) \cdots (D_A + k - 1))$, which thus approaches $k!$. Although Fig. 3(b) was done for measurements in the Z basis, the saturation of conditional probabilities at long times holds true independent of the basis

chosen for projective measurements. The saturation of conditional probabilities with projective measurements in the Z basis has been verified in the recent experiment [28] and is the first signature that the ensemble of states is approaching a Haar-random distribution.

However, a more direct approach for verifying that a certain ensemble state forms an approximate quantum k -design is to demonstrate that the trace distance $\delta_{(k)}$, which is defined in Eq. (5), approaches zero. To evaluate the trace distance [Eq. (5)], we first construct the ensemble in Eq. (3) and its k th moment in Eq. (4) using the state vector $|\Psi_A(z_B)\rangle$ and probabilities $p(z_B)$, which are derived from the time evolution of Eq. (6), followed by the simulated projective measurements on subsystem B . The construction of $\delta_{(k)}$ is validated against well-established results for different system sizes in Appendix C. The evaluation of the trace distance generally requires full knowledge of the quantum state, rendering it experimentally exhaustive. To overcome this limitation, we demonstrate in the next section how statistical learning models can be employed to estimate $\Psi_A(z_B)$, and thereby the trace distance, from a limited number of measurements.

III. RESULTS

The need for accurate quantum state tomography for ensemble states is expensive in measurement cycles. It has been proven that for an N -body quantum system, the classical data obtained from randomized measurements can approximate all reduced density matrices with fixed number of spins up to ϵ accuracy with $\ln(N)/(\epsilon^2)$ measurements [42,65]. Following this up with efficient postprocessing of the classical data with methods such as maximum-likelihood optimization or RBM can potentially allow future experiments to estimate trace distances between the fitted many-body wave function and Haar-random distribution. The results of our approach are compared against the predictions of shadow tomography, where the density matrix is reconstructed using a computationally more efficient U_1 statistics with negligible bias for the sample size considered here, and is discussed in more detail in Appendix D 4.

Figure 4 compares the performance of different approaches in terms of the trace distance with respect to the size of the measurement dataset. Estimates calculated from each of the different methods m in the j th repetition, denoted by $\delta_{(k)}^{m,j}$, are compared to the exact trace distance $\delta_{(k)}^e$, which is calculated numerically using the true state vector $|\Psi(t)\rangle$ obtained by numerically solving the dynamics of the Rydberg system. The value of $\delta_{(k)}^e$ will in general be nonzero, since the system we are studying produces only *approximate* state designs. The comparison between $\delta_{(k)}^e$ and $\delta_{(k)}^{m,j}$ for each method m is quantified using the mean absolute error, defined as

$$\epsilon_{\text{MAPE}} = \frac{1}{(M \times T)} \sum_{j=1}^M \sum_{i=1}^T \left| \frac{\delta_{(k)}^{m,j}(t_i) - \delta_{(k)}^e(t_i)}{\delta_{(k)}^e(t_i)} \right|, \quad (7)$$

where $\delta_{(k)}^{m,j}(t_i)$ is the value of the trace distance obtained using Eq. (5) for dataset-based method m evaluated at time step t_i .

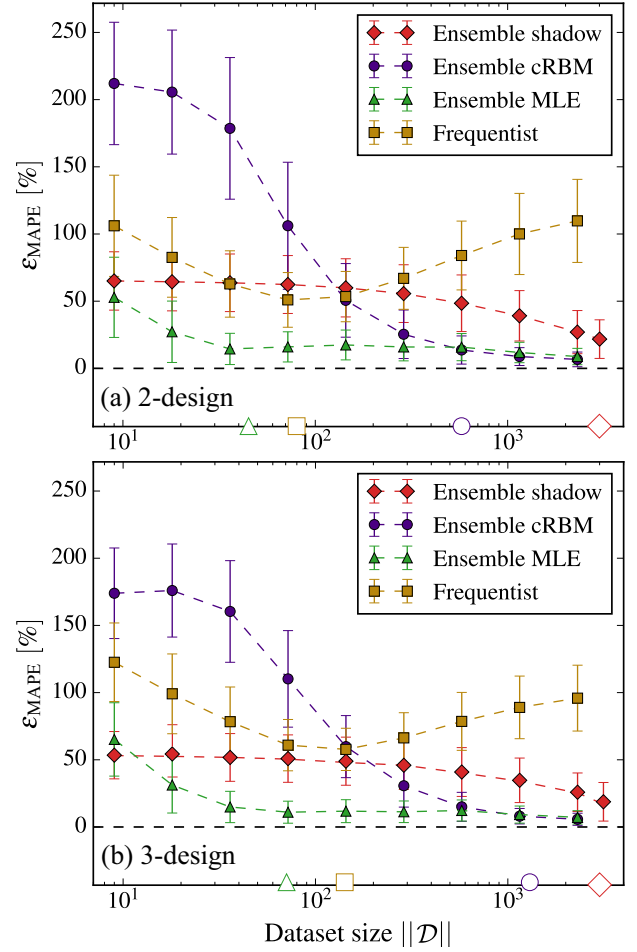


FIG. 4. Plot of the mean absolute percentage error as defined in Eq. (7) for (a) $k = 2$ and (b) $k = 3$ with each of the four methods. For statistical robustness, the mean percentage error is evaluated by averaging across 21 equally spaced time steps in the thermalized regime from 0.4 and 1.4 μs and across ten repetitions for every time step of the numerical experiment of system size $N = 10$ with $N_A = 2$. The sample size on the x axis corresponds to the number of measurements made in each repetition, and the empty markers correspond to the optimal dataset sizes for 2/3-design (Shadow: 3000/3000, CRBM: 576/1152, MLE: 45/72, Frequentist: 80/144) in the range explored.

The label j indexes the repetition of the numerical experiment. Thus, Fig. 4 shows the efficacy of the chosen method m in characterizing the ensemble of states as a quantum k -design with respect to the number of measurements needed for $k = 2$ and $k = 3$, shown in panels (a) and (b), respectively. For statistical robustness, the mean of the relative error is calculated over $T = 21$ time steps and $M = 10$ repetitions of the sampling process. In Fig. 4, we find that the best method for characterizing the ensemble of states is the MLE method, showing a relatively low value of the error ϵ_{MAPE} across the range of dataset sizes studied in this work, including in the low-sample regime. The CRBM approach shows poor results in the low-sample regime but improves rapidly as more measurements become available ($||D|| > 1000$). CRBM performs much worse than MLE in the low-sampling regime because the actual number of samples that can be used for training the CRBM is

much lower than the size of the available dataset $\mathcal{D}(z_B)$. This is related to a technical constraint where the algorithm requires at least one measurement to be in the Z basis (which is hardwired in the software implementation [66]) for subsystem A, which is sometimes missed in the random measurement scheme. Removing this constraint in future implementations of the software will potentially help in making better predictions. Shadow tomography performs better than CRBM in the low-sampling regime, but its error ϵ_{MAPE} converges slower than CRBM and MLE. It appears that reconstructing the density matrix directly from the measurements in randomly chosen bases, as performed in shadow tomography (refer to Appendix D4), requires more measurements to achieve the same level of accuracy in characterizing quantum k -designs compared to the maximum-likelihood method. Finally, the standard frequentist approach exhibits a consistent overestimation of the trace distance throughout the investigated dataset sizes, with no significant improvement with increasing sample size. This is because in the frequentist approach, we have only Z-basis measurements for subsystem A, as is typical in conventional Rydberg experiments, which implies that it does a poor job of capturing the full information of the state function. Although all approaches perform better in the high measurement regime, the MLE method has better accuracy, and it is a few orders of magnitude faster to train. The empty markers on the x axis indicate the minimum number of simulated measurements needed at each time step in order to get the best possible result for each of the algorithms within the range of dataset sizes analyzed. These are also summarized in Table I and are used to compare each of the methods in Fig. 5.

