Simulating disordered quantum Ising chains via dense and sparse restricted Boltzmann machines

S. Pilati[1] and P. Pieri[1,2]

[1]*School of Science and Technology, Physics Division, Università di Camerino, 62032 Camerino (MC), Italy*

[2]*INFN, Sezione di Perugia, 06123 Perugia (PG), Italy*

(Received 25 March 2020; accepted 2 June 2020; published 22 June 2020)

In recent years, generative artificial neural networks based on restricted Boltzmann machines (RBMs) have been successfully employed as accurate and flexible variational wave functions for clean quantum many-body systems. In this article, we explore their use in simulations of disordered quantum Ising chains. The standard dense RBM with all-to-all interlayer connectivity is not particularly appropriate for large disordered systems, since in such systems one cannot exploit translational invariance to reduce the amount of parameters to be optimized. To circumvent this problem, we implement sparse RBMs, whereby the visible spins are connected only to a subset of local hidden neurons, thus reducing the amount of parameters. We assess the performance of sparse RBMs as a function of the range of the allowed connections, and we compare it with that of dense RBMs. Benchmark results are provided for two sign-problem-free Hamiltonians, namely pure and random quantum Ising chains. The RBM *Ansätzes* are trained using the unsupervised learning scheme based on projective quantum Monte Carlo (PQMC) algorithms. We find that the sparse connectivity facilitates the training process and allows sparse RBMs to outperform their dense counterparts. Furthermore, the use of sparse RBMs as guiding functions for PQMC simulations allows us to perform PQMC simulations at a reduced computational cost, avoiding possible biases due to finite random-walker populations. We obtain unbiased predictions for the ground-state energies and the magnetization profiles with fixed boundary conditions, at the ferromagnetic quantum critical point. The magnetization profiles agree with the Fisher–de Gennes scaling relation for conformally invariant systems, including the scaling dimension predicted by the renormalization-group analysis.

DOI: 10.1103/PhysRevE.101.063308

I. INTRODUCTION

Carleo and Troyer's 2017 article [1] was the impetus for a vibrant research activity on the use of artificial neural networks as variational *Ansätzes* for ground-state wave functions. Several neural networks have been adopted, but restricted Boltzmann machines (RBMs) [2–4] have emerged as one of the most accurate and versatile [5]. Beyond approximating ground-state wave functions, they have also been used to perform quantum state tomography [6,7], to accelerate classical and quantum Monte Carlo simulations [8,9], to solve classical combinatorial optimization problems [10], as well as to simulate excited states [11], open quantum systems [12–15], and unitary dynamics [1]. Recently, a procedure to use RBMs to also represent states with non-Abelian symmetries has been introduced [16]. A plethora of competitive neural-network models have been investigated, including (possibly among others) unrestricted Boltzmann machines [17], deep Boltzmann machines [18], deep feed-forward neural networks [11,19–22], convolutional neural networks [23–25], generalized transfer-matrix states [26], neural-network backflow models [27], and neural Gutzwiller-projected wave functions [28]. More recently, autoregressive neural networks have also been employed [29–31]. Very different physical systems have already been simulated via artificial neural networks. The first studies addressed unfrustrated spin Hamiltonians [1]. They have been followed by studies on bosonic lattice models [20,22,23], frustrated spin Hamiltonians [24,25,32,33], fermionic lattice models [27], as well as on continuous-space bosonic [34,35] and fermionic systems [35], particularly electronic ones [36–39]. Topological states [40–42] and nuclear systems have been addressed as well [43]. The investigations mentioned above addressed few-body systems or clean extended models. To the best of our knowledge, Hamiltonians with random disorder have not been addressed so far. In particular, it is still unclear whether neural-network wave functions can accurately describe the ground state of disordered quantum many-body systems.

In this article, we explore the use of artificial neural networks—specifically of RBMs—to simulate the ground state of one-dimensional disordered quantum spin Hamiltonians. Among the three available schemes to train the RBMs—reinforcement learning [1], supervised learning [44], and unsupervised-learning [9]—we adopt the latter. In this scheme, the training of the RBM is performed in combination with projective quantum Monte Carlo simulations (PQMCs). On the one hand, this scheme provides one with accurate RBM approximations for the ground-state wave function. On the other hand, the use of the optimized RBMs as guiding functions for successive PQMC simulations [9] allows one to eliminate the possible bias due to the finite random-walker population [17,45–47]. Notably, this provides one with unbiased estimates of ground-state properties, even when the optimized RBM wave function is not exact.

Standard RBMs are characterized by a dense connectivity structure, whereby all visible spins are connected to all

neurons in the (unique) hidden layer. In general, this implies that the number of parameters to be optimized increases with the square of the system size. Such a large amount of parameters often leads to difficulties in the optimization of variational *Ansätzes* for many-body systems, meaning that the optimization algorithm might fail to identify optimal parameters. Furthermore, the dense connectivity increases the computational cost of any RBM-based Monte Carlo simulation, including PQMC simulations guided by RBMs. In clean systems, translational invariance can be exploited to reduce the scaling of the number of parameters with the system size from quadratic to linear [1,48]. Clearly, this is not possible in disordered systems. In this article, we investigate the use of sparse RBMs [40,41,49,50] for disordered systems. Specifically, a notion of spatial distance is embedded in the connectivity structure, and the visible spins are connected only to a certain number of nearest-neighbor hidden neurons. As suggested by studies performed in the field of machine learning [50,51], one expects the sparse connectivity to allow the model to learn the dominant correlations at a substantially reduced computational cost. This should also reduce the risk of overfitting. Notably, the sparse connectivity reduces the computational cost of evaluating wave-function ratios, substantially reducing the computational complexity of variational and projective quantum Monte Carlo simulations. The main testbed we consider in this article is the random Ising chain, beyond the pure ferromagnetic model, which has been addressed also in previous studies. We analyze the accuracy of sparse RBMs as a function of the range of the allowed connections, exploring both short-range and midrange connectivities. We also compare sparse RBMs with their dense counterparts. Furthermore, the accuracy of PQMC simulations guided by optimized sparse RBMs is verified. The observables we analyze are the ground-state energy and the magnetization profile in setups with fixed boundary conditions.

Our analysis indicates that the local connectivity facilitates the training process, allowing sparse RBMs to approximate the ground state of the one-dimensional random spin model more accurately than dense RBMs with a comparable number of variational parameters. The PQMC simulations guided by optimized sparse RBMs provide unbiased results at a substantially reduced computational cost compared to the case of dense guiding functions. In particular, the magnetization profiles computed by the PQMC algorithm satisfy the Fisher–de Gennes scaling relation [52]. The scaled profiles agree with the predictions for conformally invariant systems [53], even for the (non-conformally-invariant) random model, indicating that the corrections to the conformal scaling are essentially negligible, as previously found also in Ref. [54]. The scaling dimension agrees with the results of the renormalization-group treatment [55].

The rest of the article is organized as follows: Section II introduces the model Hamiltonian as well as the dense and the sparse RBMs. The unsupervised-learning scheme and the PQMC algorithm are also briefly described. Our results for ground-state energies and magnetization profiles are reported in Sec. III. They are compared with Jordan-Wigner predictions and with the Fisher–de Gennes scaling relation, respectively. Our conclusions and some future perspectives are given in Sec. IV.

