

# Reconstructing quantum states with generative models

Juan Carrasquilla,<sup>1</sup> Giacomo Torlai,<sup>2,3</sup> Roger G. Melko,<sup>2,3</sup> and Leandro Aolita<sup>4,5</sup>

<sup>1</sup>*Vector Institute for Artificial Intelligence,*

*MaRS Centre, Toronto, ON, Canada M5G 1M1*

<sup>2</sup>*Department of Physics and Astronomy,*

*University of Waterloo, Ontario, N2L 3G1, Canada*

<sup>3</sup>*Perimeter Institute for Theoretical Physics,*

*Waterloo, Ontario N2L 2Y5, Canada*

<sup>4</sup>*Instituto de Física, Universidade Federal do Rio de Janeiro,*

*Caixa Postal 68528, Rio de Janeiro, RJ 21941-972, Brazil*

<sup>5</sup>*ICTP South American Institute for Fundamental Research, Instituto de Física Teórica,*

*UNESP-Universidade Estadual Paulista R. Dr. Bento T. Ferraz 271,*

*Bl. II, São Paulo 01140-070, SP, Brazil*

A major bottleneck in the quest for scalable many-body quantum technologies is the difficulty in benchmarking their preparations, which suffer from an exponential “curse of dimensionality” inherent to their quantum states. We present an experimentally friendly method for density matrix reconstruction based on deep neural-network generative models. The learning procedure comes with a built-in approximate certificate of the reconstruction and makes no assumptions on the state under scrutiny, making it both reliable and unconditional. It can efficiently handle a broad class of complex systems including prototypical states in quantum information, as well as ground states of local spin models common to condensed matter physics. The key insight is to reduce the state tomography task to an unsupervised learning problem of the statistics of an informationally complete set of quantum measurements. This constitutes a modern machine learning approach to the validation of large quantum devices, which may prove relevant as a neural-network ansatz over mixed states suitable for variational optimization.

Since the turn of the century, advances in several competing quantum technologies have demonstrated control and measurements sufficiently accurate to enable devices of up to tens, or soon hundreds, of qubits [2–7]. A comprehensive characterization of these devices entails the reconstruction of their quantum state from measurements on identically prepared copies, a task known as quantum state tomography [8–10]. In their familiar conception, exact tomographic techniques become impractical [11] on large quantum systems due to the exponential complexity inherent to the description of generic quantum many-body systems.

Physical systems of interest – such as those generated by the dynamics of a local Hamiltonian – are not generic. Their particular structure guarantees that the full complexity of Hilbert space is in principle not required for their accurate description [12, 13]. In a wide range of physical situations, a priori structural information about the state under scrutiny can help alleviate exponential scaling. In such cases, a reconstruction can be achieved by introducing a plausible parametrization of the state, whose computational manipulation and storage scale favourably with system size. In fact, success requires that all computational resources scale well with system size. This includes the number of measurements needed for the reconstruction, and the time required for learning the parameters that best fit the measurement statistics. Some examples following this spirit are permutationally invariant tomography [14, 15], quantum compressed sensing [16], and tomographic schemes based on tensor networks [17–20]. Alternatively, if the ideal target state is known, one can try to directly certify the fidelity between the experimental and ideal states, without attempting a reconstruction of the former [17, 21–24].

In a quest for efficiency and versatility, methodologies inspired by undirected graphical models such as the restricted Boltzmann machine (RBM) have recently appeared [25, 26]. All of these methods are effective for different classes of states, but they all share the drawback of limited versatility. Perhaps most notably, even though matrix-product-state (MPS) tomography [17, 19, 20] has led to impressive progress in the theoretical and experimental reconstruction of states of large spin chains, generalizations to higher-dimensional lattices rapidly become computationally intractable. Furthermore, even for one-dimensional lattices,

the limitation of low entanglement in practice restricts MPS tomography to only short-time dynamics [20]. Among machine learning strategies, a scalable unsupervised learning formulation for density matrix reconstruction remains elusive. In current treatments that purify RBMs to generalize neural quantum state tomography to mixed states [26], training and manipulation introduces exponential scaling in any dimension.

In this paper we combine two state-of-the-art classes of algorithms to alleviate these scaling issues. The first is the tensor-network paradigm, designed using well-understood underlying principles of quantum entanglement to efficiently represent quantum states. The second is powerful generative models, a key ingredient in modern deep learning research. Generative models can be tractably defined and trained in any spatial dimension, display an extraordinary expressive power [27], and have the potential to represent highly entangled states [28–31]. They are used to understand probabilistic distributions defined over high-dimensional data with rich structure, typically in tasks such as density estimation, denoising, missing value imputation, and sampling [32]. We show how to reduce unconditional state tomography to an explicit, unsupervised learning problem using these powerful probabilistic models. The reduction consists of directly parametrizing the outcome probabilities of tomographically complete measurements on an arbitrary state. The resulting method is experimentally friendly since it only requires routinely available single-particle measurements. We show that this strategy can efficiently learn a variety of complex states, from paradigmatic multi-qubit states undergoing local noise, to ground states of local spin models in both one and two spatial dimensions. Provided that the measurements are tomographically complete, our approach is also reliable since the performance of the reconstruction can be approximately certified efficiently by sampling from the reconstructed distribution.

We consider experimental measurements given by informationally complete (IC) positive-operator valued measures (POVMs) (see Methods). POVMs describe the most general type of measurements allowed by quantum theory, beyond the usual notion of von Neumann projective measurements [33]. Importantly, informational completeness means that their measurement statistics contains all of the information about the state, thus specifying it

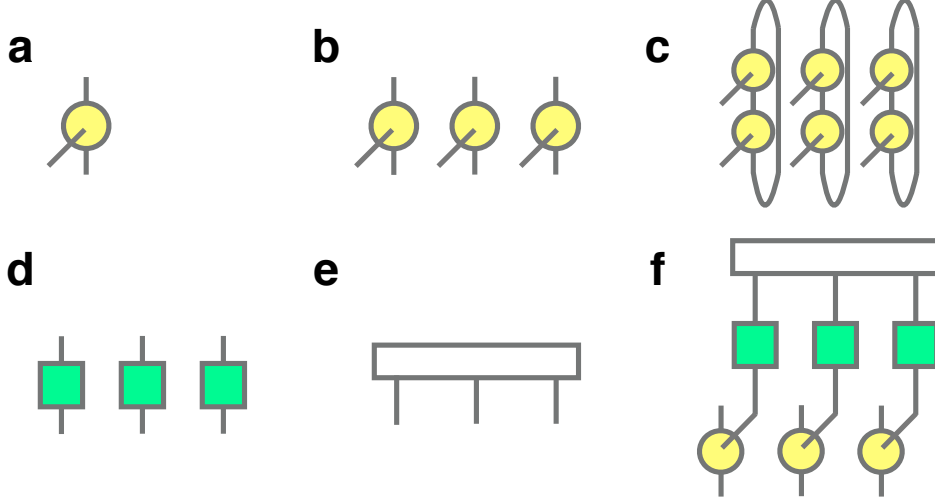


FIG. 1. Tensor-network schematics of our formalism. (a) The single-qubit measurement  $\mathbf{M} = \{M^{(a)}\}_a$  is a three-index tensor represented by a yellow circle with three lines emerging from it. Vertical indices act on the physical degrees of freedom while the outgoing one labels the measurement outcome  $a$ . (b) The  $N$ -qubit measurement  $\mathbf{M} = \{M^{(a_1)} \otimes M^{(a_2)} \otimes \dots M^{(a_N)}\}_{a_1, \dots, a_N}$  corresponds to the same local measurement on each qubit. (c) The components  $T_{\mathbf{a}, \mathbf{a}'} = \text{Tr} [M^{(\mathbf{a})} M^{(\mathbf{a}')}]$  of the overlap matrix  $\mathbf{T}$  correspond to the physical-index contraction of the measurement operators. (d) Consequently, when  $\mathbf{T}$  is invertible, its inverse  $\mathbf{T}^{-1}$  factorizes into the tensor product of single-qubit matrices  $\mathbf{T}_{1\text{-qubit}}^{-1}$  (green squares). (e) We represent the outcome distribution  $\mathbf{P}$  as an  $N$ -index tensor where the indices encode the measurement outcome of each qubit. (f) A generic density matrix  $\varrho$  is given by a contraction of  $\mathbf{P}$ ,  $\mathbf{T}^{-1}$ , and  $\mathbf{M}$  over the outcome indices:  $\varrho = \sum_{\mathbf{a}, \mathbf{a}'} P(\mathbf{a}) T_{\mathbf{a}, \mathbf{a}'}^{-1} M^{(\mathbf{a}')} = \mathbb{E}_{\mathbf{a} \sim \mathbf{P}} \left( \sum_{\mathbf{a}'} T_{\mathbf{a}, \mathbf{a}'}^{-1} M^{(\mathbf{a}')} \right)$ . Expressed this way, all non-local correlations of  $\varrho$  are encoded explicitly into  $\mathbf{P}$ , as both  $\mathbf{T}$  and  $\mathbf{M}$  are single-qubit factorable. Our state ansatz parameterizes the outcome distribution with a neural-network generative model, i.e.  $\mathbf{P} = \mathbf{P}_{\text{model}}$ .

univocally. For concreteness, we consider physical systems composed of  $N$  qubits. We build our measurements starting from an  $m$ -outcome single-qubit POVM  $\mathbf{M} = \{M^{(a)}\}_a$  [see Fig.1 (a)], defined by positive semi-definite operators  $M^{(a)} \geq 0$ , each one labeled by a measurement outcome  $a = 0, 1, \dots, m-1$ , that satisfy the normalization condition  $\sum_a M^{(a)} = \mathbb{1}$ . Our  $N$ -qubit measurement is given by the tensor product of single-qubit POVM elements

$\mathbf{M} = \{M^{(a_1)} \otimes M^{(a_2)} \otimes \dots M^{(a_N)}\}_{a_1, \dots, a_N}$ , which are graphically depicted in Fig 1 (b). That is, we measure the same fixed IC POVM on each qubit.