Figure 5 shows the trace distances $\delta_{(2)}$ and $\delta_{(3)}$ as a function of time in order to verify the quantum state 2- and 3-designs, respectively. The black line is the trace distance calculated using $|\Psi_A(t)\rangle$, which corresponds to the case where there is complete knowledge of the state. The other lines show the estimations obtained with each of the three algorithms using simulated measurements. The algorithms were applied to $M = 10$ independent simulated datasets at each time point, and the plot shows the mean (solid line) and standard deviation (shaded area) across the repetitions. In the initial times, all four algorithms display severe underestimation of the $\delta_{(2,3)}$ value. However, in the steady-state regime (after about 0.25 μs), the estimates of the trace distance from the MLE, CRBM, and shadow methods approach the true value (solid black line), clearly outperforming the frequentist approach. Apart from predicting the average value for $\delta_{(k)}(t)$, it is interesting to find that these four methods also capture some of the fluctuations in the trace distance observed using the exact method. The shadow tomography does not perform as well as MLE, despite the fact that the classical shadow method has a 2 orders of magnitude larger number of measurements for the trace distance estimation compared to MLE (see also Fig. 4). We suggest an explanation for this, which is that CRBM and MLE attempt the reconstruction of a state vector, whereas shadow tomography reconstructs a density matrix. The increased number of parameters that need to be estimated may be responsible for the higher number of samples required for the characterization of the system using the classical shadow method.

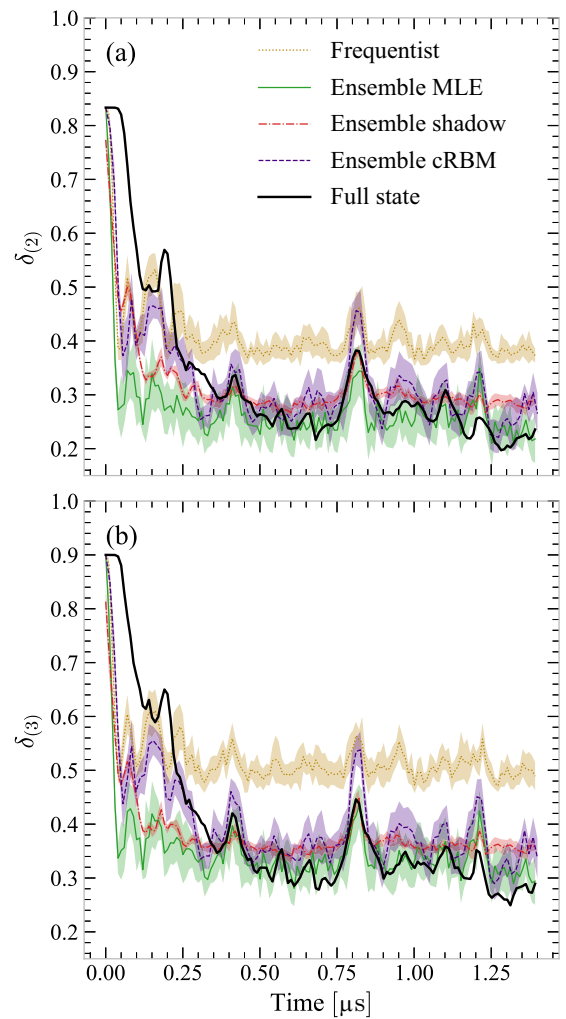


FIG. 5. The figure shows trace distances estimated for 2-design ($\delta_{(2)}$) and 3-design ($\delta_{(3)}$) using different algorithms as a function of time. Each algorithm was performed on a dataset of size marked on the x axis of Fig. 4. For statistical robustness, at each time, the simulated experiment was repeated 10 times. The mean reconstructed trace distance is shown with a line, and the standard deviation is shown as a shaded region. The black line shows the target values, calculated using the simulated time-dependent wave function without approximation.

IV. DISCUSSION

The complexity of large quantum systems makes them difficult to characterize accurately, and full state tomography is inefficient as it requires an exponentially large number of experimental runs, which also amounts to an exponentially large amount of data to process. Improving the efficiency of quantum state tomography, in particular for large systems, is a constant challenge and an active area of research [38,39]. Characterizing the randomness of an ensemble of states is no exception to this plight. Full characterization of a quantum state density matrix in a Hilbert space of dimensionality D necessitates at least the preparation of $O(D^2)$ independent copies of the system [43,69]. However, the Hilbert space of N interacting spins, such as that of the Rydberg setup, has dimensionality $D = 2^N$, in which case the full

characterization of a quantum state density matrix ρ will require $O(D^2) = O(2^{2N})$ independent copies of the system, which is still an exhaustively large number of copies.

Often, a far less complete description of the state is adequate in certain cases, such that the number of experimental runs and the amount of data needed is drastically reduced. The key idea is to translate the information about the many-body state of a system into classical data by performing measurements in the random bases. We sacrifice some information to make characterization feasible, aiming to strike a balance between data requirements and accuracy. In this regard, classical shadow tomography has a relatively reduced scaling to $O(\log^4(\mathcal{M}) \log(D))$ [70], where $\mathcal{M}$ are the two-outcome measurements [70]. Measurements on a randomized basis can reduce this scaling further to $O(\log(\mathcal{M}))$ independent copies [40,42] and are at the heart of the modern classical shadow methods [40]. However, depending on the many-body quantum state that is being characterized, relevant information can be lost, leading to an inaccurate description of the quantum state. In such cases, statistical learning methods can be used to make accurate predictions with low estimation error, as outlined in Table I and, in certain cases, more efficient predictions of the quantum state, as demonstrated in this work on characterization of emergent random ensembles of pure states.

This work uniquely accomplishes the efficient characterization of ensembles of random states by combining the benefits of random basis state measurements and statistical learning methods. The ability to identify information in reduced datasets that would otherwise be missed by naive observations is one of the appealing features of this kind of learning [71]. In contrast to the shadow tomography, which requires more than 10^3 measurements to converge even with efficient U_1 statistics, MLE (~ 50 – 70) and CRBM (~ 500 – 1000) require significantly fewer measurements. While both methods, MLE and CRBM, yield more accurate estimations, the current implementations of CRBM are computationally more time-consuming by orders of magnitude when compared to the MLE approach. This may change in the future with better and more flexible versions of CRBM routines; however, currently, this work indicates that MLE offers the most practical balance between accuracy and efficiency, and it is applicable to a broad class of quantum simulators that realize spin models.

Future works can involve noise modeling that takes into account imperfections in measurements, perhaps using Bayesian methods [72–74], and could make comparisons between the approach used in this work with other similar methods, such as shadow estimation [40], as well as the case of mixed states occurring due to nonunitary processes in the dynamics. Moreover, our findings resonate with complexity-theoretic arguments such as anticoncentration theorems, which establish the necessity of approximate designs for demonstrating quantum speedups and validating randomness in near-term devices [75].

ACKNOWLEDGMENT

R.M. acknowledges support from the U.S. National Institute of Standards and Technology (NIST) through the CIPP program under Award No. 60NANB24D218.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: EXPRESSION FOR k th MOMENT FOR RANDOM HAAR ENSEMBLE

Consider a Hilbert space with dimensionality d , then the Haar ensemble is defined as the continuous random probability distribution over pure states defined over that space. Using the Schur-Weyl duality, an analytical expression to calculate the k th moment of the Haar ensemble [76] is given as

$$\hat{\rho}_{\text{Haar}}^{(k)} = \frac{\sum_{\pi \in S_k} \hat{P}_d(\pi)}{d(d+1) \cdots (d+k-1)}. \quad (\text{A1})$$

