

ARTICLE OPEN



Certification of quantum states with hidden structure of their bitstrings

O. M. Sotnikov¹, I. A. Iakovlev¹, A. A. Iliasov², M. I. Katsnelson¹, A. A. Bagrov^{2,3} and V. V. Mazurenko¹✉

The rapid development of quantum computing technologies already made it possible to manipulate a collective state of several dozens of qubits, which poses a strong demand on efficient methods for characterization and verification of large-scale quantum states. Here, we propose a numerically cheap procedure to distinguish quantum states which is based on a limited number of projective measurements in at least two different bases and computing inter-scale dissimilarities of the resulting bit-string patterns via coarse-graining. The information one obtains through this procedure can be viewed as a ‘hash function’ of quantum state—a simple set of numbers which is specific for a concrete wave function and can be used for certification. We show that it is enough to characterize quantum states with different structure of entanglement, including the chaotic quantum states. Our approach can also be employed to detect phase transitions in quantum magnetic systems.

npj Quantum Information (2022)8:41 | <https://doi.org/10.1038/s41534-022-00559-7>

INTRODUCTION

Theoretical description of objects invisible to human eye represents one of the challenging but, at the same time, most intriguing problems in physics through its history. For example, despite incessant improvement of optical instruments and the ability to look into more and more distant corners of the Universe, in many cases one can conclude on the existence of a planet only in an indirect way by analyzing its tiny influence on the orbits of neighboring visible planets¹ and stellar brightness².

In the opposite limit of the atomic scale, the situation is even more complicated. When the object of our principle interest is a many-body quantum state—wave function or density matrix—we should conclude on its existence and properties indirectly on the basis of measurements. Moreover, in contrast to observation of celestial objects whose collective motion could be completely described with laws of classical mechanics, a measurement in quantum world does not provide a complete information about a system due to the uncertainty principle³, and characterizing quantum matter from such limited probes represents a non-trivial methodological and technical problem.

The conventional technique to analyze quantum state of a multi-component physical system is quantum tomography, which is based on the idea of complete⁴ or partial⁵ reconstruction of the wave function or density matrix from a number of measurements. Complexity of the tomographic procedure is mainly related to the number of qubits involved and the complexity of the quantum state itself, about which one might or might not have some prior expectations. In many cases, it could be non-trivial to choose a set of observables which is tomographically complete (or sufficient for partial reconstruction) and, at the same time, experimentally accessible⁵.

The main fundamental limitation of quantum tomography is that one needs to store and manipulate the to-be-reconstructed quantum state on a classical computer, which makes characterization of systems that comprise more than a few dozens of qubits unfeasible. Taking into account that quantum states of 53 qubits can already be generated on modern quantum devices⁶, and a

significant increase of this number is expected in the coming years, seeking an approach that overcomes this limitation appears to be a problem of high importance.

A natural way to reduce the memory required for state reconstruction is to store it in an implicit form of a compact variational ansatz. One of the most promising approaches of this kind is the recently proposed neural-network version of quantum tomography^{7,8}, which represents the wave function as a Neural Quantum State⁹ and reconstructs it via the learning procedure. While this approach has many benefits such as very high expressibility of neural-network ansätze^{10,11}, it does not resolve all the problems of quantum tomography. Some quantum states, such as defined by wave functions with random or uniform distributions of amplitudes over the Hilbert space basis, require exponentially large number of measurements (of the order of the Hilbert space dimension) for reconstruction. The situation cannot be improved by employing neural networks, since there are no features that the neural network can detect in the measured data, learn and generalize⁷.

Thus, it is important to find a way to bypass the resource-consuming routine of conventional quantum tomography at least in certain contexts. A typical problem, when there is a chance not to get engaged in this procedure, is certification of a state prepared on a quantum information processing device. In this case, there are strong prior expectations of what this state should be. Thus, instead of its complete reconstruction, one could hope to read out simple signature serving as a fingerprint of the many-body state—in a spirit similar to hash functions in computer science^{12,13}—to make sure that the state is, with high probability, indeed the correct one (see ref. ¹⁴ for the usage of hash functions in quantum tomography).

In this paper, we introduce such a signature that can be constructed by means of a reasonable number of simple von Neumann measurements of the quantum state and does not require computing correlation functions. Ideologically, this can be viewed as going along the line of the very recent approach of classical shadow tomography^{15,16}, though the signature we

¹Theoretical Physics and Applied Mathematics Department, Ural Federal University, Mira Str. 19, 620002 Ekaterinburg, Russia. ²Institute for Molecules and Materials, Radboud University, 6525 AJ Nijmegen, The Netherlands. ³Department of Physics and Astronomy, Uppsala University, SE-75120 Uppsala, Sweden. ✉email: vmazurenko2011@gmail.com

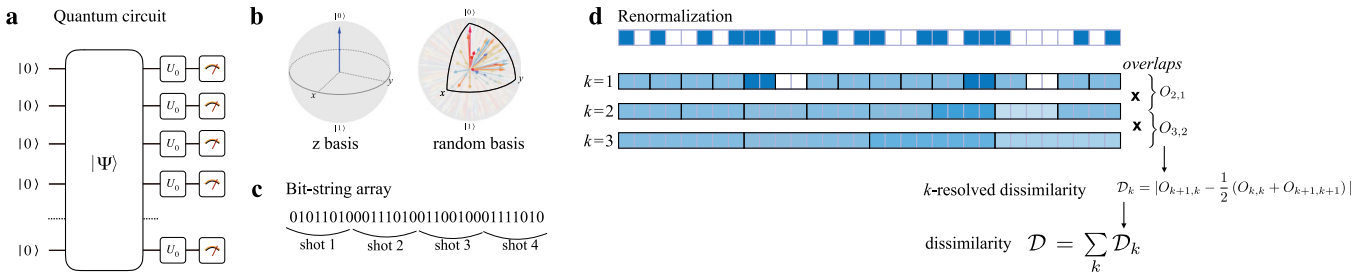


Fig. 1 Protocol for computing dissimilarity of a quantum state. **a** First, one prepares a state on a quantum device and chooses the measurement basis by applying rotational gates U_θ to individual qubits. **b** In this paper, we work with σ^z and random bases whose Bloch sphere representations are shown in the picture. We say that the set of measurements is performed in a random basis if, for each shot of measurement, a random vector belonging to the highlighted sector of the Bloch sphere is uniformly sampled and the corresponding parameters of gate U_θ are applied. **c** A number of measurements is performed and their outcomes — bitstrings of length N — are then stacked together in a one-dimensional binary array of length $N \times N_{\text{shots}}$ that serves as a classical representation of the quantum state. **d** The array is coarse-grained in several steps (indexed with k). Different schemes can be employed, but here we use plain averaging with fixed filter size Λ . In the picture, blue and white squares in the top line correspond to '0' and '1' bits in the array shown in (c), and black rectangles depict the blocks where averaging occurs at every step of coarse-graining. Overlap-based dissimilarities \mathcal{D}_k between subsequent arrays are computed and summed up to the overall dissimilarity \mathcal{D} . See Methods section for more details.

employ is different. To accomplish that, we heavily rely on the concept of multi-scale structural complexity of classical patterns that has been recently defined by some of the authors of this paper¹⁷. To avoid possible terminological confusions with the well-established notion of quantum complexity, here we call it dissimilarity (since it is based on counting how much different spatial scales of an object differ from each other). The detailed description of the protocol is given in the Methods section, and here we outline the main idea.

Assume, we have access to a many-body quantum state. To do benchmark tests, in this paper we use both numerical wave functions (e.g., resulting from exact diagonalization) and physical quantum states generated on the IBM quantum simulator¹⁸. With no loss of generality, we will be considering spin-1/2 systems. A single-shot projective measurement of such a state results in a string of bits of length N —measured spin projections on a chosen direction: $|S_i\rangle = |0110\dots 010\rangle$ (0 for spin-down and 1 for spin-up)—where N is the number of qubits. Performing the measurement many times (denote this number with N_{shots}) and collecting the outcomes in a string, we obtain a bit-string array of length $L = N \times N_{\text{shots}}$. This array can then be viewed as a one-dimensional pattern, and its inter-scale dissimilarity can be computed. For that, we do several steps of coarse-graining (we label the steps with index k) and for each pair of subsequent scales compute how distinct the corresponding coarse-grained strings are. The distinction is assumed to be large if overlap of arrays at two subsequent scales is small. For two neighboring scales, we call these measures partial dissimilarities, \mathcal{D}_k , and their sum over all scales $\mathcal{D} = \sum_k \mathcal{D}_k$ gives the total inter-scale dissimilarity. Different schemes of coarse-graining can be employed, and here we resort to the simplest option: we fix filter of width Λ (usually $\Lambda = 2$), and at step k we substitute all the pixels within a window of size Λ^k with the average value of pixels in this window at the previous step, Fig. 1. Despite probabilistic nature of the measurements, in all the tested cases dissimilarity turns out to be a statistically robust signature of the state.

If this procedure is performed in a single basis, it does not reveal any information on the phase structure of the quantum state, since measurement outcomes are defined solely by probability distribution on the Hilbert space basis $|\Psi(S_i)\rangle^2$. Also, unique characterization of a many-body quantum state with a single number is clearly impossible. However, if such bit-string arrays are constructed in two or more different Hilbert space bases, one obtains a sequence of numbers that implicitly contains information on both amplitude and phase structure of the state. The more bases are involved, the less it is likely that two different quantum

states would share the same dissimilarity signature (in a different context, the tomographic advantage of using several bases was discussed in ref. 19).

In this paper, we do not go beyond measurements in two bases, and this seems enough to characterize several important families of quantum states. As a warm up, we consider the families of Dicke and Schrödinger cat states which have compact analytical representations, and demonstrate how the concept of bit-string inter-scale dissimilarity can be used for dimensional reduction and visualization of specific signatures of wave functions. We also reveal the connection between the dissimilarity measure and the von Neumann bipartite entanglement entropy which plays a central role in quantum information theory. Then, we test our approach by using it for certification of random quantum states characterized by complete delocalization in the Hilbert space, which we do both numerically and analytically. We also show that the proposed approach scales nicely and requires the same experimental efforts to certify 16-qubit and 53-qubit states. Using the transverse-field Ising, the Shastry-Sutherland, and the bond-alternating XXZ Heisenberg models as examples, we show that the inter-scale dissimilarity can be used as a universal tool for detecting quantum phase transitions in many-body systems, including the topological ones. Finally, we discuss how the concept of inter-scale dissimilarity can be used for dimensional reduction and visualization of many-body quantum states.