A theoretical bedrock of quantum physics provides a link between quantum theory and experiment in terms of a linear relation between the statistics of the measurement outcomes  $\mathbf{a} = (a_1, a_2, \dots, a_N)$ ,  $\mathbf{P} = \{P(\mathbf{a})\}_{\mathbf{a}}$  with  $P(\mathbf{a}) \geq 0$  and  $\sum_{\mathbf{a}} P(\mathbf{a}) = 1$ , and the quantum state  $\varrho$ . Known as Born's rule, this relationship can be written as  $P(\mathbf{a}) = \text{Tr}[M^{(\mathbf{a})} \varrho]$ . Provided that the measurement is informationally complete, this relation can formally be inverted. In other words, the density matrix can be unambiguously inferred from the probability distribution of measurement outcomes. This can be explicitly expressed in a particularly concise and simple way when the overlap matrix  $\mathbf{T}$ , of elements  $T_{\mathbf{a}, \mathbf{a}'} = \text{Tr}[M^{(\mathbf{a})} M^{(\mathbf{a}')}]$ , is invertible:  $\varrho = \sum_{\mathbf{a}, \mathbf{a}'} P(\mathbf{a}) T_{\mathbf{a}, \mathbf{a}'}^{-1} M^{(\mathbf{a}')} = \mathbb{E}_{\mathbf{a} \sim \mathbf{P}} (\sum_{\mathbf{a}'} T_{\mathbf{a}, \mathbf{a}'}^{-1} M^{(\mathbf{a}')} )$ , with  $\mathbb{E}_{\mathbf{a} \sim \mathbf{P}}$  representing the (classical) expected value over  $\mathbf{a}$  distributed according to  $\mathbf{P}$  [see Figs. 1 (c-f) for a graphical representation of these relations].

Given a collection of experimental measurement outcomes  $E = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{N_s}\}$ , with  $N_s$  samples, our strategy to infer the state of the system begins with learning a model  $P_{\text{model}}(\mathbf{a})$  that describes the measurement statistics in terms of expressive neural generative models. This task can be carried out using a wide variety of models and training strategies, including variational autoencoders, generative adversarial networks, restricted Boltzmann machines, and powerful autoregressive models based upon recurrent neural networks (RNN), among others. To demonstrate our approach, we first consider the prototypical  $N$ -qubit Greenberger-Horne-Zeilinger (GHZ) state, which is a highly non-classical state specified by  $|\Psi\rangle = \frac{1}{\sqrt{2}} (|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$ . We examine mixed states arising from GHZ states subject to local depolarizing noise on each qubit independently. Each qubit is depolarized (i.e., all its information lost) with probability  $p$ , while it is left untouched with probability  $1-p$  [33]. We measure the so-called tetrahedral POVM  $\mathbf{M}_{\text{tetra}}$  on each qubit. This is an IC POVM with  $m = 4$  outcomes, which is the minimum required on qubit measurements for informational completeness, and an invertible overlap matrix. Each measurement operator in  $\mathbf{M}_{\text{tetra}}$  is proportional to a rank-1 projector pointing to a different vertex of a regular tetrahedron in

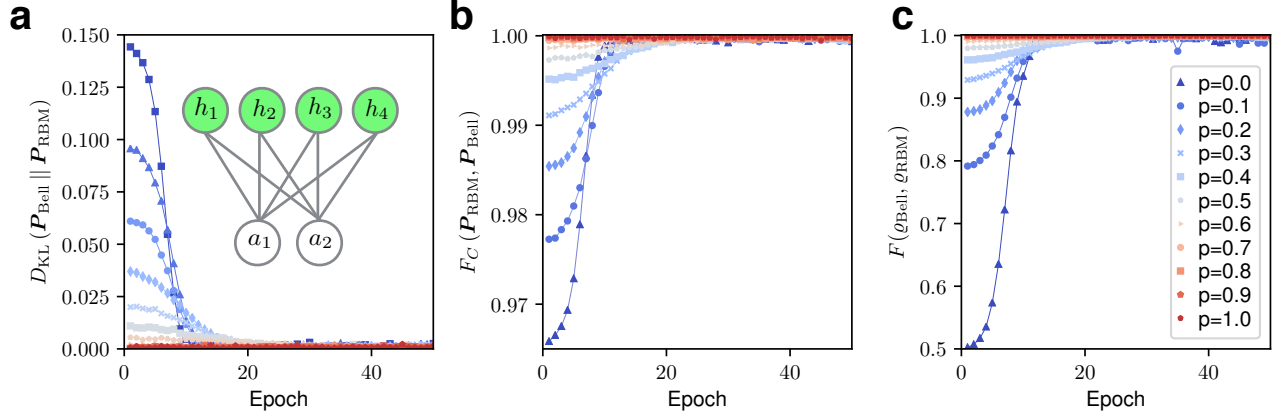


FIG. 2. Learning a Bell state under local depolarizing noise. (a) The Kullback-Liebler divergence as a function of the training epochs for different depolarization probabilities  $p$ . One epoch refers to a training cycle that exposes the learning algorithm to the entire training dataset. The sample size of the dataset used during the training is  $N_s = 6 \times 10^4$  observations. The inset displays a graphical representation of the RBM employed in the reconstruction of the state. This architecture features a visible layer with multinomial variables  $a_i = 0, 1, 2, 3$  (white circles) and binary hidden units  $h_i = 0, 1$  (green) connected via a set of weight parameters represented by the grey connections. (b) The classical fidelity  $F_C$  between the reconstructed distribution  $\mathbf{P}_{\text{RBM}}$  and  $\mathbf{P}_{\text{GHZ}}$ . (c) The quantum fidelity between the depolarized Bell state and its RBM reconstruction.

the Bloch sphere, which explains its name (see Methods).

To begin the implementation of our method, we require a parameterization of  $P_{\text{model}}(\mathbf{a})$ . Because of their extensive familiarity as a tool to represent quantum states [25, 26, 29–31, 34–37], we parametrize  $P_{\text{model}}(\mathbf{a})$  in terms of an RBM as a first demonstration. The main extension over more familiar RBMs is the need for four-dimensional multinomial visible units, which were previously introduced in the context of collaborative filtering [38], for applications such as the famous Netflix Prize [39]. Using such RBMs, we learn the statistics of numerically simulated measurements on the GHZ state with  $N = 2$  qubits, i.e. the Bell state  $|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle)$ , under local depolarization with different values of noise strengths  $0 \leq p \leq 1$ . Our results are shown in Fig. 2.

The RBM model is trained using a standard contrastive divergence learning procedure [32], which aims at maximizing the log-likelihood of the the data with respect to

the parameters in the model. In Fig. 2 (a) we show the Kullback-Leibler (KL) divergence  $D_{\text{KL}}(\mathbf{P}_{\text{Bell}} \parallel \mathbf{P}_{\text{RBM}}) = \mathbb{E}_{\mathbf{a} \sim \mathbf{P}_{\text{Bell}}} \left[ \log \frac{P_{\text{RBM}}(\mathbf{a})}{P_{\text{Bell}}(\mathbf{a})} \right]$ , which measures how much the RBM model distribution  $\mathbf{P}_{\text{RBM}}$  diverges from the exact distribution  $\mathbf{P}_{\text{Bell}}$ . As the training progresses, the KL divergence successfully decreases to values near zero for all values of noise  $p$ . Likewise, the classical fidelity  $F_C(\mathbf{P}_{\text{RBM}}, \mathbf{P}_{\text{Bell}}) = \mathbb{E}_{\mathbf{a} \sim \mathbf{P}_{\text{Bell}}} \left[ \sqrt{\frac{P_{\text{RBM}}(\mathbf{a})}{P_{\text{Bell}}(\mathbf{a})}} \right] = \mathbb{E}_{\mathbf{a} \sim \mathbf{P}_{\text{RBM}}} \left[ \sqrt{\frac{P_{\text{GHZ}}(\mathbf{a})}{P_{\text{RBM}}(\mathbf{a})}} \right]$  is a standard measure of proximity between two distributions. Indeed, it is such that  $0 \leq F_C(\mathbf{P}_{\text{model}}, \mathbf{P}_{\text{Bell}}) \leq 1$  for all  $\mathbf{P}_{\text{RBM}}$  and  $\mathbf{P}_{\text{Bell}}$ , with  $F(\mathbf{P}_{\text{RBM}}, \mathbf{P}_{\text{Bell}}) = 1$  if and only if  $\mathbf{P}_{\text{RBM}} = \mathbf{P}_{\text{Bell}}$ . As depicted in Fig. 2(b), this quantity approaches unity as the model distribution  $\mathbf{P}_{\text{model}}$  converges toward the distribution  $\mathbf{P}_{\text{Bell}}$ . Finally, in Fig. 2(c) we consider the quantum fidelity  $F(\varrho_{\text{Bell}}, \varrho_{\text{RBM}}) = \text{Tr} \left[ \sqrt{\sqrt{\varrho_{\text{Bell}}} \varrho_{\text{RBM}} \sqrt{\varrho_{\text{Bell}}}} \right]$ , the standard measure of proximity between states that generalizes  $F_C$  from probability distributions to density matrices. In general, the bound  $F_C(\mathbf{P}_{\text{RBM}}, \mathbf{P}_{\text{Bell}}) \geq F(\varrho_{\text{Bell}}, \varrho_{\text{RBM}})$  holds. However, in our numerical experiments we observe that both the classical and quantum fidelities collectively approach unity with remarkably similar behavior as a function of the training epoch. This suggests that the classical fidelity serves as good figure of merit for the quality of the reconstructed quantum state. We argue that informational completeness together with the physical nature of the state should – for most practical situations – imply that if  $F_C$  is close to one, then so is  $F$ . From an experimental perspective it is particularly natural to consider the fidelity between  $\mathbf{P}$  and  $\mathbf{P}_{\text{model}}$  as an indication for reconstruction quality, given that the experimental observables correspond to samples drawn from  $\mathbf{P}$ . Whereas  $F$  is in practice inaccessible for anything but modest  $N$ ,  $F_C$  can be efficiently estimated accurately for large  $N$  via Monte Carlo, as an average over samples from  $\mathbf{P}_{\text{model}}$ , if  $\mathbf{P}$  is known.