The sum in the expression above iterates through the elements π of the permutation group S_k . These represent all the ways of rearranging an array of k distinct objects. For example, we explicitly report the elements of the group S_3 below, which are needed to construct the third moment of the ensemble $\rho_{\text{Haar}}^{(3)}$:

$$\pi_1 : \begin{cases} \pi_1(1) = 1, \\ \pi_1(2) = 2, \\ \pi_1(3) = 3, \end{cases} \quad \pi_2 : \begin{cases} \pi_2(1) = 2, \\ \pi_2(2) = 1, \\ \pi_2(3) = 3, \end{cases} \quad (\text{A2})$$

$$\pi_3 : \begin{cases} \pi_3(1) = 3, \\ \pi_3(2) = 2, \\ \pi_3(3) = 1, \end{cases} \quad \pi_4 : \begin{cases} \pi_4(1) = 1, \\ \pi_4(2) = 3, \\ \pi_4(3) = 2, \end{cases} \quad (\text{A3})$$

$$\pi_5 : \begin{cases} \pi_5(1) = 3, \\ \pi_5(2) = 1, \\ \pi_5(3) = 2, \end{cases} \quad \pi_6 : \begin{cases} \pi_6(1) = 2, \\ \pi_6(2) = 3, \\ \pi_6(3) = 1. \end{cases} \quad (\text{A4})$$

The operator $\hat{P}_d(\pi)$ acts as a swap operator that acts on k copies of d -dimensional qudits and is defined as follows [76]:

$$\hat{P}_d(\pi) = \sum_{i_1, \dots, i_k \in [1, 2, \dots, d]} |i_{\pi(1)}, \dots, i_{\pi(k)}\rangle \langle i_1, \dots, i_k|. \quad (\text{A5})$$

APPENDIX B: SIMULATING MEASUREMENT DATA

Repeated measurements are taken to obtain statistics by randomly sampling the measurement directions between the three mutually orthogonal directions. For spin systems, and, in particular, for Rydberg systems, this can be experimentally implemented by randomly rotating each spin using a unitary operator followed by a state measurement in the Z basis [24]. The dataset $\mathcal{D}$ of size $|\mathcal{D}|$ is constructed by measuring the subsystem B in the Z basis and the system A in all bases. It is subsequently split into subsets that share the same outcome z_B of the subsystem B measurement, which we label as $\mathcal{D}(z_B)$. Each $\mathcal{D}(z_B)$ contains multiple simulated measurements of subsystem A for a particular z_B , as shown in Fig. 6. The i th measurement in A is denoted by $|r_A^{(i)}(z_B)\rangle_{\mathbf{b}} = |b_{A,1}^{(i)}\rangle \otimes \cdots \otimes |b_{A,N_A}^{(i)}\rangle$, where $|b_{A,j}^{(i)}\rangle \in \{|0\rangle, |1\rangle\}_{X,Y,Z}$ are the eigenstates of the $\sigma_{b_j}^{(i)}$ Pauli operator with eigenvalue ± 1 . It should be noted that the amount of data available to train the state $|\Psi(z_B)\rangle$ for a particular z_B fluctuates for each run of numerically simulated

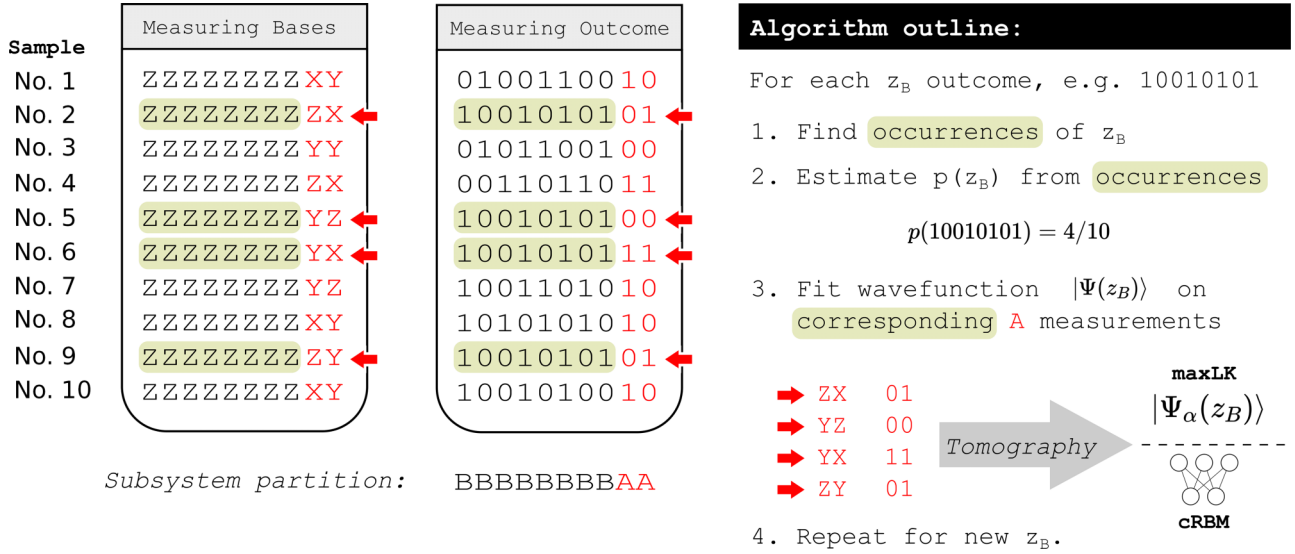


FIG. 6. The figure illustrates the scheme for generating the simulated measurement datasets $\mathcal{D}(z_B)$. The numerically generated bitstrings are collected and sorted in subsets containing the common z_B . A probability is assigned to each unique z_B string by counting the number of occurrences of that bitstring in the dataset. The corresponding measurements of A associated with a particular z_B determine the state $|\Psi_A(z_B)\rangle$. The tuple $(p(z_B), |\Psi(z_B)\rangle)$ accordingly defines the ensemble state $\mathcal{E}$.

measurements. In general, z_B bitstrings with a higher probability $p(z_b)$ are more likely to be sampled, and the size of the corresponding dataset $\mathcal{D}(z_b)$ will be, on average, larger. This also implies that subsets $\mathcal{D}(z_b)$ with low $p(z_b)$ may have measurements for A that are missing in certain directions. For example, in the illustrative set of measurements provided in Fig. 6, the dataset $\mathcal{D}(10010101)$ contains measurements of subsystem A in the basis $\{ZX, YZ, YX, ZY\}$ for a given outcome z_B . Here, the first spin for A was never measured in the X basis, which may then affect the quality of the training. However, the states with particularly low $p(z_B)$ do not contribute significantly to the $\rho_{\mathcal{E}}^{(k)}$, thereby reducing the impact of inaccurate state characterization due to the limited data.

APPENDIX C: TRACE DISTANCE WITH SYSTEM SIZE

To verify that a certain ensemble state forms an approximate quantum k -design is to demonstrate that the trace distance $\delta_{(k)}$, which is defined in Eq. (5), approaches zero. To evaluate the trace distance [Eq. (5)], we first construct the ensemble in Eq. (3) and its k th moment in Eq. (4) using the state vector $|\Psi_A(z_B)\rangle$ and probabilities $p(z_B)$, which are derived from the time evolution of Eq. (6), followed by the simulated projective measurements on subsystem B . The steady-state value of the trace distance, $\langle \delta_{(k)} \rangle$, is then obtained by averaging $\delta_{(k)}(t)$ over 130 uniformly spaced time points within the interval $(1.5, 5.0) \mu\text{s}$. Figure 7 shows the numerical results for the emergence of approximate k -designs in subsystem sizes ($N_A = 1, 2$, and 3) as a function of the subsystem B size N_B . In this analysis, we consider various possible ways of bipartitioning the system, demonstrating that the characterization of a k -design is independent of the specific choice of subsystem A . Furthermore, the dependence of $\langle \delta_{(k)} \rangle$ on N_B follows a simple scaling law, $\langle \delta_{(k)} \rangle \propto \frac{1}{(N_B)^{1/k}}$, consistent with the recent numerical observations of Ref. [28].

The exponents γ_k are obtained using linear fitting of the data on a log-log scale. Moreover, a comparison across the three subplots reveals that γ_k increases with N_A for a fixed k . Since larger values of k correspond to more uniform approximate designs, this implies that increasing the subsystem size N_A leads to a more Haar-like approximate design, provided that $N_A \ll N_B$.

APPENDIX D: ESTIMATORS FOR RANDOM ENSEMBLE STATE

In this Appendix, we outline the details of the four statistical learning methods employed in this work, namely, (1) CRBM, (2) maximum-likelihood estimation, (3) the frequentist approach, and (4) shadow tomography, as depicted in Fig. 8.