RESULTS

Notable entangled quantum states

To demonstrate the idea of bit-string arrays and inter-scale dissimilarity, we begin with the Schrödinger cat states defined by superposition of merely two basis vectors in the Hilbert space

$$|\Psi_\theta\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle^{\otimes N} + \sin\left(\frac{\theta}{2}\right)|1\rangle^{\otimes N}. \quad (1)$$

Parametrized by angle θ , this family of states interpolates between trivial product state $|0\rangle^{\otimes N}$ at $\theta = 0$ and the famous Greenberger-Horne-Zeilinger (GHZ) state $\Psi_{\text{GHZ}} = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$ at $\theta = \frac{\pi}{2}$. These states can be realized with quantum circuit²⁰ shown in Fig. 2a. First, with rotational gate U_θ one prepares $\cos(\frac{\theta}{2})|0\rangle + \sin(\frac{\theta}{2})|1\rangle$ state of one of the qubits in the system and takes it as a control qubit to perform controllable-NOT operation on the second qubit. This operation results in a two-qubit entangled state $\cos(\frac{\theta}{2})|00\rangle + \sin(\frac{\theta}{2})|11\rangle$. Repeating it $N - 1$ times, one eventually entangles all the qubits and obtains the target Schrödinger cat state.

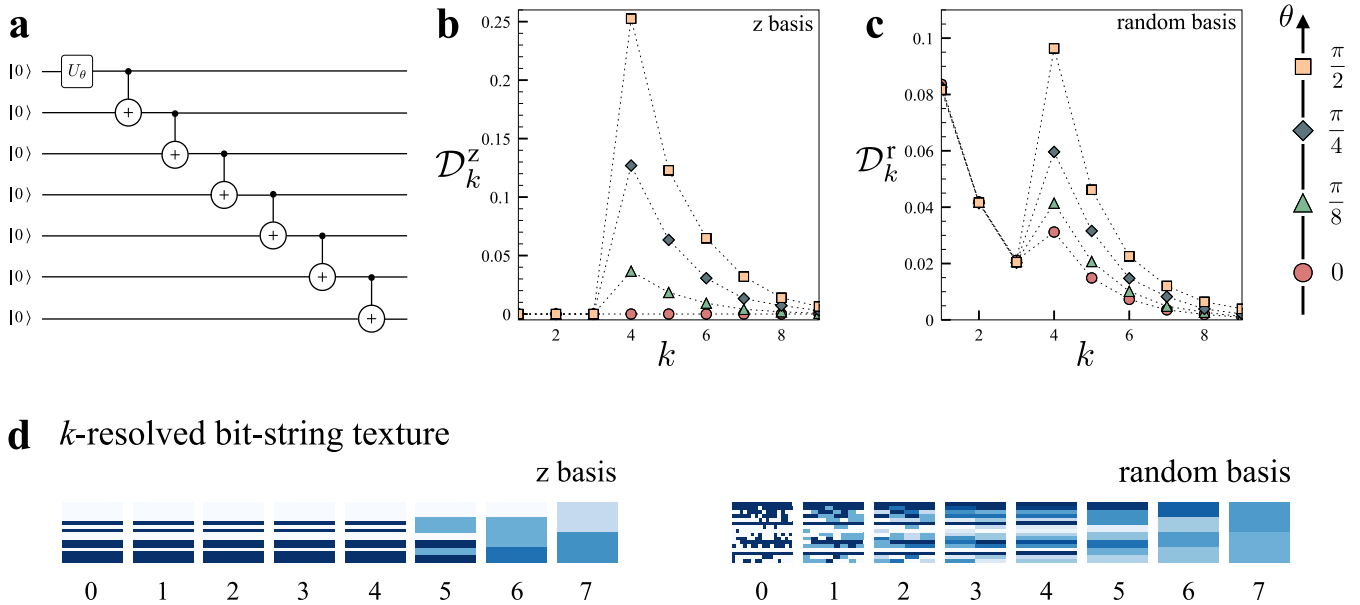


Fig. 2 Dissimilarity results for Schrödinger cat states. **a** Quantum circuit generating Schrödinger cat states. **b, c** Partial dissimilarities \mathcal{D}_k of 16-qubit Schrödinger cat states calculated in the σ^z and the random bases correspondingly. Here, $\Lambda = 2$. The peak at $k = 4$ is due to the fact that the coarse-graining window becomes of the size of the system at this scale, $\Lambda^k = N$. **d** Visualization of bit-string arrays. In these images, individual bitstrings are horizontal lines of 16 bits that are concatenated to form a long string (stacked here vertically for the purpose of presentation). Left picture shows an example of array sampled from a cat state with $\theta = \frac{\pi}{2}$ in the σ^z basis, and the right one — measured in the random basis. Here $k = 0$ represents texture of the measured array per se, and $k > 0$ show its evolution upon coarse-graining.

In the σ^z -basis, projective measurements of such states can only result in either 0000...0 or 1111...1 bitstring. Clearly, first steps of coarse-graining (Fig. 2b) affect only internal content of individual bitstrings of length N , where it simply maps 0000...0 \rightarrow 0000...0 and 1111...1 \rightarrow 1111...1. Thus the randomly assembled array of bitstrings remains intact, and partial dissimilarities $\mathcal{D}_k \equiv 0$ for k such that $\Lambda^k < N$ (for $k < 4$ when we take $N = 16$ and $\Lambda = 2$). At $\Lambda^k \geq N$, the coarse-graining flow starts mixing individual bitstrings, and non-trivial contributions to the dissimilarity emerge. In random basis (Fig. 2c), \mathcal{D}_k take finite values at all scales k , though due to the trivial structure of basis vectors defining Ψ_θ partial dissimilarities do not depend on θ at $\Lambda^k < N$.

Importantly, each state presented in Fig. 2b, c reveals a distinct set of \mathcal{D}_k which can be used to distinguish states from each other. Schrödinger cat states are the simplest example of many-body entangled wave functions, but in what follows we will show that the same idea can be exploited when dealing with much more complex states. It has to be stressed out one more time that, while individual bitstrings are assembled into array in a random order set by the outcomes of consecutive projective measurements, the partial dissimilarities and their total sum are robust upon repeatedly performing the set of measurements.

Another type of entangled states that are instructive to consider is the family of Dicke states²¹,

$$|\Psi_D\rangle = \frac{1}{\sqrt{C_D^N}} \sum_j P_j(|0\rangle^{\otimes N-D} \otimes |1\rangle^{\otimes D}), \quad (2)$$

where the sum goes over all possible permutations of qubits. By increasing D from 1 to $\frac{N}{2}$, one increases the number of basis vectors involved into the quantum state. Recently, these states have been experimentally realized²², and their verification is a challenging task if the number of qubits is large²³. As a proof of concept, in this paper we study Dicke states of 16 qubits, and initialize them on quantum simulator using the Least Significant Bit procedure²⁴.

Partial dissimilarities of the 16-spin Dicke states computed in σ^z -basis and in the random basis with filter size $\Lambda = 2$ are shown in Fig. 3. One can see that two different bases encode information

about two ranges of scales. For any given parameter D , when bit-string arrays are constructed from measurement in the σ^z basis, \mathcal{D}_k take non-zero values only for $k < 4$, which follows from the fact that all the Hilbert space basis vectors possessing non-zero amplitudes have equal amount of spin-up entries, and after 4 steps of averaging every bit string reduces to exactly the same number, and all the patterns are destroyed. Contrary, in the random basis, states with different D can be distinguished from \mathcal{D}_k at larger spatial scales, $k \geq 4$.

Since both families of states smoothly interpolate between the regimes of low and high entanglement, it is interesting to study if there are any relations between the introduced measure of inter-scale dissimilarity and quantum correlations. To do that, we consider the von Neumann entanglement entropy

$$S(\rho_A) = -\text{Tr}_A \rho_A \log_2(\rho_A), \quad (3)$$

$$\rho_A = \text{Tr}_B \rho_{AB},$$

where the system is divided into two equal parts A and B of $N/2$ qubits, compute its dependence on either θ or D (depending on the family), and plot it alongside the inter-scale dissimilarity of bit-string arrays computed in the σ^z basis.

The result is shown in Fig. 4. While the Dicke and the Schrödinger cat states are quite different in the regard that variation of parameter D modifies the structure of the wave function support in the Hilbert space basis, and θ only changes the balance between two basis vectors bearing non-zero amplitudes, in both cases dissimilarity nicely captures dependence of entropy on the parameters labeling the state within the family. Although the precise analytical correspondence between these two concepts is still to be revealed, it could be a good indication that it is possible to employ dissimilarity to estimate entanglement entropy, which is generally very difficult to reconstruct from experimental measurements, especially when dealing with multi-qubit systems inaccessible to quantum tomography. In a certain way, it is similar to the approach proposed in ref. ²⁵, where it was shown that, with the help of neural networks, entanglement can

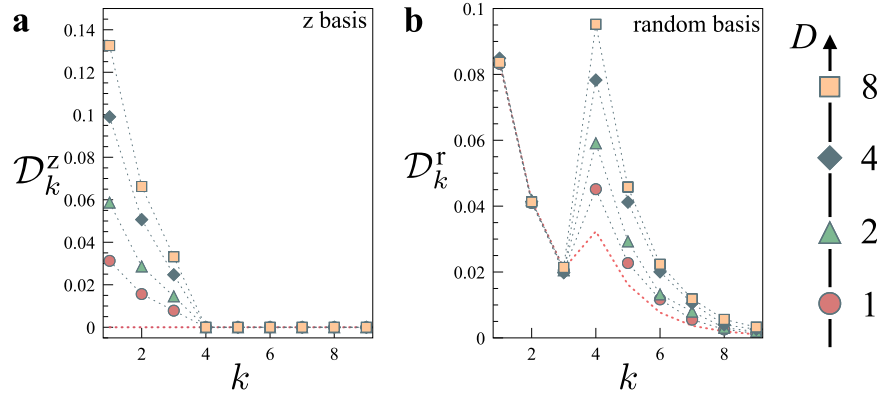


Fig. 3 Dissimilarity results for Dicke states. Partial dissimilarities of Dicke states with different D index calculated in the σ^z (a) and the random (b) bases. The trivial state $(|0\rangle^{\otimes 16})$ profiles (dashed red lines) are given for comparison.

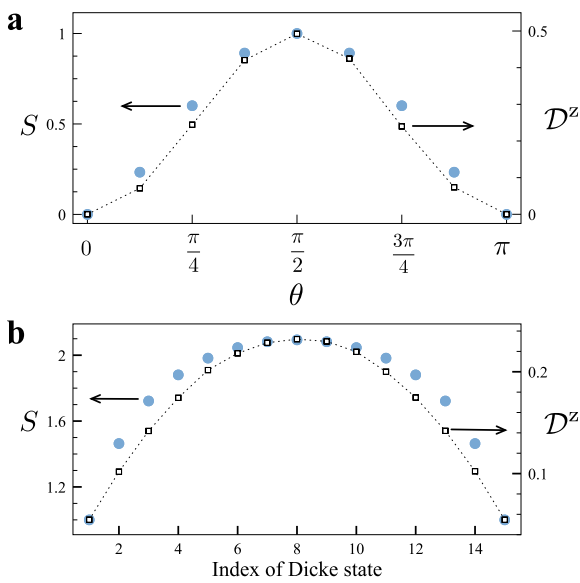


Fig. 4 Comparison of entanglement and dissimilarity. a Entanglement entropy S (blue circles) and overall dissimilarity \mathcal{D}^z (white squares) of the Schrödinger cat states as functions of angle θ . b The same characteristics of the Dicke states as functions of index D .

be reconstructed from visual pattern representations of quantum states.