We now focus on the scaling of the resources required for the learning procedure as a function of the number of qubits  $N$ . First we observe numerically that, for qubit numbers  $N > 4$ , the training becomes increasingly difficult for the RBM in the regime of small noise  $p$ . We therefore opt for a different probabilistic model based upon autoregressive recurrent neural networks (see Appendix). These models are ordinarily used in end-to-end sequence learning tasks [40], and are the state-of-the-art engines behind machine translation and

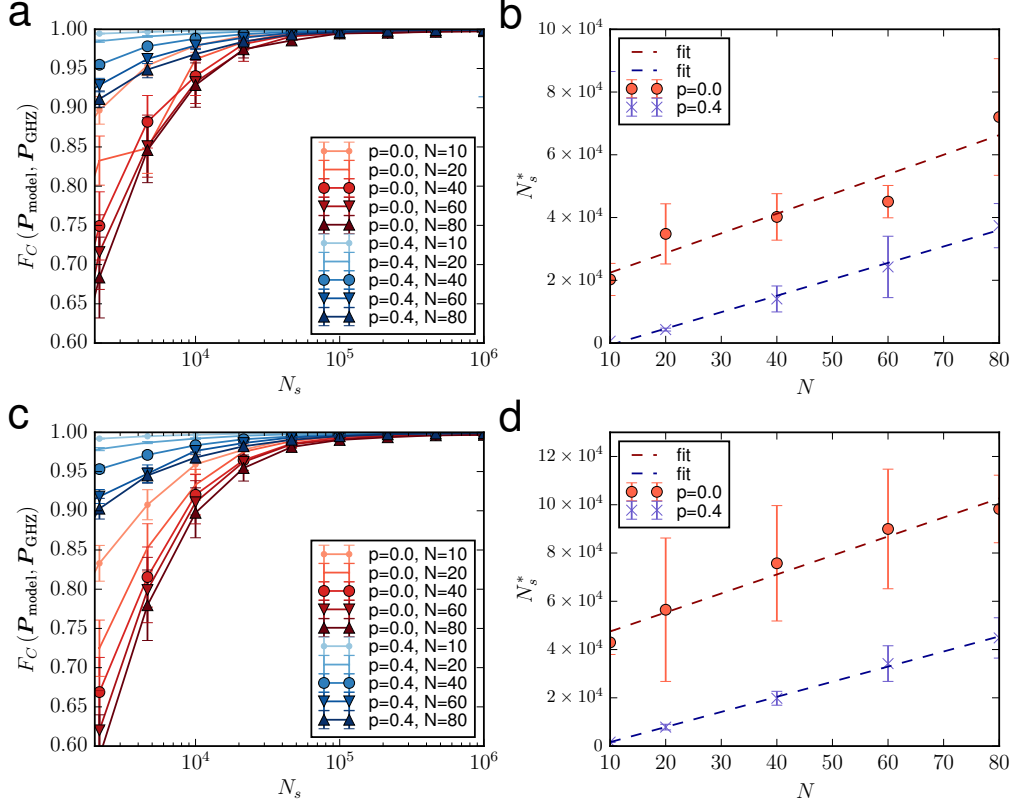


FIG. 3. Sample complexity of learning locally depolarized GHZ states with recurrent neural network (RNN) models. (a) Estimated average classical fidelity as a function of the number of measurements performed on the GHZ state for two different values of noise  $p = 0$  (red) and  $p = 0.4$  (blue) using the 4-outcome tetrahedral POVM. The reported classical fidelity is an average over a few ( $N_m = 30$ ) models taken around the optimal model during training and the error bars represent the one s.d. statistical uncertainty calculated over the different models. The classical fidelity for each model during training is estimated based upon  $10^5$  samples drawn from the RNN model. (b) The number of samples  $N_s^*$  necessary to attain an average classical fidelity  $F_C(\mathbf{P}_{\text{RNN}}, \mathbf{P}_{\text{GHZ}}) = 0.99$ . (c) and (d) are analogous to (a) and (b) for the 6-outcome Pauli POVM. Dashed lines represent linear fits, all of which exhibit a correlation coefficient  $r > 0.94$ .

speech recognition systems [41, 42]. We discovered that, in our setting, RNN models are faster to train compared to RBMs.

To study noisy GHZ states for large  $N$ , we produce synthetic datasets mimicking experimental measurements of the exact distribution  $\mathbf{P}_{\text{GHZ}}$  (see Appendix for details). For the



tetrahedral POVM, we investigate the classical fidelity as a function of  $N_s$  used for training a RNN model  $\mathbf{P}_{\text{RNN}}$  for different values of  $N$  and  $p$ . We plot the results in Fig. 3(a). As in the RBM case, we find that  $F_C(\mathbf{P}_{\text{RNN}}, \mathbf{P}_{\text{GHZ}})$  quickly approaches unity for all of the states that we consider. We also find that learning the noiseless ( $p = 0$ ) states takes significantly more effort in terms of training set size. This reflects the larger amount of information contained in a pure state relative to its depolarized counterparts [33, 43]. Remarkably, as shown in Fig. 3(b), the number  $N_s^*$  of samples required to learn the state up to  $F_C(\mathbf{P}_{\text{RNN}}, \mathbf{P}_{\text{GHZ}}) = 0.99$  is found to scale approximately linearly with  $N$ . To investigate if this scaling is a just a peculiarity of the particular POVM chosen, we implement the same learning protocol with a different measurement. We use a single-qubit IC POVM with  $m = 6$  outcomes, each one described by a POVM element proportional to the rank-1 projector onto one of the 6 eigenstates of the three usual Pauli matrices (see Methods). We refer to this as the Pauli-6 POVM and denote it by  $\mathbf{M}_{\text{Pauli-6}}$ . Even though the  $N$ -qubit sample space is now exponentially larger than the tetrahedral POVM, we find out that  $N_s^*$  still scales linearly with  $N$ , with only a slightly larger slope than for  $\mathbf{M}_{\text{tetra}}$ . While we again define  $N_s^*$  based on an average value of  $F_C(\mathbf{P}_{\text{RNN}}, \mathbf{P}_{\text{GHZ}}) = 0.99$ , we remark that in all cases, the maximum classical fidelity we find is at least  $F_C(\mathbf{P}_{\text{RNN}}, \mathbf{P}_{\text{GHZ}}) \geq 0.999$  for a number of samples  $N_s = 10^6$  for all system sizes and values of noise  $p$ . Finally, we note that this sample complexity is consistent with arguments that, in a setting where measurements are chosen probabilistically, quantum states can be approximately learned using only a linear number of measurements [44, 45].

We now turn our attention to the reconstruction of states arising from ground states of Hamiltonians with local interactions, relevant for modern condensed matter physics, quantum chemistry, atomic and molecular optics, and quantum computing. We begin with the paradigmatic example of the antiferromagnetic transverse-field Ising model in one dimension, with Hamiltonian  $H = J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z + h \sum_i \sigma_i^x$ . The transverse-field Ising model is the testbed par excellence for state-of-the-art quantum simulators in both strongly out-of-equilibrium dynamics [7, 20, 46] as well as quasi-adiabatic regimes [47–51]. It has been

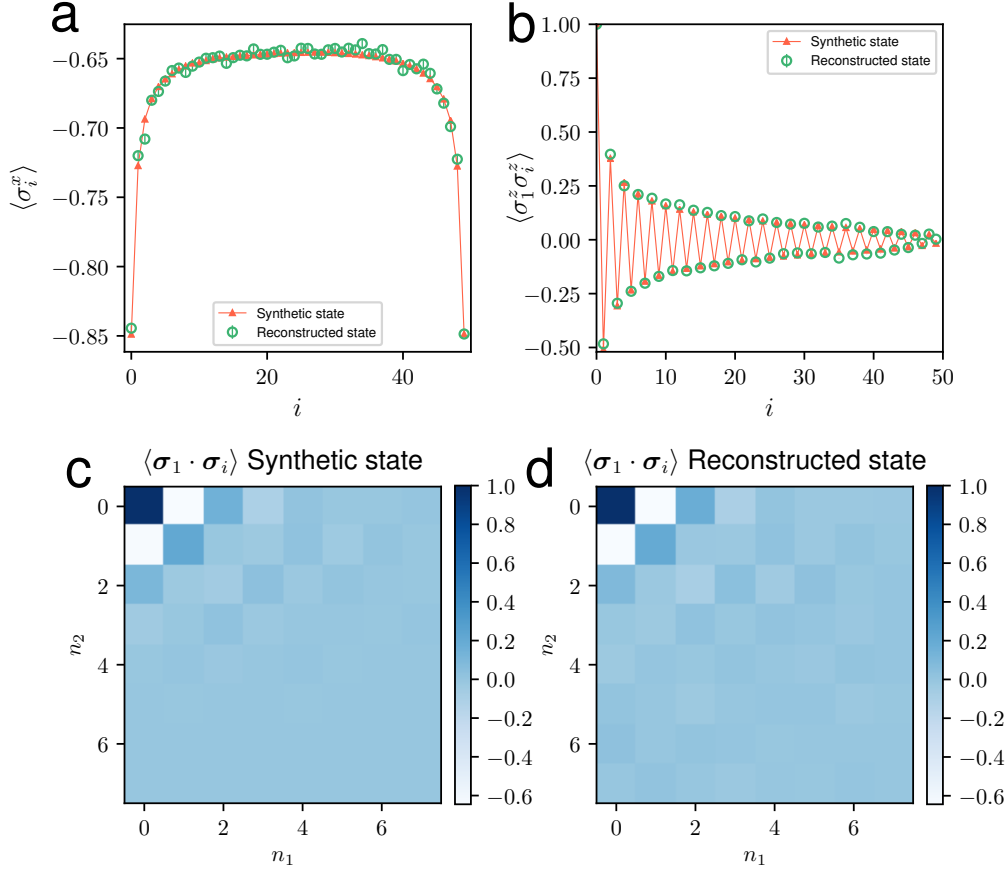


FIG. 4. Learning ground states of local Hamiltonians in one and two dimensions with RNN models for the Pauli-4 POVM ( $m = 4$  outcomes per qubit). Similar results are obtained for the tetrahedral POVM. Antiferromagnetic transverse-field Ising model in 1D for  $N = 50$  spins: (a) 1-body  $\langle \sigma_i^x \rangle$  and (b) 2-body  $\langle \sigma_1^z \sigma_i^z \rangle$  correlators calculated from the exact synthetic state and from the RNN reconstruction as a function of the lattice distance  $i$ . Antiferromagnetic translationally-invariant Heisenberg model on the triangular lattice with  $N = 8 \times 8$  spins: 2-body correlators  $\langle \sigma_1 \cdot \sigma_i \rangle$  calculated from (c) the exact synthetic state and (d) from the RNN reconstruction as a function of the position of the different lattice sites  $i$ . Here,  $i = n_1 \times L + n_2$  labels the  $i$ -th spin located at position  $\mathbf{r}_i = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2$ , where  $\mathbf{a}_1 = (1, 0)$  and  $\mathbf{a}_2 = \frac{1}{2}(1, \sqrt{2})$  are primitive lattice vectors of the triangular lattice, and  $n_{i=1,2} = 0, 1, \dots, L - 1$ .