1. Complex restricted Boltzmann machines

An RBM is a type of bipartite neural network with two layers, which are referred to as visible and hidden layers, as shown in Fig. 9. It is used to learn the probability distribution over binary data in an unsupervised approach. The network shows no intralayer connections but just interlayer ones, thus motivating the name *restricted*. The nodes in the visible and hidden layers have bias vectors $\mathbf{b}$ and $\mathbf{c}$ associated with them, whereas a matrix $\mathbf{W}$ provides weights for the connections. The nodes z_i and h_j can only take binary values 0 or 1 and are thus used to represent a probability distribution $p(\mathbf{z})$, where $\mathbf{z}$ is a vector of N binary entries $\{0, 1\}$, i.e., $\mathbf{z} \in \{0, 1\}^N$. The state of the RBM is completely defined using the two vectors:

$$\mathbf{z} = (z_1, z_2, \dots, z_V), \quad z_i \in \{0, 1\}, \quad (\text{D1a})$$

$$\mathbf{h} = (h_1, h_2, \dots, h_H), \quad h_j \in \{0, 1\}, \quad (\text{D1b})$$

where V and H are the total numbers of nodes in the visible and hidden layers, respectively. Each state of the RBM has an

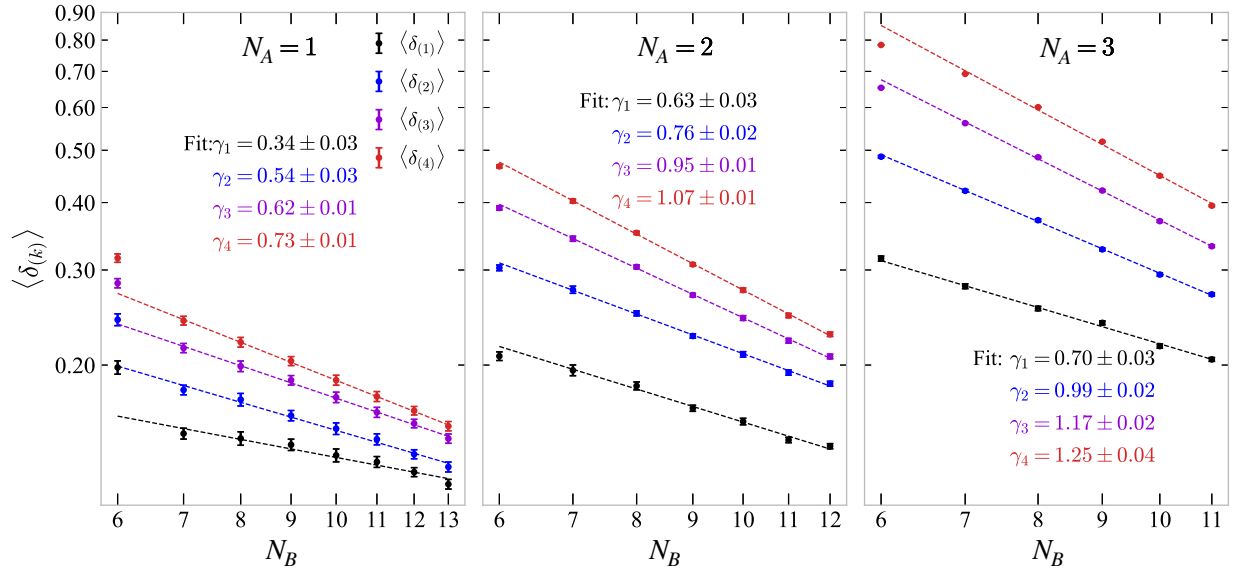


FIG. 7. Shows the time average and standard deviation of the value of the trace distance in the steady state (y axis) as a function of different subsystem B sizes (x axis). The system considered has a total size of $N = 14$ with varying subsystems and subsystem B sizes. The expected scaling law of the trace distance is established, where the γ_k values are shown in the panels corresponding to different quantum state k -designs.

associated energy given by

$$E_\theta(\mathbf{z}, \mathbf{h}) = -\mathbf{b}^T \mathbf{z} - \mathbf{c}^T \mathbf{h} - \mathbf{z}^T \mathbf{W} \mathbf{h}, \quad (\text{D2})$$

which is parametrized by the weights and biases of the model $\theta = \{\mathbf{b}, \mathbf{c}, \mathbf{W}\}$. The probability distribution $p_\theta(\mathbf{z}, \mathbf{h})$ (chosen to be a Boltzmann distribution) and the partition function $Z_\theta(\mathbf{z}, \mathbf{h})$ associated with the state energy are given as

$$p_\theta(\mathbf{z}, \mathbf{h}) = \frac{1}{Z_\theta} e^{-E_\theta(\mathbf{z}, \mathbf{h})}, \quad (\text{D3})$$

$$Z_\theta = \sum_{\mathbf{z}} \sum_{\mathbf{h}} e^{-E_\theta(\mathbf{z}, \mathbf{h})}. \quad (\text{D4})$$

In our work, we are interested in the probability $p_\theta(\mathbf{z})$, which is the marginal of $p_\theta(\mathbf{z}, \mathbf{h})$ and has the following analytical expression:

$$p_\theta(\mathbf{z}) = \sum_{\mathbf{h}'} p_\theta(\mathbf{z}, \mathbf{h}') = \frac{1}{Z_\theta} \mathbf{e}^{-\mathcal{E}_\theta(\mathbf{z})}, \quad (\text{D5})$$

$$\mathcal{E}_\theta(\mathbf{z}) = -\mathbf{b}^T \mathbf{z} - \sum_{j=1}^H \ln \left[1 + \exp \left(c_j + \sum_{i=1}^V W_{ij} z_i \right) \right]. \quad (\text{D6})$$

Using this formalism, a particular state $\mathbf{z} = 010 \dots 1$ is associated with a probability $p_\theta(\mathbf{z})$ that is parametrized by $\theta = \{\mathbf{b}, \mathbf{c}, \mathbf{W}\}$. The goal is to find the optimal parameters θ that give the approximation $p_\theta(\mathbf{z}) \approx \mathbf{q}(\mathbf{z})$, given some data

$$\mathcal{D} = [z^{(1)}, z^{(2)}, \dots, z^{(D)}], \quad \text{where } z^{(i)} \in \{0, 1\}^V, \quad (\text{D7})$$

sampled according to the target distribution $q(\mathbf{z})$ from which the data are sampled. This is referred to as training of the RBM with respect to the data.

Training the RBM with data: The commonly used cost function that needs to be minimized while training the RBM is the Kullback-Leibler divergence [68,77], which is

defined as

$$C_\theta = \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{q(\mathbf{z})}{p_\theta(\mathbf{z})} = \sum_{\mathbf{z}} q(\mathbf{z}) \log q(\mathbf{z}) - q(\mathbf{z}) \log p_\theta(\mathbf{z}), \quad (\text{D8})$$

where the term $\sum_{\mathbf{z}} q(\mathbf{z}) \log p_\theta(\mathbf{z}) = \langle \log p_\theta \rangle_q$ is the averaged value of $\log p_\theta(\mathbf{z})$ with respect to $q(\mathbf{z})$. Given the function C_θ , the parameters are updated to minimize this using the following rule:

$$\theta \leftarrow \theta - \eta \nabla_\theta C_\theta. \quad (\text{D9})$$

Here, η is the *learning rate*, and it affects the step size of the parameter update. The dependence of the cost function on the parameters is captured by the following expression [68,77]:

$$\nabla_\theta C_\theta \approx \langle \nabla_\theta \mathcal{E}_\theta \rangle_{\mathcal{D}} - \langle \nabla_\theta \mathcal{E}_\theta \rangle_{p_\theta}. \quad (\text{D10})$$