Random quantum states

Our next goal is to demonstrate that the dissimilarity measures can serve as a signature not only of highly structured states with simple analytical representations, but of rather generic many-body states. To do that, we consider Haar-random wave functions uniformly sampled from the Hilbert space and characterized by the Porter-Thomas distribution of bit-string probabilities $p = |\langle x_1, \dots, x_N | \psi \rangle|^2$ that have recently been used to demonstrate quantum supremacy⁶. These states play an important role in studying quantum chaos theory²⁶, quantum information theory²⁷ and information processing, including research domains of superdense coding of quantum states²⁸ and data hiding²⁹, and even transport phenomena³⁰. While complete tomography of a given random state is an extremely complicated task since the minimal number of measurements to be performed to reconstruct a random quantum state should be of order of the Hilbert space dimension^{4,7}, here we show that to certify if a state belongs to the

Haar-random class one can resort to computing inter-scale dissimilarities of relatively short bit-string arrays.

As it was shown in refs. ^{6,30,31}, random quantum states can be initialized with shallow pseudo-random circuits that can differ in the number and types of gates, and practical realization of these circuits on a real quantum device depends on its architecture. In this work, we generate random quantum states of a 16-qubit system on the IBM quantum simulator with the protocol proposed in ref. ³⁰, which guarantees an accurate approximation of the Haar-random state with a compact circuit shown in Fig. 5a. More specifically, the circuit is formed in cycles, each having one- and two-qubit gate layers. Within the first layer, for each qubit in system one randomly chooses from \sqrt{X} , \sqrt{Y} and T gates, where \sqrt{X} (\sqrt{Y}) are $\pi/2$ rotations around the x-axis (y-axis) of the Bloch sphere, and the non-Clifford gate $T = \text{diag}(1, e^{i\pi/4})$. In turn, the second layer comprises controlled-Z gates, $\text{diag}(1, 1, 1, -1)$, whose topology is randomly chosen from the set of configurations with fixed couplings between qubits, as described in ref. ³⁰.

In both σ^z and random bases, the inter-scale partial dissimilarities of the array generated by sampling 8192 bitstrings from a random quantum state follow the same decaying profile, Fig. 5b. Such a profile is a robust signature of typical Haar-random states. It can be shown that, for a chosen filter size Λ , the dependence of \mathcal{D}_k on the step index k obeys a simple analytical law in the averaging coarse-graining scheme:

$$\mathcal{D}_k = \frac{1}{2}(1 - \Lambda^{-1})\Lambda^{-k}. \quad (4)$$

To derive this law from Eq. (9), the central limit theorem must be employed as elaborated in the “Methods” section. This dependence is easy to reconstruct from a limited number of simple projective measurements, and it serves as a signature of the class of typical Haar-random states.

To go beyond the simple 16-qubit case and perform an ultimate test of the method, we have applied it to the real experimental data generated on the Google Sycamore quantum processor⁶. For systems of 16, 32 and 53 qubits, we have taken 8192 bitstrings measured in the σ^z basis and calculated partial dissimilarities, which turned out to perfectly fit Eq. (4). The result for the prominent example of 53-qubit system is presented in Fig. 5c.

In a real-world scenario, the bit-string arrays are clearly a subject to the gate errors and other sources of noise, and we have to understand how these imperfections are reflected in the dissimilarity signatures of the state. Previous studies^{6,31} have demonstrated that random quantum states are hypersensitive to the gate errors, which is considered to be a defining property of quantum chaos. When the error rates increase, the distribution of probabilities of the bitstrings generated by a random circuit deviates from the Porter-Thomas law $\text{Pr}(p) = 2^N e^{-2^N p}$ and converges to equal probabilities of all the bitstrings: $\text{Pr}(p) = \delta(2^{-N} - p)$.

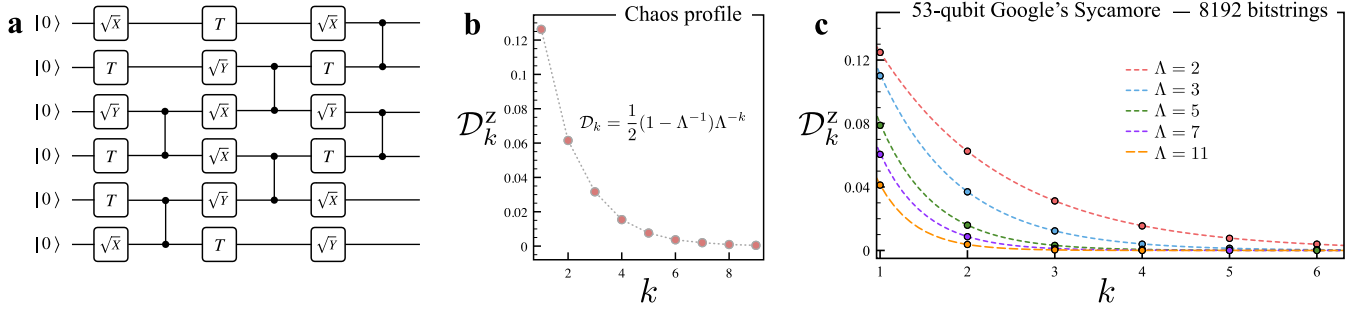


Fig. 5 Preparation of chaotic states. **a** Fragment of a quantum circuit generating chaotic quantum state according to the protocol proposed in ref. ³⁰. **b** Partial dissimilarities (red circles) of bit-string arrays resulting from 8192 projective measurements of a 19-layer-deep quantum chaotic circuit with 16 qubits in the σ^z basis (in the random basis, the profile of \mathcal{D}_k is exactly the same). Here, filter size $\Lambda = 2$. Dashed line shows the analytical fit with (4). **c** Partial dissimilarities of bit-string arrays in the σ^z basis resulting from 8192 projective measurements of the state produced by the 53-qubit Sycamore quantum processor by Google. These data were taken from ref. ⁶, and different filter sizes Λ were used to compute \mathcal{D}_k . Dashed lines correspond to Eq. (4). Random basis measurements are not available in this case.

To quantify this deviation, the authors of refs. ^{6,31} have introduced the cross-entropy benchmarking procedure. It allows to estimate with a limited number of measurements how close a sampler—a given quantum circuit—to one of the two limiting cases: the ideal random quantum circuits with Porter-Thomas distribution of probabilities and uniform sampler with identical probabilities $p(x_1, \dots, x_N) = 2^{-N}$. In this respect, it is important to distinguish between outputs of quantum circuits with the Porter-Thomas and the uniform probability distributions by calculating the inter-scale dissimilarity.

To elaborate on this point, we prepared a quantum circuit consisting of only the Hadamard gates that generates a 16-qubit state with uniform probabilities in the σ^z basis: $|X\rangle = (H|0\rangle)^{\otimes 16}$. Each qubit is then in the superposition $(|0\rangle + |1\rangle)/\sqrt{2}$. The obtained dissimilarity profile of the generated uniform state fully coincides with that obtained for the chaotic quantum circuits (Fig. 5b), with the overall dissimilarity $\mathcal{D}^z = 0.25$. Thus, from σ^z basis measurements we cannot distinguish these two states that are fully delocalized in the Hilbert space. However, in the random basis they have different profiles of \mathcal{D}_k and overall \mathcal{D} . While the chaotic quantum circuit is characterized by an isotropic character of the dissimilarity that is independent on the measurement basis (so, the \mathcal{D}_k profile is exactly the same in the random and the σ^z bases), the $|X\rangle$ state in the random basis reveals its trivial nature. Namely, its $\{\mathcal{D}_k^r\}$ and the resulting dissimilarity $\mathcal{D}^r = 0.204$ coincide with that obtained for $|0\rangle^{\otimes 16}$ in the random basis (as shown in Fig. 3b). This suggests that the inter-scale dissimilarity can be used to quantify deviations from a truly chaotic quantum states, which would be interesting to verify experimentally.

Phase transitions in magnetic systems

Since the inter-scale bit-string dissimilarity appears to be a rather unique signature of many-body state, it is natural to expect that it should be sensitive to crossing phase boundaries in the parametric spaces of many-body quantum systems. If so, one can hope that it can be used as a sensitive indicator of phase transitions and directly used for constructing quantum phase diagrams, which is a crucial task in understanding phenomenology of correlated materials and designing materials. The common practice is to distinguish different phases of a quantum or classical many-body system by calculating the order parameter³² and low-order correlation functions such as susceptibility, scalar chirality and others. However, in many cases devising the order parameter is a non-trivial analytical problem, especially in the case of topological phases³³. Besides that, a quantum system may have a rich variety of different electronic and magnetic phases depending on internal (interactions) and external (temperature, pressure,

magnetic field) parameters, and there could be no universal operator that can probe the whole phase diagram.

To overcome this problem, a lot of effort has been put into designing alternative approaches based on neural networks^{34–36}, unsupervised machine learning techniques³⁷, and quantum information theory concepts^{38,39}. These methods usually rely on manipulating eigenstates of the quantum system on a classical computer, which puts natural limitations on the size of systems that can be studied in this way. Also, it can be time- and resource-demanding to conduct, e.g., the learning procedure.

The progress in developing quantum simulators and quantum computing devices suggests a distinct way for large-scale representation of a quantum systems and analysis of their phase diagrams. Instead of solving the Hamiltonian numerically, one can imitate it in an, e.g., optical experiment. For example, by varying depth of the potential in optical lattices, one can change the ratio between hopping integrals and on-site Coulomb interaction in the simulated strongly-correlated electronic or bosonic system, and scan through its parametric space in this way. Recent advances in this field include simulation of the electronic metal-to Mott insulator transition^{40,41} and destruction of the antiferromagnetic long-range order with temperature and doping⁴². Analysis of such experiments is then conducted by means of a limited set of site-resolved measurements performed on the system, and the relevant information should be extracted from these measurements, whose number is much smaller than the Hilbert space dimension. We refer the reader to refs. ^{43,44} for an interesting machine learning-based approach to the analysis of optical lattice experiments, and in what follows we discuss how the concept of bit-string arrays and their inter-scale dissimilarity can enter the game and aid reconstruction of phase diagrams of simulated quantum matter.

As some of us have shown in ref. ¹⁷, the classical prototype of inter-scale dissimilarity—the structural complexity of patterns—can be used to detect phase transitions in classical systems without any prior knowledge of the order parameter, and in an extremely numerically cheap unsupervised manner. Now, we will show how it can be extended onto the quantum case and help reconstruct quantum phase diagrams of many-body systems from simple projective measurements. We will be using the transverse-field Ising and the Shastry-Sutherland models as examples.