demonstrated in ion traps [20, 47–49], cold atoms in optical lattices [7, 50], superconducting qubit circuits [6, 46] of up to 1800 qubits [52], and Rydberg-atom platforms of up to 51 qubits [51]. We focus on the most challenging parameter regime for state reconstruction,

the quantum critical point at  $h/J = 1$ . For a system of  $N = 50$  qubits, we first obtain synthetic data mimicking  $10^6$  experimental POVM measurements on the ground state of  $H$ , then perform a reconstruction of the state from these measurements using the same RNN model used for the GHZ state discussed in Fig. 3. Here, we test our method with two different IC POVMs, the tetrahedron and a modified Pauli POVM (termed Pauli-4 POVM,  $\mathbf{M}_{\text{Pauli-4}}$ ). The latter can be experimentally implemented with the same ease as Pauli-6 but with the advantages of having only  $m = 4$  outcomes and an invertible overlap matrix (see Methods). Invertibility of  $\mathbf{T}$  is important because it gives us the remarkable ability to efficiently estimate expectation values of local observables on the reconstructed state  $\varrho_{\text{RNN}}$  directly from  $\mathbf{P}_{\text{RNN}}$ , i.e. without an explicit construction of  $\varrho_{\text{RNN}}$ . This is done stochastically via sampling from  $\mathbf{P}_{\text{RNN}}$  (see Methods). The resulting reconstructions attain a fidelity  $F_C(\mathbf{P}_{\text{RNN}}, \mathbf{P}_{\text{Ising}}) \approx 0.998$ . Furthermore, we find that the 1- and 2-body correlations functions (and, therefore, also the total energy) of  $\varrho_{\text{RNN}}$  display an excellent agreement with the synthetic state, consistent with the exact values to within error bars (Figs. 4 (a) and (b)).

The results we have presented so far demonstrate the power and scalability of our technique on some states with simple structure, which are also amenable to previous tomography approaches [17, 19, 20, 25, 26]. We now expand our investigation to a case that lies outside of conventional approaches. To do so we study a model of frustrated magnetism in two spatial dimensions whose ground state possesses a highly non-trivial sign structure in the computational basis: the antiferromagnetic Heisenberg model in the triangular lattice with Hamiltonian  $H = \sum_{\langle ij \rangle} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j$ , where  $\boldsymbol{\sigma}_i = (\sigma_i^x, \sigma_i^y, \sigma_i^z)$  is the Pauli vector at site  $i$ . We use a tensor network approximation to the ground state of the model for a lattice of  $N = L \times L = 8 \times 8 = 64$  spins to produce synthetic samples from the tetrahedral and Pauli-4 POVM measurements (see Appendix for details). We find that, surprisingly, the RNN model learns the state with a similar success rate as in the one-dimensional case. Apart from the classical fidelity, which, e.g., for the Pauli-4 measurement, reaches  $F_C(\mathbf{P}_{\text{RNN}}, \mathbf{P}_{\text{Heisenberg}}) \approx 0.98$ , we observe a remarkable agreement between the correlation function  $\langle \boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_i \rangle$  measured from our reconstruction  $\varrho_{\text{RNN}}$  and the exact value in the synthetic

state (see Figs. 4 (c) and (d)).

We take advantage of this example to emphasize that our state ansatz corresponds to a contraction between a neural function approximator parametrizing  $\mathbf{P}$  and a Kronecker factorized tensor network composed of local, complex-valued tensors, which are efficiently contractible. The sign structure of the state is provided by the factorized tensor, weighted by the distribution  $P(\mathbf{a})$ . All of the entanglement, and – more generally – any potential classical intractability in the state, can be directly traced back to  $\mathbf{P}$ . One can readily apply any other powerful state-of-the-art probabilistic model and training strategy, such as, e.g., the variational autoencoder or generative adversarial networks, all of which have natural definitions in higher spatial dimensions. Compared to other direct parametrizations of the state which rely on complex-valued generalizations of graphical models [25, 26, 34], our neural networks contain only real parameters, which makes them faster to train. Thus, our ansatz may prove applicable in the study of ground and thermal states of quantum many-body systems via variational energy minimization.

In conclusion, these results demonstrate a scalable machine learning procedure for reconstructing pure and mixed states with structure described by a wide range of generative models in conjunction with easily available measurements. The procedure includes a certification scheme based upon the classical fidelity of the measurement statistics. We demonstrated our method for prototypical states in quantum information as well as ground states of local Hamiltonians relevant to condensed matter, cold atomic systems, and quantum simulators. Most importantly, this works demonstrates how state-of-the-art algorithms in machine learning, combined with established tensor network frameworks in quantum physics, will be necessary in order to make headway in validating and characterizing quantum simulators and commercially available quantum devices in the near-term era of approximate quantum computing [53].

## METHODS

### I. INFORMATIONALLY COMPLETE GENERALIZED MEASUREMENTS

Generalized (non von Neumann) measurements are described by positive-operator valued measures (POVMs). These are defined by decompositions  $\mathbf{M} = \{M^{(a)}\}_a$  of the identity  $\mathbb{1}$ , i.e.  $\sum_a M^{(a)} = \mathbb{1}$ , in terms of non-negative operators  $M^{(a)} \geq 0$ . Additionally, a POVM  $\mathbf{M}$  is said to be informationally complete (IC) if its elements  $M^{(a)}$  span the whole space of bounded-norm, linear operators on the Hilbert in question, where the density matrix to reconstructs lives. Here we consider 3 different single-qubit IC POVMs.

The first one is the tetrahedral POVM  $\mathbf{M}_{\text{tetra}} = \{M^{(a)} = \frac{1}{4}(\mathbb{1} + \mathbf{s}^{(a)} \cdot \boldsymbol{\sigma})\}_{a \in \{0,1,2,3\}}$ , whose outcomes correspond to sub-normalized rank-1 projectors along the directions  $\mathbf{s}^{(0)} = (0, 0, 1)$ ,  $\mathbf{s}^{(1)} = (\frac{2\sqrt{2}}{3}, 0, -\frac{1}{3})$ ,  $\mathbf{s}^{(2)} = (-\frac{\sqrt{2}}{3}, \sqrt{\frac{2}{3}}, -\frac{1}{3})$ , and  $\mathbf{s}^{(3)} = (-\frac{\sqrt{2}}{3}, -\sqrt{\frac{2}{3}}, -\frac{1}{3})$  in the Bloch sphere. These define a regular tetrahedron, which explains the name and also renders  $\mathbf{M}_{\text{tetra}}$  symmetric. A POVM is symmetric if the overlap between any two different elements is the same, i.e. if its overlap matrix has constant non-diagonal elements. The overlap matrix

$$\mathbf{T}_{\text{tetra}} = \begin{bmatrix} \frac{1}{8} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{8} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{8} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{8} \end{bmatrix} \quad (1)$$

of  $\mathbf{M}_{\text{tetra}}$  is invertible. The experimental implementation of  $\mathbf{M}_{\text{tetra}}$  relies on Neumark's dilation theorem [54]. This states that any POVM composed of  $m$  rank-1 operators on a  $d$ -dimensional Hilbert space, for  $d \leq m$ , can be accomplished by a properly crafted projective measurement in an extended Hilbert space with dimension  $m$ . This implies that  $\mathbf{M}_{\text{tetra}}$  can be physically realized by coupling the system qubit to an ancillary qubit and performing a von Neumann measurement on the two qubits (see, e.g., Ref. [55] for explicit constructions).

The second example is the Pauli-6 POVM  $\mathbf{M}_{\text{Pauli-6}}$  with  $m = 6$  outcomes. Each one of its elements is, again, a sub-normalised rank-1 projector:  $\mathbf{M}_{\text{Pauli-6}} = \{M^{(0)} = \frac{1}{3} \times |0\rangle\langle 0|, M^{(1)} = \frac{1}{3} \times |1\rangle\langle 1|, M^{(2)} = \frac{1}{3} \times |+\rangle\langle +|, M^{(3)} = \frac{1}{3} \times |-\rangle\langle -|, M^{(4)} = \frac{1}{3} \times |r\rangle\langle r|, M^{(5)} = \frac{1}{3} \times |l\rangle\langle l|\}$ ,

where  $\{|0\rangle, |1\rangle\}$ ,  $\{|+\rangle, |-\rangle\}$ , and  $\{|r\rangle, |l\rangle\}$  stand for the eigenbases of the Pauli operators  $\sigma^z$ ,  $\sigma^x$ , and  $\sigma^y$ , respectively. Hence, this POVM encapsulates into a single (generalized) measurement all three usual (von Neumann) Pauli measurements. Therefore, it can be physically implemented directly (with no ancillas) by first randomly choosing  $x$ ,  $y$ , or  $z$ , and then measuring the respective Pauli operator. Clearly, any positive probabilities other than  $\frac{1}{3}$  are possible too. This POVM is not symmetric, and its overlap matrix is

$$\mathbf{T}_{\text{Pauli-6}} = \frac{1}{9} \begin{bmatrix} 1 & 0 & 1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 1 & 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 1 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1 & 0 \\ 1/2 & 1/2 & 1/2 & 1/2 & 0 & 1 \end{bmatrix}, \quad (2)$$

which is not invertible. This means that the linear inversion of Figs. 1 (d) and (f) is not possible. It also implies that estimations of expectation values of local observables are difficult even for moderate  $N$ . Luckily, however, non-invertibility poses no particular challenge for the estimation of the classical fidelity, as discussed in the main text.