II. MODELS AND METHODS

The models we consider are defined by the following one-dimensional quantum Ising Hamiltonian:

$$\hat{H} = -\sum_{j=1}^{N} J_j \sigma_j^z \sigma_{j+1}^z - \Gamma \sum_{j=1}^{N} \sigma_j^x. \tag{1}$$

σ_j^x and σ_j^z are conventional Pauli matrices at the lattice sites $j = 1, \ldots, N$. N is the total number of spins, and we consider periodic boundary conditions, i.e., $\sigma_{N+1}^\alpha = \sigma_1^\alpha$, for $\alpha = x, z$. The parameters $J_j \geqslant 0$ fix the strength of the (ferromagnetic) interactions between the spins at the sites j and $j+1$. Γ is the intensity of the (uniform) transverse magnetic field. In the following, the eigenstates of the Pauli matrix σ_j^z with eigenvalues $x_j = \pm 1$ are denoted as $|x_j\rangle$. The quantum states of N spins $|\boldsymbol{x}\rangle = |x_1 x_2, \ldots, x_N\rangle$, with $\boldsymbol{x} = (x_1, \ldots, x_N)$, form the computational basis considered in this article. With $|\psi\rangle$ we denote the quantum state corresponding to the wave function $\langle \boldsymbol{x}|\psi\rangle = \psi(\boldsymbol{x})$.

If one chooses uniform couplings $J_j = J > 0$ for all $j = 1, \ldots, N$, the Hamiltonian (1) describes the pure ferromagnetic Ising chain. This model undergoes a quantum phase transition from a paramagnetic phase for $\Gamma > J$ to a ferromagnetic phase for $\Gamma < J$. Beyond the pure model, we consider random Ising chains with couplings J_j sampled from a probability distribution $\mathcal{P}(J_j)$. In general, one can always perform a gauge transformation in the Hamiltonian (1) such that all couplings J_j except one, at most, are non-negative. For sufficiently large system sizes, a single negative coupling does not change the bulk properties of the system. Therefore, without loss of generality, we may restrict ourselves to distributions with support on positive couplings. Specifically, we consider the flat distribution $\mathcal{P}(J) = \theta(J)\theta(1-J)$, where $\theta(x)$ is the unit step function: $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ otherwise. We consider also the binary distribution $\mathcal{P}(J) = \delta(J-2)/2 + \delta(J-1/2)/2$, where $\delta(x)$ is the Dirac delta function. In the random Ising chain, the ferromagnetic quantum phase transition occurs at the critical transverse field $\Gamma = \exp(\overline{\ln(J)})$ [56]; the overline indicates the average over the chosen probability distribution. In the case of the flat distribution described above, one obtains $\Gamma \cong 0.367\,92$, while for the binary distribution the critical point is $\Gamma = 1$.

A. Dense and sparse restricted Boltzmann machines

As first shown in Ref. [1], the ground-state wave function of a quantum spin Hamiltonian like (1) can be approximated using Boltzmann machines. Reference [1] addressed pure quantum spin models via standard dense RBMs. Here we address also random Ising models, using RBMs with dense and with sparse connectivities. Boltzmann machines are generative stochastic neural networks commonly employed for density estimation, i.e., to infer the probability distribution underlying a given (typically large) dataset of samples. RBMs are formed by the visible layer with the N spin variables $\boldsymbol{x}$, and by a layer of hidden neurons, which includes N_h additional spin variables $h_i = \pm 1$, with $i = 1, \ldots, N_h$. The set of hidden variables will be indicated as $\boldsymbol{h} = (h_1, \ldots, h_{N_h})$. The probability associated with each configuration $(\boldsymbol{x}, \boldsymbol{h})$ of the two-layer system is written in the form of the Boltzmann weight $P(\boldsymbol{x}, \boldsymbol{h})$

corresponding to a classical Ising Hamiltonian. RBMs are characterized by the absence of intralayer visible-visible and hidden-hidden interactions. In the standard dense RBMs, all visible spins interact with all hidden spins. The corresponding coupling parameters are denoted J_{ij}. In the sparse RBMs, instead, each visible spin j interacts only with a subset of hidden neurons. This subset is denoted in the following as $\mathcal{N}_j$. The classical Ising Hamiltonian associated with the sparse RBM reads

$$H_{\mathrm{RBM}}(\boldsymbol{x},\boldsymbol{h}) = -\sum_j \sum_{i\in\mathcal{N}_j} J_{ij} h_i x_j - \sum_j a_j x_j - \sum_i b_i h_i. \quad (2)$$

The parameters a_j and b_i, called biases, play the role of local longitudinal magnetic fields. Together with the couplings J_{ij}, they define the RBM. The parameters will be collectively indicated as $\mathbf{W} \equiv (\{J_{ij}\}, \{a_j\}, \{b_i\})$. The dense RBMs correspond to the choice $\mathcal{N}_j = \{1, \ldots, N_h\}$ for all $j = 1, \ldots, N$. In the sparse RBM considered in this article, the size of the hidden layer matches that of the visible layer, i.e., $N_h = N$. The visible spins are connected only to a number $N_c < N_h$ of nearest-neighbor hidden neurons, corresponding to the set of hidden-spin indices $\mathcal{N}_j = \{j - (N_c - 1)//2, \ldots, j - (N_c - 1)//2 + N_c - 1\}$, where the symbol // indicates integer division. Periodic boundary conditions are applied to the hidden-neuron indices, analogously to the visible-spin indices. For dense RBMs, one has $N \times N_h$ interlayer couplings J_{ij}, besides $N + N_h$ biases. For sparse RBMs, the number of interlayer couplings is $N \times N_c$ while the number of bias terms remains $N + N_h$.

The connectivity structures corresponding to the two architectures are visualized in Fig. 1. Notice that the connections across the periodic boundary are not shown. For comparison, it is also worth mentioning that in the case of the unrestricted Boltzmann machine considered in Ref. [17], intralayer visible-visible and hidden-hidden direct interactions are included, but they are limited to the nearest-neighbor couplings. The corresponding connectivity structure is also shown in Fig. 1. This structure is analogous to that of the shadow wave function introduced in Refs. [57,58] to describe the liquid and solid phases of helium-4.

Sparse RBMs with short-range interlayer connections have been previously considered also in Ref. [41]. In that reference, they have been show to be equivalent to a specific type of entangled plaquette states. Due to the (exponentially) more favorable scaling of the computation cost, they allow the use of larger plaquettes. [41]. It has also been shown that certain topological states, namely the toric code and the one-dimensional symmetry-protected topological cluster state, admit an exact representation via RBM with local inter-layer connections [40]. In practical applications, accurately approximating the ground-state wave function via dense RBMs requires a number of hidden units $N_h \gtrsim N$. Thus, the number of parameters to be optimized scales, to leading order (without counting the bias terms), at least as N^2. In clean periodic systems, translational invariance can be exploited to reduce this quadratic scaling to a linear scaling with system size [1,48]. However, this is not possible in disordered models, which are the main focus of this article. As shown in Sec. III, sparse RBMs can reach high accuracy already for $N_c \ll N$, implying a substantial reduction of the number of parameters to be optimized. It is worth mentioning that one could implement sparse RBMs with larger hidden layers $N_h > N$. This would provide the same flexibility one has with dense RBMs. In this respect, it is worth recalling that dense RBMs with an arbitrarily large number of hidden neurons can, in principle, approximate any binary distribution [59,60]. They share this propriety with deep neural networks with many hidden layers. It follows that a sparse RBM with sufficiently large N_h and N_c must also be a universal approximator. However, in practice the number of neurons in the unique hidden layer of an RBM might have to increase exponentially with the system size [61], making the training process and RBM-based Monte Carlo simulations computationally impractical. This highlights the importance of identifying RBM architectures that reach a sufficient accuracy with a limited number of parameters.