The simplest example of a quantum phase transition is the paramagnet-to-ferromagnet transitions in the ferromagnetic Ising model in the transverse magnetic field given by the Hamiltonian

$$H = J \sum_{ij} \hat{S}_i^z \hat{S}_j^z + h \sum_i \hat{S}_i^x, \quad (5)$$

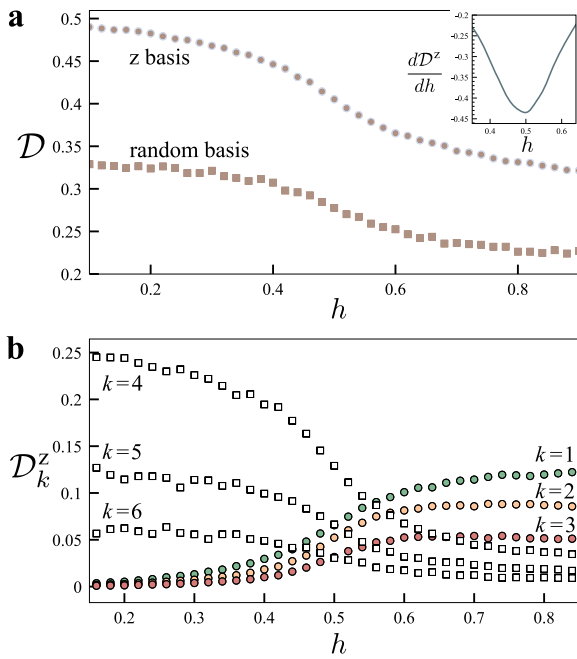


Fig. 6 Ising model results. **a** Dissimilarity of the Ising model ground state as a function of the transverse magnetic field in the σ^z and the random bases; the inset shows derivative of the dissimilarity in the σ^z basis with respect to h . **b** Partial dissimilarities \mathcal{D}_k in the σ^z basis at different coarse-graining steps $k = 1 \dots 6$. The global change in the trend between $k = 3$ and $k = 4$ is related to the coarse-graining window reaching the size of the system, $\Lambda^k = N$, so that the averaging starts mixing different bitstrings.

where J and h are the exchange interaction between nearest neighbor spins and the external magnetic field along x-axis, respectively, and we consider the case of one-dimensional chain of 16 spins with periodic boundaries. The critical value of magnetic field is known to be $h_c = 0.5|J|$, and to reproduce this value is the first benchmark test for our method before we consider more sophisticated examples.

In the regime of weak magnetic field, the 16-spin system's ground state obtained with the exact diagonalization approach⁴⁵ is a superposition of two fully polarized states $|\uparrow\rangle^{\otimes N}$ and $|\downarrow\rangle^{\otimes N}$, which is nothing but the entangled GHZ state discussed above. In the σ^z basis, the bit-string array generated by projective measurements is a random sequence of 000...0 and 111...1 blocks. In turn, at very high magnetic fields the qubits are pointing in the same direction along x axis, and the state is just a trivial product state that can be obtained from $|0000\dots 0\rangle$ by rotating all the qubits with the same Hadamard gate.

Figure 6 shows the overall dissimilarity as a function of the magnetic field. One can see that in both σ^z - and random bases, the dissimilarity steadily decreases with increasing h , and the corresponding derivative $\mathcal{D}'(h)$ reveals the well-known transition point at $h = 0.5$ (we take $J = -1$). The phase transition is also reflected in the partial dissimilarities \mathcal{D}_k corresponding to individual renormalization steps. At low magnetic fields, the state is close to GHZ and there is clearly little inter-scale dissimilarity at small k : on the fine scale, coarse-graining of $|0000\dots 0\rangle$ does not bring any dissimilarity,—and the main contributions to \mathcal{D} come from larger k , i.e. from the spatial scales covering several N -qubit blocks. Contrary to that, at larger fields finer scales start playing more important role. For each k , the phase transition at $h = 0.5$ is visible in the derivative $\mathcal{D}'_k(h)$.

A much less trivial test of the method is to check whether it can reveal transition points in highly-frustrated spin systems with

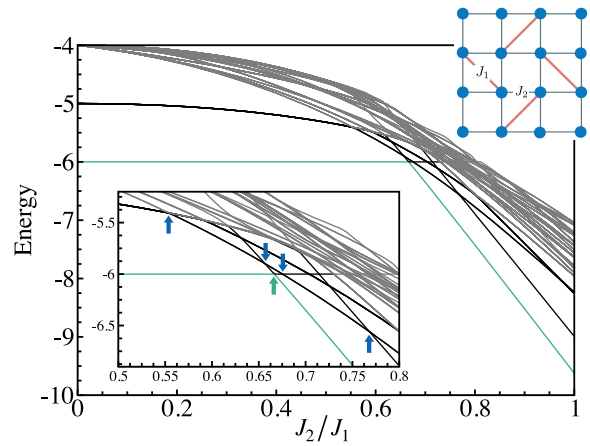


Fig. 7 Exact diagonalization results for the Shastry-Sutherland model. Upper right inset: schematic representation of the Shastry-Sutherland model 16-spin supercell used in this work. Main plot and the inner inset: low-energy part of its spectrum as a function of the inter-dimer exchange interaction J_2/J_1 . Arrows denote transitions between quantum states. The green line represents the ground state.

richer phase diagrams. For that, we consider the Shastry-Sutherland model⁴⁶ with competing antiferromagnetic interactions on the orthogonal dimer lattice, which plays a crucial role in understanding physical properties of the $\text{SrCu}(\text{BO}_3)_2$ system^{47–49}. The corresponding Hamiltonian contains intra- and inter-dimer interactions, which are denoted J_1 and J_2 correspondingly (Fig. 7):

$$H = \sum_{\text{dimer}} J_1 \hat{\mathbf{S}}_i \hat{\mathbf{S}}_j + \sum_{\text{inter-dimer}} J_2 \hat{\mathbf{S}}_i \hat{\mathbf{S}}_j. \quad (6)$$

As it was previously shown, the system features a gapped singlet ground state at $J_2 = 0$, gapless long-range antiferromagnetic Néel state at $J_2 \gg J_1$, but also a plaquette phase in-between, in the range of $0.67 < J_2/J_1 < 0.76$. While, strictly speaking, the quantum phase transition is defined in the thermodynamics limit of infinite lattices, its precursor could be detected already in a small system³⁸. For example, in the case of Shastry-Sutherland model it has been suggested that by analyzing spin gap and spin-spin correlation functions one can extract the singlet-plaquette and plaquette-Néel transitions from exact diagonalization studies of small clusters⁵⁰. We are going to show that it can also be done with the inter-scale dissimilarity measure, which is agnostic about the nature of phase transition and much easier to implement on quantum simulators and quantum computers.

We have performed exact diagonalization study⁴⁵ of a 16-spin Shastry-Sutherland supercell—the smallest cluster on which the model can be defined. Its energy spectrum is presented in Fig. 7. One can see that up to $J_2 = 0.66J_1$ the ground state of the system is the singlet state separated from the first excited state with a non-zero spin gap, and its energy is independent on the inter-dimer coupling value J_2 . At $J_2 = 0.66J_1$ a quantum phase transition takes place. The previous studies⁵⁰ have shown that increasing the supercell size does not change the position of the critical point. The inter-scale dissimilarity (Fig. 8) naturally captures this transition: for $J_2 < 0.66J_1$, \mathcal{D} of the ground state computed from 8192 measurements is a constant, $\mathcal{D} = 0.25$, and an abrupt transition occurs at the critical point in both the σ^z and the random bases. The corresponding partial dissimilarities at $J_2 = 0$ and $J_2 = J_1$ are shown in Fig. 8c.

In the thermodynamic limit, the cases of $J_2 = 0$ and $J_2 = 1$ correspond to the magnetic phases with and without spin gap between the ground and the first excited state. In the finite-size system, it means that non-trivial signatures of phase transitions could be encoded not only in the ground state, but also in the

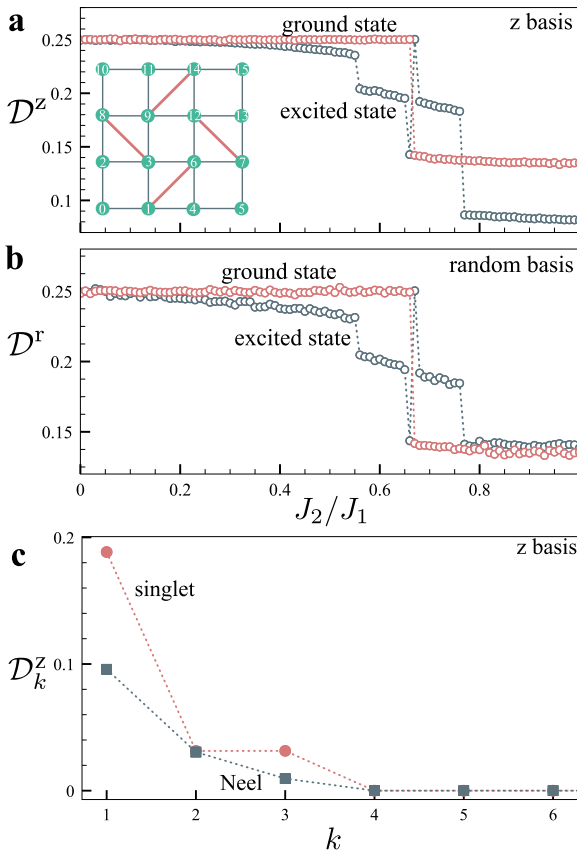


Fig. 8 Dissimilarity results for the Shastry-Sutherland model. Dissimilarity of the ground and the first excited states (with zero total spin) of the Shastry-Sutherland model as a function of the inter-dimer coupling in the σ^z (a) and the random (b) bases. The inset in panel (a) shows how spins are grouped to obtain flattened bitstrings. c Comparison of the partial dissimilarity profiles obtained for the singlet ($J_2 = 0$) and the Néel ($J_2 = 1$) states in the σ^z basis.

excitation spectrum. At $J_2 < 0.55J_1$, the first excited state has three-fold kind degeneracy: it is of triplet type with total spin values $S^z = 0, \pm 1$. Above the transition point, it is replaced with a two-fold degenerate state with zero total spin. This state reconfiguration causes the difference in magnetization profiles for the inter-dimer order parameter above and below the point of $J_2 = 0.55J_1$ when the external magnetic field is applied. According to the previous studies⁵¹, the magnetization features a plateau at 1/8 of the full moment for $J_2 = 0.65$, but not for $J_2 = 0.4$.

At the point of $J_2 = 0.76J_1$ (Fig. 7), the plaquette-Néel phase transition takes place. Stability of this point upon varying the system size was previously confirmed by different methods^{47,50,52}.