The third single-qubit POVM we consider is the Pauli-4 POVM  $\mathbf{M}_{\text{Pauli-4}} = \{M^{(0)} = \frac{1}{3} \times |0\rangle\langle 0|, M^{(1)} = \frac{1}{3} \times |1\rangle\langle 1|, M^{(2)} = \frac{1}{3} \times |+\rangle\langle +|, M^{(3)} = \frac{1}{3} \times (|-\rangle\langle -| + |r\rangle\langle r| + |l\rangle\langle l|)\}$ . Its first three elements coincide with their counterparts in  $\mathbf{M}_{\text{Pauli-6}}$ , while its fourth element equals the sum of the last three elements of  $\mathbf{M}_{\text{Pauli-6}}$ . For this reason, Pauli-4 can be physically implemented with the same ease as Pauli-6. In fact, the experimental procedure is almost the same, the only difference being that, for Pauli-4, one adds a trivial classical post-processing that identifies three different outcomes ( $a = 3$ ,  $a = 4$ , and  $a = 5$ ) of Pauli-6 with a single one ( $a = 3$ ) of Pauli-4. Pauli-4 is neither symmetric nor rank-1. However, apart from the computational advantage of having only  $m = 4$  outcomes, its overlap matrix

$$\mathbf{T}_{\text{Pauli-4}} = \frac{1}{9} \begin{bmatrix} 1 & 1/2 & 1/2 & 1 \\ 1/2 & 1 & 1/2 & 1 \\ 1/2 & 1/2 & 1 & 1 \\ 1 & 1 & 1 & 6 \end{bmatrix}. \quad (3)$$

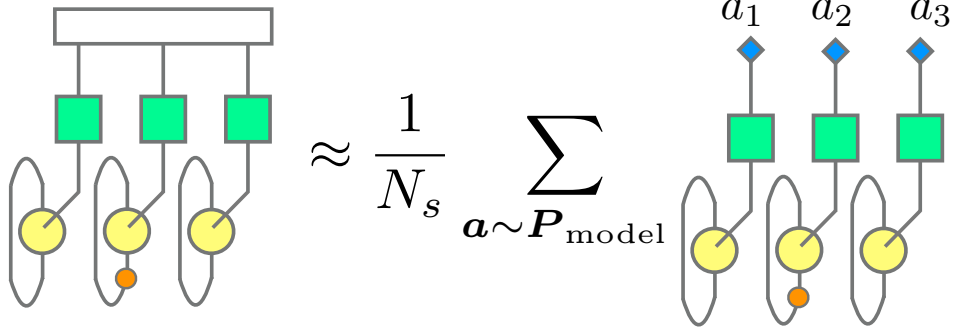


FIG. 5. Direct estimation of local observables from  $N_s$  model samples (i.e., without  $\varrho_{\text{model}}$ ), for the exemplary case of a single-qubit observable  $O$  (orange tensor) and  $N = 3$ . The expectation value  $\text{Tr}[O \varrho_{\text{model}}]$  (left-hand side) equals the expected value  $\mathbb{E}_{\mathbf{a} \sim \mathbf{P}_{\text{model}}}[Q_O(\mathbf{a})]$  of a random variable  $Q_O(\mathbf{a})$  defined by an  $\mathbf{a}$ -dependent tensor contraction (right-hand side), with the string  $\mathbf{a}$  schematically represented by the blue, rhombus-shaped single-index tensors. The expected value can in turn be efficiently Monte-Carlo estimated as an average over  $N_s$  realizations of  $Q_O(\mathbf{a})$  with  $\mathbf{a}$  sampled from  $\mathbf{P}_{\text{model}}$ .

is invertible.

## II. DIRECT STOCHASTIC ESTIMATION OF LOCAL OBSERVABLES

Given samples  $N_s$  from a model distribution  $\mathbf{P}_{\text{model}}$ , parametrizing a state  $\varrho_{\text{model}}$ , and a single-qubit factorable POVM with invertible overlap matrix, it is possible to efficiently compute the expectation value  $\langle O \rangle = \text{Tr}[O \varrho_{\text{model}}]$  of any observables  $O$  acting non-trivially only on a constant (i.e.  $N$ -independent) number of qubits. Remarkably, this can be done directly from the samples, i.e. without the reconstructed density matrix  $\varrho_{\text{model}}$ . Particular examples of such observables are the 1- and 2-body correlators studied in Fig. 4. To see this, note first that, since  $\mathbf{M}$  is IC, one can expand any arbitrary observable as  $O = \sum_{\mathbf{a}} Q_O(\mathbf{a}) M^{(\mathbf{a})}$ , where  $Q_O(\mathbf{a})$  are a-priori unknown complex coefficients. These coefficients are univocally specified by the linear system of equations  $\text{Tr}[O M^{(\mathbf{a}')} ] = \sum_{\mathbf{a}} Q_O(\mathbf{a}) T_{\mathbf{a}, \mathbf{a}'}$ , for all  $\mathbf{a}'$ , which can be straightforwardly solved because  $\mathbf{T}$  is invertible and factorable. Then, using the above expansion, note that  $\text{Tr}[O \varrho_{\text{model}}] = \sum_{\mathbf{a}} Q_O(\mathbf{a}) P_{\text{model}}(\mathbf{a}) = \mathbb{E}_{\mathbf{a} \sim \mathbf{P}_{\text{model}}}[Q_O(\mathbf{a})]$ . Now,

since  $O$  has non-trivial support only a constant number of qubits, the variance of  $Q_O(\mathbf{a})$  (as a random variable over  $\mathbf{a}$ ) is independent of  $N$ . Therefore, the expected value over  $\mathbf{a}$  sampled from  $\mathbf{P}_{\text{model}}$  can be estimated efficiently up to arbitrary constant precision  $\varepsilon$  (i.e., with the computational run-time and  $N_s$  both scaling polynomially in  $N$  and  $\varepsilon^{-1}$ ) by an average  $\frac{1}{N_s} \sum_{\mathbf{a}} Q_O(\mathbf{a})$  over  $N_s$  samples of the random variable  $Q_O(\mathbf{a})$ . See Fig. 5.

Analogously, apart from local observables, one can also directly estimate efficiently the expectation value of non-local observables  $O_{\text{MPO}}$  that admit an efficient matrix-product operator decomposition, provided that the variance of  $Q_{O_{\text{MPO}}}(\mathbf{a})$  scales polynomially with  $N$ . This includes the quantum fidelity between  $\varrho_{\text{model}}$  and a target MPS of constant bond dimension, again provided that the corresponding variance scales well (see the appendix).

- 
- [1] Banaszek, K., Cramer, M. & Gross, D. Focus on quantum tomography. *New Journal of Physics* **15**, 125020 (2013).
  - [2] Cirac, J. I. & Zoller, P. Goals and opportunities in quantum simulation. *Nature Physics* **8**, 264–266 (2012).
  - [3] Aspuru-Guzik, A. & Walther, P. Photonic quantum simulators. *Nat. Phys.* **8**, 285–291 (2012).
  - [4] Bloch, I., Dalibard, J. & Nascimbène, S. Quantum simulations with ultracold quantum gases. *Nat. Phys.* **8**, 267–276 (2012).
  - [5] Blatt, R. & Roos, C. F. Quantum simulations with trapped ions. *Nat. Phys.* **8**, 277–284 (2012).
  - [6] Houck, A. A., Türeci, H. E. & Koch, J. On-chip quantum simulation with superconducting circuits. *Nat. Phys.* **8**, 292–299 (2012).
  - [7] Gross, C. & Bloch, I. Quantum simulations with ultracold atoms in optical lattices. *Science* **357**, 995 (2017).
  - [8] Vogel, K. & Risken, H. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A* **40**, 2847 (1989).
  - [9] James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001). URL <https://link.aps.org/doi/10.1103/PhysRevA.64.052312>.



- [10] Roos, C. F. *et al.* Bell states of atoms with ultralong lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* **92**, 220402 (2004).
- [11] Haffner, H. *et al.* Scalable multiparticle entanglement of trapped ions. *Nature* **438**, 643–646 (2005).
- [12] Poulin, D., Qarry, A., Somma, R. & Verstraete, F. Quantum Simulation of Time-Dependent Hamiltonians and the Convenient Illusion of Hilbert Space. *Phys. Rev. Lett.* **106**, 170501 (2011).
- [13] Kliesch, M., Barthel, T., Gogolin, C., Kastoryano, M. & Eisert, J. A dissipative quantum church-turing theorem. *Phys. Rev. Lett.* **107**, 120501 (2011).
- [14] Tóth, G. *et al.* Permutationally invariant quantum tomography. *Phys. Rev. Lett.* **105**, 250403 (2010).
- [15] Moroder, T. *et al.* Permutationally invariant state reconstruction. *New J. Phys.* **14**, 105001 (2012).
- [16] Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S. & Eisert, J. Quantum State Tomography via Compressed Sensing. *Phys. Rev. Lett.* **105**, 150401 (2010).
- [17] Cramer, M. *et al.* Efficient quantum state tomography. *Nature communications* **1**, 149 (2009). URL <http://www.nature.com/articles/ncomms1147>.
- [18] Baumgratz, T., Gross, D., Cramer, M. & Plenio, M. B. Scalable reconstruction of density matrices. *Phys. Rev. Lett.* **111**, 020401 (2013).
- [19] Han, Z.-Y. *et al.* Efficient Quantum Tomography with Fidelity Estimation. *ArXiv e-prints* (2017). 1712.03213.
- [20] Lanyon, B. P. *et al.* Efficient tomography of a quantum many-body system. *Nature Physics* **13**, 1158 EP – (2017). URL <http://dx.doi.org/10.1038/nphys4244>.
- [21] Flammia, S. T. & Liu, Y.-K. Direct fidelity estimation from few pauli measurements. *Phys. Rev. Lett.* **106**, 230501 (2011).
- [22] da Silva, M. P., Landon-Cardinal, O. & Poulin, D. Practical characterisation of quantum devices without tomography. *Phys. Rev. Lett.* **107**, 210404 (2011).
- [23] Aolita, L., Gogolin, C., Kliesch, M. & Eisert, J. Reliable quantum certification of photonic state preparations. *Nature Communications* **6**, 8498 (2015).
- [24] Gluza, M., Kliesch, M., Eisert, J. & Aolita, L. Fidelity witnesses for fermionic quantum simulations. *Phys. Rev. Lett.* **120**, 190501 (2018).