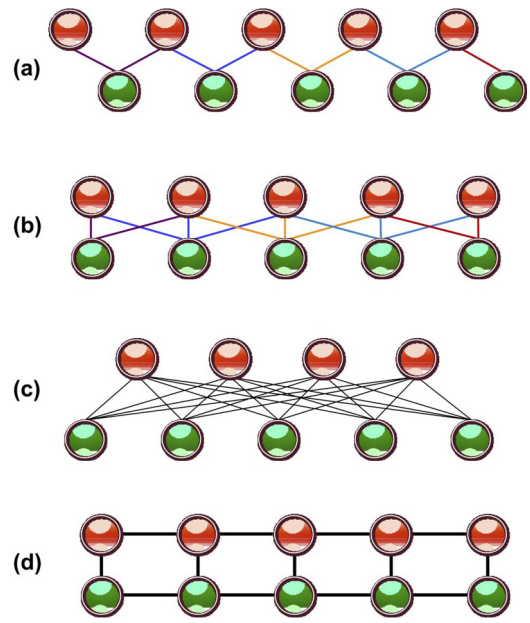

FIG. 1. Representation of the connectivity structures of four different artificial neural networks. Segments indicate the allowed interactions. Diagrams (a) and (b) represent examples of the sparse restricted Boltzmann machine employed in this article, whereby visible spins interact with a limited number of local hidden neurons; in the cases visualized here, this number is $N_c = 2$ (a) and $N_c = 3$ (b). Diagram (c) represents the standard dense restricted Boltzmann machine with all-to-all inter-layer connectivity [1]. Diagram (d) represents a type of unrestricted Boltzmann machine [17], alias a shadow wave function [57,58]. This is characterized by interlayer connections only among spins corresponding to the same index, plus nearest-neighbor intralayer connections.

For the RBM, the probability associated with the visible-spin configuration $\boldsymbol{x}$ is computed as the marginal distribution over all possible hidden-spin configurations $\boldsymbol{h}$:

$$P(\boldsymbol{x}) = \sum_{\boldsymbol{h}} P(\boldsymbol{x},\boldsymbol{h}) = \frac{1}{Z}\sum_{\boldsymbol{h}} \exp\left[-H_{\mathrm{RBM}}(\boldsymbol{x},\boldsymbol{h})\right]. \quad (3)$$

The normalization factor $Z = \sum_{\boldsymbol{x},\boldsymbol{h}} \exp\left[-H_{\mathrm{RBM}}(\boldsymbol{x},\boldsymbol{h})\right]$ is the partition function. The absence of intralayer interactions allows one to analytically trace out the hidden-spin

configurations, resulting in the (unnormalized) marginal distribution $P(\boldsymbol{x}) \propto \exp\left(\sum_j a_j x_j\right) \prod_i F_i(\boldsymbol{x})$, where $F_i(\boldsymbol{x}) = 2\cosh\left[b_i + \sum_{j\in\mathcal{N}_i} J_{ij} x_j\right]$, and $\mathcal{N}_i$ indicates the set of visible-spin indices j connected to the hidden neuron i. For the dense RBM, this set is $\mathcal{N}_i = \{1, \ldots, N\}$ for any i. For the sparse RBM we implement, one has $\mathcal{N}_i = \{i - (N_c//2), \ldots, i - (N_c//2) + N_c - 1\}$. As proposed in Ref. [1], the function $P(\boldsymbol{x})$ can be used to define an (unnormalized) ground-state wave function, $\psi_{\mathrm{RBM}}(\boldsymbol{x}) \propto P(\boldsymbol{x})$.

In the general case, the RBM parameters $\mathbf{W}$ should be complex-valued in order to describe both amplitude and phase. Otherwise, one could consider two real-valued RBMs, one for the amplitude and the other for the phase. In this article, we consider only models whose ground-state wave function can be assumed to be real and non-negative in the chosen computational basis. Therefore, the RBM parameters can be restricted to real values. The RBM could be trained via the variational minimization of the energy expectation value for the state $|\psi_{\mathrm{RBM}}\rangle$ [1]. This procedure is analogous to the reinforcement learning schemes employed in the field of machine learning, with the (negative) energy expectation value playing the role of a reward function. In many machine-learning applications of RBMs, the parameters $\mathbf{W}$ are determined instead via unsupervised machine-learning [2]. In these schemes, the RBM is trained to mimic the unknown distribution underlying a dataset of stochastic samples, as mentioned above. To generate the training dataset, we use PQMC algorithms, as explained in the following.

B. Projective Quantum Monte Carlo simulations

PQMC algorithms are designed to simulate the ground state of quantum systems by stochastically evolving the imaginary-time Schrödinger equation. In the practical implementation, the algorithm efficiency is boosted by introducing a suitable *Ansatz* for the ground-state wave function. This is usually referred to as a guiding function. It is indicated in the following as $\psi_G(\boldsymbol{x})$. In the PQMC simulation, one allows the product $f(\boldsymbol{x}, \tau) = \psi(\boldsymbol{x}, \tau)\psi_G(\boldsymbol{x})$ to evolve, where $\psi(\boldsymbol{x}, \tau)$ is the time-dependent wave function, according to the modified imaginary-time Schrödinger equation:

$$f(\boldsymbol{x}, \tau + \Delta\tau) = \sum_{\boldsymbol{x}'} \tilde{G}(\boldsymbol{x}, \boldsymbol{x}', \Delta\tau) f(\boldsymbol{x}', \tau). \tag{4}$$

Here, $\tilde{G}(\boldsymbol{x}, \boldsymbol{x}', \Delta\tau) = G(\boldsymbol{x}, \boldsymbol{x}', \Delta\tau)\frac{\psi_G(\boldsymbol{x})}{\psi_G(\boldsymbol{x}')}$, where $G(\boldsymbol{x}, \boldsymbol{x}', \Delta\tau) = \langle \boldsymbol{x}| \exp\left[-\Delta\tau(\hat{H} - E_{\mathrm{ref}})\right]|\boldsymbol{x}'\rangle$ is the imaginary-time Green's function for a time step $\Delta\tau$, E_{ref} is a reference energy introduced to stabilize the dynamics, and we set $\hbar = 1$ throughout this article.

The dynamics is simulated by stochastically evolving a large population of system configurations, usually called random walkers. These random walkers are subjected to stochastic configuration updates and to a branching process whereby they are replicated or annihilated. Specifically, we employ the continuous-time algorithm of Refs. [17,62,63]. This avoids finite time-step errors. To sample the configuration updates, one has to compute the ratio of guiding-function values $\frac{\psi_G(\boldsymbol{x})}{\psi_G(\boldsymbol{x}')}$ for all possible transitions $\boldsymbol{x}' \to \boldsymbol{x}$ induced by the Hamiltonian. In the case of (1), the allowed transitions correspond to the N possible spin flips. If an RBM wave function is chosen as a guiding function, i.e., if $\psi_G(\boldsymbol{x}) = \psi_{\mathrm{RBM}}(\boldsymbol{x})$, the ratio for the flip of spin j is evaluated as

$$\frac{\psi_G(\boldsymbol{x})}{\psi_G(\boldsymbol{x}')} = \exp(-2a_j x_j') \prod_{i\in\mathcal{N}_j} \frac{F_i(\boldsymbol{x})}{F_i(\boldsymbol{x}')}, \tag{5}$$

where $\boldsymbol{x} = \{x_1', \ldots, -x_j', \ldots, x_N'\}$. Due to the reduced number of factors, sparse RBMs with $N_c \ll N_h$ lead to a substantial reduction of the computational cost of evaluating Eq. (5). The evaluation of wave-function ratios represents the computational bottleneck of both variational and projective quantum Monte Carlo simulations. Therefore, by using sparse RBMs instead of dense RBMs, one obtains a substantial reduction in the computational complexity. Specifically, one obtains a speed-up of order N_h/N_c. It is worth mentioning that for both dense and sparse RBMs, the computational cost of evaluating the arguments $b_i + \sum_{j\in\mathcal{N}_i} J_{ij} x_j$ of the cosh function appearing in $F_i(\boldsymbol{x})$ can be substantially reduced by bookkeeping their values, updating them only when a visible-spin flip is performed.