The first term is often called the *positive phase*, and it represents the average of the gradient of $\mathcal{E}_\theta$ [cf. Eq. (D5)] evaluated at the values of $\mathbf{z}$ in the dataset $\mathcal{D}$. The second term is the *negative phase*, and it represents the average of the gradient of $\mathcal{E}_\theta$ taken over the learned probability distribution $p_\theta(\mathbf{z})$. This last term can be computationally expensive to evaluate during training. Indeed, the expression for $p_\theta(\mathbf{z})$ involves the calculation of the partition function Z_θ in Eq. (D4) and is a task that scales like $\mathcal{O}(2^H \times 2^V)$, which becomes intractable for large systems. The standard approach to calculate the value $\langle \nabla_\theta \mathcal{E}_\theta \rangle_{p_\theta}$ is using an algorithm called *contrastive divergence*. This algorithm allows to sample $\mathbf{z}$ bitstrings according to p_θ without calculating the partition function of the system explicitly. A detailed explanation of the optimization algorithm is out of the scope of this work, but it can be found in Refs. [68,77]. For a system size of two qubits, the wave function is described as

$$|\Psi\rangle = \psi_{00}|00\rangle + \psi_{01}|01\rangle + \psi_{10}|10\rangle + \psi_{11}|11\rangle, \quad (\text{D11})$$

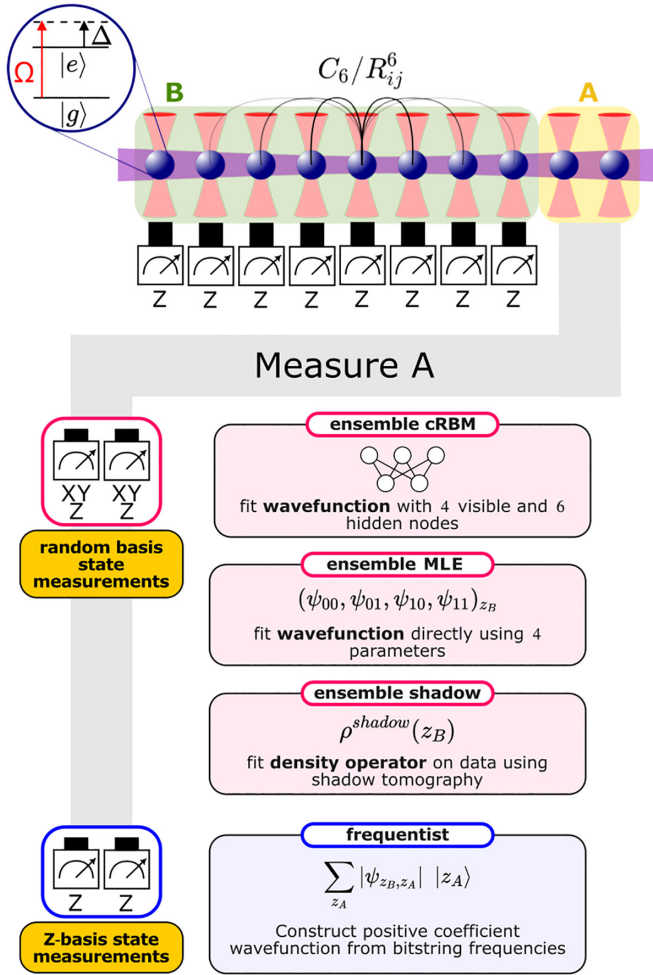


FIG. 8. This figure schematically summarizes the four approaches used to verify approximate quantum k -design. Numerical simulations provide measurement data for subsystem B in Z basis with probabilities $p(z_B)$ for all four methods. In the case of ensemble MLE, CRBM, and shadow ensemble approaches, the measurement basis for subsystem A is chosen randomly in one of the three directions. Applying MLE when the system A is measured only in the Z basis becomes equivalent to the frequentist method.

and following Born's rule, the probability distribution is given by $p(z) = |\psi_z|^2$. This defines the data that will be used to train the RBM, which in turn implies optimizing the parameters θ such that the probability $p_\theta(z) \approx p(z)$. Wave functions can be reconstructed using positive real-valued coefficients $\psi_{z,\theta} := \sqrt{p_\theta(z)}$, but this approach has the disadvantage that phase information is discarded. A more general approach suggested by Torlai [78] is to consider two RBMs, one that describes the amplitude of the coefficients ψ_z and the other that describes the phase $\phi(z)$ of the coefficients ψ_z . The two RBMs are parametrized with their own weights and biases separately, which are labeled as θ and μ . They both have associated effective energies $\mathcal{E}_\theta(z)$ and $\mathcal{E}_\mu(z)$, described by Eq. (D5). The coefficients of the wave function are now parametrized as

$$\begin{aligned} \psi_{z,\theta\mu} &= \sqrt{p_\theta(z)} e^{i\phi_\mu(z)} = Z_\theta^{-1/2} e^{-\mathcal{E}_\theta(z)/2} e^{-i\mathcal{E}_\mu(z)/2} \\ &= Z_\theta^{-1/2} e^{-(\mathcal{E}_\theta(z) + i\mathcal{E}_\mu(z))/2}. \end{aligned} \quad (\text{D12})$$

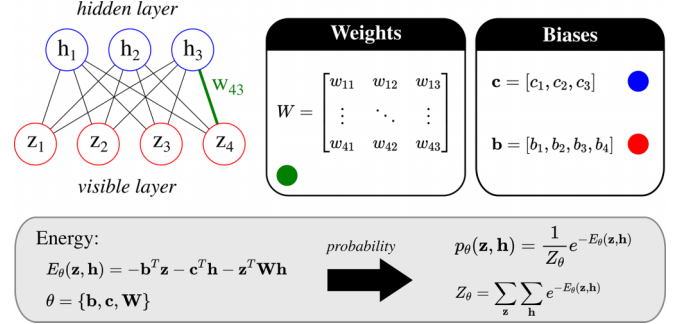


FIG. 9. Network diagram of RBM. The network is divided into a hidden and visible layer, each one consisting of nodes h and z that take binary values 0 and 1. RBM is parametrized by θ that groups the weights associated with the interlayer connections and biases associated with the hidden and visible nodes. An energy $E_\theta(\mathbf{h})$ can be calculated for each network configuration. The energy allows defining a probability $p_\theta(\mathbf{h})$ that is inspired by the Boltzmann distribution. Z_θ is the normalizing partition function.

One can reconstruct the phase ϕ_z of the quantum state by measuring along different directions. In this manner, a complex quantum wave function can be represented by an RBM, and its parameters (θ, μ) can be trained to learn the coefficients $\psi_{z,\theta\mu} \approx \psi_z$.

The CRBM is optimized via a hyperparameter search that minimizes the wave function reconstruction error, quantified by the trace distance with respect to numerically obtained states. This procedure yields an optimal architecture of two nodes in the visible layer and three nodes in the hidden layer for each RBM, with a total of two RBMs within a CRBM and 12 800 training epochs. This is implemented using a Python library QuCumber [66]. The resulting trained model is then applied directly within the proposed measurement framework, enabling reconstruction of arbitrary two-qubit states generated by the Hamiltonian.

2. Maximum-likelihood method

Another way to efficiently estimate $|\Psi_A(z_B)\rangle$ from the dataset is to apply MLE. Similar to the previous approach, the coefficients in Eq. (D11) are fitted in a manner that maximizes the *likelihood* of the dataset $\mathcal{D}(z_B)$. This is achieved by minimizing the following cost function [78]:

$$\mathcal{C}_\Psi^{\mathcal{D}(z_B)} = -\frac{1}{\|\mathcal{D}(z_B)\|} \sum_{i=1}^{\|\mathcal{D}(z_B)\|} \log |\langle \Psi_A(z_B; \Psi) | r_A^{(i)}(z_B) \rangle|^2, \quad (\text{D13})$$

over all the possible vectors of complex coefficients Ψ in Eq. (D11). The optimization is numerically implemented using the Python library NumPy [79,80]. Interestingly, if the measurements performed on A are restricted to the Z basis, then minimizing the above cost function is equivalent to implementing the frequentist method, which uses the following ansatz:

$$|\Psi_A(z_B)\rangle = \sum_{z'_A} \sqrt{p(z'_A|z_B)} |z'_A\rangle = \sum_{z'_A} \sqrt{\frac{p(z'_A, z_B)}{p(z_B)}} |z'_A\rangle. \quad (\text{D14})$$

The joint probability $p(z_A, z_B)$ is estimated from the outcome frequencies in the datasets. The frequentist analysis is the approach used in the Rydberg experiment [28] and served as a benchmark for the two methods proposed in this work.