From Fig. 8, one can see that all three transitions—at $J_2 = 0.55J_1$, $J_2 = 0.66J_1$, and $J_2 = 0.76J_1$ —are accurately reflected in the inter-scale dissimilarity of bit-string array sampled in σ^z and random bases from the first excited state of the Shastry-Sutherland model. We also show that the partial dissimilarities of the ground state calculated for $J_2 = 0$ and $J_2 = 1$ have specific distinguishable profiles. We believe this to be a strong argument in favour of universality of the suggested approach to automatic construction of phase diagrams of many-body systems simulated on quantum devices.

While dissimilarity is computed for one-dimensional bitstrings, to account for the two-dimensional nature of the system, we group the spins in 2d clusters before flattening, as shown in the inset of Fig. 8a. We have also tested that several ways to enumerate sites and form bitstrings do not reveal any additional

phase transitions in comparison with results presented in Fig. 8, while the concrete shape of the dissimilarity curve as a function of the model parameters can be different. We admit that, in a general case, the approach to flattening and concatenation can in principle affect the results.

So far we have been computing inter-scale dissimilarity of arrays composed out of 8192 measured bitstrings. However, it can be shown that in fact a much smaller number of measurements would suffice to complete the task of detecting phase transition points in many-body quantum systems. We found that, in the σ^z basis, partial dissimilarities \mathcal{D}_k of the Ising model ground states remain almost the same when we do 256 measurements instead of 8192. In the random basis, the minimal number of measurements that allows to reveal the ferromagnetic-paramagnetic transition is about 1024. In turn, the abrupt changes in the inter-scale dissimilarity of the Shastry-Sutherland model states could be revealed with mere 16 measurements. Thus, the method we propose allows one to accurately reconstruct phase diagrams of quantum spin Hamiltonians by using small-size supercells and a limited number of measurements.

Topological quantum phases

While the interscale dissimilarity seems to be sensitive enough to detect quantum phase transitions, it should be tested whether it also can be used to identify boundaries of topologically non-trivial phases. This task is more challenging since transitions of this type are governed by changes in the global properties of the system and often cannot be probed on the level of local correlation functions. An archetypal example of a system hosting topological order is the one-dimensional bond-alternating XXZ model^{53,54}, which exhibits time-reversal, inversion, and rotational symmetries. The corresponding Hamiltonian has the form

$$H_{\text{XXZ}} = J \sum_{ij \in \mathbf{O}} (\hat{S}_i^x \hat{S}_j^x + \hat{S}_i^y \hat{S}_j^y + \delta \hat{S}_i^z \hat{S}_j^z) + J' \sum_{ij \in \mathbf{E}} (\hat{S}_i^x \hat{S}_j^x + \hat{S}_i^y \hat{S}_j^y + \delta \hat{S}_i^z \hat{S}_j^z), \quad (7)$$

where \mathbf{O} and \mathbf{E} , correspondingly, denote the odd and the even bonds of the one-dimensional spin chain with open boundaries (denoted with red and blue in Fig. 9a). Depending on the value of the anisotropy parameter δ and the ratio of the exchange interactions, J'/J , the XXZ model was shown to host three different phases⁵³—trivial, topological and antiferromagnetic. Each phase is defined by the corresponding value of the partial-reflection many-body topological invariant. The latter is a non-trivial correlation function, which can be constructed only with some prior knowledge of the properties of the system. In this regard, the recent development of unsupervised approaches⁵⁴ to detect different phases of topological quantum systems is of timely interest.

Here, we consider the 16-spin chain, sample bitstrings of length 8 from the central part of the system, and use them to compute the dissimilarities. In this way, we manage to reproduce the phase diagram of the model, which has been previously obtained by calculating the topological invariant introduced in ref. ⁵³. The measurements of only the central spins are aimed to minimize the boundary effects caused by the small chain length. In contrast to the Ising and the Shastry-Sutherland models, for which \mathcal{D}^z and \mathcal{D}^r were found to behave similarly at all critical points, in the case of the bond-alternating XXZ model we observe that each of the two dissimilarities reveals its own phase transition. The random basis dissimilarity \mathcal{D}^r indicates the transition between the trivial (I) and the antiferromagnetic phase (III). In turn, the σ^z -basis dissimilarity \mathcal{D}^z dissimilarity senses the boundary between the antiferromagnetic (III) and the topological (II) phases. Here, to obtain the smooth and accurate phase boundaries, we took $N_{\text{shots}} \sim 10^6$. It is much larger than what was required in the previously considered

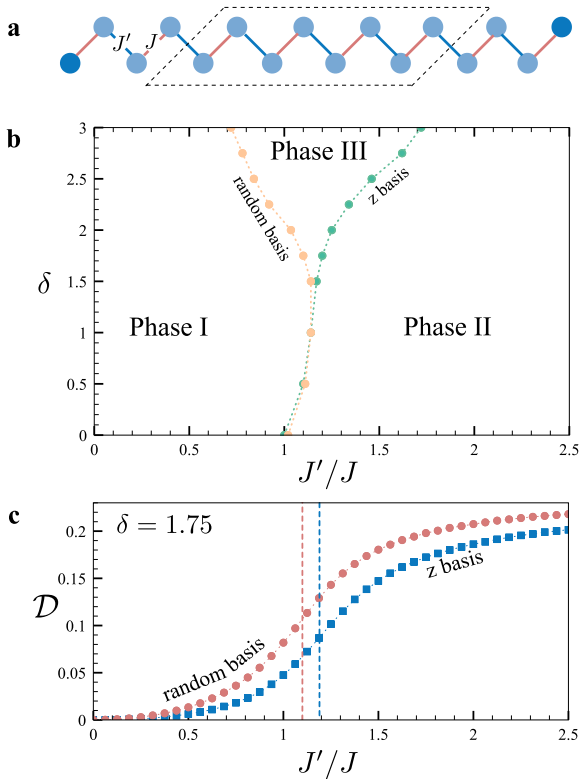


Fig. 9 Phase diagram of XXZ model. **a** Schematic representation of the bond-alternating 16-spin XXZ model. The spins in the central part of the chain denoted with the dashed line are measured to calculate the dissimilarities. **b** Phase diagram constructed by using the dissimilarity calculated in the σ^z and the random bases. The phase boundaries are defined as maximal values of the dissimilarity derivatives with respect to J' for a given δ . **c** Dissimilarities in the σ^z (blue squares) and the random (red circles) bases calculated at the anisotropy value $\delta = 1.75$. The vertical lines denote the positions of the maximum of dissimilarity derivatives with respect to J' .

cases. Still, from the point of view of real experiments⁶, such number of measurements is absolutely realistic.

Interestingly, to reconstruct the phase diagram using the classical shadow tomography, which is in principle a more complete protocol than the inter-scale dissimilarity approach, additional steps of post-processing are required (such as the Principle Component Analysis), and it is still problematic to precisely identify the phase boundaries⁵⁴. It indicates that the dissimilarity approach could be more advantageous in the context of studying phase transitions in experiments with programmable quantum simulators.

Multi-basis dissimilarity map

So far, we have analyzed a number of distinct examples of quantum states and demonstrated that their inter-scale dissimilarities (both overall and partial) computed in different measurement bases can be regarded as easily measurable signatures. To make this discussion more concise, it is natural to consider all the states within a single unifying context.

To accomplish that, we shall introduce the concept of dissimilarity map. For the sake of nicer visualization, assume that we characterize each quantum state with only two numbers—its overall dissimilarities \mathcal{D}^z and \mathcal{D}^r measured in the σ^z and random bases correspondingly. Each state is then represented by a point in two-dimensional space. Figure 10 shows several classes of states plotted on such a map. One can see that states belonging to different families nicely group in recognizable lines. The

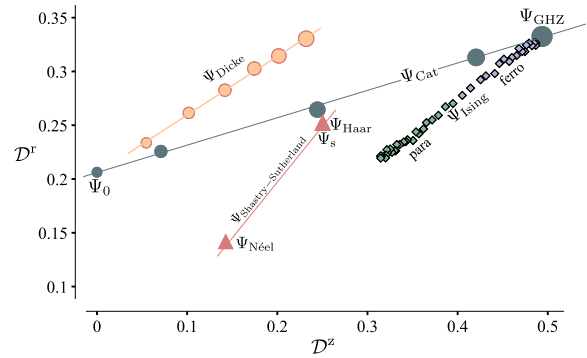


Fig. 10 Dissimilarity map. Low-dimensional representation of the 16-qubit quantum states studied in this work with respect to their dissimilarity calculated in σ^z and random bases. Ψ_0 , Ψ_s , Ψ_{Haar} denote the trivial $|0\rangle^{\otimes N}$, the singlet and the random quantum states, respectively.

dissimilarity map can be then thought of as an approach to dimensional reduction that embeds higher-dimensional data in a plane (if more bases were used, it would be a three- or four-dimensional space instead). Some states still share the same location on the map, like the singlet state of the Shastry-Sutherland model Ψ_s and the chaotic state Ψ_{Haar} both having $\mathcal{D}^r = \mathcal{D}^z = 0.25$. This is not unexpected, since a many-body state cannot be uniquely represented with only two numbers. However, taking into account also their partial dissimilarity profiles (Figs. 5 and 8) we can distinguish the states. This way, \mathcal{D} and \mathcal{D}_k computed in several (two or more) different bases altogether form a hash of quantum state.

DISCUSSION

In this paper, we have shown that bit-string arrays resulting from projective measurements of many-body quantum systems should be viewed as objects possessing internal hidden structure that contains important information about the measured quantum state. By computing inter-scale dissimilarities of the arrays, it is possible to define a specific characteristic of the state which serves as its ‘hash’ that can be then used to certify the state and to estimate its closeness to the desired target state.

Two measures have been introduced: the overall dissimilarity \mathcal{D} of the array in a chosen measurement basis, and the scale-dependent set of partial dissimilarities \mathcal{D}_k , which are building blocks of the quantum state signature. Since the bit-string array in a fixed basis is defined only by the probability distribution over the Hilbert space basis $|\psi(S_i)|^2$, it does not distinguish between two wave functions with the same set of amplitudes but different structure of phase. Thus it is important to compute \mathcal{D} and \mathcal{D}_k in two or more different bases. Since the procedure of performing projective measurements and computing dissimilarities is experimentally simple and numerically cheap, it is easy to repeat this procedure in several bases and construct a hash consisting of several numbers.

We would like to stress out that, in fact, the use of at least two measurement bases to characterize quantum system is not only practical, but also an important conceptual requirement directly related to Bohr’s complementarity principle^{55,56}. According to this principle, when observing a quantum system one gains information not about the quantum state per se but rather about the results of its interaction with a classical measuring device. Formally, the result of this interaction is described by the von Neumann theory of measurements⁵⁷ as a projection of the system density matrix with only diagonal elements surviving in the basis dictated by the device. The use of at least two noncommutative

projection operators corresponding to two complementary measurement devices is a necessary prerequisite of quantumness, as follows from a general ‘separation-of-conditions principle’⁵⁸. The latter dictates a description of quantum quantities by, at least, two-index matrices rather than ‘classical’ strings.