To implement the branching process, and for the tuning of E_{ref}, we use the textbook recipe of Ref. [64]. The tuning of E_{ref} is designed to stabilize the average number of random walkers close to a target value N_{w}. In the long imaginary-time limit, the random walkers sample spin configurations with a probability distribution proportional to $f(\boldsymbol{x}, \tau \to \infty) \cong \psi_G(\boldsymbol{x})\psi_0(\boldsymbol{x})$, where $\psi_0(\boldsymbol{x})$ is the ground-state wave function. In the large random-walker population limit, $N_{\mathrm{w}} \to \infty$, the above approximate identity becomes exact. Therefore, one obtains unbiased estimates of the ground-state energy E via Monte Carlo integration of the following formula:

$$E = \frac{\sum_{\boldsymbol{x}} f(\boldsymbol{x}, \tau \to \infty) E_{\mathrm{loc}}(\boldsymbol{x})}{\sum_{\boldsymbol{x}} f(\boldsymbol{x}, \tau \to \infty)}, \tag{6}$$

where the local energy is $E_{\mathrm{loc}}(\boldsymbol{x}') = \sum_{\boldsymbol{x}} H_{\boldsymbol{x}',\boldsymbol{x}} \frac{\psi_G(\boldsymbol{x})}{\psi_G(\boldsymbol{x}')}$, with $H_{\boldsymbol{x}',\boldsymbol{x}} = \langle \boldsymbol{x}|\hat{H}|\boldsymbol{x}'\rangle$. Analogous formulas can be written for other operators that commute with the Hamiltonian. The forward-walking technique provides unbiased estimates also for other operators with a diagonal representation in the chosen computational basis (see, e.g., Ref. [65]). For finite N_{w} a systematic bias might occur [45–47,66–68]. This is known to be the major deficiency of PQMC algorithms. In fact, if one does not introduce a guiding function, i.e., if one sets $\psi_G(\boldsymbol{x}) = 1$, the random-walker population N_{w} required to keep this systematic bias below a given threshold increases exponentially with the system size [47]. If $\psi_G(\boldsymbol{x})$ is a reasonably good approximation for the ground-state wave function $\psi_0(\boldsymbol{x})$, the convergence to the exact $N_{\mathrm{w}} \to \infty$ limit is drastically accelerated [17]. This allows one to reduce the bias to negligible values, thus obtaining unbiased predictions with a computationally affordable random-walker population. As shown in Ref. [9] and briefly summarized below, one can use configurations sampled from PQMC simulations to train RBM wave functions in an unsupervised learning scheme. The trained RBMs can be used in turn as a guiding function for a subsequent PQMC run, eliminating the finite N_{w} bias. This scheme is briefly described in Sec. II D.

C. Unsupervised learning for RBMs

In unsupervised learning, the RBM is trained by maximizing the log-likelihood $L(\mathbf{W}) = \sum_l \ln P(\boldsymbol{x}_l)$ of a training set; the index l labels the instances in the training set. This is equivalent to the minimization of the Kullback-Leibler divergence [4]. For two generic distributions $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$, the Kullback-Leibler divergence reads

$$\mathrm{KL}(q||p) = \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})}. \tag{7}$$

In the case discussed here, $q(\boldsymbol{x})$ is identified with the distribution of the random walkers $f(\boldsymbol{x}, \tau \to \infty)$ obtained via PQMC simulations, while $p(\boldsymbol{x})$ corresponds to the RBM marginal distribution $P(\boldsymbol{x})$. The optimization of the RBM parameters $\mathbf{W}$ can be performed using the stochastic gradient ascent algorithm. One performs many updates of the RBM parameters following the log-likelihood gradient computed on small, randomly sampled minibatches of training instances. The gradients are computed via the k-step contrastive divergence algorithm [69]. This involves k iterations of alternated Gibbs sampling of hidden and visible variables, starting from the visible values of the training instances. In this article, the plain vanilla stochastic gradient ascent algorithm is augmented only by adding a momentum term proportional to the gradient computed at the previous update. Also learning-rate annealing is adopted, as explained below. All details of this algorithm are described in Ref. [4]. Notice that, for the values of the binary spin-variables, we follow the physics convention $x_j = \pm 1$ and $h_i = \pm 1$, instead of the convention $x_j = \{0, 1\}$ and $h_i = \{0, 1\}$, which is more common in the machine-learning literature. The modified algorithm corresponding to the physics notation is provided in Ref. [9]. See also Refs. [7,8]. The use of sparse RBMs implies simple algorithmic modifications with respect to the standard dense models. In particular, one has to compute only $N \times N_c$ gradients with respect to the coupling parameters, as opposed to $N \times N_h$. Gibbs sampling is performed with the following binary probability distributions:

$$p_{h_i=1}(\boldsymbol{x}) = \frac{1}{1 + \exp\big(-2\sum_{j\in\mathcal{N}_i} x_j J_{ij} - 2b_i\big)}$$

and

$$p_{x_j=1}(\boldsymbol{h}) = \frac{1}{1 + \exp\big(-2\sum_{i\in\mathcal{N}_j} h_i J_{ij} - 2a_j\big)}$$

for hidden and visible variables, respectively. Clearly, $p_{h_i=-1}(\boldsymbol{x}) = 1 - p_{h_i=1}(\boldsymbol{x})$ and $p_{x_j=-1}(\boldsymbol{h}) = 1 - p_{x_j=1}(\boldsymbol{h})$.

D. Self-learning PQMC algorithm

Reference [9] presented a self-learning PQMC algorithm whereby an unsupervised learning scheme is used to train the RBM. Specifically, the RBM learns the probability distribution corresponding to a large random-walker population produced by a PQMC simulation after a long imaginary-time projection. The self-learning protocol is actually implemented by performing several consecutive PQMC runs, hereby labeled by the index s, for an imaginary-time τ_s. Each run is guided by a different guiding function $\psi_{G_s}(\boldsymbol{x})$, and it is followed by an unsupervised learning stage that provides an RBM such that $P_s(\boldsymbol{x}) \propto \psi_{G_s}(\boldsymbol{x})\psi_0(\boldsymbol{x})$. The guiding function for the next PQMC run is $\psi_{G_{s+1}}(\boldsymbol{x}) = \sqrt{P_s(\boldsymbol{x})}$. For the initial run, a guiding function based on the square root of an RBM with random parameters is used. The use of the square root of the RBM probability distribution implies straightforward modifications to the formulas for the guiding-function ratios Eq. (5) and for $E_{\mathrm{loc}}(\boldsymbol{x})$. If N_{w} and τ_s are large enough, and if the training of the RBM succeeds, one has a fast convergence $\psi_{G_s}(\boldsymbol{x}) \to \psi_0(\boldsymbol{x})$ for $s \to \infty$ [9]. Even if these assumptions are not exactly satisfied, this self-learning scheme provides remarkably accurate approximations for the ground-state wave function. This accuracy is quantified in Sec. III by comparing the energy expectation value corresponding to the optimized RBM guiding function, defined as

$$E = \frac{\langle \psi_{G_{s\to\infty}} | \hat{H} | \psi_{G_{s\to\infty}} \rangle}{\langle \psi_{G_{s\to\infty}} | \psi_{G_{s\to\infty}} \rangle}, \tag{8}$$

with the exact ground-state energy. These expectation values can be determined via standard Monte Carlo integration performed with the single spin-flip METROPOLIS algorithm.

Even more importantly, the RBM guiding function turns out to be sufficiently accurate to eliminate the possible bias in the PQMC results due to the finite N_{w}. This is demonstrated by the data reported in Sec. III.

E. Simulation details

The results reported in Sec. III are obtained with the following simulation parameters. The target random-walker population is $N_{\mathrm{w}} = 10^4$. The time step is $\Delta\tau = 0.04$.

The final imaginary time of each PQMC run is $\tau_s = 20$. The training dataset is accumulated by storing $N_{\mathrm{w}}/20$ randomly selected walkers at each imaginary-time step, excluding the initial time segment of each PQMC run corresponding to $\tau \in [20s, 20s + 8]$. The number of PQMC runs ranges from 20 to 50.