3. Equivalence between maximum-likelihood and frequentist approaches

In this Appendix, we mathematically show that the frequentist approach is just a particular case of the maximum-likelihood when considering a positive coefficient wave function. There are $M = 4$ possible outcomes for a two-qubit wave function $|\Psi\rangle$ that is measured only in the Z basis, which are (00, 01, 10, 11). The probability of the m th outcome is $p_m = |\psi_m|^2$, where $\psi_m = \langle m|\Psi\rangle$ is the coefficient of the wave function in the computational basis. Consider a case where the measurement has been repeated N times, giving the following outcomes [00, 01, 00, ...]. The number of times the m th outcome occurs in the dataset is labeled with n_m where trivially $\sum_m n_m = N$. The likelihood function defined in Eq. (D13) for our set of measurements of the two-qubit wave function is rewritten as

$$\begin{aligned} L &= \frac{1}{N} \sum_{i=1}^N \log |\langle m^{(i)} | \Psi \rangle|^2 = \frac{1}{N} \sum_{i=1}^N \log p_{m^{(i)}} \\ &= \frac{1}{N} (n_{00} \log p_{00} + n_{01} \log p_{01} + n_{10} \log p_{10} + n_{11} \log p_{11}) \\ &= \frac{1}{N} \log \left(\prod_m p_m^{n_m} \right) = \frac{1}{N} \log \mathcal{L}, \end{aligned} \quad (\text{D15})$$

where $m^{(i)}$ is the outcome of the i th measurement in the set of N measurements. Using the technique of Lagrange multipliers, L can be maximized under the constraint $\sum_m p_m = 1$, which is stated below,

$$\frac{\partial(\mathcal{L} - \alpha \sum_m p_m)}{\partial p_{m'}} = \frac{n_{m'}}{p_{m'}} \mathcal{L} - \alpha = 0 \quad \forall m'. \quad (\text{D16})$$

The factor $\mathcal{L}/\alpha$ is a constant that can be fixed by enforcing the condition $\sum_m p_m = 1$, giving

$$p_m = \frac{n_m}{N}. \quad (\text{D17})$$

Thus, L is maximized when $p_m = n_m/N$, implying that the maximum-likelihood approach reduces to the frequentist method when measurements only along one direction are considered.

4. Shadow tomography

The shadow tomography on the ensemble of states is performed similarly to the procedure described in Refs. [40,43,44]. For every simulated measurement, the subsystem B is measured in the Z basis, whereas each qubit A is measured in a random basis uniformly sampled from the set $\{X, Y, Z\}$. Each qubit in subsystem A is collapsed into some state a_i , where i indexes the measured qubit, which depends both on the chosen basis and the measurement outcome. The dataset $\mathcal{D}(z_B)$ contains measurements of the state $|\Psi(z_B)\rangle$ in different bases. We label with $|a^{(n)}\rangle$ the state that subsystem A is

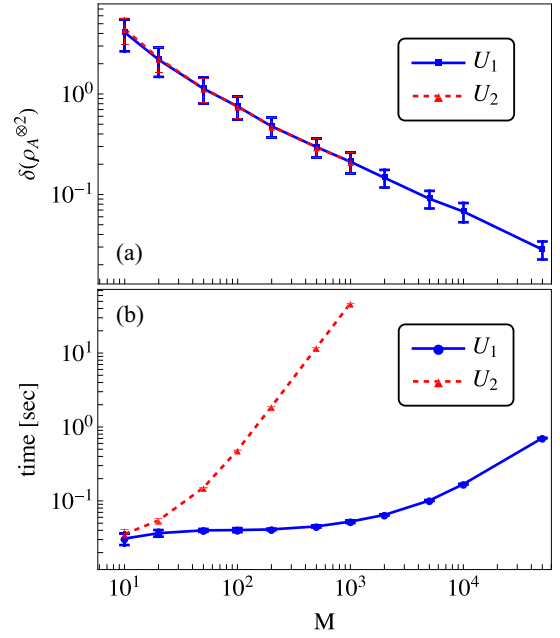


FIG. 10. (a) Comparing trace distance from the exact density matrix from quantum evolution $[(\rho_A^{\text{exact}})^{\otimes k}]$ with estimate in Eq. (D22) (U_2 statistics) and Eq. (D24) (U_1 statistics) shown with red triangle and blue circle, respectively. (b) Comparison of computation time of U_1 (blue circle) and U_2 (red triangle) statistics for estimation of $\rho_A^{\otimes 2}$ with increasing sample size. Note: Results for computation time are based on calculations performed on an Intel Core Ultra 7 265 K.

collapsed into upon n th measurement, and with $|a_i^{(n)}\rangle$ the collapsed state of qubit i . We consider a total M measurements, and for each outcome z_B , the dataset $\mathcal{D}(z_B)$ contains many shots of subsystem A measured in random local bases. More specifically,

$$\mathcal{D}(z_B) := \{(s_n, b_n)\}, \quad (\text{D18})$$

where s_n records the random Pauli basis selected on each qubit and b_n denotes the measurement outcome of ± 1 . From a single element in $\mathcal{D}(z_B)$, we build a single shadow $\hat{\rho}_A(z_B)$ via the inverse measurement map

$$\hat{\rho}_A^{(n)}(z_B) = \bigotimes_i (3|a_i^{(n)}\rangle\langle a_i^{(n)}| - I), \quad (\text{D19})$$

derived in Ref. [40].

In general, $\rho_A^{\otimes k}(z_B)$ can be typically estimated using U -statistics of order k , which provides an unbiased approach to estimate a parameter, defined as

$$U_k = \binom{M}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq M} h(\mathbf{X}_{i_1}, \mathbf{X}_{i_2}, \dots, \mathbf{X}_{i_k}), \quad (\text{D20})$$

where $\mathbf{X}_{i_k} \in \{X_1, X_2, \dots, X_n\}$ are samples from a distribution and h is a symmetric kernel function. Here, $X_n = \hat{\rho}_A^{(n)}(z_B)$ and for simplicity, let us consider two cases: $k = 1$ and $k = 2$, for which the estimates of $\rho_A(z_B)$ and $\rho_A^{\otimes 2}(z_B)$ are given by

$$\rho_A(z_B) = \frac{1}{M} \sum_{n=1}^M \hat{\rho}_A^{(n)}(z_B), \quad (\text{D21})$$

$$\rho_A^{\otimes 2}(z_B) = \frac{1}{M(M-1)} \sum_{n \neq m}^M \hat{\rho}_A^{(n)}(z_B) \otimes \hat{\rho}_A^{(m)}(z_B), \quad (\text{D22})$$

which can be generalized to the k th moment. However, with an increase in k , the computational complexity scales as M^k .