It has to be admitted that uniqueness of this signature is not guaranteed, and one can not exclude the possibility that two distinct quantum states have similar sets of \mathcal{D} and \mathcal{D}_k . However, if the number of involved measurement bases is large enough, such a coincidence seems highly unlikely. Here, we have constructed merely two-dimensional dissimilarity maps for bit-string arrays obtained from measurements in the random and σ^z bases, and this was already enough to characterize several important families of many-body quantum states. In the cases, when two different wave functions were indistinguishable on the map (like the singlet and the chaotic states), they could be distinguished by their \mathcal{D}_k sets. If one is concerned about issue of non-uniqueness, the method can be used as a cheap preprocessing scheme within a larger framework of certification. First the dissimilarity signature is computed, and if it strongly deviates from the target state signature, the prepared state can be discarded right away. And only if the two states appear close enough, more advanced analysis should be performed.

An important advantage of the proposed approach is its scalability. Due to simplicity of computing the inter-scale dissimilarities, this procedure can be conducted for a large number of qubits. By using a classical computer, one could potentially characterize states of quantum systems of several thousands qubits which goes far beyond the abilities of available intermediate-scale quantum devices. For example, if one uses 128 Gb RAM, the estimated sizes of quantum systems that can be characterized in this way lie in the range from 8192 to 1048576 qubits, if the number of bitstrings in the array is taken to be 2^{20} or 2^{13} , correspondingly.

In this paper, we have analyzed two potential applications of the inter-scale dissimilarity signature—certification of quantum states and construction of phase diagrams. However, other research lines can be initiated, and we would like to briefly discuss them.

An important problem in quantum computing is to devise a quantum circuit that represents the desired target state. Usually, it is accomplished by optimization of the circuit architecture (topology, choice of gates) with overlap between the circuit and the target wave function being the objective function. For a large number of qubits, computing overlap at every iteration of optimization could be quite costly. Instead, one can aim at achieving the desired dissimilarity signature $\mathcal{D}_{\text{target}}$ and minimize the norm $\|\mathcal{D}_{\text{target}} - \mathcal{D}_{\text{circuit}}\|$ which, as discussed before, does not require significant resources to be computed even for a large system.

Another possible application of this concept could be in the domain of quantum optics experiments in which observer’s eyes play the role of photons detector^{59,60} with a minimal detection threshold of single photon⁶¹. Such a fascinating sensitivity of human eyes to the light has already become a basis for different scenarios of experiments⁶² aimed at detecting entanglement. Such experiments require accumulation of statistics over ‘seen’ and ‘not seen’ events. Since human eyes are much slower in counting light pulses than real photon detectors, collecting large amounts of data in such a setting is challenging, and a method that allows to harvest information from limited data could come handy. Representing two possible outcomes of a single measurement, ‘seen’ or ‘not seen’, as binary digits, one can construct an array that can be analyzed from the inter-scale dissimilarity point of view. As has been exemplified with Dicke and Schrödinger cat states, the latter can be used to estimate entanglement entropy of the state.

Finally, it should be highlighted that by constructing the low-dimensional dissimilarity map for a number of quantum states one, in fact, performs automatic dimensional reduction and visualization of a high-dimensional dataset—a common task in machine learning which is often solved in unsupervised manner by employing such methods as self-organized Kohonen map, t-distributed stochastic neighbor embedding (t-SNE)^{63,64}, principal component analysis or uniform manifold approximation and projection algorithms⁶⁵ (see ref. ⁶⁶ for a primer of how the latter can be used in the context of many-body quantum physics). These algorithms usually require some notion of distance between the original higher-dimensional data points and try to approximately preserve the relative distances when projecting points onto a lower-dimensional space (usually, two- or three-dimensional). By computing and visualizing dissimilarity signatures using two or three complementary measurement bases, one effectively solves the same problem for a dataset consisting of many-body quantum states. While it is possible to use the conventional dimensional reduction methods to classify and visualize quantum states by defining fidelity-based distance between them³⁷, this would require storing and manipulating many-body states on a classical computer. Thus, using dissimilarity maps could be an easy to implement alternative that does not require much resources. Although it is not directly related to the distance between quantum states in the Hilbert space, it nevertheless consistently and neatly clusters quantum states belonging to different families without even relying on any optimization scheme.

METHODS

Calculating inter-scale dissimilarity of bit-string arrays

To assign a characteristic hash function to a quantum state we perform three steps (Fig. 1): (i) initialization of the quantum state on a real quantum device or simulator, (ii) a number of projective measurements in at least two different bases, and (iii) computing the inter-scale dissimilarities of the resulting bit-string arrays.

The *initialization* of a quantum state may be done by different means. For instance, one can use variational approaches^{67–69} and adiabatic algorithms^{70–72} to approximate the target state on a quantum device. When dealing with a some small-scale quantum system, like the 16-qubit states studied in this paper, it is possible to initialize a state by taking the wave function coefficients obtained with exact diagonalization and employing the Least Significant Bit procedure²⁴ that features one-by-one disentanglement of qubits. Some particular quantum states can be directly generated with known quantum circuits, which is the case for the quantum chaos and the Schrödinger cat states. In this work, all the manipulations with quantum states were performed with the Qiskit package¹⁸.

Once a quantum state is initialized on a device, we measure it in two or more bases. Here, we refrained to projective measurements in the σ^z basis and the random basis, though using more bases can be beneficial for constructing unique hashes of many-body states. In other words, we sample N_{shots} basis vectors represented by bitstrings $\{x_i\}$ from the probability distribution $p(x_i) = |\psi(x_i)|^2$, where N_{shots} is a reasonably small number of measurements (16–8192 in the studied cases), and by doing this in two bases we should have access not only to the amplitudes, but also to the phases of the wave function. The measurement outputs in each basis are then arranged into one-dimensional sequence of bitstrings which can be regarded as a binary array of length $L = N \times N_{\text{shots}}$. Random basis measurements are performed in the following way. Prior to every shot i of measurement, rotational gate $U_0^{(i)}$ parametrized by randomly generated angles θ_i , ϕ_i and λ_i is applied to each qubit (Fig. 1a). For the next shot, values θ_{i+1} , ϕ_{i+1} and λ_{i+1} are sampled, and a rotational gate $U_0^{(i+1)}$ is applied. The angles are generated in such a way that, once the procedure is repeated many times, the single-shot gates uniformly cover a segment of the Bloch sphere: $\theta \in [0, \frac{\pi}{2}]$, $\phi \in [0, \frac{\pi}{2}]$ and $\lambda \in [0, \frac{\pi}{2}]$. The reason why we choose one of the bases to be random in the aforesaid sense is that it is expected to be the most unbiased one if we apply this protocol to diverse quantum states with completely different structures.

Having constructed the bit-string arrays, we analyze their structure using the concept of inter-scale dissimilarity. Recently¹⁷, some of us have suggested a notion of structural complexity of classical patterns based on

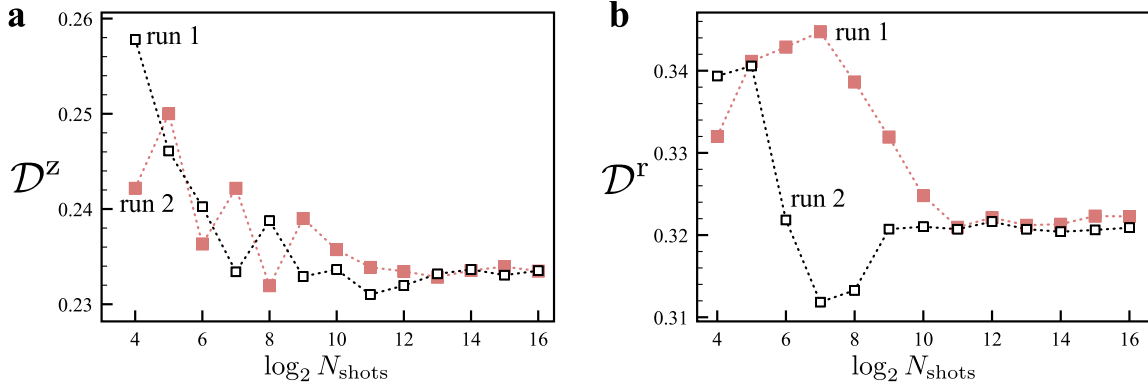


Fig. 11 Dissimilarity convergence as a function of the number of measurements. (a) and (b) represent dissimilarities obtained for σ^z and random bases. N_{shots} is number of measurements. The $D = 8$ 16-spin Dicke state is considered. In each basis, two independent runs are performed.

the idea of quantifying differences between distinct spatial scales of a pattern obtained with a multi-step renormalization (coarse-graining) protocol. Here, we formally apply this procedure to the bit-string arrays viewing them as one-dimensional patterns.

Let us denote such an array as vector \mathbf{b}^0 of length L . At every step of coarse-graining k , a vector of the same length is constructed as

$$b_i^k = \frac{1}{\Lambda^k} \sum_{l=1}^{\Lambda^k} b_{\Lambda^k \lfloor (i-1)/\Lambda^k \rfloor + l}^{k-1}, \quad (8)$$

where square brackets denote taking integer part. This means that at each iteration the whole array is divided into blocks of Λ^k size, and elements within a block are substituted with the same value resulting from averaging all elements of the block. Initially those elements are either 0 or 1, and for $k > 0$ they take real values (in fact, for the sake of nicer normalization in our calculations we assumed that '0' bits have values equal to -1). Index l enumerates elements belonging to the same block. For simplicity, we usually assume that the bit-string length is an integer power of filter size Λ : $\log_2 N \in \mathbb{N}$.

Dissimilarity between scales k and $k+1$ is then defined as

$$\mathcal{D}_k = |O_{k+1,k} - \frac{1}{2}(O_{k,k} + O_{k+1,k+1})|, \quad (9)$$

where $O_{m,n}$ is the overlap between vectors at scales m and n :

$$O_{m,n} = \frac{1}{L} (\mathbf{b}^m \cdot \mathbf{b}^n). \quad (10)$$

There are two quantities of our principal interest: \mathcal{D}_k that contains scale-resolved information on the pattern structure of the generated bit-string array and overall dissimilarity, $\mathcal{D} = \sum \mathcal{D}_k$, where the sum goes over all the renormalization steps. \mathcal{D} and $\{\mathcal{D}_k\}$ computed in several bases together comprise the hash function of quantum state that can be used for its certification.

Convergence of dissimilarity with the number of measurements

The number of measurements required for dissimilarity to converge depends on the state to be characterized. As mentioned in the main text, for the random Haar-typical states, it is enough to make about $2^{13} = 8192$ measurements even if the Hilbert space dimension is really large (53 qubits). If the state is more structured, a larger number of measurements (relative to the Hilbert space dimension) might be needed.