In each unsupervised-learning stage, the number of stochastic gradient ascent steps is $N_{\mathrm{steps}} = 5 \times 10^4$. The minibatch size is $N_b = 20$. The learning rate η is kept fixed within each learning stage, but it is annealed stage after stage, following the simple empirical protocol $\eta(s) = \eta_0 c^s$; the annealing rate is between $c = 0.65$ and 0.85, and the learning rate at the first learning training stage $s = 0$ is $\eta_0 = 0.01$. The coefficient of the momentum term is $\nu = \eta/10$. The one-step contrastive divergence algorithm, corresponding to $k = 1$, is found to suffice. The initial guiding function $\psi_{G_{s=0}}(\boldsymbol{x})$ is the square root of an RBM with random couplings sampled from a flat distribution in the range $J_{ij} \in [-0.025 : 0.025]$. For the simulations performed in the absence of a longitudinal magnetic field at the boundaries, the bias terms a_j and b_i are initialized to zero, and they are not updated during the training processes. This choice reflects the spin-flip symmetry of the ground state. For the simulations performed with a nonzero longitudinal field at the boundaries, the bias terms are initialized to random values with a flat distribution $a_j, b_i \in [0 : 0.05]$. They are then optimized during the training stage.

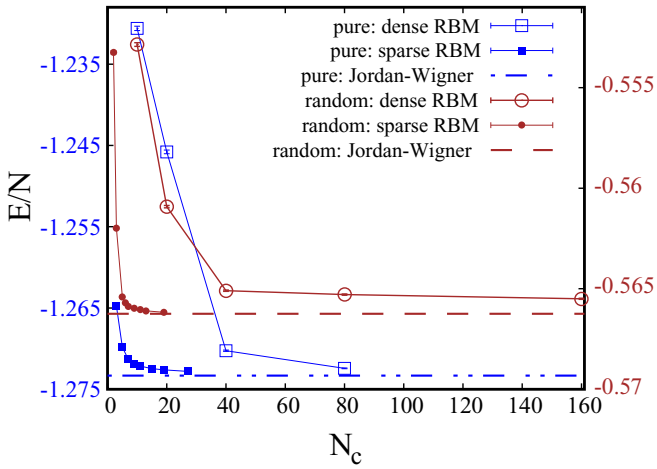

FIG. 2. Energy per spin E/N corresponding to two optimized neural-network *Ansätzes* as a function of the number of connected hidden spins N_c. For the dense RBMs, which are represented by the empty symbols, N_c coincides with the number of hidden spins N_h. The full symbols correspond to the sparse RBMs, for which N_c is the number of allowed local connections. The (blue) squares represent the results for the clean ferromagnetic Ising model, and they refer to the left vertical axis. The (brown) circles correspond to one realization of the random Ising model with couplings sampled from the flat distribution (see text), and they refer to the right vertical axis. Both models include $N = 80$ spins and they are simulated at the coupling strength corresponding to the respective ferromagnetic quantum critical points. The horizontal lines indicate the exact ground-state energies computed through the Jordan-Wigner transformation for $N = 80$.

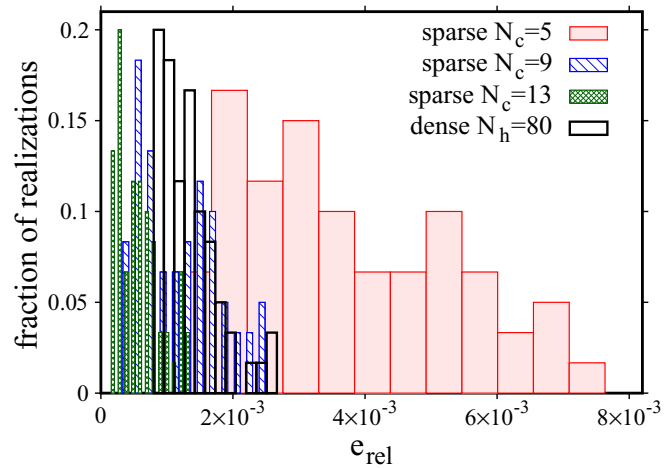

FIG. 3. Histogram of the relative errors $e_{\rm rel} = (E - E_{\rm JW})/|E_{\rm JW}|$ of the ground-state energies E corresponding to the optimized RBM *Ansätzes*, with respect to the (exact) Jordan-Wigner result $E_{\rm JW}$. A total of 60 realizations of the random couplings are considered. The transverse field intensity is set at the ferromagnetic quantum critical point. The full columns correspond to three sparse RBMs different connection numbers N_c. The empty columns correspond to a dense RBM with $N_h = 80$ hidden neurons. The width of some columns is reduced for visibility.

III. RESULTS

To analyze the accuracy of the optimized RBM wave functions, we compare the corresponding energy expectation value (8) with the exact ground-state energy. The comparison is performed for the pure and the random ferromagnetic quantum Ising chains. In the latter case, the couplings are sampled from the flat distribution defined at the beginning of Sec. II. The exact ground-state energies, hereafter denoted with $E_{\rm JW}$, are determined via the Jordan-Wigner transformation. We follow the formalisms reported in Refs. [70,71] and in Ref. [72] for the pure and the random models, respectively. The comparison for the system size $N = 80$ is shown in Fig. 2. This analysis is performed at the respective quantum critical points for the two models (see Sec. II). The motivation is that, in this regime, it is more challenging to approximate ground-state wave functions with neural-network *Ansätzes* [1,17,41,73].

For the pure model, the dense RBM reaches a relative error $e_{\rm rel} = (E - E_{\rm JW})/|E_{\rm JW}| \simeq 7 \times 10^{-4}$ with $N_h = 80$ hidden neurons. The sparse RBM reaches a similar relative error with just $N_c = 15$ allowed connections per visible spins, while with $N_c = 27$ one has $e_{\rm rel} \simeq 4 \times 10^{-4}$. For the specific realization of the random chain considered in Fig. 2, the dense RBM is not particularly accurate, reaching $e_{\rm rel} \simeq 1.3 \times 10^{-3}$ with $N_h = 160$. Training even larger dense RBMs becomes computationally challenging. Instead, the sparse RBM with $N_c = 19$ provides $e_{\rm rel} \simeq 1.3 \times 10^{-4}$, corresponding to an accuracy improvement of one order of magnitude. These findings suggest that sparse RBMs should be preferred to the standard dense RBMs since for pure models they achieve comparable accuracy at a reduced computational cost and, most importantly, they outperform dense RBMs for random models. This is due to the better optimization process that can be attained with a reduced number of variational parameters.

With disordered systems, it is crucial to verify that high accuracy can be achieved with any realization of the random couplings. To verify this, we analyze in Fig. 3 the histogram of the relative errors for 60 realizations. The four datasets correspond to three sparse RBMs with different N_c, and to one dense RBM with $N_h = 80$. For the sparse RBMs, one observes a systematic improvement of the accuracy with increasing N_c for both average and maximum errors. For the dense RBM, the average relative error is $\overline{e}_{\rm rel} = 1.4(4) \times 10^{-3}$ (the standard deviation is in parentheses), which is comparable to the result for the sparse RBM with $N_c = 9$, namely $\overline{e}_{\rm rel} = 1.2(6) \times 10^{-3}$. This is remarkable, given that the sparse RBM has only a fraction N_c/N_h of the variational parameters included in the dense RBM. This confirms that, at least in one dimension, sparse RBMs are particularly suitable to describe random spin systems.