- [25] Torlai, G. *et al.* Neural-network quantum state tomography. *Nature Physics* **14**, 447–450 (2018). URL <https://doi.org/10.1038/s41567-018-0048-5>.
- [26] Torlai, G. & Melko, R. G. Latent Space Purification via Neural Density Operators. *ArXiv e-prints* (2018). 1801.09684.
- [27] Raghu, M., Poole, B., Kleinberg, J., Ganguli, S. & Sohl-Dickstein, J. On the expressive power of deep neural networks. In Precup, D. & Teh, Y. W. (eds.) *Proceedings of the 34th International Conference on Machine Learning*, vol. 70 of *Proceedings of Machine Learning Research*, 2847–2854 (PMLR, International Convention Centre, Sydney, Australia, 2017). URL <http://proceedings.mlr.press/v70/raghu17a.html>.
- [28] Levine, Y., Sharir, O., Cohen, N. & Shashua, A. Bridging Many-Body Quantum Physics and Deep Learning via Tensor Networks. *ArXiv e-prints* (2018). 1803.09780.
- [29] Deng, D.-L., Li, X. & Das Sarma, S. Quantum entanglement in neural network states. *Phys. Rev. X* **7**, 021021 (2017). URL <https://link.aps.org/doi/10.1103/PhysRevX.7.021021>.
- [30] Gao, X. & Duan, L.-M. Efficient representation of quantum many-body states with deep neural networks. *Nature Communications* **8**, 662 (2017). 1701.05039.
- [31] Chen, J., Cheng, S., Xie, H., Wang, L. & Xiang, T. Equivalence of restricted boltzmann machines and tensor network states. *Phys. Rev. B* **97**, 085104 (2018). URL <https://link.aps.org/doi/10.1103/PhysRevB.97.085104>.
- [32] Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016). <http://www.deeplearningbook.org>.
- [33] Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, New York, NY, USA, 2011), 10th edn.
- [34] Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017). URL <http://science.sciencemag.org/content/355/6325/602>. <http://science.sciencemag.org/content/355/6325/602.full.pdf>.
- [35] Deng, D.-L., Li, X. & Das Sarma, S. Machine learning topological states. *Phys. Rev. B* **96**, 195145 (2017). URL <https://link.aps.org/doi/10.1103/PhysRevB.96.195145>.
- [36] Nomura, Y., Darmawan, A. S., Yamaji, Y. & Imada, M. Restricted boltzmann machine learning for solving strongly correlated quantum systems. *Phys. Rev. B* **96**, 205152 (2017). URL <https://link.aps.org/doi/10.1103/PhysRevB.96.205152>.
- [37] Clark, S. R. Unifying neural-network quantum states and correlator product states via tensor

- networks. *Journal of Physics A: Mathematical and Theoretical* **51**, 135301 (2018). URL <http://stacks.iop.org/1751-8121/51/i=13/a=135301>.
- [38] Salakhutdinov, R., Mnih, A. & Hinton, G. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th International Conference on Machine Learning, ICML '07*, 791–798 (ACM, New York, NY, USA, 2007). URL <http://doi.acm.org/10.1145/1273496.1273596>.
- [39] Bennett, J., Lanning, S. & Netflix, N. The netflix prize. In *In KDD Cup and Workshop in conjunction with KDD* (2007).
- [40] Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to sequence learning with neural networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. & Weinberger, K. Q. (eds.) *Advances in Neural Information Processing Systems 27*, 3104–3112 (Curran Associates, Inc., 2014). URL <http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf>.
- [41] Wu, Y. *et al.* Google’s neural machine translation system: Bridging the gap between human and machine translation. *CoRR* **abs/1609.08144** (2016). URL <http://arxiv.org/abs/1609.08144>.
- [42] Chiu, C.-C. *et al.* State-of-the-art speech recognition with sequence-to-sequence models (2018). URL <https://arxiv.org/pdf/1712.01769.pdf>.
- [43] Preskill, J. Lectures notes on quantum computation (2018). URL <http://www.theory.caltech.edu/~preskill/ph219/index.html>.
- [44] Aaronson, S. The learnability of quantum states. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **463**, 3089–3114 (2007). URL <http://rspa.royalsocietypublishing.org/content/463/2088/3089>. <http://rspa.royalsocietypublishing.org/content/463/2088/3089.full.pdf>.
- [45] Rocchetto, A. *et al.* Experimental learning of quantum states. *ArXiv e-prints* (2017). 1712.00127.
- [46] Barends, R. *et al.* Digital quantum simulation of fermionic models with a superconducting circuit. *Nature Communications* **6**, 7654 (2015).
- [47] Friedenauer, H., Schmitz, H., Glueckert, J., Porras, D. & Schaetz, T. Simulating a quantum magnet with trapped ions. *Nature Physics* **4**, 757 (2008).
- [48] Kim, K. *et al.* Quantum simulation of frustrated ising spins with trapped ions. *Nature* **465**,

- 590 (2010).
- [49] Islam, R. *et al.* Onset of a quantum phase transition with a trapped ion quantum simulator. *Nature Communications* **2**, 377 (2011).
  - [50] Simon, J. *et al.* Quantum simulation of antiferromagnetic spin chains in an optical lattice. *Nature* **472**, 307 (2011).
  - [51] Bernien, H. *et al.* Probing many-body dynamics on a 51-atom quantum simulator. *Nature (London)* **551**, 579–584 (2017). 1707.04344.
  - [52] King, A. D. *et al.* Observation of topological phenomena in a programmable lattice of 1,800 qubits. *Nature (London)* **560**, 456–460 (2018). 1803.02047.
  - [53] Preskill, J. Quantum Computing in the NISQ era and beyond. *ArXiv e-prints* (2018). 1801.00862.
  - [54] Peres, A. *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1995).
  - [55] Tabia, G. N. M. Experimental scheme for qubit and qutrit symmetric informationally complete positive operator-valued measurements using multiport devices. *Phys. Rev. A* **86**, 062107 (2012). URL <https://link.aps.org/doi/10.1103/PhysRevA.86.062107>.
  - [56] Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling (2014). URL <http://arxiv.org/abs/1412.3555>. Cite arXiv:1412.3555Comment: Presented in NIPS 2014 Deep Learning and Representation Learning Workshop.
  - [57] Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural computation* **9**, 1735–1780 (1997).
  - [58] Cho, K., van Merriënboer, B., Bahdanau, D. & Bengio, Y. On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111 (Association for Computational Linguistics, 2014). URL <http://www.aclweb.org/anthology/W14-4012>.
  - [59] Orus, R. A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States. *Annals Phys.* **349**, 117–158 (2014). 1306.2164.
  - [60] Ferris, A. J. & Vidal, G. Perfect sampling with unitary tensor networks. *Phys. Rev. B* **85**, 165146 (2012). URL <https://link.aps.org/doi/10.1103/PhysRevB.85.165146>.
  - [61] Itensor. <https://github.com/ITensor/ITensor>.

*Acknowledgments.* We would like to thank Guifre Vidal and Lukasz Cincio for discus-

sions and encouragement. We thank Nathan Berkovits, Alexandre Reily Rocha, and Pedro Vieira for organizing the ICTP-SAIFR/IFT-UNESP Minicourse on Machine Learning for Many-Body Physics, where this work was started. This research was supported by NSERC of Canada, the Perimeter Institute for Theoretical Physics, and the Shared Hierarchical Academic Research Computing Network (SHARCNET). Research at Perimeter Institute is supported through Industry Canada and by the Province of Ontario through the Ministry of Research & Innovation. J.C acknowledges financial and computational support from the AI grant. LA acknowledges financial support from the Brazilian agencies CNPq, FAPERJ, CAPES, and FAPESP. This work was supported by the Serrapilheira Institute (grant number Serra-1709-17173).

## Appendix A: Generative models

### 1. RBM with softmax units

The number of measurement outcomes at each qubit is  $m = 6$  for the Pauli-6 and  $m = 4$  tetrahedral and Pauli-4 POVMs. To represent these measurement distributions  $\mathbf{P}_{\text{model}}$  for a system with  $N$  qubits we first employ an RBM with  $N$   $m$ -index visible units [38]. We encode the states of each unit through one-hot vectors with  $m$  components on each qubit. A one-hot vector is a  $1 \times m$  vector used to distinguish each measurement outcome. The vector consists of zeros in all cells with the exception of a single 1 in a cell used uniquely to identify the measurement. The sampling strategy we follow uses block Gibbs sampling, where groups of two or more variables are sampled together from their joint distribution conditioned on all other variables, rather than sampling from each one individually. The conditional probabilities are given by

$$p(v_i^k = 1 \mid \mathbf{h}) = \frac{e^{\sum_{j=1}^{n_H} \sum_{l=1}^L W_{ij}^{kl} h_j^l + b_i^k}}{\sum_{k'=1}^m e^{\sum_{j=1}^{n_H} \sum_{l=1}^L W_{ij}^{k'l} h_j^l + b_i^{k'}}}. \quad (\text{A1})$$

Here  $n_H$  is the number of  $L$ -dimensional hidden units.  $W_{ij}^{kl}$  is the  $ijkl$ -th element of the weight matrix  $\mathbf{W}$ , which is a 4-index array encoding the interaction between the different

states  $k$  and  $l$  in the visible and hidden units  $i$  and  $j$ , respectively. The  $b_i^k$  are the biases in the visible unit and there is an analog expression for the hidden units  $a_j^l$ . The energy of the RBM is given by  $E(\mathbf{v}, \mathbf{h}) = -\sum_{ijkl} W_{ij}^{kl} v_i^k h_j^l - \sum_{ik} b_i^k v_i^k - \sum_{jk} a_j^k h_j^k$ , while the joint probability over the visible and hidden units is given by  $P(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v}, \mathbf{h})}}{Z}$ , where  $Z = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}$ . We train our RBMs using standard block Gibbs sampling and contrastive divergence [32].