For example, for the $D = 8$ 16-qubit Dicke state, the dissimilarity converges after about 2^{10} measurements, Fig. 11. In general, we expect that dissimilarity tends to converge with the size of the sampled set as a typical observable computed by Monte Carlo sampling from a sign-definite wave function (though it is not an observable itself), meaning that about $\sim N \cdot 10^4$ samples drawn from the probability distribution should be enough.

Dissimilarity of the random quantum state: analytical derivation

Inter-scale dissimilarity of bit-string arrays resulting from projective measurements of random quantum states Eq. (4) can be estimated analytically. First, let us note that $O_{k,k} = O_{k,k-1}$ if the averaging-based

coarse-graining scheme (8) is adopted. Indeed, within n -th window of size Λ^k :

$$\begin{aligned} & \frac{1}{\Lambda^k} \sum_{i=1}^{\Lambda^k} b_{\Lambda^k(n-1)+i}^k \cdot b_{\Lambda^k(n-1)+i}^k \\ &= b_{\Lambda^k(n-1)+i}^k \cdot b_{\Lambda^k(n-1)+i}^k \\ &= b_{\Lambda^k(n-1)+i}^k \cdot \frac{1}{\Lambda^k} \sum_{i=1}^{\Lambda^k} b_{\Lambda^k(n-1)+i}^{k-1} \\ &= \frac{1}{\Lambda^k} \sum_{i=1}^{\Lambda^k} b_{\Lambda^k(n-1)+i}^k \cdot b_{\Lambda^k(n-1)+i}^{k-1}, \end{aligned} \quad (11)$$

where $b_{(n-1)\Lambda^k+i}^k$ are equal to each other for all i within the window, and thus this multiplier can be taken out of the sum over i . Once summed up over all windows, l.h.s. of this identity gives $O_{k,k}$, and the r.h.s.— $O_{k,k-1}$.

Thus, the expression for partial dissimilarity \mathcal{D}_k can be rewritten as

$$\mathcal{D}_k = \frac{1}{2} |O_{k+1,k+1} - O_{k,k}|. \quad (12)$$

For a random state, $O_{k,k}$ can be evaluated in the assumption that binary elements in the bit-string array \mathbf{b}_i^0 are sampled from some random distribution $p_0(x)$ (with $x = 0$ or 1) and not correlated. In this case, the coarse-graining procedure can be viewed as follows. In step $k = 1$, the renormalized probability distribution at every position in the array is defined over $x_1 = 0, 0.5, 1$ with $p_1(0) = p_0^2(0)$, $p_1(0.5) = 2p_0(0)p_0(1)$, $p_1(1) = p_0^2(1)$. Repeating this for several steps, one can notice that probability distribution $p_k(x_k)$ is defined over random variables which are obtained by averaging of the original uncorrelated random variables x , and according to the central limit theorem $p_k \rightarrow \mathcal{N}(\mu, \sigma^2 \Lambda^{-k})$ as $k \rightarrow \infty$. Here $\mathcal{N}(\mu, \sigma^2 \Lambda^{-k})(x)$ is a normal distribution with μ and σ^2 being the mean and variance of the original distribution $p_0(x)$ correspondingly, and normalization factor Λ^{-k} is due to the used scheme of averaging.

Noticing that, on average, product of a site value on itself is

$$\langle (b_i^k)^2 \rangle_i = \frac{1}{L} \sum_{i=1}^L (b_i^k)^2 \simeq \int x^2 p_k(x) dx, \quad (13)$$

where the integral symbolically denotes discrete finite sum at finite k , we can approximately rewrite $O_{k,k}$ as:

$$O_{k,k} = \frac{1}{L} \sum_{i=1}^L (b_i^k)^2 \simeq \int x^2 p_k(x) dx, \quad (14)$$

which leads us to

$$O_{k,k} \simeq \int x^2 \mathcal{N}(\mu, \sigma^2 \Lambda^{-k})(x) dx = \mu^2 + \frac{\sigma^2}{\Lambda^k}. \quad (15)$$

In this way, we obtain for $k > 0$:

$$\mathcal{D}_k = \frac{1}{2} [O_{k,k} - O_{k+1,k+1}] = \frac{\sigma^2}{2\Lambda^k} (1 - \Lambda^{-1}). \quad (16)$$

Although the central limit theorem formally holds for $k \rightarrow \infty$, it turns out that this estimate reproduces the numerically computed partial dissimilarities already starting with $k = 1$.

For $k=0$ it should be computed separately. Given $O_{0,0} \simeq \langle x^2 \rangle$, we obtain:

$$D_0 \simeq \frac{1}{2} \left(\langle x^2 \rangle - \mu^2 - \frac{\sigma^2}{\Lambda} \right) = \frac{\sigma^2}{2} (1 - \Lambda^{-1}). \quad (17)$$

On the continuity of dissimilarity upon deformations of the quantum state

Here, we would like to provide an idea of why the partial dissimilarities $\{\mathcal{D}_k\}$ are continuous functions of the parental quantum state. As shown in the previous section, in order to calculate the dissimilarity, the only object needed is the overlap $O_{k,k}$:

$$O_{k,k} = \frac{1}{L} \sum_{i=1}^L (b_i^k)^2, \quad (18)$$

where b_i^k is the average of the spin values in the i -th block.

In turn, overlap $O_{k,k}$ can be approximated as a certain expectation of the coarse-grained probability distribution $p_k(x)$:

$$O_{k,k} \simeq \int x^2 p_k(x) dx, \quad (19)$$

which turns exact in the limit of infinite number of measurements.

In order to demonstrate that $\{\mathcal{D}_k\}$ change smoothly upon deformations of the quantum state, we only need to show that the probability distribution at the k -th step of the coarse-graining, $p_k(x)$, depends continuously on the probability distribution $|\psi(s)|^2$ set by the original wave function (note that $p_k(x)$ is defined on the values of individual symbols, while $|\psi(s)|^2$ is defined on the basis vectors).

However, this is rather obvious. To compute dissimilarity, one needs to carry out several steps of averaging. The latter is a linear operation, and so is the composition of a number of such operations. To be more precise, possible values of b_i^k at the k -th step are $b_n = n\Lambda^{-k}$, where $n = 0 \dots \Lambda^k$ gives the number of 1s within the i -th window in the original array. Bitstrings s are divided into windows of length Λ^k , and the averaged values of entries within each window is computed. They are then replaced by string s_k consisting of b_n 's. The probability $p_k(b_n)$ is then given by a linear combination

$$p_k(b_n) = \sum_{s: b_n \in s_k} w_s |\psi(s)|^2, \quad (20)$$

where the sum goes over bitstrings s such that the corresponding strings s_k (derived from the parental s) contain the value b_n . Weights w_s of this linear combination are equal to the proportions of b_n among all the entries of s_k . Thus, at every step of coarse-graining up to $k = \log_\Lambda N$, the probability distribution $p_k(x)$ depends on the squared amplitudes of the wave function linearly, which implies that the partial (and thus overall) dissimilarities depend continuously on the quantum state.

To illustrate this, consider a simple example of a 4-spin state:

$$\psi = a|0011\rangle + b|0101\rangle + c|0010\rangle, \quad (21)$$

with $|a|^2 + |b|^2 + |c|^2 = 1$. Then $N = 4$, and the window size is $\Lambda^k = 2$ ($\Lambda = 2$ and $k = 1$). $p_0(x)$ has two probabilities:

$$p_0(0) = \frac{1}{2}|a|^2 + \frac{1}{2}|b|^2 + \frac{3}{4}|c|^2 \quad (22)$$

$$p_0(1) = \frac{1}{2}|a|^2 + \frac{1}{2}|b|^2 + \frac{1}{4}|c|^2. \quad (23)$$

After the coarse-graining step, string '0011' becomes '01', '0101' becomes ' $\frac{1}{2}$ ', and '0010' becomes ' $0\frac{1}{2}$ '. The corresponding probabilities of individual entries are then:

$$p_1(0) = \frac{1}{2}|a|^2 + \frac{1}{2}|c|^2, \quad (24)$$

$$p_1(1) = \frac{1}{2}|a|^2, \quad (25)$$

$$p_1\left(\frac{1}{2}\right) = |b|^2 + \frac{1}{2}|c|^2. \quad (26)$$

And after the second step of coarse-graining that brings the window size to the length of an individual measured bitstring, we obtain two

possibilities:

$$p_2\left(\frac{1}{2}\right) = |a|^2 + |b|^2, \quad (27)$$

$$p_2\left(\frac{1}{4}\right) = |c|^2. \quad (28)$$

If $\Lambda^k \geq N$, and N divides Λ^k , the linearity can also be shown. This case is even simpler since it reproduces the logic of the previous section. The averaged spin values within the window of size Λ^k are obtained by averaging $|\psi(s)|^2$, and the central limit theorem can be applied to show that $p_k(x)$ can be approximated by a normal distribution.

DATA AVAILABILITY

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

CODE AVAILABILITY

Codes for generating bit-string arrays for considered quantum states and for calculating the corresponding dissimilarities have been deposited in Zenodo, <https://doi.org/10.5281/zenodo.6257657>, <https://doi.org/10.5281/zenodo.5256316>.

Received: 15 September 2021; Accepted: 24 March 2022;

Published online: 22 April 2022

REFERENCES

- Batygin, K. & Brown, M. E. Evidence for a distant giant planet in the solar system. *Astron. J.* **151**, 22 (2016).
- Vanderburg, A. et al. A habitable-zone Earth-sized planet rescued from false positive status. *Astrophys. J. Lett.* **893**, L27 (2020).
- Heisenberg, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Phys.* **43**, 172–198 (1927).
- Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S. & Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105**, 150401 (2010).
- Tóth, G. et al. Permutationally invariant quantum tomography. *Phys. Rev. Lett.* **105**, 250403 (2010).
- Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
- Torlai, G. et al. Neural-network quantum state tomography. *Nat. Phys.* **14**, 447–450 (2018).
- Carrasquilla, J., Torlai, G., Melko, R. G. & Aolita, L. Reconstructing quantum states with generative models. *Nat. Mach. Intell.* **1**, 155–161 (2019).
- Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
- Deng, D.-L., Li, X. & Sarma, S. D. Quantum entanglement in neural network states. *Phys. Rev. X* **7**, 021021 (2017).
- Westerhout, T., Astrakhantsev, N., Tikhonov, K. S., Katsnelson, M. I. & Bagrov, A. A. Generalization properties of neural network approximations to frustrated magnet ground states. *Nat. Commun.* **11**, 1593 (2020).
- Dumey, A. I. Indexing for rapid random access memory. *Comput. Autom.* **5**, 6–9 (1956).
- Peterson, W. W. Addressing for random-access storage. *IBM J. Res. Dev.* **1**, 130–146 (1957).
- Cotler, J. & Wilczek, F. Quantum overlapping tomography. *Phys. Rev. Lett.* **124**, 100401 (2020).
- Aaronson, S. Shadow tomography of quantum states. *SIAM J. Comput.* **49**, STOC18-368-STOC18-394 (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).
- Bagrov, A. A., Iakovlev, I. A., Iliasov, A. A., Katsnelson, M. I. & Mazurenko, V. V. Multi-scale structural complexity of natural patterns. *Proc. Natl Acad. Sci. USA* **117**, 30241–30251 (2020).
- IBM Quantum. <https://quantum-computing.ibm.com/> (2021).
- Smith, A. W. R., Gray, J. & Kim, M. S. Efficient quantum state sample tomography with basis-dependent neural networks. *PRX Quantum* **2**, 020348 (2021).
- Kuzmak, A. R. & Tkachuk, V. M. Detecting entanglement by the mean value of spin on a quantum computer. *Phys. Lett. A* **384**, 126579 (2020).
- Dicke, R. H. Coherence in spontaneous radiation processes. *Phys. Rev.* **93**, 99–110 (1954).