One of our main goals is to obtain sufficiently accurate guiding functions to eliminate the bias in the PQMC simulations due to the finite random-walker population. Figure 4 shows the histogram of the energy predictions from PQMC simulations guided by the optimized sparse RBMs. These predictions are computed via Monte Carlo integration of Eq. (6). Again, 60 realizations of the random Ising chain are considered, and the analysis is performed at criticality, since this is the regime where eliminating the bias requires larger random-walker populations [17,47]. The PQMC results agree precisely with the Jordan-Wigner predictions, displaying only

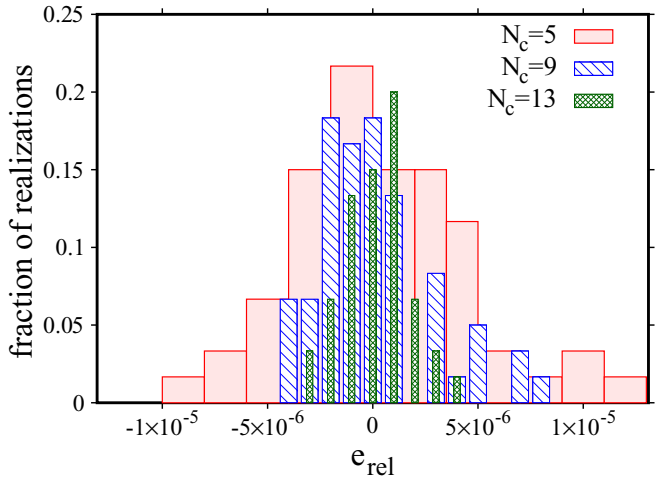

FIG. 4. Histogram of the relative errors $e_{\rm rel}$ of the ground-state energies E predicted by projective quantum Monte Carlo simulations guided by optimized sparse RBMs. A total of 60 realizations of the random couplings of the quantum Ising chain are considered. The system parameters are as in Fig. 3. The three datasets correspond to different connection numbers N_c. The width of some columns is reduced for visibility.

random fluctuations around $e_{\rm rel} = 0$. These fluctuations are compatible with the statistical uncertainty of the Monte Carlo integration (not shown in the figure). As expected, these fluctuations decrease for increasing N_c, since the guiding function becomes more accurate. This reduces the fluctuations in the local energy $E_{\rm loc}(\boldsymbol{x})$ and, therefore, in the result of the Monte Carlo integration.

As discussed above, the sparse RBMs allow us to perform efficient PQMC simulations of random and, more generally, of inhomogeneous quantum spin systems. As a further showcase, we consider the inhomogeneity due to fixed boundary conditions. In this setup, the average magnetizations of the extremal spins are fixed at their maximum values, i.e., $m(j) \equiv \langle\psi_0|\sigma_j^z|\psi_0\rangle/\langle\psi_0|\psi_0\rangle = 1$ for $j = 1$ and $j = N$. The fixed boundary conditions are conveniently implemented by applying strong longitudinal magnetic fields on the extremal spins. Specifically, we add to the Hamiltonian (1) a border term $-\lambda(\sigma_1^z + \sigma_N^z)$ with $\lambda = 500$. Notice that the fixed boundary conditions break the spin-flip symmetry of the Hamiltonian. We determine via PQMC simulations the average magnetization profiles $m(j)$ of the pure ferromagnetic model and of a random chain with couplings sampled from the binary distribution defined in Sec. II. The guiding function is the optimized sparse RBM with $N_c = 9$. The PQMC predictions are compared with the Fisher–de Gennes scaling theory $m(j) = N^{-\chi} f(j/N)$, where χ is the scaling dimension of the magnetization operator and $f(j/N)$ is its scaling function. For conformally invariant two-dimensional models (that is, 1+1 in the present quantum model), and assuming equal boundary conditions at the two extremes, as we do here, the scaling relation can be written as

$$m(j) = A\left[\frac{N}{\pi}\sin\left(\pi\frac{j}{N}\right)\right]^{-\chi}, \tag{9}$$

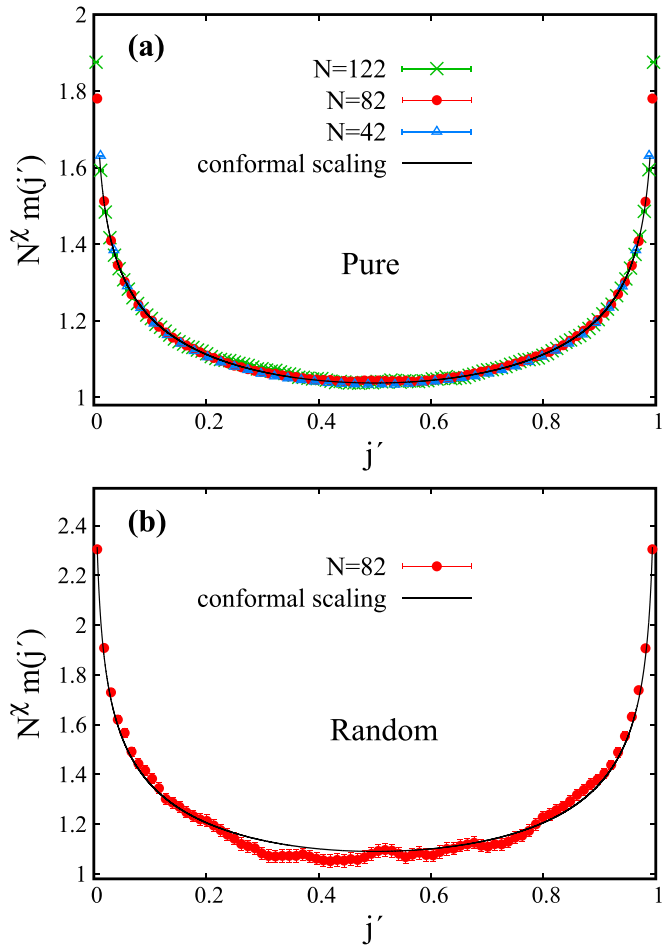

FIG. 5. Profile of the rescaled local magnetization $N^\chi m(j)$ as a function of the rescaled spin index $j\prime = (j - 1/2)/N$, where $j = 1, \ldots, N$ is the site index and N is the system size. $m(j)$ is the average magnetization at the site j, and χ is the scaling dimension of the magnetization. The different symbols represent the PQMC predictions for different N. The system has fixed boundary conditions $m(1) = m(N) = 1$. The transverse field intensity corresponds to the ferromagnetic quantum critical point. Panel (a) corresponds to a pure ferromagnetic Ising model. Panel (b) corresponds to a random Ising chain, with the couplings sampled from a binary distribution (see the text). The results are averaged over 1200 realizations. The continuous (black) curves represent fitting functions based on the Fisher–de Gennes scaling relation for conformally invariant models; see Eq. (9). The fitted scaling dimensions are consistent with the expected results, $x_m = 1/8$ and $x_m = (3 - \sqrt{5})/4$, for the pure and the random models, respectively (see the text).

where A is a nonuniversal prefactor [53]. More generally, Eq. (9) represents the first term of a Fourier expansion [54]. In the upper panel of Fig. 5, the PQMC prediction for the pure model is compared with the conformal-invariance scaling, considering three different system sizes. The agreement is very precise already for the intermediate size we consider, namely $N = 82$. This indicates that, for comparable or even larger sizes, possible corrections to the universal scaling due to finite-size effects are negligible. By using the scaling dimension χ as a fitting parameter on the $m(j)$ data for $N = 122$, besides the prefactor A, we obtain $\chi = 0.131(5)$,

in agreement with the expected result $\chi = 1/8$. The error bar takes into account the fluctuations observed when removing up to three extremal points from the fit. The agreement with the conformal-invariance scaling is expected, since the pure model is conformally invariant. Instead, the random Ising chain breaks conformal invariance. However, it was found in Ref. [54] that the Fourier terms beyond the first one provide almost negligible contributions. In the lower panel of Fig. 5, the conformal scaling is compared with the PQMC results for the random model. To reduce the fluctuations originating from the large variations of the magnetization profiles for different realizations of the random couplings, the PQMC data are averaged over as many as 1200 instances. Only the system size $N = 82$ is considered, since this is expected to suffice to essentially eliminate finite-size corrections, as demonstrated above for the pure model. We find agreement with the conformal scaling, thus confirming the finding of Ref. [54]. By fitting the conformal scaling relations to the PQMC results, we obtain $\chi = 0.19(1)$, in agreement with the renormalization-group prediction for random Ising chains, $\chi = (3 - \sqrt{5})/4 \cong 0.190\,98$ [55,74].