## 2. Recurrent neural network models

Recurrent neural networks (RNN) are extensions of the traditional feedforward neural networks designed to process sequential data [32, 56]. In our context, the data corresponds to a measurement outcome string  $\mathbf{a} = (a_1, a_2, \dots, a_N)$ , which we understand as a sequence of single-qubit outcomes according to the given ordering of the qubits. The RNNs process sequential data by updating a recurrent hidden state  $\mathbf{h}_i$  whose value at each time is dependent on that of the previous steps, i.e.

$$\mathbf{h}_i = \phi(\mathbf{h}_{i-1}, a_i; \boldsymbol{\theta}), \quad (\text{A2})$$

where  $\phi$  is a non-linear function with parameters  $\boldsymbol{\theta}$ . In its original form, RNNs process  $\mathbf{h}_{i-1}$  via

$$\mathbf{h}_i = f(\mathbf{W}[\mathbf{a}_i; \mathbf{h}_{i-1}]), \quad (\text{A3})$$

where  $\mathbf{z} = [\mathbf{x}; \mathbf{y}]$  concatenates vectors a  $k$ -dimensional vector  $\mathbf{x}$  and an  $q$ -dimensional vector  $\mathbf{y}$ , i.e.  $\mathbf{z} = [x_1, x_2, \dots, x_k, y_1, y_2, \dots, y_q]$ . The matrix  $\mathbf{W}$  and bias vector  $\mathbf{b}$  are the trainable parameters of the model. The function  $f$  is a sigmoid function or a hyperbolic tangent, and  $\mathbf{a}_i$  is a one-hot vector encoding of the integer outcome  $a_i$ .

A generative RNN model outputs a probability distribution over the next element in a sequence given all previously observed measurements, i.e.

$$P(a_i | a_1, \dots, a_{i-1}) = S(\mathbf{h}_{i-1}; \mathbf{U}) \quad (\text{A4})$$

where  $S$  is a softmax and  $\mathbf{U}$  are further parameters optimized during training. Using the

chain rule of probability, the full model is given by

$$P_{\text{model}}(a_1, a_2 \dots a_N) = P(a_1) P(a_2|a_1) P(a_3|a_1, a_2) \dots P(a_N|a_1, a_2 \dots a_{N-1}). \quad (\text{A5})$$

In its original form, training RNNs is challenging since capturing long-term dependencies between the qubit measurements tends to make the gradients of the cost function with respect to the parameters in the RNN either explode or vanish. Furthermore, because of the recurrent structure, the long-term dependencies tend to be masked exponentially with respect to the effect of short-term dependencies. To overcome these limitations, a possible direction is to modify the recurrent unit using a long short-term memory (LSTM) unit [57], which are explicitly developed to avoid the long-term dependency problem. Similarly, the gated recurrent unit (GRU) [58], which was originally introduced in the context of neural machine translation, is designed to adaptively capture dependencies of different scales. The GRU processes the sequential data through

$$\mathbf{z}_i = \sigma(\mathbf{W}_z[\mathbf{h}_{i-1}; \mathbf{a}_i]) \quad (\text{A6})$$

$$\mathbf{r}_i = \sigma(\mathbf{W}_r[\mathbf{h}_{i-1}; \mathbf{a}_i])$$

$$\hat{\mathbf{h}}_i = \tanh(\mathbf{W}_c[\mathbf{r}_i \odot \mathbf{h}_{i-1}; \mathbf{a}_i]). \quad (\text{A7})$$

$$\mathbf{h}_i = (1 - \mathbf{z}_i) \odot \mathbf{h}_{i-1} + \mathbf{z}_i \odot \hat{\mathbf{h}}. \quad (\text{A8})$$

Here the hidden state  $\mathbf{h}_i$  is updated through a linear interpolation between the previous activation  $\mathbf{h}_{i-1}$  and the candidate hidden state  $\hat{\mathbf{h}}_i$ . The update gate  $\mathbf{z}_i$  decides how much to update the contents of the hidden state. Here  $\mathbf{r}_i$  is a set of reset gates and  $\odot$  denotes element-wise vector multiplication. When the reset gates are “off”, i.e.  $\mathbf{r}_i \approx 0$ , they cancel out the unit which then acts as if it is reading the first element of the sequence, effectively making the unit “forget” part of the sequence that has already been encoded in the state vector  $\mathbf{h}_i$ , if necessary. The matrices  $\mathbf{W}_z$  parametrize the GRU and are optimized using standard maximum likelihood estimation. The GRU Eq. A6 are graphically represented in Fig. 6. For Fig. 6, we emphasize that the meaning of lines, circles, and squares is different from the tensor-network notation used throughout the text. Here, lines with arrowheads

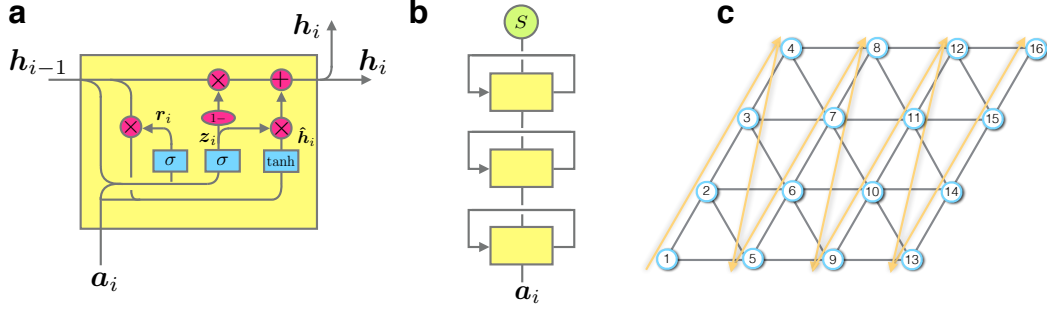


FIG. 6. Probabilistic RNN model. (a) Graphical representation of the gated recurrent unit described in Eq. A6 and used in our numerical experiments. (b) Our deep RNN model stacks three GRU units (yellow blocks) followed by a fully connected layer with a softmax activation denoted by  $S$ . (c) Enumeration and one-dimensional path along which the RNN processes the measurement outcomes for the two-dimensional Heisenberg model.

denote incoming vectors from the output of one node to the inputs of others. The magenta circles/ellipses represent pointwise operations such as vector addition or multiplication. The blue rectangles represent neural network layers labeled by the type of nonlinearity they used. Lines merging denote vector concatenation, while lines forking means the vector is copied and used in another operation.

To make our models more expressive we train a deep RNN model with three stacked GRU units (see Fig. 6(b)). In all of our experiments the dimension of the state vector  $h_i$  is set to 100. In one spatial dimensional there is a natural enumeration scheme for the lattice sites (e.g. from right to left or viceversa) along which the RNN model processes the data. In the absence of a notion of spatial dimensionality, like in the GHZ state, one can choose an arbitrary enumeration of the qubits. In two dimensional systems, such as the triangular Heisenberg model in the manuscript, we choose an enumeration of the lattice sites which, in turn, defines a one-dimensional path filling the two-dimensional lattice. The RNN then uses this path to process and learn the data. This is exemplified in Fig. 6(c) for a  $4 \times 4$  triangular lattice. An analogous enumeration is used in our experiments in the main text.



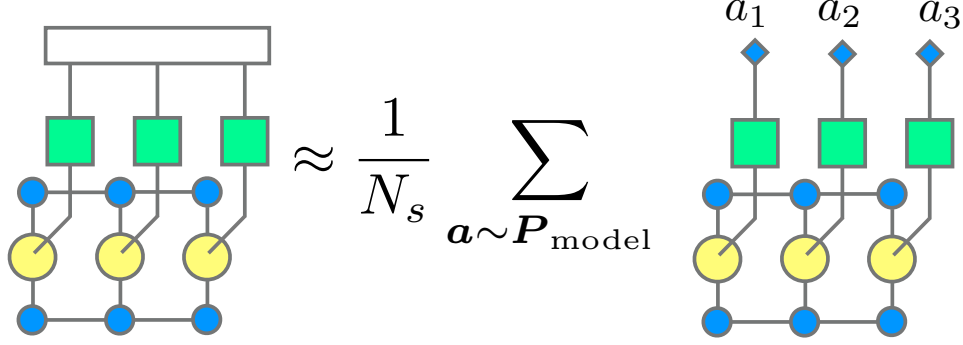


FIG. 7. Estimation of the quantum fidelity between  $\varrho_{\text{model}}$  and a target  $|\Psi\rangle$  directly from  $N_s$  samples from the model distribution  $\mathbf{P}_{\text{model}}$  for  $N = 3$ . If  $|\Psi\rangle$  is an MPS (blue circles) of constant bond dimension, it can be expressed as the virtual-index contraction of 2- and 3-index tensors of constant size. The squared fidelity corresponds to the contraction of the MPS with  $\varrho_{\text{model}}$  (left-hand side). This, in turn, can be efficiently Monte-Carlo estimated as an average (right-hand side) over  $N_s$  realizations of  $Q_{|\Psi\rangle\langle\Psi|}(\mathbf{a})$  with  $\mathbf{a}$  sampled from  $\mathbf{P}_{\text{model}}$