22. Prevedel, R. et al. Experimental realization of Dicke states of up to six qubits for multiparty quantum networking. *Phys. Rev. Lett.* **103**, 020503 (2009).
23. Liu, Y.-C., Yu, X.-D., Shang, J., Zhu, H. & Zhang, X. Efficient verification of Dicke states. *Phys. Rev. Appl.* **12**, 044020 (2019).
24. Shende, V. V., Bullock, S. S. & Markov, I. L. Synthesis of quantum logic circuits. *IEEE Trans Comput. Aid. Design* **25**, 1000–1010 (2006).
25. Khoo, J. Y. & Heyl, M. Quantum entanglement recognition. *Phys. Rev. Res.* **3**, 033135 (2021).
26. Scott, A. J., Brun, T. A., Caves, C. M. & Schack, R. Hypersensitivity and chaos signatures in the quantum baker's maps. *J. Phys. A* **39**, 13405–13433 (2006).
27. Page, D. N. Average entropy of a subsystem. *Phys. Rev. Lett.* **71**, 1291–1294 (1993).
28. Harrow, A., Hayden, P. & Leung, D. Superdense coding of quantum states. *Phys. Rev. Lett.* **92**, 187901 (2004).
29. Hayden, P., Leung, D., Shor, P. W. & Winter, A. Randomizing quantum states: constructions and applications. *Commun. Math. Phys.* **250**, 371–391 (2004).
30. Richter, J. & Pal, A. Simulating hydrodynamics on noisy intermediate-scale quantum devices with random circuits. *Phys. Rev. Lett.* **126**, 230501 (2021).
31. Boixo, S. et al. Characterizing quantum supremacy in near-term devices. *Nat. Phys.* **14**, 595–600 (2018).
32. Nomura, Y. & Imada, M. Dirac-type nodal spin liquid revealed by refined quantum many-body solver using neural-network wave function, correlation ratio, and level spectroscopy. *Phys. Rev. X* **11**, 031034 (2020).
33. Sotnikov, O. M. et al. Probing the topology of the quantum analog of a classical skyrmion. *Phys. Rev. B* **103**, L060404 (2021).
34. van Nieuwenburg, E. P. L., Liu, Y.-H. & Huber, S. D. Learning phase transitions by confusion. *Nat. Phys.* **13**, 435–439 (2017).
35. Zhang, Y. et al. Machine learning in electronic-quantum-matter imaging experiments. *Nature* **570**, 484–490 (2019).
36. Uvarov, A., Kardashin, A. & Biamonte, J. Machine learning phase transitions with a quantum processor. *Phys. Rev. A* **102**, 102415 (2020).
37. Yang, Y., Sun, Z.-Z., Ran, S.-J. & Su, G. Visualizing quantum phases and identifying quantum phase transitions by nonlinear dimensional reduction. *Phys. Rev. B* **103**, 075106 (2021).
38. Bagrov, A. A. et al. Detecting quantum critical points in the t-t' Fermi-Hubbard model via complex network theory. *Sci. Rep.* **10**, 20470 (2020).
39. Lewis-Swan, R. J., Muleady, S. R. & Rey, A. M. Detecting out-of-time-order correlations via quasi-adiabatic echoes as a tool to reveal quantum coherence in equilibrium quantum phase transitions. *Phys. Rev. Lett.* **125**, 240605 (2020).
40. Jördens, R., Strohmaier, N., Günter, K., Moritz, H. & Esslinger, T. Mott insulator of fermionic atoms in an optical lattice. *Nature* **455**, 204–207 (2008).
41. Schneider, U. et al. Metallic and insulating phases of repulsively interacting fermions in a 3D optical lattice. *Science* **322**, 1520–1525 (2008).
42. Mazurenko, A. et al. Experimental realization of a long-range antiferromagnet in the Hubbard model with ultracold atoms. *Nature* **545**, 462–466 (2017).
43. Bohrdt, A. et al. Classifying snapshots of the doped Hubbard model with machine learning. *Nat. Phys.* **15**, 921–924 (2019).
44. Khatami, E. et al. Visualizing strange metallic correlations in the two-dimensional Fermi-Hubbard model with artificial intelligence. *Phys. Rev. A* **102**, 033326 (2020).
45. Westerhout, T. lattice-symmetries: a package for working with quantum many-body bases. *J. Open Source Softw.* **6**, 3537 (2021).
46. Shastry, B. S. & Sutherland, B. Exact ground state of a quantum mechanical antiferromagnet. *Physica B* **108**, 1069–1070 (1981).
47. Corboz, P. & Mila, F. Tensor network study of the Shastry-Sutherland model in zero magnetic field. *Phys. Rev. B* **87**, 115144 (2013).
48. Mazurenko, V. V., Skornyakov, S. L., Anisimov, V. I. & Mila, F. First-principles investigation of symmetric and antisymmetric exchange interactions of SrCu₂(BO₃)₂. *Phys. Rev. B* **78**, 195110 (2008).
49. Badrtdinov, D. I., Tsirlin, A. A., Mazurenko, V. V. & Mila, F. SrCu₂(BO₃)₂ under pressure: a first-principles study. *Phys. Rev. B* **101**, 224424 (2020).
50. Nakano, H. & Sakai, T. Third boundary of the Shastry-Sutherland model by numerical diagonalization. *J. Phys. Soc. Jpn.* **87**, 123702 (2018).
51. Miyahara, S. & Ueda, K. Superstructures at magnetization plateaus in SrCu₂(BO₃)₂. *Phys. Rev. B* **61**, 3417–3424 (2000).
52. Lou, J., Suzuki, T., Harada, K. & Kawashima, N. Study of the Shastry Sutherland Model Using Multi-scale Entanglement Renormalization Ansatz. Preprint at <https://arxiv.org/abs/1212.1999> (2012).
53. Elben, A. et al. Many-body topological invariants from randomized measurements in synthetic quantum matter. *Sci. Adv.* **6**, eaaz3666 (2020).
54. Huang, H.-Y., Kueng, R., Torlai, G., Albert, V.V. & Preskill, J. Provably efficient machine learning for quantum many-body problems. Preprint at <https://arxiv.org/abs/2106.12627> (2021).
55. Bohr, N. The quantum postulate and the recent development of atomic theory. *Nature* **121**, 580–590 (1928).
56. Wheeler, J. A. & Zurek, W. H. (eds), *Quantum Theory and Measurement* (Princeton University Press, Princeton, 1983).
57. von Neumann, J. *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955).
58. De Raedt, H., Katsnelson, M. I., Willsch, D. & Michielsen, K. Separation of conditions as a prerequisite for quantum theory. *Ann. Phys. (NY)* **403**, 112–135 (2019).
59. Hecht, S., Schlaer, S. & Pirenne, M. Energy, quanta, and vision. *J. Gen. Physiol.* **25**, 819–840 (1942).
60. Rieke, F. & Baylor, D. A. Single-photon detection by rod cells of the retina. *Rev. Mod. Phys.* **70**, 1027–1036 (1998).
61. Tinsley, J. N. et al. Direct detection of a single photon by humans. *Nat. Commun.* **7**, 12172 (2016).
62. Vivoli, V. C., Sekatski, P. & Sangouard, N. What does it take to detect entanglement with the human eye. *Optica* **3**, 473–476 (2016).
63. Hinton, G., Roweis, S. Stochastic neighbor embedding. Paper presented at the 15th International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 833–840 (NIPS 2002).
64. van der Maaten, L. J. P. & Hinton, G. E. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
65. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).
66. Bagrov, A. A., Iliasov, A. A. & Westerhout, T. Kinetic samplers for neural quantum states. *Phys. Rev. B* **104**, 104407 (2021).
67. Wecker, D., Hastings, M. B. & Troyer, M. Progress towards practical quantum variational algorithms. *Phys. Rev. A* **92**, 042303 (2015).
68. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
69. Sotnikov, O. M. & Mazurenko, V. V. Neural network agent playing spin Hamiltonian games on a quantum computer. *J. Phys. A - Math. Theor.* **53**, 135303 (2020).
70. Born, M. & Fock, V. Beweis des Adiabatsatzes. *Z. Phys.* **51**, 165–180 (1928).
71. Farhi, E., Goldstone, J., Gutmann, S. & Sipser, M. A numerical study of the performance of a quantum adiabatic evolution algorithm for satisfiability (Tech. Rep. MIT-CTP N 3006, Massachusetts Institute of Technology, 2000).
72. Farhi, E. et al. A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. *Science* **292**, 472–475 (2001).

ACKNOWLEDGEMENTS

We thank Ehsan Khatami, Juan Carrasquilla, Alexander Tsirlin and Tom Westerhout for useful discussions. The work of V.V.M., O.M.S. and I.A.I. was supported by Act 211 Government of the Russian Federation, contract 02.A03.21.0006. A.A.B. acknowledges financial support from Knut and Alice Wallenberg Foundation through Grant No.2018.0060. M.I.K. acknowledges a support from European Research Council via Synergy grant 854843 (FASTCORR). Exact diagonalization and quantum simulator calculations were performed on the Uran supercomputer at the IMM UB RAS. We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

AUTHOR CONTRIBUTIONS

I.A.I., O.M.S. and V.V.M. designed the research. A.A.B., M.I.K. and V.V.M. provided general guidance. O.M.S. and V.V.M. performed quantum simulations and generated the bitstrings. I.A.I. calculated the dissimilarities of the obtained bitstrings. A.A.I. derived the analytical expression for the dissimilarity of the random states. A.A.B. developed the concept of the quantum hash function based on the bit-string dissimilarity. V.V.M., I.A.I., A.A.B. prepared the figures. V.V.M., I.A.I., A.A.I., M.I.K., A.A.B. wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to V. V. Mazurenko.

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

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022