Thus, in order to reduce the computational time, we consider a “plug-in estimator” approach, where we consider copies of Eq. (D21) following U_1 instead of U_2 statistics, which provides us with an approximate estimate $\bar{\rho}_A^{\otimes 2}(z_B) = (\rho_A(z_B))^{\otimes k}$. To identify the regime where the approximation holds, we can derive a relation between the two estimators $\bar{\rho}_A^{\otimes 2}(z_B)$ and $\rho_A^{\otimes 2}(z_B)$. To begin with, we expand $\bar{\rho}_A^{\otimes 2}(z_B)$ as follows:

$$\bar{\rho}_A^{\otimes 2}(z_B) = (\rho_A(z_B))^{\otimes k}, \quad (\text{D23})$$

$$\begin{aligned} &= \left(\frac{1}{M} \sum_{n=1}^M \hat{\rho}_A^{(n)}(z_B) \right) \otimes \left(\frac{1}{M} \sum_{m=1}^M \hat{\rho}_A^{(m)}(z_B) \right), \\ &= \frac{1}{M^2} \sum_{n,m=1}^M \hat{\rho}_A^{(n)}(z_B) \otimes \hat{\rho}_A^{(m)}(z_B), \\ &= \frac{1}{M^2} \left(\sum_{n \neq m}^M \hat{\rho}_A^{(n)}(z_B) \otimes \hat{\rho}_A^{(m)}(z_B) \right. \\ &\quad \left. + \sum_{n=1}^M \hat{\rho}_A^{(n)}(z_B) \otimes \hat{\rho}_A^{(n)}(z_B) \right). \end{aligned} \quad (\text{D24})$$

Using the expression of $\rho_A^{\otimes 2}$ from Eq. (D22), we get

$$\begin{aligned} \bar{\rho}_A^{\otimes 2}(z_B) &= \frac{M(M-1)}{M^2} \rho_A^{\otimes 2} + \frac{1}{M^2} \sum_{n=1}^M \hat{\rho}_A^{(n)}(z_B) \otimes \hat{\rho}_A^{(n)}(z_B), \\ &\approx \rho_A^{\otimes 2} + \underbrace{\frac{1}{M^2} \sum_{n=1}^M \hat{\rho}_A^{(n)}(z_B) \otimes \hat{\rho}_A^{(n)}(z_B)}_{\text{Bias}}. \end{aligned} \quad (\text{D25})$$

Thus, from Eq. (D25), the approximation is valid in the limit of sufficiently large measurements M . To demonstrate this, Fig. 10(a) compares the trace distance from the exact density matrix from quantum evolution $[(\rho_A^{\text{exact}})^{\otimes k}]$ with the estimate in Eq. (D22) (based on U_2 statistics, shown as red triangles) and Eq. (D24) (based on U_1 statistics, shown as blue circles). The two curves become indistinguishable for sufficiently large M , indicating that it is sufficient to use U_1 statistics for estimating $\rho_A^{\otimes k}$.

We further compare the computational run-time in Fig. 10(b), which highlights the computational efficiency of the U_1 estimator relative to the unbiased U_2 estimator. The added computational cost in U_2 arises from $M(M-1)$ combinations required for the tensor product in Eq. (D22), as compared to only M combinations required in U_1 statistics in Eqs. (D21) and (D24), as discussed earlier. Thus, for constructing a k -design, using U_1 statistics provides an M^{k-1} advantage over the corresponding U_k statistics.

-
- [1] B. Schneier, *Applied Cryptography*, 20th Anniversary ed. (John Wiley & Sons, Ltd., Indianapolis, IN, 2015), Chap. 1, pp. 1–18.
 - [2] L. Guang, C. M. Assi, and A. Benslimane, Enhancing IEEE 802.11 random backoff in selfish environments, *IEEE Trans. Veh. Technol.* **57**, 1806 (2008).
 - [3] D. Seetharam and S. Rhee, An efficient pseudo random number generator for low-power sensor networks [wireless networks], in *29th Annual IEEE International Conference on Local Computer Networks* (IEEE, New York, 2004), pp. 560–562.
 - [4] T. E. Hull and A. R. Dobell, Random number generators, *SIAM Rev.* **4**, 230 (1962).
 - [5] G. Marsaglia, Random number generators, *J. Mod. Appl. Stat. Methods* **2**, 2 (2003).
 - [6] G. Marsaglia, A. Zaman, and W. W. Tsang, Toward a universal random number generator, *Stat. Probab. Lett.* **9**, 35 (1990).
 - [7] F. James, A review of pseudorandom number generators, *Comput. Phys. Commun.* **60**, 329 (1990).
 - [8] S. K. Park and K. W. Miller, Random number generators: Good ones are hard to find, *Commun. ACM* **31**, 1192 (1988).
 - [9] L. Blum, M. Blum, and M. Shub, A simple unpredictable pseudo-random number generator, *SIAM J. Comput.* **15**, 364 (1986).
 - [10] E. Bannai, On tight spherical designs, *J. Comb. Theory Ser. A* **26**, 38 (1979).
 - [11] E. Bannai and E. Bannai, A survey on spherical designs and algebraic combinatorics on spheres, *Eur. J. Comb.* **30**, 1392 (2009).
 - [12] J. Emerson, Y. S. Weinstein, M. Saraceno, S. Lloyd, and D. G. Cory, Pseudo-random unitary operators for quantum information processing, *Science* **302**, 2098 (2003).
 - [13] D. R. Stinson, *Combinatorial Designs: Constructions and Analysis* (Springer-Verlag, New York, 2004), Vol. 480.
 - [14] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80**, 012304 (2009).
 - [15] M. Ohliger, V. Nesme, and J. Eisert, Efficient and feasible state tomography of quantum many-body systems, *New J. Phys.* **15**, 015024 (2013).
 - [16] M. Adam, Applications of unitary k -designs in quantum information processing, Master’s thesis, Masaryk University, Brno, Czech Republic, 2013.
 - [17] C. Lancien and C. Majenz, Weak approximate unitary designs and applications to quantum encryption, *Quantum* **4**, 313 (2020).
 - [18] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, *Phys. Rev. A* **100**, 032328 (2019).
 - [19] F. G. Brandão, W. Chemsassy, N. Hunter-Jones, R. Kueng, and J. Preskill, Models of quantum complexity growth, *PRX Quantum* **2**, 030316 (2021).
 - [20] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, *Nat. Phys.* **14**, 595 (2018).