## Appendix B: Direct stochastic estimation of quantum fidelity with respect to an MPS

For POVMs with invertible overlap matrix, the quantum fidelity between a reconstructed state  $\varrho_{\text{model}}$  and a generic pure state  $|\Psi\rangle$  can in principle be computed sampling the model distribution  $\mathbf{P}_{\text{model}}$ . Since one of the states is pure, the squared fidelity is given by

$$F^2(|\Psi\rangle\langle\Psi|, \varrho_{\text{model}}) = \mathbb{E}_{\mathbf{P}_{\text{model}}} [Q_{|\Psi\rangle\langle\Psi|}(\mathbf{a})], \quad (\text{B1})$$

with  $Q_{|\Psi\rangle\langle\Psi|}(\mathbf{a}) = \sum_{\mathbf{a}'} T_{\mathbf{a}, \mathbf{a}'}^{-1} \langle\Psi| M^{(\mathbf{a}')} |\Psi\rangle$ . Whereas the computation of each  $Q_{|\Psi\rangle\langle\Psi|}(\mathbf{a})$  is in general intractable, it can be done efficiently for any state  $|\Psi\rangle$  admitting an efficient MPS representation (see Fig. 7). In addition, however, for the Monte-Carlo estimation to converge efficiently, it is also required that the variance of the estimator  $Q_{|\Psi\rangle\langle\Psi|}(\mathbf{a})$  does not grow too fast in  $N$ . Unfortunately, we find that the variance of  $Q_{|\Psi\rangle\langle\Psi|}(\mathbf{a})$  for the  $N$ -qubit GHZ state (see Sec. C for its MPS representation) grows dramatically with  $N$ , making the estimation unfeasible for systems as small as  $N \approx 8$ . It is an open question whether there are non-trivial MPS states for which the corresponding variance grows slowly with  $N$ .

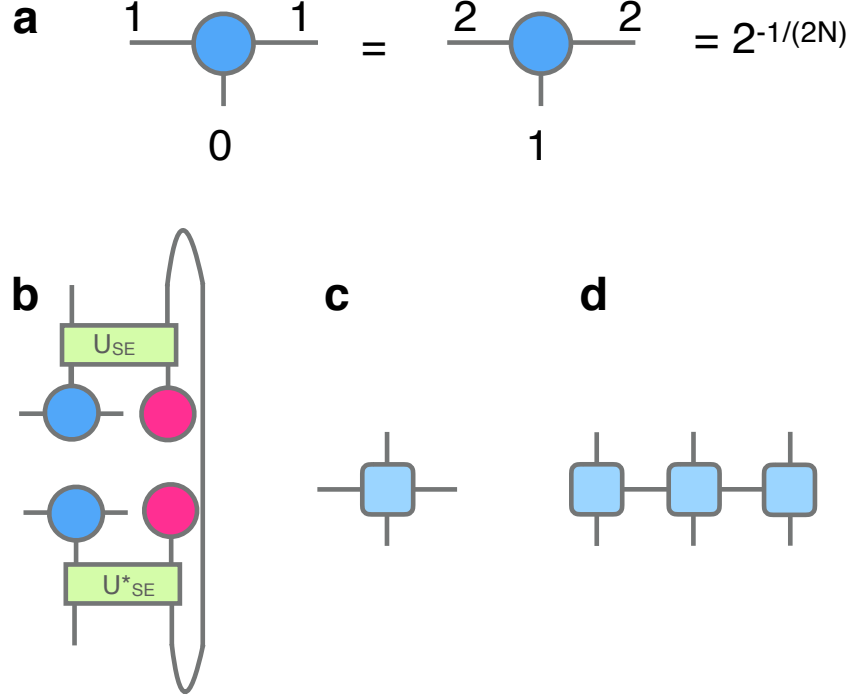


FIG. 8. Tensor-network description of the GHZ states. (a) The non-zero elements of the tensors representing the GHZ state as an MPS. (b) Applying the unitary operator  $U_{SA}$  locally to the tensors in the pure GHZ and their ancillae. (c) The tensor in (b) is reshaped to a 4 dimensional tensor making up the MPO representation of the noisy GHZ state in (d).

### Appendix C: Matrix product operator representation of the noisy GHZ and its sampling

The GHZ state in the absence of noise  $p = 0$  can be written down as an MPS of bond dimension  $\chi = 2$  [59]. The non-zero elements of the MPS tensors are displayed in Fig. 8 (a). Similarly, the locally depolarized GHZ state can be expressed as an MPO with bond dimension  $\chi = 4$  as follows. Given an  $N$ -qubit MPS representation of the pure GHZ, attach an ancillary ququart initialized in a reference state  $|0\rangle_A$  to each of the  $N$  system qubits. Then apply a local unitary operator  $U_{SA}$  acting on the system qubit (S) and ancilla (A) that represents a unitary dilation of the depolarizing channel. Once the ancilla is traced out, the desired locally depolarized state is obtained on S. On a single qubit, the depolarizing channel can be realized by an isometry mapping the state of the qubit  $|\Psi\rangle_A$  to (qubit +

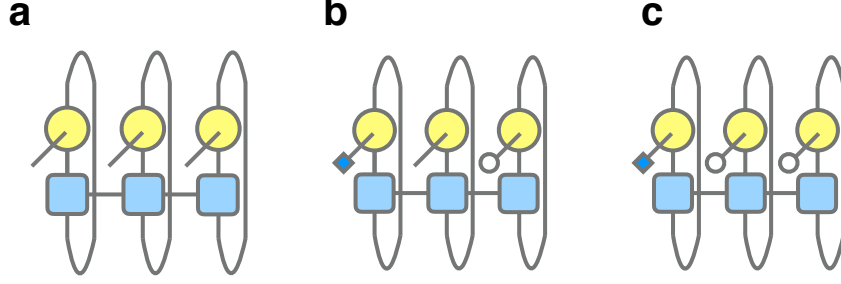


FIG. 9. Sampling an outcome string  $\mathbf{a}$  from  $\mathbf{P}_{\text{MPO}}$ . (a) Tensor-network representation of  $\mathbf{P}_{\text{MPO}}$ . The marginal conditional probability  $P(a_2|a_1)$  is calculated as the ratio of the vector (b) and normalization constant (c).

ancilla ququart)  $|\Psi\rangle_{AE}$  state acting as [43]

$$|\Psi\rangle_A \mapsto |\Psi\rangle_{AE} = \sqrt{1-p} |\Psi\rangle_S \otimes |0\rangle_A + \sqrt{\frac{p}{3}} (\sigma_S^x |\Psi\rangle_S \otimes |1\rangle_A + \sigma_S^y |\Psi\rangle_S \otimes |2\rangle_A + \sigma_S^z |\Psi\rangle_S \otimes |3\rangle_A). \quad (\text{C1})$$

On an  $N$ -qubit system, apply this operation on each of the qubits and their corresponding ancillae. In practice, this amounts to acting with  $U_{\text{SA}}$  on each of the MPS tensors of the original GHZ state supplemented with a 4-dimensional ancilla initialized in the state  $|0\rangle$ , followed by a trace over the ancilla (see Fig. 8 (b)). The resulting tensor is reinterpreted and reshaped as the factors (Fig. 8 (c)) in an MPO that represents the density matrix of the system, as depicted Fig. 8 (d).

Given a factorable POVM [as in Fig. 1 (b)], and an MPO density matrix with small bond dimension  $\chi$ , the measurement statistics  $\mathbf{P}_{\text{MPO}}$ , depicted graphically in Fig. 9 (a), can be tractably sampled using the ideas introduced in Ref. [60]. Our strategy invokes the chain rule of probability in Eq. A5 and computes sequentially a series of conditional single-qubit probabilities  $\{P(a_1), P(a_2|a_1), \dots, a_N\}$ , each of which can be computed efficiently for the tensor network representation in Fig. 9 (a). First we compute  $P(a_1)$  by contracting the entire tensor network from right to left with constant vectors (white vectors  $\mathbf{e} = (1, 1, 1, 1)$  in Fig. 9 (b)). Then we sample  $P(a_1)$  and fix its index to the obtained sample  $a'_1$  [blue diamond vector with components  $a_i^* = \delta_{i,a'_1}$  in Fig. 9 (b)] Proceed to site 2 and compute  $P(a_2|a_1)$  which is the ratio between  $P(a'_1, a_2) = \sum_{a_3, \dots, a_N} P(a'_1, a_2, a_3, \dots, a_N)$  [graphically

depicted in Fig. 9 (b)] and  $P(a'_1) = \sum_{a_2, a_3, \dots, a_N} P(a'_1, a_2, a_3, \dots, a_N)$  presented in Fig. 9 (c). Sample  $P(a_2|a_1)$  and continue doing the same for the rest of qubits in the system until the  $N$ -th qubit is reached. Note that apart from the exact sample  $\mathbf{a}' = (a'_1, a'_2, a'_3, \dots, a'_N)$ , its probability is also tractable and given by Eq. A5.

#### Appendix D: Density-matrix renormalization group calculations

The training datasets for the reconstruction of the ground states of spin Hamiltonians have been generated using the density matrix renormalization group (DMRG). The algorithm is implemented within the framework of MPSs, and it is executed using the ITensor library [61]. Given a particular spin model, with the Hamiltonian expressed as a MPO, the DMRG algorithm attempts to find the optimal MPS with the lowest energy. The elements of the  $N$  tensors in the MPS are optimized using a unit cell containing two sites. For the one-dimensional antiferromagnetic Ising model at the critical point, the ground state is obtained by performing 5 DMRG sweeps along the chain. The resulting MPS for  $N = 50$  spins has bond dimension  $\chi = 25$ , and a cut-off error of  $9.93 \times 10^{-10}$ . The second spin model we considered is the Heisenberg model on a two-dimensional triangular lattice. Given the symmetries of the Hamiltonian, we restrict the DMRG calculations to the sector corresponding to a zero total magnetization,  $\sum_i \sigma_i^z = 0$ . To find an approximation to the ground state of the model, we perform 10 DMRG sweeps with a maximum bond dimension of  $\chi = 200$ , and a resulting truncation error of  $1.8 \times 10^{-4}$ .