IV. CONCLUSIONS

We have investigated the use of artificial neural networks to approximate the ground-state wave function of one-dimensional disordered quantum spin models. Our focus was on the feasibility of reaching high accuracy and on the computational efficiency of the neural-network models. Both issues are substantially more relevant in quantum many-body physics compared to the typical applications of neural networks in computer science and in engineering. We have found that restricted Boltzmann machines with a local sparse connectivity reach higher accuracy, when trained via unsupervised learning, compared to the standard dense RBMs with all-to-all interlayer connectivity. This is a consequential finding that highlights the crucial role of the connectivity structure of the neural-network wave functions. The reduced connectivity of the sparse RBMs we implemented leads to a linear scaling with system size of the number of optimizable parameters. This has to be compared to the generally quadratic scaling of dense RBMs. The sparse connectivity facilitates the training process, allowing the RBM model to better approximate the ground state, and it also accelerates the Monte Carlo simulations of RBMs, since the computational cost to evaluate wave-function ratios is reduced compared to the case of dense connectivity. The optimized RBMs can also be used as guiding functions for projective quantum Monte Carlo simulations. In particular, we have shown that they allow one to completely eliminate the bias due to the finite random walker population in disordered spin systems. This possible bias is a weakness of the PQMC algorithms [75,76], and the lack of appropriate guiding functions has so-far limited the scope of their application in this field. Notably, the sparse connectivity accelerates also the PQMC simulations.

As a future perspective, sparse RBMs could find use in studies of combinatorial optimization problems. In that context, projective quantum Monte Carlo algorithms have emerged as a stringent benchmark for physical quantum annealers [47,77–79]. However, the lack of guiding functions appropriate for the typical instances of complex optimization problems, which can be mapped to spin-glass models, has limited their success [80]. Further studies should be performed to test the accuracy of sparse RBMs in more challenging models than the random Ising chains considered here. A relevant testbed would be higher-dimensional spin Hamiltonians, possibly with long-range interactions or frustrated interactions. In particular, frustration effects have been shown to negatively affect the generalization properties of dense RBMs [81]. In this context, it would also be interesting to investigate different sparse connectivities inspired by the theory of complex networks. Interesting candidates are RBMs with scale-free or small-world topologies [51]. We leave these endeavors to future studies.

ACKNOWLEDGMENTS

S.P. and P.P. acknowledge financial support from the FAR2018 project titled "Supervised machine learning for quantum matter and computational docking" of the University of Camerino and from the Italian MIUR under the project PRIN2017 CEnTraL 20172H2SC4. S.P. also acknowledges the CINECA award under the ISCRA initiative for the availability of high-performance computing resources and support.

[1] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[2] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, A learning algorithm for Boltzmann machines, Cogn. Sci. **9**, 147 (1985).

[3] G. E. Hinton, Training products of experts by minimizing contrastive divergence, Neural Comput. **14**, 1771 (2002).

[4] A. Fischer and C. Igel, An introduction to restricted Boltzmann machines, in *Iberoamerican Congress on Pattern Recognition* (Springer, Berlin, 2012), pp. 14–36.

[5] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac, Restricted Boltzmann machines in quantum physics, Nat. Phys. **15**, 887 (2019).

[6] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).

[7] G. Torlai and R. G. Melko, Machine learning quantum states in the NISQ era, Annu. Rev. Condens. Matter Phys. **11**, 325 (2020).

[8] L. Huang and L. Wang, Accelerated Monte Carlo simulations with restricted Boltzmann machines, Phys. Rev. B **95**, 035105 (2017).

[9] S. Pilati, E. M. Inack, and P. Pieri, Self-learning projective quantum Monte Carlo simulations guided by restricted Boltzmann machines, Phys. Rev. E **100**, 043301 (2019).

[10] J. Gomes, K. A. McKiernan, P. Eastman, and V. S. Pande, Classical quantum optimization with neural network quantum states, arXiv:1910.10675.

[11] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Symmetries and Many-Body Excitations with Neural-Network Quantum States, Phys. Rev. Lett. **121**, 167204 (2018).

[12] A. Nagy and V. Savona, Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems, Phys. Rev. Lett. **122**, 250501 (2019).

[13] M. J. Hartmann and G. Carleo, Neural-Network Approach to Dissipative Quantum Many-Body Dynamics, Phys. Rev. Lett. **122**, 250502 (2019).

[14] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Variational Neural-Network Ansatz for Steady States in Open Quantum Systems, Phys. Rev. Lett. **122**, 250503 (2019).

[15] N. Yoshioka and R. Hamazaki, Constructing neural stationary states for open quantum many-body systems, Phys. Rev. B **99**, 214306 (2019).

[16] T. Vieijra, C. Casert, J. Nys, W. De Neve, J. Haegeman, J. Ryckebusch, and F. Verstraete, Restricted Boltzmann Machines for Quantum States with Non-Abelian or Anyonic Symmetries, Phys. Rev. Lett. **124**, 097201 (2020).

[17] E. M. Inack, G. E. Santoro, L. Dell'Anna, and S. Pilati, Projective quantum Monte Carlo simulations guided by unrestricted neural network states, Phys. Rev. B **98**, 235145 (2018).

[18] G. Carleo, Y. Nomura, and M. Imada, Constructing exact representations of quantum many-body systems with deep neural networks, Nat. Commun. **9**, 5322 (2018).

[19] I. Lagaris, A. Likas, and D. Fotiadis, Artificial neural network methods in quantum mechanics, Comput. Phys. Commun. **104**, 1 (1997).

[20] H. Saito, Solving the Bose-Hubbard model with machine learning, J. Phys. Soc. Jpn. **86**, 093001 (2017).

[21] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. **8**, 662 (2017).

[22] Z. Cai and J. Liu, Approximating quantum many-body wave functions using artificial neural networks, Phys. Rev. B **97**, 035116 (2018).

[23] H. Saito and M. Kato, Machine learning technique to find quantum many-body ground states of bosons on a lattice, J. Phys. Soc. Jpn. **87**, 014001 (2018).

[24] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, Solving frustrated quantum many-particle models with convolutional neural networks, Phys. Rev. B **98**, 104426 (2018).

[25] K. Choo, T. Neupert, and G. Carleo, Two-dimensional frustrated J_1-J_2 model with neural network quantum states, Phys. Rev. B **100**, 125124 (2019).

[26] L. Pastori, R. Kaubruegger, and J. C. Budich, Generalized transfer matrix states from artificial neural networks, Phys. Rev. B **99**, 165123 (2019).

[27] D. Luo and B. K. Clark, Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions, Phys. Rev. Lett. **122**, 226401 (2019).

[28] F. Ferrari, F. Becca, and J. Carrasquilla, Neural Gutzwiller-projected variational wave functions, Phys. Rev. B **100**, 125131 (2019).

[29] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems, Phys. Rev. Lett. **124**, 020503 (2020).

[30] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, Phys. Rev. Research **2**, 023358 (2020).

[31] C. Roth, Iterative retraining of quantum spin models using recurrent neural networks, arXiv:2003.06228.

[32] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, arXiv:2002.04613.

[33] T. Duric and T. Seva, Efficient neural-network based variational Monte Carlo scheme for direct optimization of excited energy states in frustrated quantum systems, arXiv:2002.09246.

[34] H. Saito, Method to solve quantum few-body problems with artificial neural networks, J. Phys. Soc. Jpn. **87**, 074002 (2018).

[35] M. Ruggeri, S. Moroni, and M. Holzmann, Nonlinear Network Description for Many-Body Quantum Systems in Continuous Space, Phys. Rev. Lett. **120**, 205302 (2018).