- [21] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, On the complexity and verification of quantum random circuit sampling, *Nat. Phys.* **15**, 159 (2019).
- [22] J. Haferkamp, D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, and J. Bermejo-Vega, Closing gaps of a quantum advantage with short-time Hamiltonian dynamics, *Phys. Rev. Lett.* **125**, 250501 (2020).
- [23] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Probing many-body dynamics on a 51-atom quantum simulator, *Nature (London)* **551**, 579 (2017).
- [24] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, *Science* **364**, 260 (2019).
- [25] Y. S. Weinstein, W. G. Brown, and L. Viola, Parameters of pseudorandom quantum circuits, *Phys. Rev. A* **78**, 052332 (2008).
- [26] P. Hayden and J. Preskill, Black holes as mirrors: Quantum information in random subsystems, *J. High Energy Phys.* **09** (2007) 120.
- [27] L. Piroli, C. Sünderhauf, and X.-L. Qi, A random unitary circuit model for black hole evaporation, *J. High Energy Phys.* **04** (2020) 063.
- [28] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, R. Finkelstein, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale, H. Pichler, F. G. S. L. Brandão, S. Choi, and M. Endres, Preparing random states and benchmarking with many-body quantum chaos, *Nature (London)* **613**, 468 (2023).
- [29] J. S. Cotler, D. K. Mark, H.-Y. Huang, F. Hernández, J. Choi, A. L. Shaw, M. Endres, and S. Choi, Emergent quantum state designs from individual many-body wave functions, *PRX Quantum* **4**, 010311 (2023).
- [30] S. Nautiyal, Quantum chaotic dynamics and effective random state generation, Ph.D. thesis, Brandeis University, 2023.
- [31] F. G. Brandao, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, *Commun. Math. Phys.* **346**, 397 (2016).
- [32] E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, and J. Eisert, Mixing properties of stochastic quantum Hamiltonians, *Commun. Math. Phys.* **355**, 905 (2017).
- [33] Y. Nakata, C. Hirche, M. Koashi, and A. Winter, Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics, *Phys. Rev. X* **7**, 021006 (2017).
- [34] A. Elben, B. Vermersch, M. Dalmonte, J. Cirac, and P. Zoller, Rényi entropies from random quenches in atomic Hubbard and spin models, *Phys. Rev. Lett.* **120**, 050406 (2018).
- [35] A. G. White, D. F. V. James, W. J. Munro, and P. G. Kwiat, Exploring Hilbert space: Accurate characterization of quantum information, *Phys. Rev. A* **65**, 012301 (2001).
- [36] A. I. Lvovsky, H. Hansen, T. Aichele, O. Benson, J. Mlynek, and S. Schiller, Quantum state reconstruction of the single-photon Fock state, *Phys. Rev. Lett.* **87**, 050402 (2001).
- [37] J. Itatani, J. Levesque, D. Zeidler, H. Niikura, H. Pépin, J. C. Kieffer, P. B. Corkum, and D. M. Villeneuve, Tomographic imaging of molecular orbitals, *Nature (London)* **432**, 867 (2004).
- [38] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, *Nat. Commun.* **1**, 149 (2010).
- [39] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, Efficient tomography of a quantum many-body system, *Nat. Phys.* **13**, 1158 (2017).
- [40] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [41] A. W. Smith, J. Gray, and M. S. Kim, Efficient quantum state sample tomography with basis-dependent neural networks, *PRX Quantum* **2**, 020348 (2021).
- [42] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, *Nat. Rev. Phys.* **5**, 9 (2023).
- [43] R. O'Donnell and J. Wright, Efficient quantum tomography, in *Proceedings of the 48th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2016), pp. 899–912.
- [44] M. McGinley and M. Fava, Shadow tomography from emergent state designs in analog quantum simulators, *Phys. Rev. Lett.* **131**, 160601 (2023).
- [45] T. Schuster, J. Haferkamp, and H.-Y. Huang, Random unitaries in extremely low depth, *Science* **389**, 92 (2025).
- [46] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, *Commun. Math. Phys.* **291**, 257 (2009).
- [47] J. Riddell, K. Klobas, and B. Bertini, Quantum state designs from minimally random quantum circuits, *arXiv:2503.05698*.
- [48] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [49] C.-Y. Park and M. J. Kastoryano, Expressive power of complex-valued restricted Boltzmann machines for solving nonstoquastic Hamiltonians, *Phys. Rev. B* **106**, 134437 (2022).
- [50] T. Gallagher, Rydberg atoms, in *Springer Handbook of Atomic, Molecular, and Optical Physics* edited by G. W. F. Drake (Springer, New York, NY, 2006), pp. 235–245.
- [51] M. Saffman, T. G. Walker, and K. Mølmer, Quantum information with Rydberg atoms, *Rev. Mod. Phys.* **82**, 2313 (2010).
- [52] C. S. Adams, J. D. Pritchard, and J. P. Shaffer, Rydberg atom quantum technologies, *J. Phys. B: At. Mol. Opt. Phys.* **53**, 012002 (2019).
- [53] A. Browaeys and T. Lahaye, Many-body physics with individually controlled Rydberg atoms, *Nat. Phys.* **16**, 132 (2020).
- [54] D. Bluvstein, A. Omran, H. Levine, A. Keesling, G. Semeghini, S. Ebadi, T. T. Wang, A. A. Michailidis, N. Maskara, W. W. Ho, *et al.*, Controlling quantum many-body dynamics in driven Rydberg atom arrays, *Science* **371**, 1355 (2021).
- [55] A. Browaeys, D. Barredo, and T. Lahaye, Experimental investigations of dipole–dipole interactions between a few Rydberg atoms, *J. Phys. B: At. Mol. Opt. Phys.* **49**, 152001 (2016).
- [56] K. Mukherjee, H. P. Goswami, S. Whitlock, S. Wüster, and A. Eisfeld, Two-dimensional spectroscopy of Rydberg gases, *New J. Phys.* **22**, 073040 (2020).
- [57] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, T. Lahaye *et al.*, Quantum simulation of 2D antiferromagnets with hundreds of Rydberg atoms, *Nature (London)* **595**, 233 (2021).

- [58] K. Mukherjee, H. Barghathi, A. Del Maestro, and R. Mukherjee, Quantum simulation of Motzkin spin chain with Rydberg atoms, [arXiv:2603.23422](#).
- [59] M. Saffman, Quantum computing with atomic qubits and Rydberg interactions: Progress and challenges, *J. Phys. B: At. Mol. Opt. Phys.* **49**, 202001 (2016).
- [60] L. Henriët, L. Beguin, A. Signoles, T. Lahaye, A. Browaeys, G.-O. Reymond, and C. Jurczak, Quantum computing with neutral atoms, *Quantum* **4**, 327 (2020).
- [61] E. Urban, T. A. Johnson, T. Henage, L. Isenhower, D. Yavuz, T. Walker, and M. Saffman, Observation of Rydberg blockade between two atoms, *Nat. Phys.* **5**, 110 (2009).
- [62] C. Chen, G. Bornet, M. Bintz, G. Emperauger, L. Leclerc, V. S. Liu, P. Scholl, D. Barredo, J. Hauschild, S. Chatterjee, M. Schuler, A. M. Läuchli, M. P. Zaletel, T. Lahaye, N. Y. Yao, and A. Browaeys, Continuous symmetry breaking in a two-dimensional Rydberg array, *Nature (London)* **616**, 691 (2023).
- [63] S. Notarnicola, A. Elben, T. Lahaye, A. Browaeys, S. Montangero, and B. Vermersch, A randomized measurement toolbox for an interacting Rydberg-atom quantum simulator, *New J. Phys.* **25**, 103006 (2023).
- [64] G. Bornet, G. Emperauger, C. Chen, F. Machado, S. Chern, L. Leclerc, B. Gely, D. Barredo, T. Lahaye, N. Y. Yao, and A. Browaeys, Enhancing a many-body dipolar Rydberg tweezer array with arbitrary local controls, *Phys. Rev. Lett.* **132**, 263601 (2024).
- [65] B. Vermersch, A. Elben, L. M. Sieberer, N. Y. Yao, and P. Zoller, Probing scrambling using statistical correlations between randomized measurements, *Phys. Rev. X* **9**, 021061 (2019).
- [66] M. J. S. Beach, I. De Vlucht, A. Golubeva, P. Huembeli, B. Kulchytskyy, X. Luo, R. G. Melko, E. Merali, and G. Torlai, Qucumber: Wavefunction reconstruction with neural networks, *SciPost Phys.* **7**, 009 (2019).
- [67] R. D. Gill and S. Massar, State estimation for large ensembles, *Phys. Rev. A* **61**, 042312 (2000).
- [68] G. Torlai, B. Timar, E. P. van Nieuwenburg, H. Levinel, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, Integrating neural networks with a quantum simulator for state reconstruction, *Phys. Rev. Lett.* **123**, 230504 (2019).
- [69] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, in *Proceedings of the 48th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2016), pp. 913–925.
- [70] S. Aaronson, Shadow tomography of quantum states, in *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing* (ACM, New York, 2018), pp. 325–338.
- [71] M. Fujii, R. Kutsuzawa, Y. Suzuki, Y. Nakata, and M. Owari, Characterizing quantum pseudorandomness by machine learning, [arXiv:2205.14667](#).
- [72] R. Mukherjee, F. Sauvage, H. Xie, R. Löw, and F. Mintert, Preparation of ordered states in ultra-cold gases using Bayesian optimization, *New J. Phys.* **22**, 075001 (2020).
- [73] F. Sauvage and F. Mintert, Optimal quantum control with poor statistics, *PRX Quantum* **1**, 020322 (2020).
- [74] R. Mukherjee, H. Xie, and F. Mintert, Bayesian optimal control of Greenberger-Horne-Zeilinger states in Rydberg lattices, *Phys. Rev. Lett.* **125**, 203603 (2020).
- [75] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert, Anticoncentration theorems for schemes showing a quantum speedup, *Quantum* **2**, 65 (2018).
- [76] A. W. Harrow and S. Mehraban, Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates, *Commun. Math. Phys.* **401**, 1531 (2023).
- [77] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirack, Restricted Boltzmann machines in quantum physics, *Nat. Phys.* **15**, 887 (2019).
- [78] G. Torlai, Augmenting quantum mechanics with artificial intelligence., Ph.D. thesis, Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada 2018.
- [79] I. J. Myung, Tutorial on maximum likelihood estimation, *J. Math. Psychol.* **47**, 90 (2003).
- [80] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, *et al.*, Array programming with NumPy, *Nature (London)* **585**, 357 (2020).