[36] J. Han, L. Zhang, and E. Weinan, Solving many-electron Schrödinger equation using deep neural networks, J. Comput. Phys. **399**, 108929 (2019).

[37] D. Pfau, J. S. Spencer, A. G. d. G. Matthews, and W. M. C. Foulkes, *Ab-Initio* solution of the many-electron Schrödinger equation with deep neural networks, arXiv:1909.02487.

[38] J. Hermann, Z. Schätzle, and F. Noé, Deep neural network solution of the electronic Schrödinger equation, arXiv:1909.08423.

[39] J. Kessler, F. Calcavecchia, and T. D. Kühne, Artificial neural networks as trial wave functions for quantum Monte Carlo, arXiv:1904.10251.

[40] D.-L. Deng, X. Li, and S. Das Sarma, Machine learning topological states, Phys. Rev. B **96**, 195145 (2017).

[41] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Neural-Network Quantum States, String-Bond States, and Chiral Topological States, Phys. Rev. X **8**, 011006 (2018).

[42] Y. Zheng, H. He, N. Regnault, and B. A. Bernevig, Restricted Boltzmann machines and matrix product states of one-dimensional translationally invariant stabilizer codes, Phys. Rev. B **99**, 155129 (2019).

[43] J. W. T. Keeble and A. Rios, Machine learning the deuteron, arXiv:1911.13092.

[44] D. Kochkov and B. K. Clark, Variational optimization in the AI era: Computational graph states and supervised wave-function optimization, arXiv:1811.12423.

[45] J. H. Hetherington, Observations on the statistical iteration of matrices, Phys. Rev. A **30**, 2713 (1984).

[46] M. Boninsegni and S. Moroni, Population size bias in diffusion Monte Carlo, Phys. Rev. E **86**, 056712 (2012).

[47] E. M. Inack, G. Giudici, T. Parolini, G. Santoro, and S. Pilati, Understanding quantum tunneling using diffusion Monte Carlo simulations, Phys. Rev. A **97**, 032307 (2018).

[48] K. Sohn and H. Lee, Learning invariant representations with local transformations, in *Proceedings of the 29th International Conference on International Conference on Machine Learning* (Omnipress, Madison, 2012), pp. 1339–1346.

[49] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, Equivalence of restricted Boltzmann machines and tensor network states, Phys. Rev. B **97**, 085104 (2018).

[50] D. Sehayek, A. Golubeva, M. S. Albergo, B. Kulchytskyy, G. Torlai, and R. G. Melko, Learnability scaling of quantum states: Restricted Boltzmann machines, Phys. Rev. B **100**, 195125 (2019).

[51] D. C. Mocanu, E. Mocanu, P. H. Nguyen, M. Gibescu, and A. Liotta, A topological insight into restricted Boltzmann machines, Mach. Learn. **104**, 243 (2016).
[52] P.-G. de Gennes and M. E. Fisher, Phénomènes aux parois dans un mélange binaire critique, C. R. Acad. Sci. Paris B **287**, 207 (1978).
[53] T. W. Burkhardt and E. Eisenriegler, Universal order-parameter profiles in confined critical systems with boundary fields, J. Phys. A **18**, L83 (1985).
[54] F. Iglói and H. Rieger, Density Profiles in Random Quantum Spin Chains, Phys. Rev. Lett. **78**, 2473 (1997).
[55] D. S. Fisher, Critical behavior of random transverse-field Ising spin chains, Phys. Rev. B **51**, 6411 (1995).
[56] R. Shankar and G. Murthy, Nearest-neighbor frustrated random-bond model in d=2: Some exact results, Phys. Rev. B **36**, 536 (1987).
[57] L. Reatto and G. L. Masserini, Shadow wave function for many-boson systems, Phys. Rev. B **38**, 4516 (1988).
[58] S. Vitiello, K. Runge, and M. H. Kalos, Variational Calculations for Solid and Liquid He 4 with a Shadow Wave Function, Phys. Rev. Lett. **60**, 1970 (1988).
[59] Y. Freund and D. Haussler, Unsupervised learning of distributions on binary vectors using two layer networks, in *Advances in Neural Information Processing Systems 4* (Morgan-Kaufmann Publishers Inc., San Francisco, 1992), pp. 912–919.
[60] N. Le Roux and Y. Bengio, Representational power of restricted Boltzmann machines and deep belief networks, Neural Comput. **20**, 1631 (2008).
[61] J. Martens, A. Chattopadhya, T. Pitassi, and R. Zemel, On the representational efficiency of restricted Boltzmann machines, in *Advances in Neural Information Processing Systems 26* (Curran Associates Inc., Red Hook, 2013), pp. 2877–2885.
[62] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017).
[63] S. Sorella and L. Capriotti, Green function Monte Carlo with stochastic reconfiguration: An effective remedy for the sign problem, Phys. Rev. B **61**, 2599 (2000).
[64] J. Thijssen, *Computational Physics* (Cambridge University Press, Cambridge, 2007).
[65] J. Boronat, Monte Carlo simulations at zero temperature: helium in one, two, and three dimensions, in *Microscopic Approaches to Quantum Liquids in Confined Geometries*, edited by E. Krotscheck and J. Navarro (World Scientific, Singapore, 2002), Chap. 2, pp. 21–90.
[66] O. Golinelli, T. Jolicoeur, and R. Lacaze, Haldane gaps in a spin-1 Heisenberg chain with easy-plane single-ion anisotropy, Phys. Rev. B **45**, 9798 (1992).
[67] N. Nemec, Diffusion Monte Carlo: Exponential scaling of computational cost for large systems, Phys. Rev. B **81**, 035119 (2010).
[68] L. Pollet, N. V. Prokof'ev, and B. V. Svistunov, Stochastic lists: Sampling multivariable functions with population methods, Phys. Rev. B **98**, 085102 (2018).
[69] Y. Bengio and O. Delalleau, Justifying and generalizing contrastive divergence, Neural Comput. **21**, 1601 (2009).
[70] P. Pfeuty, The one-dimensional Ising model with a transverse field, Ann. Phys. (NY) **57**, 79 (1970).
[71] G. G. Cabrera and R. Jullien, Role of boundary conditions in the finite-size Ising model, Phys. Rev. B **35**, 7062 (1987).
[72] A. P. Young and H. Rieger, Numerical study of the random transverse-field Ising spin chain, Phys. Rev. B **53**, 8486 (1996).
[73] M. Collura, L. Dell'Anna, T. Felser, and S. Montangero, On the descriptive power of neural-networks as constrained tensor networks with exponentially large bond dimension, arXiv:1905.11351.
[74] D. S. Fisher, Random Transverse Field Ising Spin Chains, Phys. Rev. Lett. **69**, 534 (1992).
[75] D. M. Ceperley and B. Alder, Quantum Monte Carlo, Science **231**, 555 (1986).
[76] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).
[77] C.-W. Liu, A. Polkovnikov, and A. W. Sandvik, Quantum Versus Classical Annealing: Insights from Scaling Theory and Results for Spin Glasses on 3-Regular Graphs, Phys. Rev. Lett. **114**, 147203 (2015).
[78] C.-W. Liu, A. Polkovnikov, and A. W. Sandvik, Quasi-adiabatic quantum Monte Carlo algorithm for quantum evolution in imaginary time, Phys. Rev. B **87**, 174302 (2013).
[79] T. Parolini, E. M. Inack, G. Giudici, and S. Pilati, Tunneling in projective quantum Monte Carlo simulations with guiding wave functions, Phys. Rev. B **100**, 214303 (2019).
[80] L. Stella and G. E. Santoro, Quantum annealing of an Ising spin-glass by Green's function Monte Carlo, Phys. Rev. E **75**, 036703 (2007).
[81] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Generalization properties of neural network approximations to frustrated magnet ground states, Nat. Commun. **11**, 1593 (2020).